KLE LAW ACADEMY BELAGAVI

(Constituent Colleges: KLE Society's Law College, Bengaluru, Gurusiddappa Kotambri Law College, Hubballi, S.A. Manvi Law College, Gadag, KLE Society's B.V. Bellad Law College, Belagavi, KLE Law College, Chikodi, and KLE College of Law, Kalamboli, Navi Mumbai)

STUDY MATERIAL

for

MICRO ECONOMICS

Prepared as per the syllabus prescribed by Karnataka State Law University (KSLU), Hubballi

Compiled by

Dr. Ramesh, Asst. Prof.

Reviewed by

Ms. Sarala Vaidya, Asst. Prof.

KLE Society's Law College, Bengaluru

Micro Economics

Syllabus

Objectives:

1. To help students acquire knowledge of some of the important principle and theories of microeconomics.
2. To provide the foundation for the study of other branches of economics.
3. To develop analytical, reasoning and graphical presentation skills.
4. To enable the student to appreciate the utility of economics in day-to-day life.

Unit –I: Microeconomics and Theory of Consumption

- Scope of microeconomics, its limitations and uses, positive and normative economics,
- Problem of choice-wants and resources, basic economic problems common to all economies, role of price mechanism in a mixed economy,
- Cardinal analysis-law of diminishing marginal utility, law of equip-marginal utility,
- Consumers surplus (Marshallian),
- Ordinal utility analysis indifference curves-properties, map, price line, consumer equilibrium, price effect, income effect and substitution effect.

Unit –II: Demand and Supply

- Law of demand, reasons for the downward slope of demand curve, exception to the law, changes in demand.
- Elasticity-kinds, types price elasticity with diagram, factors determining price elasticity, methods of measurement: percentage method, arc-method, total outlay method
- Law of supply, changes in supply.

Unit –III: Theory of Production

- Production function, law of variable proportions-short-run and long-run, laws of returns, economies of scale, lso-quants, locusts, production equilibrium
- Cost-opportunity cost, real cost, types-short-run, long-run-average, marginal, fixed, variable (with diagrams), long run cost curve
- Revenue-average, marginal,

Unit –IV: Product Pricing

- Concepts of firms, industry, equilibrium,
- Perfect competition, price and output determination and role of time element in the theory of price determination,
- Monopoly, price output determination, price discrimination,
- Monopolistic competition, price and output determination, selling costs, product differentiation, wastes in monopolistic competition,

— Oligopoly features, duopoly, monophony

Unit –V: Factor Pricing

— Nature of factor markets, marginal productivity theory of distribution.
— Rent-demand and supply theories, quasi rent, transfer earning.
— Wages-reasons for wage differentials, collective bargaining.
— Interest-classical, neo-classical and Keynesian.
— Profit-dynamic, innovation, risk and uncertainty theory.

Prescribed Books:

1. M.L. Seth, Micro Economics/ Principles of Economics
2. M.L. Jhingan, Micro Economics
3. H.S. Agarwal, Micro Economics

Reference Books:

1. Henderson J and R.E Quandt (1980), Microeconomic Theory: A mathematical approach, McGraw Hill; New Delhi.
2. Koutsoyiannis, A. (1990), Modern Microeconomics, Macmillan.
3. Lipsey, R.G. and K.A. Chrystal (1999), Principles of Economics (9th Ed) Oxford University Press, Oxford.
4. Samuelson, P.A. and W.D. Nordaus (1998), Economics, Tata McGraw Hill, New Delhi.
5. Stonier, A.W. and D.C. Hague (1972), A Textbook of Economic Theory, ELBS and Lognman Group, London.

TABLE OF CONTENTS

UNIT –I: MICROECONOMICS AND THEORY OF CONSUMPTION

Scope of Microeconomics, Its Limitations and Uses

Introduction

The term *'microeconomics'* derived from the Greek word *'mikros'*, meaning 'small'.

Microeconomics is the study of the economic actions of individual and small group of individuals.

Microeconomics may be defined as that branch of economic analysis which studies the economic behaviour of the individual unit, may be a person, a particular household, or a particular firm. It is a study of one particular unit rather than all the units combined together.

Definitions

- "Microeconomics is the study of particular firms, particular households, individual prices, wages, incomes, individual industries, particular commodities." *Boulding K E.*
- Microeconomics deals with the division of total output among industries, products and firms and the allocation of resources among competing groups. It considers problems of income distribution. Its interest is in relative prices of particular goods and services. *Ackley G.*
- "Microeconomics consists of looking at the economy through a microscope, as it were, to see how the millions of cells in the body economic-the individuals or households as consumers, and the individual or firms as producers-play their part in the working of the whole economic organism." *Prof. Learner*

Types of microeconomics

- Microeconomics has been divided into following three types:

i. *Simple micro-statics*

— It studies different microeconomic variables and their relationships at a given point of time under the conditions of equilibrium.

— It assumes that the values of microeconomic variables will remain unchanged and for this reason a simple micro-static model does not involve time period.

ii. *Comparative micro-static*

— It compares the equilibrium positions at different points of time.
— In micro-statics there is no change in the demand and supply functions. But due to a change in the independent variable, the demand function or the supply function or both will shift and the market will reach a new state of equilibrium giving a new price.

Micro-dynamics (The Cobweb Model)

— It explains lagged relationship between the microeconomic variables.
— Microdynamics throw full light on the happenings in the market during the period of transition from one static equilibrium to another.
— It is a study of disequilibrium; it studies the process through which the new equilibrium in the market is established.
— The Cobweb model is used to explain the dynamics of demand, supply and the price over long periods of time.

Scope

Microeconomic theory seeks to determine the mechanism by which the different economic units attain the position of equilibrium, proceeding from the individual units to a narrowly defined group such as a single industry or a single market. Since microeconomic analysis concerns itself with narrowly defined groups such as an industry or market. However it does not study the totality of behaviour of all units in the economy for any particular economic activity. In other words, the study of economic system or economy as a whole lies outside the domain of microeconomic analysis.

I. **Microeconomic Theory Studies Resource Allocation, Product and Factor Pricing.**

Microeconomic theory takes the total quantity of resources as given and seeks to explain how they are allocated to the production of particular goods. It is the allocation of resources that

determines what goods shall be produced and how they shall be produced. The allocation of resources to the production of various goods in a free-market economy depends upon the prices of the various goods and the prices of the various factors of production. Therefore, to explain how the allocation of resources is determined, microeconomics proceeds to analyze how the relative prices of goods and factors are determined. Thus the theory of product pricing and the theory of factor pricing (or the theory of distribution) falls within the domain of microeconomics. The theory of product pricing explains how the relative prices of cotton cloth, foodgrains, jute, kerosene oil and thousands of other goods are determined. The theory of distribution explains how wages (price for the use of labour), rent (payment for the use of land), interest (price for the use of capital) and profits (the reward for the entrepreneur) are determined. Thus, the theory of product pricing and the theory of factor pricing are the two important branches of microeconomic theory.

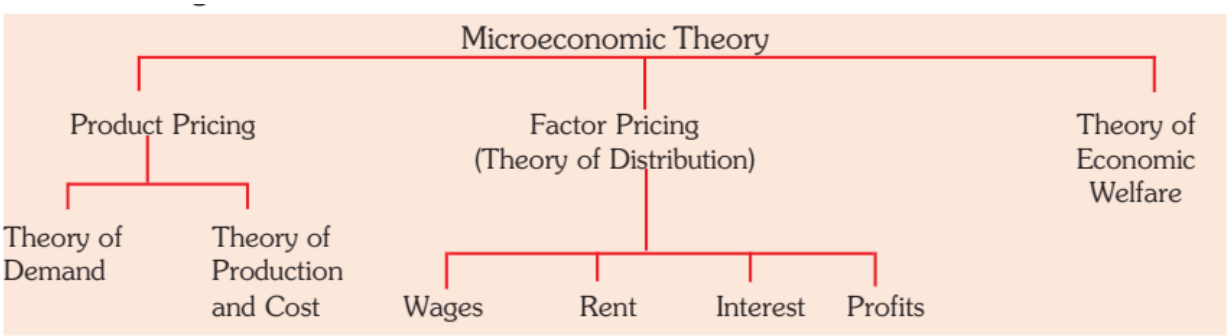
II. **Microeconomics as a Study of Economic Efficiency.**

Microeconomics also seeks to explain whether the allocation of resources determined is efficient. Efficiency in the allocation of resources is attained when the resources are so allocated that maximizes the satisfaction of the people. Economic efficiency involves three efficiencies; efficiency in production, efficiency in distribution of goods among the people (This is also called efficiency in consumption) and allocative economic efficiency, that is, efficiency in the direction of production. Microeconomic theory shows under what conditions these efficiencies are achieved. Microeconomics also shows what factors cause departure from these efficiencies and result in the decline of social welfare from the maximum possible level.

Economic efficiency in production involves minimization of cost for producing a given level of output or producing a maximum possible output of various goods from the given amount of outlay or cost incurred on productive resources. When such productive efficiency is attained, then it is no longer possible by any reallocation of the productive resources or factors among the production of various goods and services to increase the output of any good without a reduction in the output of some other good. Efficiency in consumption consists of distributing the given amount of produced goods and services among millions of the people for consumption in such a way as to maximize the total satisfaction of the society. When such efficiency is achieved it is no longer possible by any redistribution of goods among the people to make some people better off

without making some other ones worse off. Allocative economic efficiency or optimum direction of production consists of producing those goods which are most desired by the people, that is, when the direction of production is such that maximizes social welfare.

The whole content of microeconomic theory is presented in the following chart:



III. **Microeconomics and the Economy as a Whole.**

It is generally understood that microeconomics does not concern itself with the economy as a whole and an impression is created that microeconomics differs from macroeconomics in that whereas the latter examines the economy as a whole; the former is not concerned with it. But this is not fully correct. That microeconomics is concerned with the economy as a whole is quite evident from its discussion of the problem of allocation of resources in the society and judging the efficiency of the same. Both microeconomics and macroeconomics analyze the economy but with two different ways or approaches. Microeconomics examines the economy, so to say, microscopically, that is, it analyses the behaviour of individual economic units of the economy, their inter-relationships and equilibrium adjustment to each other which determine the allocation of resources in the society. This is known as general equilibrium analysis.

No doubt, microeconomic theory mainly makes particular or partial equilibrium analysis, that is, the analysis of the equilibrium of the individual economic units, taking other things remaining the same. But microeconomic theory, as stated above, also concerns itself with general equilibrium analysis of the economy wherein it is explained how all the economic units, various product markets, various factor markets, money and capital markets are interrelated and interdependent to each other and how through various adjustments and readjustments to the changes in them, they reach a general equilibrium, that is, equilibrium of each of them

individually as well as collectively to each other. Professor A. P. Lerner rightly points out, "Actually microeconomics is much more intimately concerned with the economy as a whole than is macroeconomics, and can even be said to examine the whole economy microscopically. We have seen how economic efficiency is obtained when the "cells" of the economic organism, the households and firms, have adjusted their behaviour to the prices of what they buy and sell. Each cell is then said to be 'in equilibrium.' But these adjustments in turn affect the quantities supplied and demanded and therefore also their prices. This means that the adjusted cells then have to readjust themselves. This in turn upsets the adjustment of others again and so on. An important part of microeconomics is examining whether and how all the different cells get adjusted at the same time. This is called general equilibrium analysis in contrast with particular equilibrium or partial equilibrium analysis. General equilibrium analysis is the microscopic examination of the inter-relationships of parts within the economy as a whole. Overall economic efficiency is only a special aspect of this analysis.

The Uses of Microeconomic

Microeconomics provides a theoretical framework for a systematic analysis of the economic behavior of the individual households, firms, industries and factor owners (the owners of factors of production—land, labour and capital). Economic theories bring out the nature of relationship between the interrelated economic variables and interdependence of the diverse elements of the economic system. The application of economic theories, logic and tools of analysis proves helpful in business decision making and in the formulation of economic policies of the government in many ways.

I. The study of economics contributes a great deal to the understanding of the complexity of the economic system and its working. In the words of Lerner, *'Microeconomic theory facilitates the understanding of what would be a hopelessly complicated confusion of billions of facts by constructing simplified models of behaviour which are sufficiently similar to the actual phenomena, to be of help in understanding them. These models enable the economists to examine the degree to which actual phenomena depart from certain ideal constructions that would most completely achieve individual and social objective.'* The clearer the understanding of the working of the economic system, the greater the efficiency in control and management of the economy.

II. Microeconomic theories establish *cause-and-effect relationship* between two or more interrelated and interdependent economic variables at micro level and, thereby, provide the basis for predicting the future course of economic events. Economic predictions are of great importance in planning the future course of economic activities by individuals, business firms and the government. Economic predictions may be conditional and inaccurate. For example, prediction of future price of a commodity may take the following form of a statement: If demand for a commodity increases, other things remaining the same, its price will increase. Despite the fact that a prediction of this kind is conditional, future trend of price is known more precisely than it would have been in the absence of any prediction. If one has information regarding the demand trend of a commodity and other related factors, one can predict the future trend in price with a greater accuracy. Approximate predictions are also of great importance for the consumers to make adjustment in their expenditure pattern; for the producers to plan their production; and for the public policy makers to formulate policy regarding price of the commodity.

III. Microeconomic theories contribute a great deal also in formulating economic policies and in examining the appropriateness and effectiveness of economic policies. Policy makers may, therefore, apply relevant microeconomic theories to explain the problem at hand and analyze the implications of alternative policies and select one which seems to be most appropriate. Public economic policies which go against the economic laws may not only prove ineffective but may also create more problems than they solve. For example, if the government increases the rate of excise duty for raising additional revenue without analyzing the nature of its demand and supply curves, the tax revenue may not increase, it may instead decrease. Besides, it may reduce both production and consumption and impose extra burden on the consumers. Microeconomic theories may be applied to examine the implications and effectiveness of the policies adopted by the government.

IV. Microeconomic theories, particularly price theory, can be and are, in fact, profitably used in *business decision making*. Although microeconomic theories may not offer a practicable solution to a problem of the real business world, they do help business

decision makers in building analytical models for projective future business scenario, which helps in specifying the nature of managerial problems and in determining appropriate policy actions.

V. One of the most important uses of microeconomic theories is to provide the basis for formulating propositions that maximize social welfare. Microeconomics examines how imperfect market conditions distort the allocation of resources (money, men and material), create inefficiency and lead to reduction in production, consumption and social welfare. The normative part of microeconomics, viz., welfare economics, suggests conditions for achieving 'pareto-optimality' in resource allocation with a view to maximizing social welfare. It also suggests ways and means to correct inefficient allocation of resources and to eliminate inefficiency. Although theoretical welfare propositions are of little practical importance, their analytical value is not reduced by their impracticability.

Limitations:

1. Excessive Generalization:

Despite the immense importance of macroeconomics, there is the danger of excessive generalization from individual experience to the system as a whole.

If an individual withdraws his deposits from the bank, there is no-harm in it, but if all the persons rushed to withdraw deposits, the bank would perhaps collapse.

2. Excessive Thinking in terms of Aggregates:

Again, macroeconomics suffers from excessive thinking in terms of aggregates, as it may not be always possible to have the homogeneous constituents. *Prof. Boulding* has pointed out that 2 apples + 3 apples = 5 apples is a meaningful aggregate ; 2 apples + 3 oranges = 5 fruits may be described as a fairly meaningful aggregate ; but 2 apples + 3 sky scrapers constitute a meaningless aggregate ; it is the last aggregate which brings forth the fallacy of excessive aggregative thinking.

3. Heterogeneous Elements:

It may, however, be remembered that macroeconomics deals with such aggregates as aggregate consumption, saving, investment and income, all composed of heterogeneous quantities. Money is the only measuring rod. But the value of money itself keeps on changing, rendering economic aggregates immeasurable and incomparable in real terms. As such, the sum or average of heterogeneous individual quantities loses their significance for accurate economic analysis and economic policy.

4. Differences within Aggregates:

Under this approach one is likely to overlook the differences within aggregates. For example, during the first decade of planning in India (from 1951-1961) the national income increased by 42% ; this, however, doesn't mean that the income of all the constituents, i.e., the wage earners or salaried persons increased by as much as that of entrepreneurs or businessmen. Hence, it takes no account of differences within aggregates.

5. Aggregates must be functionally related:

The aggregates forming the main body of macroeconomic theory must be significant and mutually consistent. In other words, these should be functionally related. For example, aggregate consumption and investment expenditures—which form part of the macroeconomic theory *(Y = C + I)* would have no importance, if they were not functionally related to the levels of income, interest and employment. If these composing aggregates are mutually inconsistent or are not functionally related, the study of macroeconomic theory will be of little use.

6. Limited Application:

Macroeconomics deals with positive economics in the sense of an analysis or how the aggregate theoretical models work—these are far removed from policy applications. These models explain the functioning of an economy and working of things in abstract and precise terms. Their abstraction and precision make such models unsuitable for use due to changes in significant variables from time to time and from one situation to another. But these limitations may be taken more in the nature of practical difficulties in formulating meaningful aggregates rather than factors invalidating the immense importance of macroeconomic analysis.

With the commencement of Keynes' General Theory and his basic equation, $Y = C + I$; interest in the study of macroeconomics has deepened. Significant breakthroughs in the computation of national income accounts (the study of which forms the very basis of macroeconomics) prove it beyond doubt that the limitations of macroeconomic studies are not insurmountable.

Positive and Normative Economics

Before we answer the question whether microeconomics is a positive or a normative science, let us understand what is a positive science and a normative science. According to J.N. Keynes, '… a positive science is a body of systematized knowledge concerning what is [and] a normative or regulatory science is a body of systematized knowledge relating to criteria of what ought to be and is concerned therefore with ideal as distinguished from actual.'5 Friedman has defined 'positive science' more elaborately and clearly. In his words, 'The ultimate goal of a positive science is the development of a "theory" or "hypothesis" that yields valid and meaningful (i.e., not truistic) predictions about phenomena not yet observed.'6 Judged against these definitions, economics as a social science turns out to be both a positive and a normative science as it deals with both positive and normative economic questions: 'what is' and 'what ought to be'. Thus, microeconomics is both a positive and a normative science. The positive and normative aspects of economic studies are described below.

Microeconomics as a Positive Science

Microeconomics as a positive science seeks to analyze and explain economic phenomena, i.e., economic aspects, issues or matters, as they are. As a positive science, microeconomics seeks to answer such questions as 'what is', 'why it is' and 'what will be …'. For example, look at some questions of positive nature. Why does a hungry person spend his or her first penny on food? When and why does he or she stop spending on food? Why do people buy more of a commodity when it decreases? How does a firm decide what and how much to produce? How does the firm determine the price of its product? How does a labour decide what job to take up? Microeconomics as a positive science finds the answer to such questions. These are some questions of positive nature.

Microeconomics as a positive science explains the economic behaviour of individual decision makers under given conditions; their response to change in economic conditions brings out the

relationship between the change in economic conditions and economic decision of the people. In fact, the main function of microeconomics is to establish cause-and-effect relationship, if there is any, between two or more economic variables at micro level and to provide the basis of prediction. Emphasizing the positive science character of economics, Friedman says, 'Economics as a positive science is a body of tentatively accepted generalizations about economic phenomena that can be used to predict the consequences of change in circumstances'.7 What Friedman said about economics is true for microeconomics. One of the main tasks of microeconomics is 'to provide a system of generalizations', i.e., formulating microeconomic theories capable of being used to predict economic phenomena at micro level. This makes microeconomics a positive science.

Microeconomics as a Normative Science

Microeconomics as a normative science seeks to answer the normative question 'what ought to be' on the basis of certain predetermined norms and social values. 'What people do' or 'what happens in the market' may not be desirable for the society. For example, production and sale of harmful goods like alcohol and cigarettes may be a very profitable business. But, a question arises here. 'Is production and sale of these goods desirable for the society?' This is a normative question—a question in public interest. Microeconomics as a social science examines this question from the angle of social desirability of production and sale of such goods. It examines the social costs and benefits of production of goods like alcohol and cigarettes and answers the question whether production and sales of such goods are socially desirable. Consider another microeconomic problem. Given the growth of population and the supply of residential houses in India, house rents, if not controlled, will continue to increase and have, in fact, increased exorbitantly. Here a question arises: 'Should house rents be allowed to increase depending on the market demand and supply conditions or be controlled and regulated to protect the interest of tenants?' This is a normative question—a question in public interest. Microeconomics as a normative science examines the issue in the interest of both landlords and the tenants and prescribes the reasonable rate of house rents.

Microeconomics, as a normative science, involves value judgement on 'what is good' and 'what is bad' for the society. The values are drawn from the moral, ethical, social and political aspirations of the society. Since microeconomics prescribes methods to correct undesirable

economic happenings, it is also called a prescriptive science. To have a comparative view of positive and normative character of microeconomics, recall the issue of high food grain prices in India in 2001 and in 2010. On one hand, there was surplus food grain production8 in India, on the other hand, large-scale starvation and starvation deaths were reported from different parts of the country. This was a paradoxical situation. Yet, the Food Corporation of India (FCI), responsible for fixing the food grain price, did not take any steps to bring down the price of food grains. This problem can be examined from both positive and normative angles. Examining 'how price of food grains is determined?' is a question for positive microeconomics and 'how should the prices of food grains be determined to prevent starvation?' is a question for normative microeconomics. It may, thus, be concluded that microeconomics is both a positive and normative science. However, it is important to note that microeconomics is, fundamentally, a positive science. It acquires its normative character from the application of microeconomic theories to examine the economic phenomena from their social desirability point of view; to show the need for a public policy action and; to evaluate the policy actions of the government.

Problem of choice-wants and resources

Economic theory enunciates the laws and principles which govern the functioning of an economy and its various parts. An economy exists because of two basic facts. First, human wants for goods and services are unlimited, and secondly, productive resources with which to produce goods and services are scarce. With our wants being virtually unlimited and resources scarce, we cannot satisfy all our wants and desires by producing everything we want. That being the case, a society has to decide how to use its scarce resources to obtain the maximum possible satisfaction of its members. It is this basic problem of scarcity which gives rise to many of the economic problems which have long been the concern of economists.

Since it is not possible to satisfy all wants with the limited means of production, every society must decide some way of selecting those wants which are to be satisfied. The necessity for economizing arises therefore from the fact that we have limited productive resources such as land, raw materials, skilled manpower, capital equipment etc. at our disposal. These resources being found in limited quantity (the quantity may however increase over time), the goods they can produce are also limited. Goods are thus scarce because the productive resources are scarce. Since the resources are limited in relation to our wants, we should get most out of what we have.

Thus a society is faced with the problem of choice—choice among the vast array of wants that are to be satisfied. If it is decided to use more resources in one line of production, then resources must be withdrawn from the production of some other goods. The scarcity of resources therefore compels us to choose among the different channels of production to which resources are to be devoted. In other words, we have the problem of allocating scarce resources so as to achieve the greatest possible satisfaction of wants of the people. This is the economic problem. It is also called the economizing problem.

The scarcity of resources relative to human wants gives rise to the struggle of man for sustenance and efforts by him to promote his well-being. That the scarcity of resources in relation to human wants is the fundamental economic problem can be easily understood in the context of poor and developing countries like India where quite a large number of population live at a bare subsistence level. The struggle for existence due to the scarcity of resources is too obvious in them to need any elaborate explanation. However, to say that the developed countries, such as the U.S.A., where affluence and prosperity have been brought about also confront the scarcity problem raises some doubts. But the fact is, despite their affluence and riches, developed societies too face the problem of scarcity. Of course, their possession of goods and services has enormously increased, but so have their wants. Indeed, their wants for goods and services have been multiplying during the course of economic growth so that their present wants still remain ahead of their resources and capability to produce. As has been said above, the problem of scarcity of resources is not only the result of availability of limited resources and capability to produce but also of human wants. So long as human wants for goods and services remain ahead of the resources, both natural and acquired, the economic problem of scarcity would exist. If Americans today, for example, were content to live at the level of the Indian middle class people, all their wants would probably be fully satisfied with their available resources and capacity to produce. In that situation they would face little or no scarcity and economic problem for them would disappear. However, it needs to be emphasized again that the affluent and developed countries of the U.S.A. and Western Europe face the problem of scarcity even today as their present wants run ahead of their increased resources and capability to produce.

Since all wants cannot be satisfied due to scarcity of resources we face the problem of choice–choice among multiple wants which are to be satisfied. If it is decided to use more

resources in one line of production, some resources must be withdrawn from another commodity. Thus, the problem of choice from the viewpoint of the society as a whole refers to which goods and in what quantities are to be produced and productive resources allocated for their production accordingly so as to achieve greatest possible satisfaction of the people. An eminent English economist Lord Robbins defines economics in terms of this basic economic problem. According to him, "Economics is a science which studies human behaviour as a relationship between ends and scarce resources which have alternative uses." Here ends refer to wants which are considered to be unlimited. The use and allocation of scarce resources to produce goods and services have to be such as would maximize satisfaction. This applies both to the behaviour of the individual and of the society as a whole.

The scarcity of resources also compels us to decide how the different goods should be produced, that is, what production methods should be chosen for the production of goods so as to make best possible use of the available resources. If the resources were unlimited, the problem of how goods should be produced would not have arisen. This is because with unlimited resources it would not matter whichever method, efficient or inefficient, was employed for production of goods.

Further, due to scarcity of resources goods cannot be produced in abundant quantities to satisfy all wants of all the people of a society. This raises another problem of choice, namely, who should get how much from the national output. This means how the national product is distributed among various members of a society.

Thus, problem of scarcity gives rise to some problems generally known as basic economic problems which a society has to solve so as to promote material well-being of its people. These basic economic problems relate to what commodities are to be produced, how they are to be produced, how the national product is to be distributed among the people, and how much to provide for future growth. It is with regard to these problems of resource allocation, the choice of production methods, distribution and economic growth, which have their roots in scarcity of resources, those economists have been asking questions from time to time and providing answers for them. Besides, economists have also been raising questions about the efficiency of the resource allocation for the production of goods and their distribution among them people. This question of economic efficiency is aimed at knowing whether or not a particular allocation of

resources to the production of various goods and distribution of income among them ensures maximum social welfare.

The knowledge of the scope and purpose of economic theory can be obtained from the type of relevant questions that have been asked by the economists from time to time and their mode of answering them.

Basic Economic Problems Common to All Economies

Human wants are endless as they go on increasing whereas resources available to satisfy human wants are scarce at any point of time. Though resources also increase over time—there is always a gap between demand for and supply of resources. In their effort to meet the ever-growing needs using limited resources, economies aim at (i) achieving efficiency in production and distribution of commodities; and (ii) achieving full employment, high growth in output and stability in employment and growth. In their effort to achieve these goals, societies face certain problems.

For pedagogic reasons, the problems faced by the economies can be grouped under two categories: (a) problems in achieving efficiency in production and distribution of goods and services; and (b) problems in achieving a reasonably high growth rate, full employment and stability in the economy. The nature and origin of the two kinds of problems are discussed below.

Problems in Maximizing Production and Optimizing Distribution

The problems that arise in maximizing production and optimizing distribution of goods and services, often referred to as the basic problems, are of three kinds:

i. What to produce?
ii. How much to produce?
iii. How to produce?

i. **What to Produce?**

The problem 'what to produce?' is the problem of choice between commodities to be produced. This problem arises mainly for two reasons: (i) scarcity of resources does not permit production of all the goods and services that people would like to consume; and (ii) all the goods and services are not equally important in terms of their utility for the consumers. Some commodities yield higher utility than the others. Since all the goods and services cannot be produced for lack of resources, and all that is produced may not be bought by the consumers, the problem of making choice between the commodities arises. The problem 'what to produce' is essentially the problem of efficient allocation of scarce resources so that the output is maximum and the output mix is optimum. The objective is to satisfy the maximum needs of the maximum number of people.

ii. **How Much to Produce?**

The question 'how much to produce?' is the problem of determining the quantity of each commodity and service to be produced. This problem too arises due to scarcity of resources. For surplus production would mean wastage of scarce resources. Resource wastage defeats the objective of maximizing production from the given resources.

iii. **How to Produce?**

The problem 'how to produce?' is the problem of choice of technology, *i.e.,* the production technique. Here the problem is how to determine optimum combination of inputs—labour and capital—so that production of goods or services is maximized. This problem, too, arises mainly because of scarcity of resources. If labour and capital were available in unlimited quantities, any amount of labour and capital could be combined to produce a commodity. But resources are not available in unlimited quantity. Therefore, choosing a technology that uses resources most economically becomes a necessity.

Another important factor that gives rise to this problem is that a given quantity of a commodity can be produced with a number of alternative techniques, i.e., alternative input combinations. For example, it is always technically possible to produce a given quantity of wheat with more of labour and less of capital (i.e., with a labour-intensive technology) and with more of capital and

less of labour (i.e., with a capital-intensive technology). The same is true for most commodities. In case of some commodities, however, choices are limited. For example, production of woolen carpets and other items of handicrafts are by nature labour-intensive, while production of some other goods likes machinery, aircraft, turbines, etc. are by nature capital-intensive. In case of most commodities, however, alternative technologies are available. But the alternative techniques of production involve different costs. Therefore, the problem of choices of technology arises.

Role of price mechanism in a mixed economy

Introduction

In every economy, some fundamental problems arise and the economy faces them. Since these problems are common to all economies they are known as the central problems of an economy. The root cause of all these problems is the law of scarcity. Had the resources been unlimited like sunshine and air, there would have been no problem whatsoever. In different economies these problems are attempted to be solved differently.

The central problems are:

- What to produce?
- How to produce?
- For whom to produce?
- What should be the rate of growth of an economy?
- How to secure smooth working of an economy?

How these problems solved?

All the mentioned problems are solved by the 'price mechanism'.

In fact, price mechanism is a system of economic organization in which each individual in his/her capacity as a consumer, producer, and factor-owner is engaged in economic activity with a large measure of freedom. In every society there are legal and social institutions which govern it.

The individual economic actions must confirm the system prevailing in an economy.

What is price?

The price of anything is the rate at which it can be exchanged for anything else.

What is price mechanism?

Price mechanism is the formation and operation mechanism of market price which influence and control with supply and demand in the competitive process. It includes price formation mechanism and price adjustment mechanism.

Different types of economics systems:

- Free enterprise economy,
- Socialist economy,
- Mixed economy.

What is mixed economy?

It is a mixture of capitalism and socialism. Under this system there is freedom of economic activities and government interferences for the social welfare. Hence it is a blend of both the economies. The concept of mixed economy is of recent origin. The developing countries like India have adopted mixed economy to accelerate the pace of economic development. Even the developed countries like UK, USA, etc. have also adopted 'Mixed Capitalist System'. "Mixed economy is that economy in which both public and private sectors cooperate". Prof. Paul Samuelson.

Main features of mixed economy:

i. Co-existence of private and public sector
ii. Personal freedom
iii. Private property is allowed
iv. Economic planning
v. Price mechanism and controlled price
vi. Profit motive and social welfare

vii. Check on economic inequalities

viii. Control of monopoly power

Types of mixed economy:

i. Capitalistic mixed economy
ii. Socialistic mixed economy

However, this system is again sub-divided into two parts:

i. Liberal socialistic mixed economy
ii. Centralized socialistic mixed economy

Role of price mechanism in a mixed economy:

What to produce and in what quantity?

- In a capitalist economy profit motive or price mechanism decides as to what commodity and in what quantity is to be produced. Generally, the producers are induced to produce those commodities which can be sold for higher prices and which promise them high profits.

- In a mixed capitalist economy also price mechanism solves this problem. But in such an economy, the government keeps on interfering for general welfare of the people. In these economies, the government undertakes the production and distribution of essential goods and services.

How to produce?

- The problem of how to produce is the problem of choosing the method or technique of production.

- In a free market economy or capitalist economy this problem is solved automatically. In order to maximize the profits, each producer tries to reduce the cost of production. The producer would choose that method or technique of production which is the cheapest and which minimizes the cost of production.

- The factors of production are available in the market at different prices. The producers also come to the market to purchase them. In order to maximize their profits, the producers would purchase more of those factors whose price is low and less of those whose price is relatively high. In this way the problem of the choice of method of production is solved by the price mechanism.

- Under free market economy or capitalist economy, price mechanism allocates the productive resources.

For whom to produce?

- In a mixed capitalist economy need of the poor people are kept in mind by the government.

- The government undertakes the production and distribution of certain essential goods and services for the best interest of the society. Through price control and rationing the government ensures the supply of goods to both the rich and the poor classes of the society.

- But in mixed capitalist economy, the allocation of productive resources between different uses is done partly through the *government direction* and partly through *price mechanism*.

- In the public sector, all decisions as to 'how to produce' are taken by the government whereas in the private sector price mechanism decides the allocation of the resources among different lines of production. This decision id done on the basis of profit motive. But even in private sector the working of price mechanism is regulated by the government *rules and regulations*.

- The private enterprise is also allowed to exist for all other non-essential goods and services. But this sector also functions under the supervision and control of the government.

- The government adjusts demand and supply of goods. To remove the scarcity of certain goods, the government gives to the producers of these goods incentives like bounties, tax

rebates, etc., and to discourage the production of certain goods, the government provides disincentives like imposing of higher taxes on those commodities.

- The government imposes high direct taxes and reduces the purchasing power of the rich. The government also levies high duties on the production of luxury goods and reduces their production. On the other hand, the government provides many social securities to the poor people.

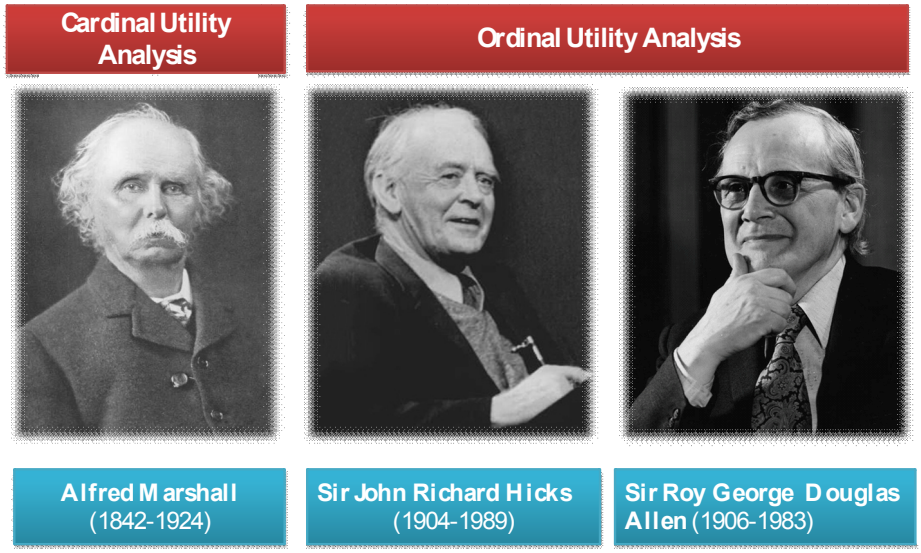
CARDINAL ANALYSIS

Basic understanding of concept of utility

The utility of a consumer is a measure of the satisfaction the consumer derives from consumption of goods and services. An individual's consumption bundle is the collection of all the goods and services consumed by that individual. An individual's utility function gives the total utility generated by his or her consumption bundle. A util is a unit of utility. Two types of utility concepts: Total Utility and Marginal Utility

Cardinal vs. Ordinal utility

Basis for Comparison	**Cardinal Utility**	**Ordinal Utility**
Meaning	Expressed numerically	Expressed by making comparison
Approach	Quantitative	Qualitative
Realistic	Less	More
Measurement	Utils	Ranks
Analysis	Marginal Utility Analysis	Indifference curve analysis, Revealed preference theory
Proposed by	Classical and Neo-classical economists like Alfred Marshall and his followers	Modern economists like J.R. Hicks, and R.G.D. Allen



Law of diminishing marginal utility

An important tenet of cardinal utility analysis relates to the behaviour of marginal utility. This familiar behaviour of marginal utility has been stated in the Law of Diminishing Marginal Utility according to which marginal utility of a good diminishes as an individual consumes more units of a good. In other words, as a consumer takes more units of a good, the extra utility or satisfaction that he derives from an extra unit of the good goes on falling. It should be carefully noted that it is the marginal utility and not the total utility that declines with the increase in the consumption of a good. The law of diminishing marginal utility means that the total utility increases at a decreasing rate. Marshall who has been a famous exponent of the cardinal utility analysis has stated the law of diminishing marginal utility as follows:

"The additional benefit which a person derives from a given increase of his stock of a thing diminishes with every increase in the stock that he already has."

This law is based upon two important facts. First, while the total wants of a man are virtually unlimited, each single want is satiable. Therefore, as an individual consumes more and more units of a good, intensity of his want for the good goes on falling and a point is reached where the individual no longer wants any more units of the good. That is, when saturation point is

reached, marginal utility of a good becomes zero. Zero marginal utility of a good implies that the individual has all that he wants of the good in question. The second fact on which the law of diminishing marginal utility is based is that the different goods are not perfect substitutes for each other in the satisfaction of various wants. When an individual consumes more and more units of a good, the intensity of his particular want for the good diminishes but if the units of that good could be devoted to the satisfaction of other wants and yielded as much satisfaction as they did initially in the satisfaction of the first want, marginal utility of the good would not have diminished.

It is obvious from above that the law of diminishing marginal utility describes a familiar and fundamental tendency of human nature. This law has been arrived at by introspection and by observing how consumers behave.

Illustration of the Law of Diminishing Marginal Utility

Consider Table 7.1. where we have presented the total and marginal utilities derived by a person from cups of tea consumed per day. When one cup of tea is taken per day, the total utility derived by the person is 12 utils. And because this is the first cup its marginal utility is also 12 utils. With the consumption of 2nd cup per day, the total utility rises to 22 utils but marginal utility falls to 10. It will be seen from the table that as the consumption of tea increases to six cups per day, marginal utility from the additional cup goes on diminishing (i.e. the total utility goes on increasing at a diminishing rate). However, when the cups of tea consumed per day increases to seven, then instead of giving positive marginal utility, the seventh cup gives negative marginal utility equal to - 2 utils. This is because too many cups of tea consumed per day (say more than six for a particular individual) may cause acidity and gas trouble. Thus, the extra cups of tea beyond six to the individual in question gives him disutility rather than positive satisfaction.

Table 7.1. Diminishing Marginal Utility

Cups of tea consumed per day (Q)	*Total Utility (utils)* TU	*Marginals Utility (utils)* $\frac{\Delta TU}{\Delta Q}$
1	12	12
2	22	10
3	30	8
4	36	6
5	40	4
6	41	1
7	39	–2
8	34	–5

Figure 7.1 illustrates the total utility and the marginal utility curves. The total utility curve drawn in Figure 7.1 is based upon three assumptions. First, as the quantity consumed per period by a consumer increases his total utility increases but at a decreasing rate. This implies that as the consumption per period of a commodity by the consumer increases, marginal utility diminishes as shown in the lower panel of Figure 7.1. Secondly, as will be observed from the figure, when the rate of consumption of a commodity per period increases to Q4, the total utility of the consumer reaches its maximum level. Therefore, the quantity Q4, of the commodity is called satiation quantity or satiety point. Thirdly, the increase in the quantity consumed of the good per period by the consumer beyond the satiation point has an adverse effect on his total utility, that is, his total utility declines if more than Q4 quantity of the good is consumed. This means beyond Q4 marginal utility of the commodity for the consumer becomes negative as will be seen from the lower panel of Figure 7.1 beyond the satiation point Q4 marginal utility curve MU goes below the X-axis indicating it becomes negative beyond quantity Q4 per period of the commodity consumed.

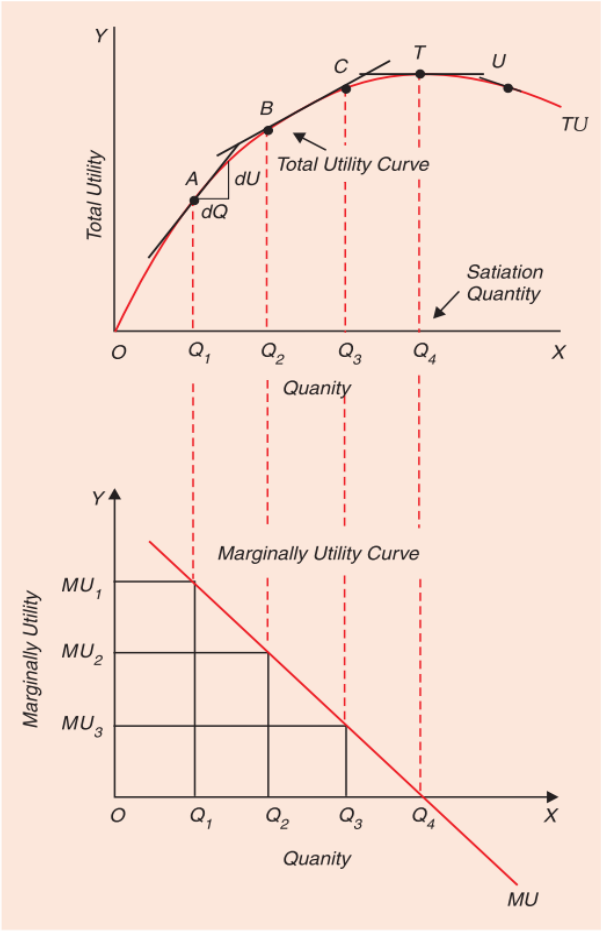


Fig. 7.1. *Total Utility and Marginal Utility*

It is important to understand how we have drawn the marginal utility curve. As stated above, marginal utility is the increase in total utility of the consumer caused by the consumption of an additional unit of the commodity per period. We can directly find out the marginal utility of the successive units of the commodity consumed by measuring the additional utility which a consumer obtains from successive units of the commodity and plotting them against their respective quantities. However, in terms of calculus, marginal utility of a commodity X is the slope of the total utility function $U = f(Qx)$. Thus, we can derive the marginal utility curve by measuring the slope at various points of the total utility curve TU in the upper panel of Figure7.1 by drawing tangents at them. For instance, at the quantity Q1 marginal utility (i.e. dU dQ = MU 1) is found out by drawing tangent at point A and measuring its slope which is then plotted against quantity Q1 in the lower panel of Figure 7.1. In the lower panel we measure marginal utility of the commodity on the Y-axis. Likewise, at quantity Q2 marginal utility of the commodity has been obtained by measuring slope of the total utility curve TU at point B and

plotting it in the lower panel against the quantity Q2. It will be seen from the figure that at Q4 of the commodity consumed, the total utility reaches at the maximum level T. Therefore, at quantity Q4 the slope of the total utility curve is zero at this point. Beyond the quantity Q4 the total utility declines and marginal utility becomes negative. Thus, quantity Q4 of the commodity represents the satiation quantity.

Another important relationship between total utility and marginal utility is worth noting. At any quantity of a commodity consumed the total utility is the sum of the marginal utilities. For example, if marginal utility of the first, second, and third units of the commodity consumed are 15, 12, and 8 units, the total utility obtained from these three units of consumption of the commodity must equals 35 units (15 + 12 + 8 = 35). Similarly, in terms of graphs of total utility and marginal utility depicted in Figure 7.1 the total utility of the quantity Q4 of the commodity consumed is the sum of the marginal utilities of the units of commodity up to point Q4. That is, the entire area under the marginal utility curve MU in lower panel up to the point Q4 is the sum of marginal utilities which must be equal to the total utility Q4T in the upper panel.

Marginal Utility and Consumer's Tastes and Preferences

The utility people derive from consuming a particular commodity depends on their tastes and preferences. Some consumers like oranges, others prefer apples and still others prefer bananas for consumption. Therefore, the utility which different individuals get from these various fruits depends on their tastes and preferences. An individual would have different marginal utility curves for different commodities depending on his tastes and preferences. Thus, utility which people derive from various goods reflect their tastes and preferences for them. However, it is worth noting that we cannot compare utility across consumers. Each consumer has a unique subjective utility scale. In the context of cardinal utility analysis, a change in consumer's tastes and preferences means a shift in his one or more marginal utility curves. However, it may be noted that a consumer's tastes and preferences do not frequently change, as these are determined by his habits. Of course, tastes and preferences can change occasionally. Therefore, in economic theory we generally assume that tastes or preferences are given and relatively stable.

Significance of Diminishing Marginal Utility

The significance of the diminishing marginal utility of a good for the theory of demand is that it helps us to show that the quantity demanded of good increases as its price falls and vice versa. Thus, it is because of the diminishing marginal utility that the demand curve slopes downward. This will be explained in detail later in this chapter. If properly understood the law of diminishing marginal utility applies to all objects of desire including money. But it is worth mentioning that marginal utility of money is generally never zero or negative. Money represents purchasing power over all other goods, that is, a man can satisfy all his material wants if he possesses enough money. Since man's total wants are practically unlimited, therefore, the marginal utility of money to him never falls to zero.

The marginal utility analysis has a good number of uses and applications in both economic theory and policy. The concept of marginal utility is of crucial significance in explaining determination of the prices of commodities. The discovery of the concept of marginal utility has helped us to explain the paradox of value which troubled Adam Smith in "The Wealth of Nations." Adam Smith was greatly surprised to know why water which is so very essential and useful to life has such a low price (indeed no price), while diamonds which are quite unnecessary, have such a high price. He could not resolve this water-diamond paradox. But modern economists can solve it with the aid of the concept of marginal utility. According to the modern economists, the total utility of a commodity does not determine the price of a commodity and it is the marginal utility which is crucially important determinant of price. Now, the water is available in abundant quantities so that its relative marginal utility is very low or even zero. Therefore, its price is low or zero. On the other hand, the diamonds are scarce and therefore their relative marginal utility is quite high and this is the reason why their prices are high. Prof. Samuelson explains this paradox of value in the following words:—"The more there is of a commodity, the less the relative desirability of its last little unit becomes, even though its total usefulness grows as we get more of the commodity. So, it is obvious why a large amount of water has a low price Or why air is actually a free good despite its vast usefulness. The many later units pull down the market value of all units."

Besides, the Marshallian concept of consumer's surplus is based on the principle of diminishing marginal utility.

Law of equi-marginal utility

Law of equi-marginal utility occupies an important place in cardinal utility analysis. It is through this principle that consumer's equilibrium is explained. A consumer has a given income which he has to spend on various goods he wants. Now, the question is how he would allocate his given money income among various goods, that is to say, what would be his equilibrium position in respect of the purchases of the various goods. It may be mentioned here that consumer is assumed to be 'rational', that is, he carefully calculates utilities and substitutes one good for another so as to maximize his utility or satisfaction. Suppose there are only two goods X and Y on which a consumer has to spend a given income. The consumer's behaviour will be governed by two factors: first, the marginal utilities of the goods and secondly, the prices of two goods.

Suppose the prices of the goods are given for the consumer. The law of equi-marginal utility states that the consumer will distribute his money income between the goods in such a way that the utility derived from the last rupee spent on each good is equal. In other words, consumer is in equilibrium position when marginal utility of money expenditure on each good is the same. Now, the marginal utility of money expenditure on a good is equal to the marginal utility of a good divided by the price of the good. In symbols,

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

where *MU* m is marginal utility of money expenditure and *MUx* is the marginal utility of *X* and *Px* is the price of *X*. The law of equi-marginal utility can therefore be stated thus: the consumer will spend his money income on different goods in such a way that marginal utility of money expenditure on each good is equal. That is, consumer is in equilibrium in respect of the purchases of two goods X and Y when

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

Now, if $\frac{MU_x}{P_x}$ and $\frac{MU_y}{P_y}$ are not equal and $\frac{MU_x}{P_x}$ is greater than $\frac{MU_y}{P_y}$, then the consumer will substitute good X for good Y. As a result of this substitution, the marginal utility of good X will fall and

marginal utility of good Y will rise. The consumer will continue substituting good X for good Y

$\frac{MU_x}{P_x}$ becomes equal to $\frac{MU_y}{P_y}$. When $\frac{MU_x}{P_x}$ becomes equal to $\frac{MU_y}{P_y}$ the consumer is in equilibrium.

But the equality of $\frac{MU_x}{P_x}$ and $\frac{MU_y}{P_y}$ can be achieved not only at one level but at different levels of expenditure. The question is how far a consumer goes in purchasing the goods he wants. This is determined by the size of his money income. With a given income and money expenditure a rupee has a certain utility for him: this utility is the marginal utility of money to him. Since the law of diminishing marginal utility applies to money income also, the greater the size of his money income the smaller the marginal utility of money to him. Now, the consumer will go on purchasing goods until the marginal utility of money expenditure on each good becomes equal to the marginal utility of money to him. Thus, the consumer will be in equilibrium when the following equation holds good:

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

Where MU_m is marginal utility of money expenditure (that is, the utility of the last rupee spent on each good).

If there are more than two goods on which the consumer is spending his income, the above equation must hold good for all of them. Thus

$$\frac{MU_x}{P_x} = \frac{MU_y}{P_y} = \cdots \ldots \ldots \frac{MU_n}{P_n} = MU_m$$

Let us illustrate the law of equi-marginal utility with the aid of an arithmetical table given below:

Table 7.2. Marginal Utility of Goods X and Y

Units	MU_x *(Utils)*	MU_y *(Utils)*
1	20	24
2	18	21
3	16	18
4	14	15
5	12	9
6	10	3

Let the prices of goods X and Y be Rs. 2 and Rs. 3 respectively. Reconstructing the above table by dividing marginal utilities *(MUx)* of X by Rs. 2 and marginal utilities *(MUy)* of Y by Rs. 3 we get the Table 7.3.

Table 7.3. Marginal Utility of Money Expenditure

Units	$\frac{MU_x}{P_x}$	$\frac{MU_y}{P_y}$
1	10	8
2	9	7
3	8	6
4	7	5
5	6	3
6	5	1

Suppose a consumer has money income of Rs. 24 to spend on the two goods. It is worth noting that in order to maximize his utility the consumer will not equate marginal utilities of the goods because prices of the two goods are different. He will equate the marginal utility of the last rupee (i.e. marginal utility of money expenditure) spent on these two goods. In other words, he will equate $\frac{MU_x}{P_x}$ with $\frac{MU_y}{P_y}$ while spending his given money income on the two goods. By looking at the Table 7.3 it will become clear that $\frac{MU_x}{P_x}$ is equal to 5 utils when the consumer purchases 6 units of good X and $\frac{MU_y}{P_y}$ is equal to 5 utils when he buys 4 units of good Y. Therefore, consumer will be in equilibrium when he is buying 6 units of good X and 4 units of good Y and will be spending (Rs. 2 × 6 + Rs. 3 × 4) = Rs. 24 on them that are equal to consumer's given income. Thus, in the equilibrium position where the consumer maximizes his utility,

$$\frac{MU_x}{P_x} = \frac{MU_y}{P_y} = \cdots \ldots \ldots \frac{MU_n}{P_n} = MU_m$$

$$\frac{10}{2} = \frac{15}{3} = 5$$

Thus, marginal utility of the last rupee spent on each of the two goods he purchases is the same, that is, 5 utils.

Consumers' equilibrium is graphically portrayed in Fig. 7.2. Since marginal utility curves of goods slope downward, curves depicting $\frac{MU_x}{P_x}$ and $\frac{MU_y}{P_y}$ also slope downward. Thus, when the consumer is buying OH of X and OK of Y, then

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

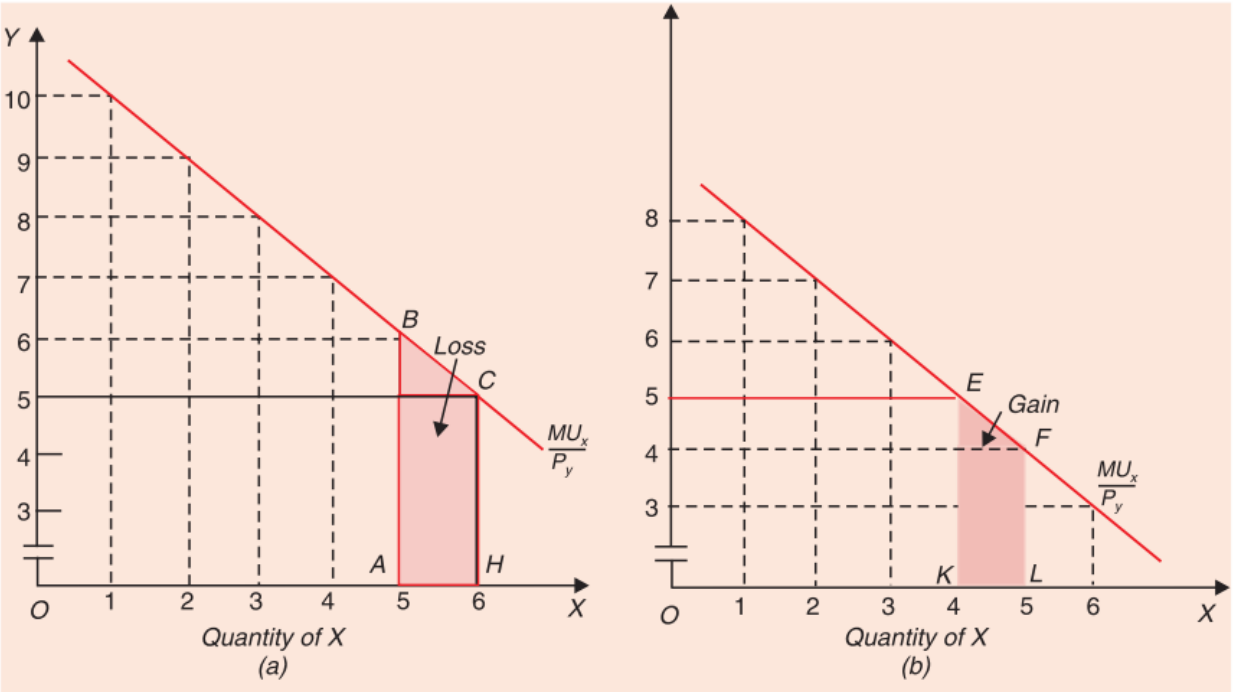


Fig. 7.2. *Equi-Marginal Utility Principle and Consumer's Equilibrium*

Therefore, the consumer is in equilibrium when he is buying 6 units of X and 4 units of Y. No other allocation of money expenditure will yield him greater utility than when he is buying 6 units of commodity X and 4 units of commodity Y. Suppose the consumer buys one unit less of good X and one unit more of good Y. This will lead to the decrease in his total utility. It will be observed from Figure 7.2 (a) that the consumption of 5 units instead of 6 units of commodity X means a loss in satisfaction equal to the shaded area ABCH and from Fig. 7.2(b) it will be seen that consumption of 5 units of commodity Y instead of 4 units will mean gain in utility equal to the shaded area KEFL. It will be noticed that with this rearrangement of purchases of the two goods, the loss in utility ABCH exceeds gain in utility KEFL. Thus, his total satisfaction will fall as a result of this rearrangement of purchases. Therefore, when the consumer is making

purchases by spending his given income in such a way that $\frac{MU_x}{P_x} = \frac{MU_y}{P_y}$, he will not like to make any further changes in the basket of goods and will therefore be in equilibrium situation by maximizing his utility.

Limitations of the Law of Equi-Marginal Utility

Like other laws of economics, law of equi-marginal utility is also subject to various limitations. This law, like other laws of economics, brings out an important tendency among the people. This is not necessary that all people exactly follow this law in the allocation of their money income and therefore all may not obtain maximum satisfaction. This is due to the following reasons:-

1) For applying this law of equi-marginal utility in the real life, consumer must weigh in his mind the marginal utilities of different commodities. For this he has to calculate and compare the marginal utilities obtained from different commodities. But it has been pointed out that the ordinary consumers are not so rational and calculating. Consumers are generally governed by habits and customs. Because of their habits and customs they spend particular amounts of money on different commodities, regardless of whether the particular allocation maximizes their satisfaction or not.

2) For applying this law in actual life and equate the marginal utility of the last rupee spent on different commodities, the consumers must be able to measure the marginal utilities of different commodities in cardinal terms. However, this is easier said than done. It has been said that it is not possible for the consumer to measure utility cardinally. Being a state of psychological feeling and also there being no objective units with which to measure utility, it is cardinally immeasurable. It is because of the immeasurability of utility in cardinal terms that the consumer's behaviour has been explained with the help of ordinal utility by J.R. Hicks and R.G.D. Allen. Ordinal utility analysis involves the use of indifference curves which we shall explain in the next chapter.

3) Another limitation of the law of equi-marginal utility is found in case of indivisibility of certain goods. Goods are often available in large indivisible units. Because the goods are indivisible, it is not possible to equate the marginal utility of money spent on them. For

instance, in allocating money between the purchase of car and foodgrains, marginal utilities of the last rupee spent on them cannot be equated. An ordinary car costs about Rs. 300,000 and is indivisible, whereas foodgrains are divisible and money spent on them can be easily varied. Therefore, the marginal utility of rupee obtained from cars cannot be equalized with that obtained from foodgrains. Thus, indivisibility of certain goods is a great obstacle in the way of equalization of marginal utility of a rupee from different commodities.

Consumer surplus (Marshallian)

Meaning of Consumer Surplus

The concept of consumer surplus was first formulated by Dupuit in 1844 to measure social benefits of public goods such as canals, bridges, national highways. Marshall further refined and popularized this in his 'Principles of Economies' published in 1890. The concept of consumer surplus became the basis of old welfare economics. Marshall's concept of consumer's surplus was based on the cardinal measurability and interpersonal comparisons of utility. According to him, every increase in consumer's surplus is an indicator of the increase in social welfare. As we shall see below, consumer's surplus is simply the difference between the price that 'one is willing to pay' and 'the price one actually pays' for a particular product.

Concept of consumer's surplus is a very important concept in economic theory, especially in theory of demand and welfare economics. This concept is important not only in economic theory but also in formulation of economic policies such as taxation by the Government and price policy pursued by the monopolistic seller of a product. The essence of the concept of consumer's surplus is that a consumer derives extra satisfaction from the purchases he daily makes over the price he actually pays for them. In other words, people generally get more utility from the consumption of goods than the price they actually pay for them. It has been found that people are prepared to pay more price for the goods than they actually pay for them. This extra satisfaction which the consumers obtain from buying a good has been called consumer surplus. Thus, Marshall defines the consumer's sruplus in the following words: "excess of the price which a consumer would be willing to pay rather than go without a thing over that which he actually does pay is the economic measure of this surplus satisfaction....it may be called consumer's surplus."

The amount of money which a person is willing to pay for a good indicates the amount of utility he derives from that good; the greater the amount of money he is willing to pay, the greater the utility he obtains from it. Therefore, the marginal utility of a unit of a good determines the price a consumer will be prepared to pay for that unit. The total utility which a person gets from a good is given by the sum of marginal utilities *(ΣMU)* of the units of a good purchased and the total price which he actually pays is equal to the price per unit of the good multiplied by the number of units of it purchased. Thus:

Consumer's surplus = What a consumer is willing to pay minus what he actually pays.

= Σ Marginal utility – (Price × Number of units of a commodity purchased)

The concept of consumer surplus is derived from the law of diminishing marginal utility. As we purchase more units of a good, its marginal utility goes on diminishing. It is because of the diminishing marginal utility that consumer's willingness to pay for additional units of a commodity declines as he has more units of the commodity. The consumer is in equilibrium when marginal utility becomes equal to the given price. In other words, consumer purchases the number of units of a commodity at which marginal utility is equal to price. This means that at the margin what a consumer will be willing to pay (i.e., marginal utility) is equal to the price he actually pays. But for the previous units which he purchases, his willingness to pay (or the marginal utility he derives from the commodity) is greater than the price he actually pays for them. This is because the price of the commodity is given and constant for him.

Marshall's Measure of Consumer Surplus

Consumer surplus measures extra utility or satisfaction which a consumer obtains from the consumption of a certain amount of commodity over and above the utility of its market value. Thus the total utility obtained from consuming water is immense while its market value is negligible. It is due to the occurrence of diminishing marginal utility that a consumer gets total utility from the consumption of a commodity greater than the utility of its market value. Marshall tried to obtain the monetary measure of this surplus, that is, how many rupees this surplus of utility is worth to the consumer. It is the monetary value of this surplus that Marshall called consumer surplus. To determine this monetary measure of consumer surplus we are required to measure two things. First, the total utility in terms of money that a consumer expects to get from

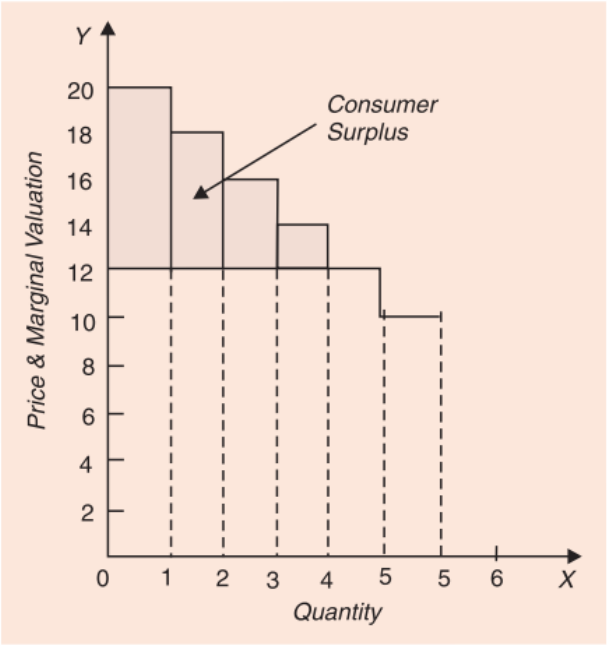


Fig. 15.1. *Consumer Surplus*

nount of a commodity. Second, the total market value of the . It is quite easy to measure the total market value as it is equal multiplied by its quantity purchased (i.e. P.Q.). An important n the way he devised to determine the monetary measure of the from the commodity. Consider Table 15.1 which has been .

ommodity is ` 20 per unit. At price of ` 20, the consumer is ıe commodity. This implies that utility which the consumer gets from this first unit is at least worth ` 20 to him otherwise he would not have purchased it at this price. When the price falls to ` 18, he is prepared to buy the second unit also. This again implies that the second unit of the commodity is at least worth ` 18 to him. Further, he is prepared to buy third unit at price ` 16. which means that it is at least worth ` 16 to him.

Table 15.1. Marginal Valuation and Consumer Surplus

No of Units	Marginal Valuation	Price		Net Marginal Benefit
1	₹ 20	₹ 12		₹ 8
2	₹ 18	₹ 12		₹ 6
3	₹ 16	₹ 12		₹ 4
4	₹ 14	₹ 12		₹ 2
5	₹ 12	₹ 12		₹ 0
6	₹ 10	₹ 12		– 2
			Total Consumer Surplus	20

Likewise, the fourth and fifth units of the commodity are at least worth ` 14 and ` 12 as he is prepared to pay these prices for the fourth and fifth units respectively, otherwise he would not have demanded them at these prices.

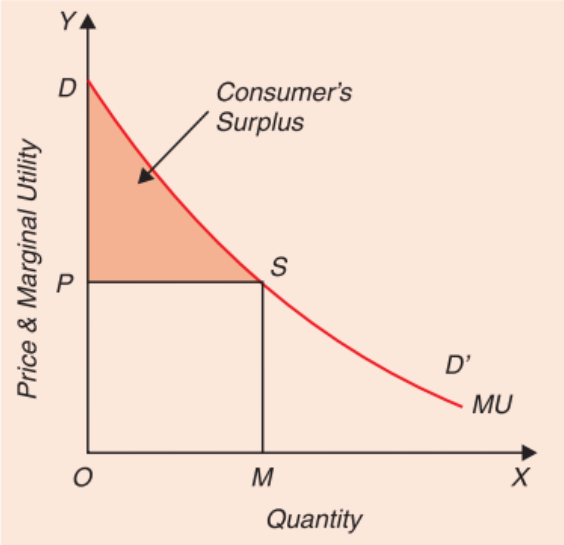


Fig. 15.2. *Marshall's Measure of Consumer's Surplus*

mand prices of these units in a slightly different way. The prices
d to pay for various units of the commodity means the marginal
these units of the commodity demanded by him. This marginal
nmodity shows the willingness of the individual to pay for it.
to pay the sum of money equal to the marginal valuation he places
he commodity he has to pay the current market price. Suppose the current market price of the commodity is Rs. 12. It will be seen from the Table 15.1 and Figure 15.1 that the consumer will buy 5 units of the commodity at this price because his marginal valuation of the fifth unit just equals the market price of Rs. 12. This shows that his marginal valuation of the first four units is greater than the market price which he factually pays for them. He will therefore obtain surplus or, net marginal benefit of Rs. 8 (Rs. 20 – 12) from the first unit, Rs. 6 (= Rs. 18 – 12) from the second unit, Rs. 4 on the third unit and Rs. 2 from the fourth unit and zero on the fifth unit. He thus obtains total consumer surplus or total net benefit) from 5 units equal to Rs. 20.

Measurement of Consumer Surplus as Area under the Demand Curve

The analysis of consumer surplus made above is based on discrete units of the commodity. If we assume that the commodity is perfectly divisible, which is usually made in economic theory, the consumer surplus can be represented by the area under the demand curve.

The measurement of consumer surplus from a commodity from the demand curve is illustrated in Fig.15.2 which along the X-axis the amount of the commodity has been measured and on the Y-axis the marginal utility (or willingness to pay for the commodity) and the price of the commodity is measured. DD is the demand or marginal utility curve which is sloping downward, indicating that as the consumer buys more units of the commodity, his willingness to pay for the additional units of the commodity or, in other words, marginal utility which he gets from the commodity falls. As said above, marginal utility shows the price which a person will be willing to pay for the different units rather than go without them. If OP is the price that prevails in the market, then the consumer will be in equilibrium when he buys OM units of the commodity, since at OM units, marginal utility is equal to the given price OP. The Mth unit of the commodity does not yield any consumer' s surplus to the consumer since this is the last unit purchased and for this price paid is equal to the marginal utility which indicates the price he will be prepared to pay rather than go without it. But for the intra-marginal units e.i. units before Mth, marginal utility is greater than the price and therefore, these units yield consumer's surplus to the consumer. The total utility of a certain quantity of a commodity to a consumer can be known by summing up the marginal utilities of the various units purchased.

In Figure 15.2, the total utility derived by the consumer from OM units of the commodity will be equal to the area under the demand or marginal utility curve up to point M. That is, the total utility of OM units in Fig. 15.2 is equal to ODSM. In other words, for OM units of the good the consumer will be prepared to pay the sum equal to Rs. ODSM. But, given the price OP, the

consumer will actually pay for OM units of the good the sum equal to Rs. OPSM. It is thus clear that the consumer derives extra utility equal to ODSM minus OPSM = DPS, which has been shaded in Figure 15.2. To conclude when we draw a demand curve, the monetary measure of consumer surplus can be obtained by the area under the demand curve over and above the rectangular area representing the total market value (i.e. P. Q) of the amount of the commodity purchased.

If the market price of the commodity rises above OP, the consumer will buy fewer units of the commodity than OM. As a result, consumer's surplus obtained by him from his purchase will decline. On the other hand, if the price falls below OP, the consumer will be in equilibrium when he is purchasing more units of the commodity than OM. As a result of this, the consumer's surplus will increase. Thus, given the marginal utility curve of the consumer, the higher the price, the smaller the consumer's surplus and the lower the price, the greater the consumer's surplus.

It is worth noting here that in our analysis of consumer's surplus, we have assumed that perfect competition prevails in the market so that the consumer faces a given price, whatever the amount of the commodity he purchases. But if the seller of a commodity discriminates the prices and charges different prices for the different units of the good, some units at a higher price and some at a lower price, then in this case consumer's surplus will be smaller. Thus, when the seller makes price discrimination and sells different units of a good at different prices, the consumer will obtain smaller amount of consumer's surplus than under perfect competition. If the seller indulges in perfect price discrimination, that is, if he charges price for each unit of the commodity equal to what any consumer will be prepared to pay for it, then in that case no consumer's surplus will accrue to the consumer.

ORDINAL UTILITY ANALYSIS

Indifference curves

The basic tool of Hicks-Allen ordinal utility analysis of demand is the indifference curve which represents all those combinations of goods which give same satisfaction to the consumer.
Since all the combinations on an indifference curve give equal satisfaction to the consumer, he will be indifferent between them, that is, it will not matter to him which one he gets. In other

words, all combinations of two goods lying on a consumer's indifference curve are equally desirable to or equally preferred by him. To understand indifference curves, it is better to start with indifference schedules. In Table 8.1, two indifference schedules are given. In each schedule the amounts of goods *X* and *Y* in each combination are so much that the consumer is indifferent among the combinations in each schedule. In schedule 1, the consumer has to start with 1 unit of *X* and 12 units of *Y*. Now, the consumer is asked to tell how much of good *Y* he will be willing to give up for the gain of an additional unit of *X* so that his level of satisfaction remains the same. If the gain of one unit of *X* compensates him fully for the loss of 4 units of *Y*, then the next combination of 2 units of *X* and 8 units of *Y* (2*X* + 8*Y*) will give him as much satisfaction as the initial combination (1*X* + 12*Y*). Similarly, by asking the consumer further how much of *Y* he will be prepared to forgo for successive increments in his stock of *X* so that his level of satisfaction remains unaltered, we get combinations 3*X* + 5*Y*, 4*X* + 3*Y*, and 5*X* + 2*Y*, each of which provides him same satisfaction as combination 1*X* + 12*Y* or 2*X* + 8*Y*. Since his satisfaction is the same whichever combination of goods in the schedule is offered to him, he will be indifferent among the combinations of two goods included in the schedule.

Table 8.1. Two Indifference Schedules

I		II	
Good X	*Good Y*	*Good X*	*Good Y*
1	12	2	14
2	8	3	10
3	5	4	7
4	3	5	5
5	2	6	4

In schedule II, the consumer has initially 2 units of *X* and 14 units of *Y*. By asking the consumer how much of *Y* he will be prepared to abandon for the successive additions of X in his stock so that his satisfaction remains equal to what he derives from the initial combination (2*X* + 14*Y*), we get combinations 3*X* + 10*Y*, 4*X* + 7*Y*, 5*X* + 5*Y* and 6*X*+ 4*Y*. Thus, each of the combinations in schedule II will be equally desirable to the consumer and he will be indifferent among them. But it should be borne in mind that the consumer will prefer any combination in schedule II to any combination in schedule I. That is, any combination in schedule II will give him more satisfaction than any combination in schedule I. This is because it

is assumed that more of a commodity is preferable to less of it (in other words, the greater quantity of a good gives an individual more satisfaction than the smaller quantity of it), the quantities of other goods with him remaining the same. Initial combination in schedule II contains more of both the goods than the initial combination in schedule I, therefore the former will give greater satisfaction to the consumer than the latter. Now, since each of the other combinations in indifference schedule II provides the consumer same satisfaction as the initial combination (2X + 14Y) of this schedule and also each of other combinations in indifference schedule I gives the same satisfaction as the initial combination (1X + 12Y), any combination of the schedule II will be preferred to (will yield greater satisfaction) than any combination of schedule I.

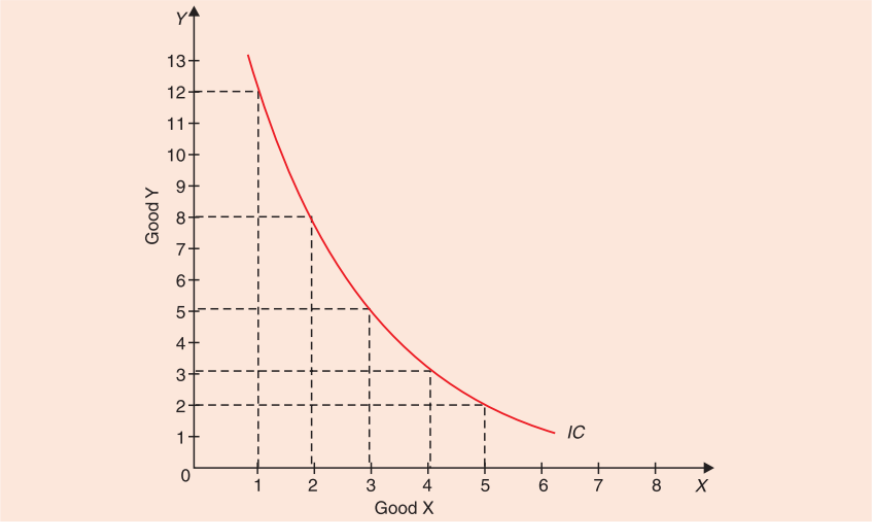


Fig. 8.1. *An Indifference Curve*

Now, we can convert the indifference schedules into indifference curves by plotting the various combinations on a graph paper. In Fig. 8.1 an indifference curve *IC* is drawn by plotting the various combinations of the indifference schedule I. The quantity of good *X* is measured on the horizontal axis, and the quantity of the good *Y* is measured on the vertical axis. As in an indifference schedule, combinations lying on an indifference curve will also be equally desirable to the consumer, that is, will give him the same satisfaction. The smoothness and continuity of an indifference curve mean that goods in question are assumed to be perfectly divisible. If the indifference schedule *II* is also converted into indifference curve, this will lie above the indifference curve *IC*.

Any combination on a higher indifference curve will be preferred to any combination on a lower indifference curve. It is thus clear that the indifference curve lying above and to the right of an indifference curve will indicate a higher level of satisfaction. It may be noted that while an indifference curve shows all those combinations of two goods which provide equal satisfaction to the consumer, it does not indicate *exactly how much* satisfaction is derived by the consumer from those combinations. This is because the concept of ordinal utility does not involve the quantitative measurability of utility. Therefore, no attempt is made to label an indifference curve by the amount of satisfaction it represents.

Marginal Rate of Substitution (MRS)

The concept of marginal rate of substitution is an important tool of indifference curve analysis of demand. The rate at which the consumer is prepared to exchange goods *X* and *Y* is known as marginal rate of substitution. In our indifference schedule I above, which is reproduced in Table 8.2, in the beginning the consumer gives up 4 units of *Y* for the gain of one additional unit of *X* and in this process his level of satisfaction remains the same. It follows that one unit gain in *X* fully compensates him for the loss of 4 units of *Y*. It means that at this stage he is prepared to exchange 4 units of *Y* for one unit of *X*. Therefore, at this stage consumer's marginal rate of substitution of *X* for Y is 4. Thus, we may define the marginal rate of substitution of X for Y as the amount of Y whose loss can just compensate the consumer for one unit gain in X. In other words, marginal rate of substitution of X for Y represents the amount of Y which the consumer has to give up for the gain of one additional unit of *X so that his level of satisfaction remains the same.*

Table 8.2. Indifference Schedule

Combination	*Good X*	*Good Y*	MRS_{xy}
A	1	12	
			4
B	2	8	
			3
C	3	5	
			2
D	4	3	
			1
E	5	2	

In Table 8.2, when the consumer moves from combination *B* to combination *C* on his indifference schedule he forgoes 3 units of *Y* for additional one unit gain in *X*. Hence, the marginal rate of substitution of *X* for *Y* is 3. Likewise, when the consumer moves from *C* to *D*, and then from *D* to *E* in his indifference schedule, the marginal rate of substitution of *X* for *Y* is 2 and 1 respectively.

Reasons for Diminishing MRSxy

Now, the question is what accounts for the diminishing marginal rate of substitution. In other words, why is it that the consumer is willing to give up less and less of *Y* for a given increment in *X* as he slides down on the curve? The following three factors are responsible for diminishing marginal rate of substitution.

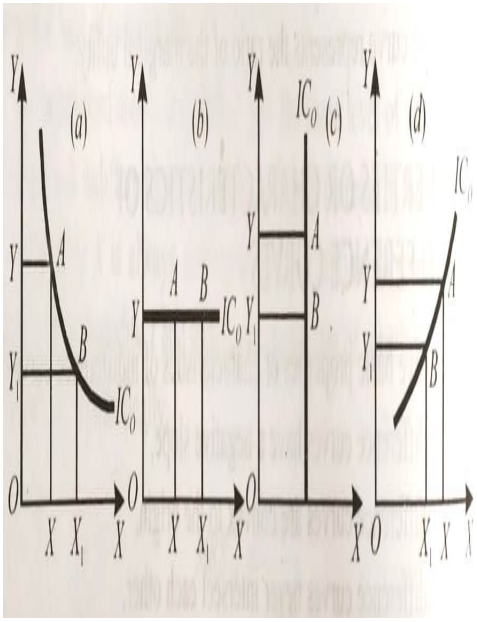
First, *the want for a particular good is satiable* so that as the consumer has more and more of a good the intensity of his want for that good goes on declining. It is because of this fall in the intensity of want for a good, say *X*, that when its stock increases with the consumer, he is prepared to forego less and less of good *Y* for every increment in *X*. In the beginning, when the consumer's stock of good *Y* is relatively large and his stock of good *X* is relatively small, consumer's marginal significance for good *Y* is low, while his marginal significance for good *X* is high. Owing to higher marginal significance of good *X* and lower marginal significance of good *Y* in the beginning the consumer will be willing to give up a larger amount of *Y* for one unit increase in good X. But as the stock of good X increases and intensity of desire for it falls, his marginal significance of good X will diminish and, on the other hand, as the stock of good Y decreases and the intensity of his desire for it increases, his marginal significance for good Y will go up. As a result, therefore, as the individual substitutes more and more of X for Y, he is prepared to give up less and less of Y for one unit increase in X.

The second reason for the decline in marginal rate of substitution is that the goods are imperfect substitutes of each other. If two goods are perfect substitutes of each other, then they are to be regarded as one and the same good, and therefore increase in the quantity of one and decrease in the quantity of the other would not make any difference in the marginal significance of the goods. Thus, in case of perfect substitutability of goods, the increase and decrease will be

virtually in the same good which cancel out each other and therefore the marginal rate of substitution remains the same and does not decline.

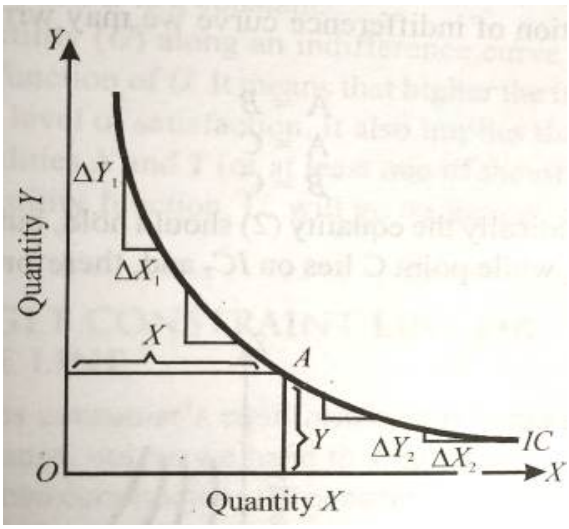
Properties

i. **Indifference curves slope downwards from left to right.**



This property implies that an indifference curve has a negative slope. This property follows from assumption I. Indifference curve being downward sloping means that when the amount of one good in the combination is increased, the amount of the other good is reduced. This must be so if the level of satisfaction is to remain the same on an indifference curve. If, for instance, the amount of good X is increased in the combination, while the amount of good Y remains unchanged, the new combination will be preferable to the original one and the two combinations will not therefore lie on the same indifference curve provided more of a commodity gives more satisfaction.

ii. **Indifference curves are convex to the origin.**



Another important property of indifference curves is that they are usually convex to the origin. In other words, the indifference curve is relatively flatter in its right-hand portion and relatively steeper in its left-hand portion.

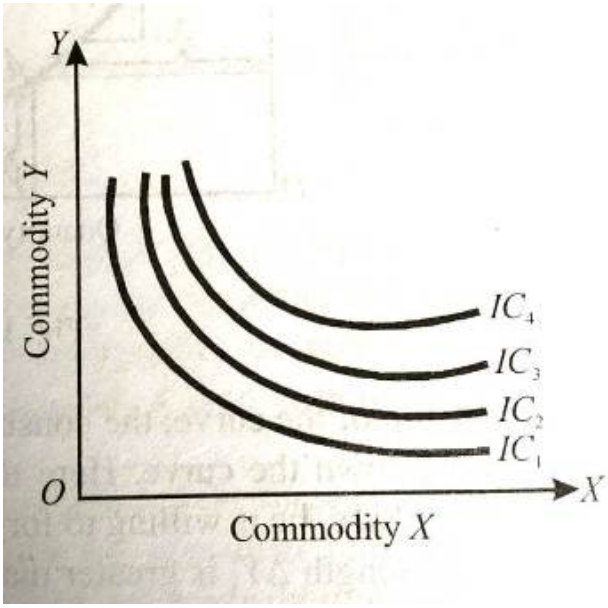
Only a convex indifference curve can mean a diminishing marginal rate of substitution of X for K If indifference curve was concave to the origin it would imply that the marginal rate of substitution of X for y increased as more and more of X was substituted, for Y.

When the indifference curve is convex to the origin, MRS diminishes as more of X is substituted for K. We therefore conclude that indifference curves are generally convex to the origin. Our assumption regarding diminishing MRSxy and the convexity of indifference curves is based upon the observation of actual behaviour of the normal consumer. If indifference curves were concave or straight lines, the consumer would succumb to monomania, that is, he would buy and consume only one good. We know that consumers in actual world do not generally buy and consume one good. It is for this reason that we reject indifference curves of concave or straight-line shapes and assume that indifference curves are normally convex to the origin.

The degree of convexity of an indifference curve depends on the rate of fall in the marginal rate of substitution of X for Y. As stated above, when two goods are perfect substitutes of each other, the indifference curve is a straight line on which marginal rate of substitution remains constant. The better substitutes the two goods are for each other, the closer the indifference curve

approaches to the straight line so that when the two goods are perfect substitutes the indifference curve is a straight line.

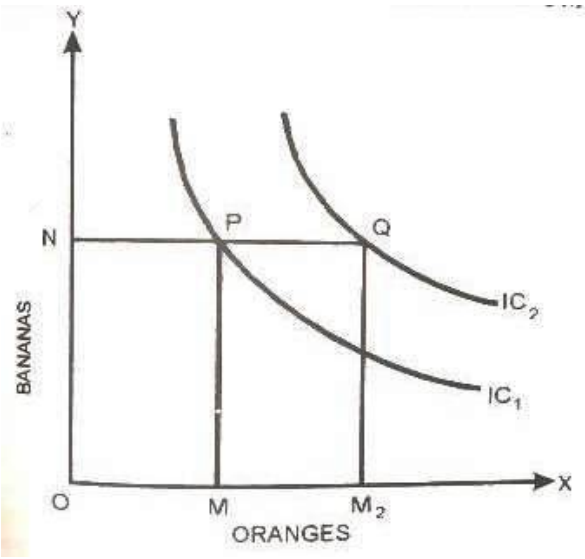
iii. No two indifference curves can intersect (cut) each other.



Indifference curves cannot even meet or touch each other or be tangent to each other at a point. The meeting of two indifference curves at a point will lead to an absurd conclusion. One of the distinctive properties of the indifference curves states that they never cut across each other. This is because each indifference curve provides a particular level of satisfaction or utility to the consumer.

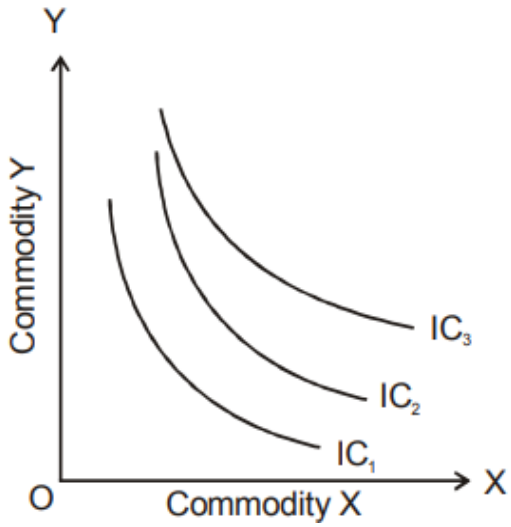
No two indifference curves can give the same utility to the consumers. If they intersect, it will signify that there is at least one combination in both the curves which provides an equivalent utility which is not possible.

iv. Every indifference curve to the right represents higher level of satisfaction than to the left.



A higher indifference curve will represent a higher level of satisfaction than a lower indifference curve. In other words, the combinations which lie on a higher indifference curve will be preferred to the combinations which lie on a lower indifference curve. Consider indifference curves IC_1 and IC_2 in Fig. The curve IC_2 is a higher indifference curve than IC_1.

v. Indifference curves need not necessarily be parallel to each other.



Indifference curves are not necessarily parallel to each other. Although, they are falling and negatively inclined to the right, yet the rate of the fall will not be the same for all Indifference Curves.

This is due to two reasons:

Firstly, the Indifference Curves are not based on the cardinal measurability of utility. *Secondly,* the rate of substitution between the two commodities need not be the same in all the indifference

schedules. It is therefore not necessary that the Indifference Curves should be parallel to each other.

Indifference Map

A complete description of consumer's tastes and preferences can be represented by an indifference map which consists of a set of indifference curves. Because the field in a two-dimensional diagram contains an infinite number of points, each representing a combination of goods *X* and *Y*, there will be an infinite number of the indifference curves each passing through combinations of goods that are equally desirable to the consumer.

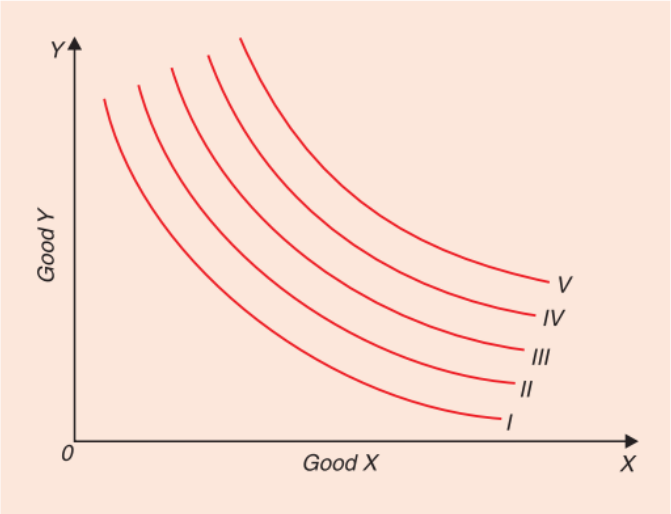


Fig. 8. 2. *Indifference Map*

In Fig.8.2 an indifference map of a consumer is shown which consists of five indifference curves. The consumer regards all combinations on the indifference curve *I* as giving him equal satisfaction. Similarly, all the combinations lying on indifference curve *II* provide the same satisfaction but the level of satisfaction on indifference curve *II* will be greater than the level of satisfaction on indifference curve *I*. Likewise, all higher indifference curves, *III*, *IV* and *V* represent progressively higher and higher levels satisfaction. It is important to remember that while the consumer will prefer any combination on a higher indifference curve to any combination on a lower indifference curve, but by *how much he prefers* one combination to another cannot be said. In other words, a higher indifference curve represents a higher level of satisfaction than a lower indifference curve but by *"how much higher*" cannot be indicated. This

is because the indifference curve system is based upon the concept of ordinal utility according to which the consumer is able to state only the '*qualitative*' differences in his various levels of satisfaction. It is not possible for the consumer to specify '*quantitative* 'differences in his various levels of satisfaction (*i.e*., by how much more or by how much less cannot be stated by him). Therefore, in an indifference map successively higher indifference curves can be denoted by any ascending series, 1,3,7, 9...; or 1, 4, 6, 8, 13...; or 1, 2, 5, 8, 10...; etc., the magnitude of these various numbers and the quantitative differences among them having no relevance. It is more usual to label the indifference curves by ordinal numbers as *I*, *II*, *III*, *IV*, *V* as is done in Fig. 8.2.

An indifference map of a consumer represents, as said earlier, his tastes and preferences for the two goods and his preferences between different combinations of them. In other words, *an indifference map portrays consumer's scale of preferences*. Scale of preferences of indifference curve analysis replaces Marshall's utility schedule. So long as consumer's tastes and preferences remain unchanged, the whole indifference map will remain the same. If the consumer's tastes and preferences undergo a change, then a new indifference map corresponding to new tastes and preferences will have to be drawn. If, for instance, good Y is eggs and good *X* is bread, and if the doctor advises our consumer to take more of eggs to overcome some diseases, the shapes of all his indifference curves will change and his indifference map will have to be redrawn. Since the doctor's advice will intensify our consumer's desire for eggs, now a smaller quantity of eggs than before will be given up by him for a given increment in bread.

The Budget (Price) Line:

The knowledge of the concept of *budget line*1 or what is also called *budget constraint* is essential for understanding the theory of consumer's equilibrium. As explained above, a higher indifference curve shows a higher level of satisfaction than a lower one. Therefore, a consumer in his attempt to maximize his satisfaction will try to reach the highest possible indifference curve. But in his pursuit of buying more and more goods and thus obtaining more and more satisfaction he has to work under two constraints: first, he has to pay the prices for the goods and, secondly, he has a limited money income with which to purchase the goods. Thus, how far he would go in for his purchases depends upon the prices of the goods and the money income which he has to spend on the goods. As explained above, indifference map represents

consumer's scale of preferences between the two goods. Now, in order to explain consumer's equilibrium there is also the need for introducing into the indifference curve analysis the budget line which represents the prices of the goods and consumer's money income.

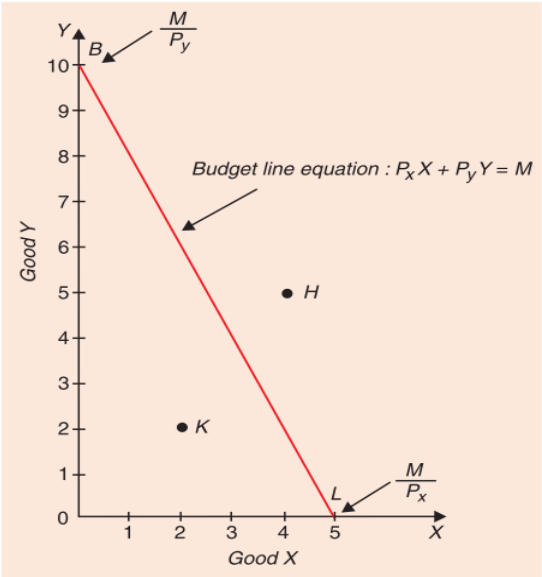


Fig. 8.14. *Budget Line or Budget Constraint*

Suppose our consumer has got income of ` 50 to spend on two goods X and Y. Let price of good X in the market be ` 10 per unit and that of Y ` 5 per unit. If the consumer spends his whole income of ` 50 on good X, he would buy 5 units of X; if he spends his whole income of ` 50 on good Y he would buy 10 units of Y. If a straight line joining 5X and l0Y is drawn, we will get what is called the price line or the budget line. Thus budget line shows all those combinations of two goods which the consumer can buy by spending his given money income on the two goods at their given prices. A look at Fig. 8.14 shows that with ` 50 and the prices of X and Y being Rs 10 and ` 5 respectively the consumer can buy 10Y and 0X, or 8Y and 1X; or 6Y and 2X, or 4Y and 3X etc. In other words, he can buy any combination that lies on the budget line with his given money income and given prices of the goods. It should be carefully noted that any combination of the two goods such as H (5Yand 4X) which lies above and outside the given budget line will be beyond the reach of the consumer. But any combination lying within the budget line such as K (2X and 2Y) will be well within the reach of the consumer, but if he buys any such combination he will not be spending all his income of ` 50. Thus, with the assumption that whole of the given income is spent on the given goods and at given prices of them, the consumer has to choose from all those combinations which lie on the budget line.

It is clear from above that budget line graphically shows the *budget constraint.* The combinations of commodities lying to the right of the budget line are *unattainable* because income of the consumer is not sufficient to buy those combinations. Given consumer's income and prices of the two goods, the combinations of goods lying to the left of the budget line are *attainable*, that is, the consumer can buy any one of them. It is also important to remember that the intercept *OB* on the *Y*-axis in Fig. 8.14 equals the amount of his entire income (*M*) divided by the price *(PY)* of commodity *Y*. That is, $OB = M/Py$. Likewise, the intercept *OL* on the *X*-axis measures the total income divided by the price of commodity *X*. Thus $OL = M/Px$.

Consumer equilibrium: the ordinal utility approach

Having introduced the necessary tools of analysis—the indifference curve and the budget line—used in the ordinal utility approach, we proceed now to discuss consumer behaviour under this approach. We begin with the analysis of consumer equilibrium.

As noted earlier, consumers attain equilibrium when they maximize the total utility of their commodities, given their income and the prices of goods and services they consume. According to the ordinal utility approach, two conditions must be satisfied for the consumer to be in equilibrium. These are

1. **Necessary** or **first-order conditions**
2. **Supplementary** or **second-order condition**.

The necessary condition for maximum utility requires that the *MRS* must be equal to the price ratio. The second-order or supplementary condition requires that the necessary condition must be fulfilled at the highest possible indifference curve.

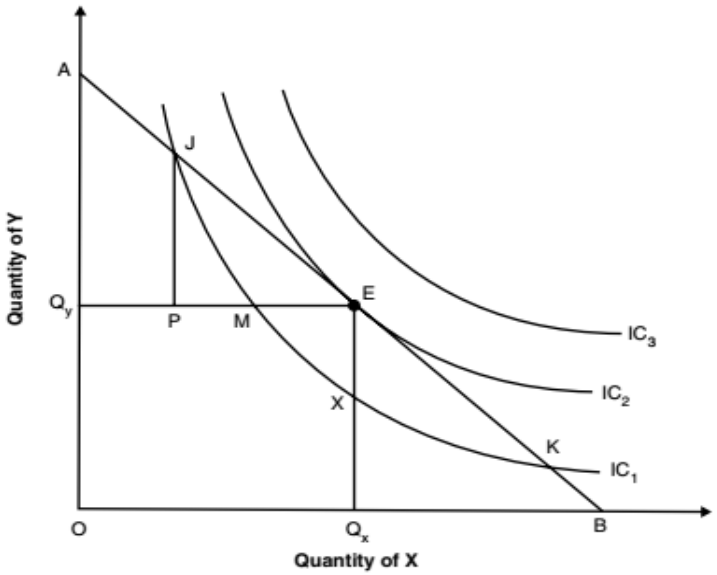


Figure 7.14 *Equilibrium of the Consumer*

In Figure 7.14, point *E* marks the consumer's equilibrium because at point *E, MRSxy = Py/Px*. This satisfies the necessary condition. Therefore, the consumer is in equilibrium at point *E*. The tangency of *IC*2 with reference to the budget line, *AB*, indicates that *IC*2 is the highest possible indifference curve that the consumer can reach, given their budgetary constraint. Point *E* satisfies, therefore, also the second-order condition. At the equilibrium point *E*, the consumer consumes *OQx* of *X* and *OQy* of *Y*, which yield the maximum satisfaction for the consumer, given the constraints.

Note that the necessary condition is satisfied with reference to two other points also—points *J* and *K* (i.e., the points of intersection between the budget line *AB* and the indifference curve *IC*1). These points do not satisfy the supplementary or second-order condition because the indifference curve *IC*1 is not the highest possible curve on which the necessary condition is fulfilled. Given the budget line *AB*, the indifference curve *IC*2 is the highest possible indifference curve that the consumer can reach. Therefore, as long as utility-maximizing consumers have the opportunity to reach curve *IC*2, they would not like to settle for a lower indifference curve.

From the information contained in Figure 7.14, it can be proved that the level of satisfaction at point *E* is greater than that on any point on *IC*1. Suppose that the consumer is at point *J*. If the consumer moves to point *M*, they will be equally well-off because points *J* and *M* are on the

same indifference curve. If they move from point *J* to *M*, they will have to sacrifice *JP* of *Y* and take *PM* of *X*. However, in the market, they can exchange *JP* of *Y* for *PE* of *X*. That is, they can get extra *ME* (= *PE* – *PM)* of *X*. Because *ME* gives the consumer extra utility, they reach point *E*. Point *E* yields a utility higher than that at point *M* or *J*. Therefore, point *E* is preferable to points *M* and *J*. The consumer will, therefore, have a tendency to move to point *E* from any point on *IC*1 to reach the highest possible indifference curve, all other factors (taste, preference, and prices of goods) remaining the same.

Income Effect: Income Consumption Curve

With a given money income to spend on goods, given prices of the two goods and given an indifference map (which portrays given tastes and preferences of the consumers), the consumer will be in equilibrium at a point in an indifference map. We are now interested in knowing how the consumer will react in regard to his purchases of the goods when his money income changes, prices of the goods and his tastes and preferences remaining unchanged. Income effect shows this reaction of the consumer. Thus, *the income effect means the change in consumer's purchases of the goods as a result of a change in his money income*. Income effect is illustrated in Fig. 8.28.

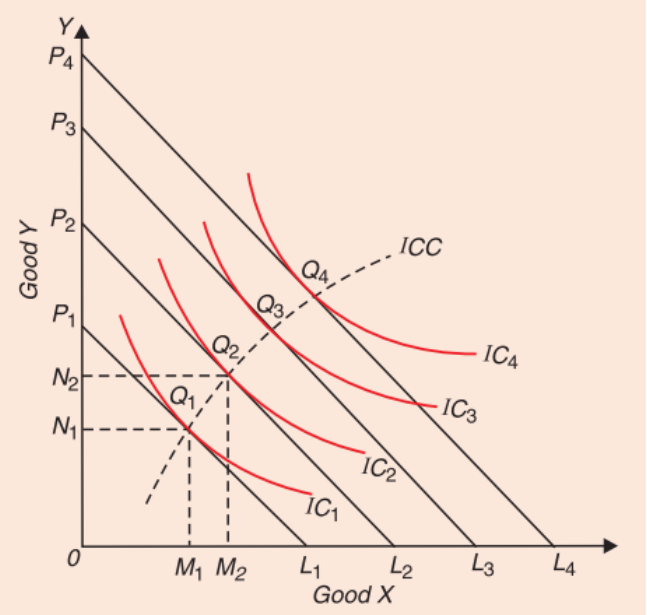


Fig. 8.28. *Income Consumption Curve: Income Effect*

With given prices and a given money income as indicated by the budget line P1Ll, the consumer is initially in equilibrium at point Q1 on the indifference curve IC1 and is having OM1 of X and ON1 of Y. Now suppose that income of the consumer increases. With his increased income, he would be able to purchase larger quantities of both the goods. As a result, budget line will shift upward and will be parallel to the original budget line P1L1. Let us assume that the consumer's money income increases by such an amount that the new budget line is P2L2 (consumer's income has increased by L1L2 in terms of X or P1P2 in terms of Y). With budget line P2L2, the consumer is in equilibrium at point Q2 on indifference curves IC 2 and is buying OM2 of X and ON 2 of Y. Thus, as a result of the increase in his income the consumer buys more quantity of both the goods. Since he is on the higher indifference curve IC2 he will be better off than before i.e., his satisfaction will increase. If his income increases further so that the budget line shifts to P3L3, the consumer is in equilibrium at point Q3 on indifference curve IC3 and is having greater quantity of both the goods than at Q2. Consequently, his satisfaction further increases. In Fig. 8.28 the consumer's equilibrium is shown at a still further higher level of income and it will be seen that the consumer is in equilibrium at Q4 on indifference curves IC4 when the budget line shifts to P4L4. As the consumer's income increases, he switches to higher indifference curves and as a consequence enjoys higher levels of satisfaction.

If now various points Q1, Q2, Q3 and Q4 showing consumer's equilibrium at various levels of income are joined together, we will get what is called Income Consumption Curve (ICC). Income consumption curve is thus the locus of equilibrium points at various levels of consumer's income. Income consumption curve traces out the income effect on the quantity consumed of the goods. Income effect can either be positive or negative. Income effect for a good is said to be positive when with the increase in income of the consumer, his consumption of the good also increases. This is the normal good case. When the income effect of both the goods represented on the two axes of the figure is positive, the income consumption curve (ICC) will slope upward to the right as in Fig. 8.28. Only the upward-sloping income consumption curve can show rising consumption of the two goods as income increases.

However, for some goods, income effect is negative. Income effect for a good is said to be negative when with the increases in his income, the consumer reduces his consumption of the good. Such goods for which income effect is negative are called Inferior Goods. This is because

the goods whose consumption falls as income of the consumer rises are considered to be some way 'inferior' by the consumer and therefore he substitutes superior goods for them when his income rises. When with the increase in his income, the consumer begins to consume superior goods, the consumption or quantity purchased by him of the inferior goods falls. When the people are poor, they cannot afford to buy the superior goods which are often more expensive. Hence as they become richer and can afford to buy more expensive goods they switch to the consumption of superior and better quality goods. For instance, most of the people in India consider cheaper common foodgrains such as maize, jawar, bajra as inferior goods and therefore when their income rises, they shift to the consumption of superior varieties of foodgrains like wheat and rice. Similarly, most of the Indian people regard Vanaspati Ghee to be inferior and therefore as they become richer, they reduce its consumption and use 'Desi Ghee' instead.

In case of inferior goods, indifference map would be such as to yield income consumption curve which either slopes backward (*i.e.*, toward the left) as in Fig. 8.29, or downward to the right as in Fig. 8.30. It would be noticed from these two figures that income effect becomes negative only after a point. It signifies that only at higher ranges of income, some goods become inferior goods and up to a point their consumption behaves like those of normal goods. In Fig. 8.29 income consumption curve (*ICC*) slopes backward *i.e.*, bends toward the *Y*axis. This shows good *X* to be an inferior good, since beyond point *Q*2, income effect is negative for good *X* and as a result its quantity demanded falls as income increases. In Fig.8.30 income consumption curve (ICC) slopes downward to the right beyond point Q2 i.e., bends towards the X-axis. This signifies that good Y is an inferior good because beyond point Q2, income effect is negative for good Y and as a result its quantity demanded falls as income increases. It follows from above that the income consumption curve can have various possible shapes.

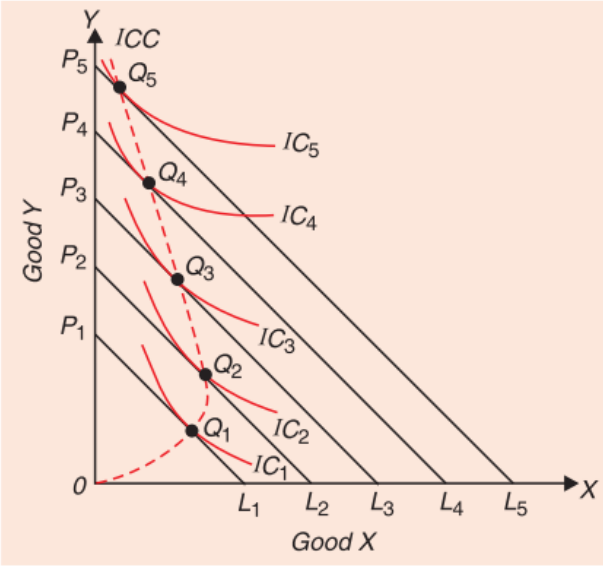


Fig. 8.29. *Income Consumption Curve in Case of Good X being Inferior Good*

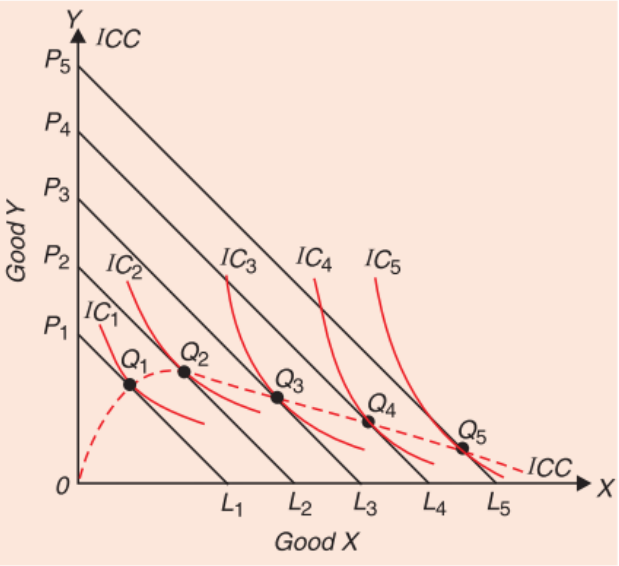


Fig. 8.30. *Income Consumption Curve in Case of Good Y being Inferior Good*

But normal goods can be either necessities or luxuries depending upon whether the quantities purchased of the goods by the consumers increase less than or more than proportionately to the increases in income. If the quantity purchased of a commodity rises less than proportionately to the increases in consumer's income, the commodity is known as a *necessity*. On the other hand, if the quantity purchased of a commodity increases more than proportionately to the increases in income, it is called a *luxury*. In Fig. 8.31, the slope of income consumption curve *ICC*1 is increasing which implies that the quantity purchased of the commodity *X* increases less than proportionately to the increases in consumer's income. Therefore, in this case of *ICC*1, good *X* is a necessity and good *Y* is luxury. On the other hand, the slope of income consumption curve *ICC*3 is decreasing which implies that the quantity purchased of good *X* increases more than proportionately to increases in income and therefore in this case good *X* is luxury and good *Y* is necessity. It will be seen from Fig. 8.31 that the income consumption curve *ICC*2 is a *linear* curve passing through the origin which implies that the increases in the quantities purchased of both the goods are rising in proportion to the increase in income and therefore neither good is a luxury or a necessity.

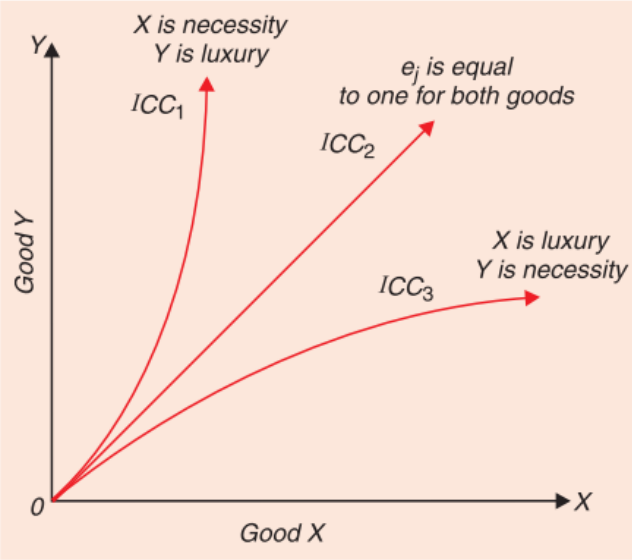


Fig. 8.31. *Income Consumption Curves of Normal Goods*

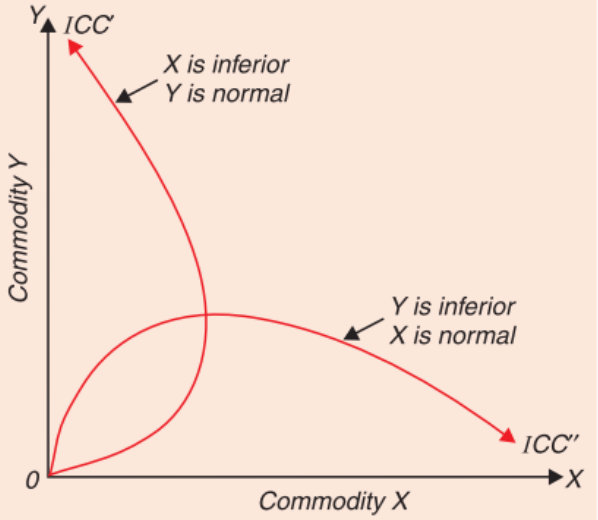


Fig. 8.32. *Income Consumption Curves of Inferior Goods*

If income effect is positive for both the goods *X* and *Y*, the income consumption curve will slope upward to the right as in Fig. 8.28 given earlier. But upward-sloping income consumption curves to the right for various goods may be of different slopes as shown in Fig. 8.31 in which income consumption curves, with varying slopes, are all sloping upward and therefore indicate both goods to be normal goods having positive income effect. *If income effect for good X is negative, income consumption curve will slope backward to the left as ICC'in Fig. 8.32. If good Y happens to be an inferior good and income consumption curve will bend towards X-axis as shown by ICC'' in Fig. 8.32.* In Figs. 8.31 and 8.32, various possible shapes which income consumption curve can take are shown *bereft of indifference curves and budget lines* which yield them. It may however be pointed out that given an indifference map and a set of budget lines there will be one income consumption curve.

A noteworthy point is that it is not the indifference curves which explain why a good happens to be an inferior good. In other words, indifference curves do not explain why income effect for a good is negative. Indifference curves can only illustrate the inferior-good phenomenon.

Substitution effect

We have explained above the effect of changes in income on purchases or consumption of a good. Another important factor responsible for the changes in consumption of a good is the substitution effect. Whereas the income effect shows the change in the quantity purchased of a good by a consumer as a result of change in his income, prices of goods remaining constant, substitution effect means the change in the quantity purchased of a good as a consequence of a change in its relative price alone, real income or level of satisfaction remaining constant. When the price of a good changes, he goes to a different indifference curve and his level of satisfaction changes. The Consumer goes to a different indifference curve as a result of a change in price because with this the real income or purchasing power of a consumer also changes. To keep the real income of the consumer constant so that the effect due to a change in the relative price alone may be known, price change is compensated by a simultaneous change in income. For example, when price of a good, say X, falls, real income of the consumer would increase and he would be in equilibrium at a higher indifference curve showing a higher level of satisfaction. In order to find out the substitution effect i.e., change in the quantity of X purchased which has come about due to the change only in its relative price, the consumer's money income must be reduced by an amount that cancels out the gain in real income that results from the decrease in price. Now, two slightly different concepts of substitution effect have been developed; one by J.R. Hicks and the other by E. Slutsky. These two concepts of substitution effect have been named after their authors. Thus, the substitution effect which is propounded by Hicks and Allen is called the Hicksian Substitution Effect and that developed by E. Slutsky is known as Slutsky Substitution Effect. The two concepts differ in regard to the magnitude of the change in money income which should be affected so as to neutralize the change in real income of the consumer which results from a change in the price. We shall explain here the Hicksian substitution effect.

In the Hicksian substitution effect price change is accompanied by a so much change in money income that the consumer is neither better off nor worse off than before, that is, he is brought to the original level of satisfaction. In other words, money income of the consumer is changed by

an amount which keeps the consumer on the same indifference curve on which he was before the change in the price. Thus the Hicksian substitution effect takes place on the same indifference curve. *The amount by which the money income of the consumer is changed so that the consumer is neither better off nor worse off than before is called compensating variation in income.* In other words, compensating variation in income is a change in the income of the consumer which is just sufficient to compensate the consumer for a change in the price of a good.

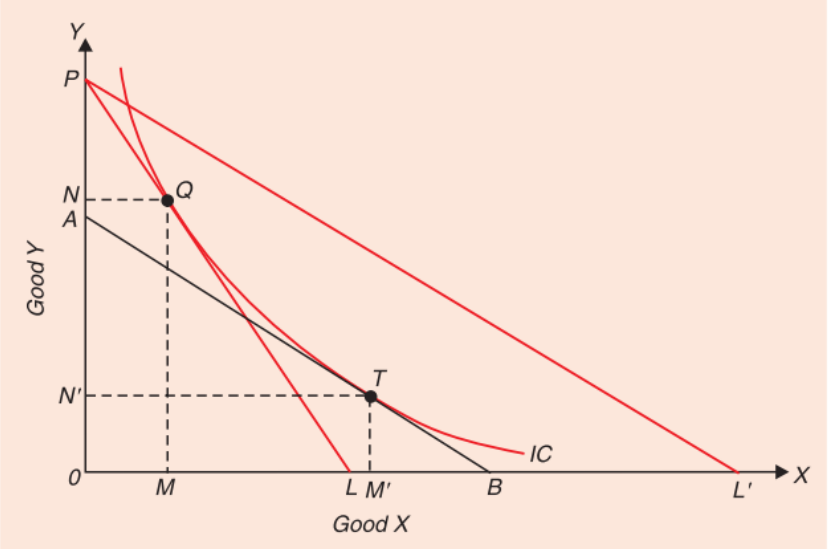


Fig. 8.37. *Hicksian Substitution Effect*

Thus, in the Hicksian type of substitution effect, income is changed by the magnitude of the *compensating variation in income.* Hicksian substitution effect is illustrated in Fig. 8.37. With a given money income and given prices of the two goods as represented by the budget line *PL*, the consumer is in equilibrium at point *Q* on the indifference curve *IC* and is purchasing *OM* of good *X* and *ON* of good *Y*. Suppose that the price of good *X* falls (price of *Y* remaining unchan-ged) so that the budget line now shifts to *PL'*. With the fall in price of *X* the consumer's real income or purchasing power would increase. In order to find out the substitution effect, this gain in real income should be wiped out by reducing the money income of the consumer by such an amount that forces him to remain on the same indifference curve *IC* on which he was before the change in price of the good *X*. When some money is taken away from the consumer to cancel out the gain in real income, then the budget line which shifted to position PL′ will now shift downward but will be parallel to PL′. In Fig.8.37 a budget line AB parallel to PL′ has been drawn at such a

distance from PL′ that it touches the indifference curve IC. It means that reduction of consumer's income by the amount PA (in terms of Y) or L′B (in terms of X) has been made so as to keep him on the same indifference curve. PA or L′B is thus just sufficient to cancel out the gain in the real income which occurred due to the fall in the price of X. PA or L′B is therefore compensating variation in income.

Now, budget line *AB* represents the new relative prices of goods *X* and *Y* since it is parallel to the budget line *PL′* which was obtained when price of good *X* had fallen. In comparison to the budget line *PL*, *X* is now relatively cheaper. The consumer would therefore rearrange his purchases of *X* and *Y* and will substitute *X* for *Y.* That is, since *X* is now relatively cheaper and *Y* is now relatively dearer than before, he will buy more of *X* and less of *Y.* It will be seen from Fig. 8.37 that budget line *AB* represents the changed relative prices but a lower money income than that of *PL,* since consumer's income has been reduced by compensating variation in income.

It will seen from Fig. 8.37 that with budget line *AB* the consumer is in equilibrium at point *T* and is now buying *OM′* of *X* and *ON′* of *Y.* Thus in order to buy *X* more he moves on the same indifference curve *IC* from point *Q* to point *T*. This increase in the quantity purchased of good *X* by *MM′* and the decrease in the quantity purchased of good *Y* by *NN′* is due to the change only in the relative prices of goods *X* and *Y*, since effect due to the gain in real income has been wiped out by making a simultaneous reduction in consumer's income. Therefore, movement from *Q* to *T* represents the substitution effect. Substitution effect on good *X* is the increase in its quantity purchased by *MM′* and substitution effect on *Y* is the fall in its quantity purchased by *NN′*. It is thus clear that as a result of the Hicksian substitution effect the consumer remains on the same indifference curve; he is however in equilibrium at a different point from that at which he was before the change in price of good *X*. The less the convexity of the indifference curve, the greater will be the substitution effect. As is known, the convexity of indifference curve is less in the case of those goods which are good substitutes. It is thus clear that the substitution effect in case of good substitutes will be large.

It is thus clear that a fall in relative price of a commodity always leads to the increase in its quantity demanded due to the substitution effect, the consumer's satisfaction or indifference

curve remaining the same. Thus the *substitution effect is always negative*. The negative substitution effect implies that the relative price of a commodity and its quantity demanded change in opposite direction, that is, the *decline* in relative price of a commodity always causes *increase* in its quantity demanded. It is this negative substitution effect which lies at the root of the famous law of demand stating inverse relationship between price and quantity demanded.

Price Effect: Price Consumption Curve

We will now explain how the consumer reacts to charges in the price of a good, his money income, tastes and prices of other goods remaining the same. Price effect shows this reaction of the consumer and measures the full effect of the change in the price of a good on the quantity purchased since no compensating variation in income is made in this case. When the price of a good charges, the consumer would be either better off or worse off than before, depending upon whether the price falls or rises. In other words, as a result of change in price of a good, his equilibrium position would lie at a higher indifference curve in case of the fall in price and at a lower indifference curve in case of the rise in price.

Price effect is shown in Fig. 8.38. With given prices of goods X and Y, and a given money income as represented by the budget line PL1, the consumer is in equilibrium at Q on indifference curve IC1. In this equilibrium position at Q, he is buying OM1 of X and ON1 of Y. Let price of good X fall, price of Y and his money income remaining unchanged. As a result of this price change, budget line shifts to the position PL2. The consumer is now in equilibrium at R on a higher indifference curve IC2 and is buying OM2 of X and ON2 of Y. He has thus become better off, that is, his level of satisfaction has increased as a consequence of the fall in the price of good X. Suppose that price of X further falls so that PL3 is now the relevant budget line. With budget line PL3 the consumer is in equilibrium at S on indifference curve IC3 where he has OM 3 of X and ON3 of Y. If price of good X falls still further so that budget line now takes the position of PL4, the consumer now attains equilibrium at T on indifference curve IC4 and has OM 4 of X and ON4 of Y. When all the equilibrium points such as Q, R, S, and T are joined together, we get what is called Price Consumption Curve (PCC). Price consumption curve traces out the price effect. It shows how the changes in price of good X will affect the consumer's purchases of X, price of Y, his tastes and money income remaining unaltered,

In Fig.8.38 price consumption curve (PCC) is sloping downward. Downward-sloping price consumption curve for good X means that as price of good X falls, the consumer purchases a larger quantity of good X and a smaller quantity of good Y. This is quite evident from Fig. 8.38. As we shall discuss in detail in the chapter concerning elasticity of demand, we obtain downward sloping price consumption curve for good X when demand for it is elastic (i.e., price elasticity is greater than one). But downward sloping is one possible shape of price consumption curve. Price consumption curve can have other shapes also. In Fig. 8.39 upward-sloping price consumption curve is shown. Upward-sloping price consumption curve for X means that when the price of good X falls, the quantity demanded of both goods X and Y rises. We obtain the upward-sloping price consumption curve for good X when the demand for good is inelastic, (i.e., price elasticity is less than one).

Price consumption curve can also have a backward-sloping shape, which is depicted in Fig. 8.40. Backward-sloping price consumption curve for good X indicates that when price of X falls, after a point smaller quantity of it is demanded or purchased. We shall see later in this chapter that this is true in case of exceptional type of goods called Giffen Goods.

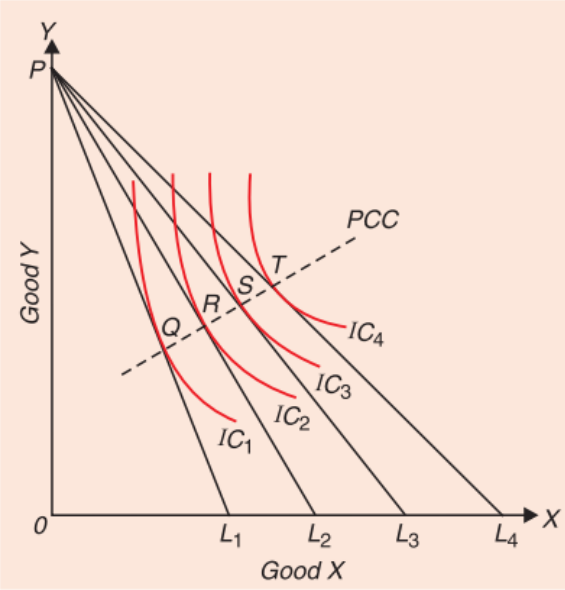


Fig. 8.39. *Upward-Sloping Price Consumption Curve*

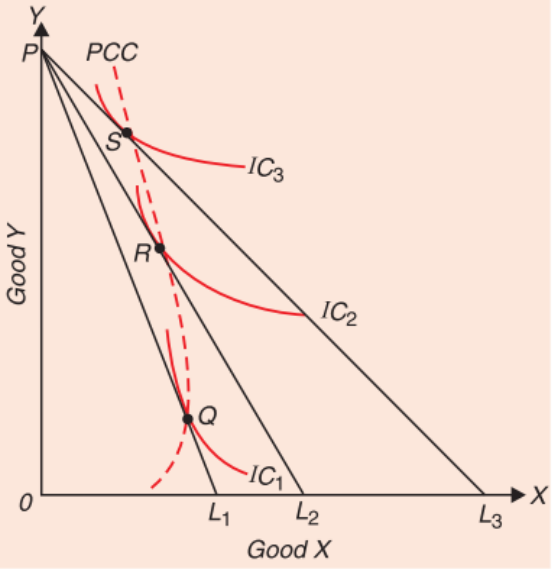


Fig. 8.40. *Backward-Sloping Price Consumption Curve in Case of Giffen Goods*

Price consumption curve for a good can take horizontal shape too. It means that when the price of the good X declines, its quantity purchased increases proportionately but quantity purchased

of good Y remains the same. Horizontal price consumption curve is shown in Fig. 8.41. We obtain horizontal price consumption curve of good X when the price elasticity of demand for good X is equal to unity. But it is rarely found that price consumption curve slopes downward throughout or slopes upward throughout or slopes backward throughout. More generally, price consumption curve has different slopes at different price ranges. At higher price levels it generally slopes downward, and it may then have a horizontal shape for some price ranges but ultimately it will be sloping upward. For some price ranges it can be backward sloping as in case of Giffen goods. A price consumption curve which has different shapes or slopes at different price ranges is drawn in Fig. 8.42. Such a type of price consumption curve means that price elasticity of demand varies at different price ranges.

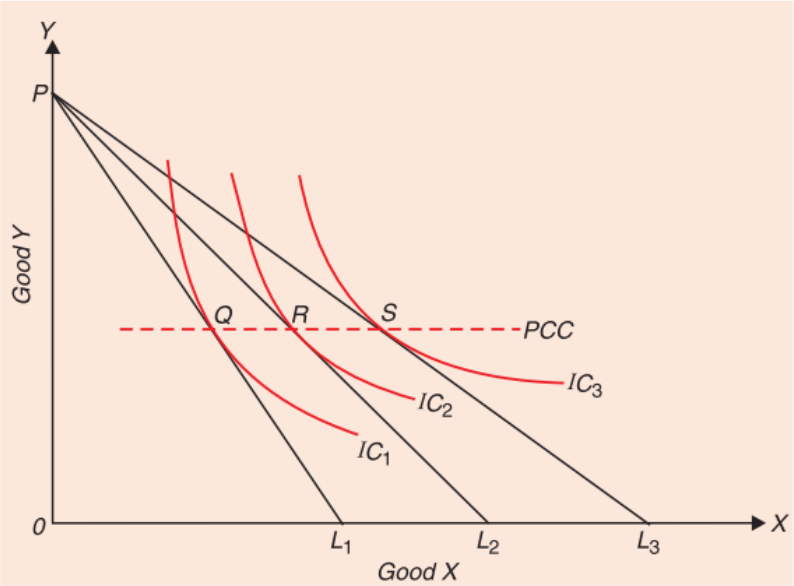


Fig.8.41. *Horizontal Price Consumption Curve*

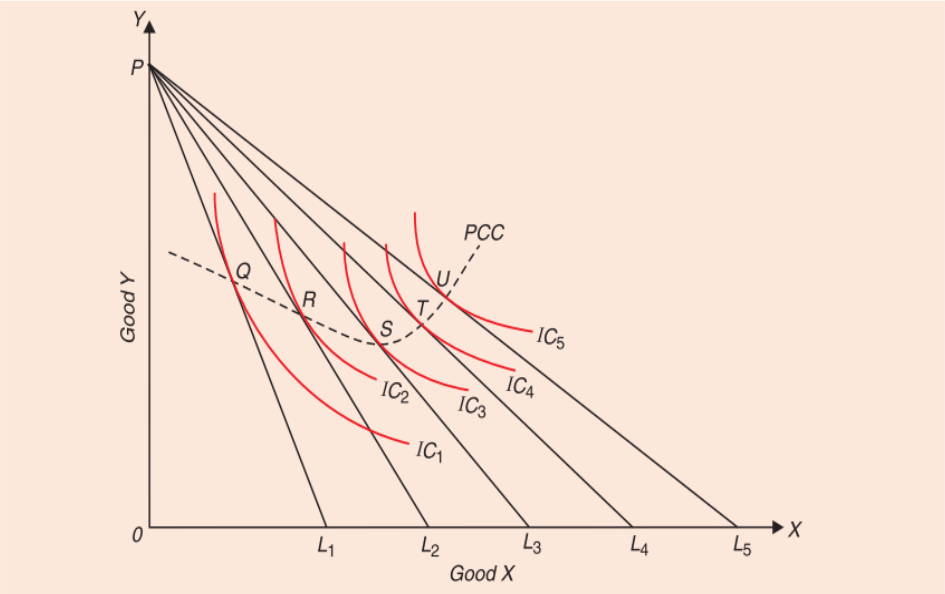


Fig. 8.42. *Price Consumption Curve with Varying Slopes*

UNIT –II: DEMAND AND SUPPLY

Concept of demand:

Conceptually, demand can be defined as the *desire to buy* a good for which the demander has *ability* and *willingness to pay*. In simple words, *demand is a desire for a good, backed by ability and willingness to pay.* A desire without ability to pay is merely a wish. A desire with ability to pay but without willingness to pay is only a *potential demand.* A desire accompanied by ability and willingness to pay makes a *real* or *effective demand.*

Individual and market demand.

For the purpose of demand analysis, a distinction is often made between the individual demand and the market demand—individual demand for analyzing consumer behaviour and market demand for analyzing market behaviour.

Individual demand

Individual demand can be defined as the quantity of a commodity that a person is willing to buy at a given price over a specified period of time, say per day, per week, per month, etc.

Market demand

Refers to the total quantity that all the users of a commodity are willing to buy at a given price over a specific period of time. In fact, market demand is the sum of individual demands for a product.

Factors determining demand

The purpose of the theory of demand is to determine the various factors that affect demand.

Demand is a multivariate relationship, that is, it is determined by many factors simultaneously

Some of the most important determinants of market demand (Sum of total consumers demand) for a particular product are:

1) *Price of the given commodity:*

It is the most important factor affecting demand for the given commodity. Generally, there exists an inverse relationship between price and quantity demanded. It means, as price increases, quantity demanded falls due to decrease in the satisfaction level of consumers. For example, If price of given commodity (say, tea) increases, its quantity demanded will fall as satisfaction derived from tea will fall due to rise in its price. Demand (D) is a function of price (P) and can be expressed as: $D = f(P)$. The inverse relationship between price and demand, known as law of demand.

2) *Tastes and preferences of the consumers:*

An important factor which determines the demand for a good is the tastes and preferences of the consumers for it. A good for which consumers' tastes and preferences are greater, its demand would be large and its demand curve will therefore lie at a higher level. People's tastes and

preferences for various goods often change and as a result there is change in demand for them. The changes in demand for various goods occur due to the changes in fashion and also due to the pressure of advertisements by the manufacturers and sellers of different products. On the contrary, when certain goods go out of fashion or people's tastes and preferences no longer remain favorable to them, the demand for them decreases.

3) *Income of the People:*

The demand for goods also depends upon the incomes of the people. The greater the incomes of the people, the greater will be their demand for goods. In drawing the demand schedule or the demand curve for a good we take income of the people as given and constant. When as a result of the rise in the income of the people, the demand increases, the whole of the demand curve shifts upward and vice versa. The greater income means the greater purchasing power. Therefore, when incomes of the people increase, they can afford to buy more. It is because of this reason that increase in income has a positive effect on the demand for a good. When the incomes of the people fall, they would demand less of a good and as a result the demand curve will shift downward. For instance, as a result of economic growth in India the incomes of the people have greatly increased owing to the large investment expenditure on the development schemes by the Government and the private sector. As a result of this increase in incomes, the demand for goodgrains and other consumer goods has greatly increased. Likewise, when because of drought in a year the agriculture production greatly falls, the incomes of the farmers decline. As a result of the decline in incomes of the farmers, they will demand less of the cotton cloth and other manufactured products.

4) *Changes in prices of the related goods.*

The demand for a good is also affected by the prices of other goods, especially those which are related to it as substitutes or complements. When we draw the demand schedule or the demand curve for a good we take the prices of the related goods as remaining constant. Therefore, when the prices of the related goods, substitutes or complements, change, the whole demand curve would change its position; it will shift upward or downward as the case may be. When the price of a substitute for a good falls, the demand for that good will decline and when the price of the substitute rises, the demand for that good will increase. For example, when price of tea and

incomes of the people remain the same but the price of coffee falls, the consumers would demand less of tea than before. Tea and coffee are very close substitutes. Therefore, when coffee becomes cheaper, the consumers substitute coffee for tea and as a result the demand for tea declines. The goods which are complementary with each other, the fall in the price of any of them would favourably affect the demand for the other. For instance, if price of milk falls, the demand for sugar would also be favourably affected. When people would take more milk, the demand for sugar will also increase. Likewise, when the price of cars falls, the quantity demanded of them would increase whichin turn will increase the demand for petrol.

5) *Advertisement expenditure.*

Advertisement expenditure made by a firm to promote the sales of its product is an important factor determining demand for a product, especially of the product of the firm which gives advertisements. The purpose of advertisement is to influence the consumers in favour of a product. Advertisements are given in various media such as newspapers, radio, television. Advertisements for goods are repeated several times so that consumers are convinced about their superior quality. When advertisements prove successful they cause an increase in the demand for the product.

6) *The number of consumers in the market:*

The greater the number of consumers of a good, the greater the market demand for it. Now, the question arises on what factors the number of consumers for a good depends. If the consumers substitute one good for another, then the number of consumers for the good which has been substituted by the other will decline and for the good which has been used in place of the others, the number of consumers will increase. Besides, when the seller of a good succeeds in finding out new markets for his good and as a result the market for his good expands, the number of consumers for that good will increase. Another important cause for the increase in the number of consumers is the growth in population. For instance, in India the demand for many essential goods, especially foodgrains, has increased because of the increase in the population of the country and the resultant increase in the number of consumers for them.

7) *Consumers' expectations with regard to future prices:*

Another factor which influences the demand for goods is consumers' expectations with regard to future prices of the goods. If due to some reason, consumers expect that in the near future prices of the goods would rise, then in the present they would demand greater quantities of the goods so that in the future they should not have to pay higher prices. Similarly, when the consumers expect that in the future the prices of goods will fall, then in the present they will postpone a part of the consumption of goods with the result that their present demand for goods will decrease.

Law of demand

Important information about demand is described by the law of demand. This law of demand expresses the functional relationship between price and quantity demanded. The law of demand or functional relationship between price and quantity demanded of a commodity is one of the best known and most important laws of economic theory. According to the law of demand, other things being equal, if the price of a commodity falls, the quantity demanded of it will rise, and if the price of the commodity rises, its quantity demanded will decline. Thus, according to the law of demand, there is inverse relationship between price and quantity demanded, other things remaining the same. These other things which are assumed to be constant are the tastes and preferences of the consumer, the income of the consumer, and the prices of related goods. If these other factors which determine demand also undergo a change at the same time, then the inverse price-demand relationship may not hold good. Thus, the constancy of these other things which is generally stated as ceteris paribus is an important qualification of the law of demand.

Demand curve and the law of demand.

The law of demand can be illustrated through a demand schedule and a demand curve. A demand schedule of an individual consumer is presented in Table 6.1. It will be seen from this demand schedule that when the price of a commodity is Rs. 12 per unit, the consumer purchases 10 units of the commodity. When the price of the commodity falls to Rs. 10, he purchases 20 units of the commodity. Similarly, when the price further falls, quantity demanded by him goes on rising until at price Rs. 2, the quantity demanded by him rises to 60 units. We can convert this demand schedule into a demand curve by graphically plotting the various price-quantity combinations,

and this has been done in Fig. 6.1 where along the X-axis, quantity demanded is measured and along the Y-axis price of the commodity is measured.

Table 6.1. Demand Schedule of an Individual Consumer

Price (Rs.)	*Quantity Demanded*
12	10
10	20
8	30
6	40
4	50
2	60

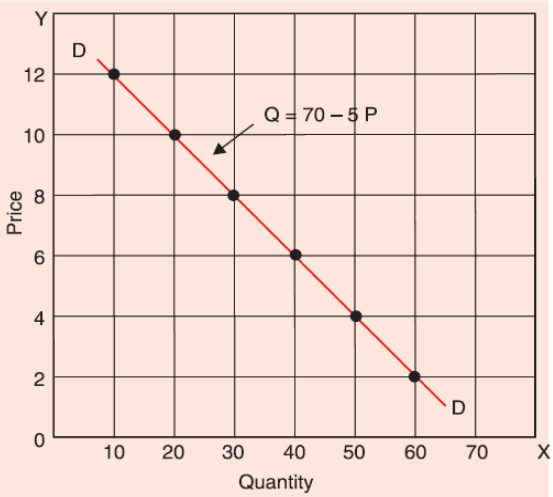


Fig. 6.1. *The Demand Curve of an Individual* ***(Q^d = 70 — 5P)***

By plotting 10 units of the commodity against price 12, we get a point in Fig.6.1. Likewise, by plotting 20 units of the commodity demanded against price 10, we get another point in Fig. 6.1. Similarly, other points are plotted representing other combinations of price and quantity demanded of the commodity and are shown in Fig. 6.1. By joining these various points, we get a curve DD, which is known as the demand curve. Thus, this demand curve is a graphic representation of quantities of a good which are demanded by the consumer at various possible prices in a given period of time.

It should be noted that a demand schedule or a demand curve does not tell us what the price is; it only tells us how much quantity of the good would be purchased by the consumer at a various possible prices. Further, it will be seen from both the demand schedule and the demand curve

that as the price of a commodity falls, more quantity of it is purchased or demanded. Since more is demanded at a lower price and less is demanded at a higher price, the demand curve slopes downward to the right. Thus, the downward-sloping demand curve is in accordance with the law of demand which, as stated above, describes inverse price-demand relationship. It is important to note here that behind this demand curve or price-demand relationship always lie the tastes and preferences of the consumer, his income, the prices of substitutes and complementary goods, all of which are assumed to be constant in drawing a demand curve. If any change occurs in any of these other factors, the whole demand schedule or demand curve will change and new demand schedule or a demand curve will have to be drawn. Further, in drawing a demand curve, we assume that the buyer or consumer does not exercise any influence over the price of a commodity, that is, he takes the price of the commodity as given and constant for him.

Exceptions to the law of demand

The law of demand is one of the fundamental laws of economics. The law of demand, however, does not apply to the following cases.

1) *Expectations regarding future prices.*

When consumers expect a continuous increase in the price of a durable commodity, they buy more of it, despite the increase in its price, to avoid the pinch of still higher price in future. Similarly, when consumers anticipate a considerable decrease in the price in future, they postpone their purchases and wait for the price to fall further, rather than buy the commodity when its price initially falls. Such decisions of the consumers are contrary to the law of demand.

2) *Prestigious Goods.*

The law does not apply to the commodities which are used as a 'status symbol', which enhance social prestige or display wealth and richness, e.g., gold, precious stones, rare paintings and antiques. Rich people buy such goods mainly because their prices are high.

3) *Giffen Goods.*

A classic exception to the law of demand is the case of Giffen goods named after a British economist, Sir Robert Giffen (1837–1910). A Giffen good does not mean any specific

commodity. It may be any inferior but essential commodity much cheaper than its substitutes, consumed mostly by the poor households and claiming a large part of their income. If the price of such goods increases (price of its substitute remaining constant), its demand increases instead of decreasing. For instance, let us suppose that the monthly minimum consumption of food grains by a poor household is 30 kg including 20 kg of bajra (an inferior good) and 10 kg of wheat (a superior good). Suppose also that bajra sells at Rs 5/kg and wheat at Rs 10/kg. At these prices, the household spends Rs 200 per month on food grains. That is the maximum it can afford. Now, if price of bajra increases to Rs 6 per kg, the household will be forced to reduce its consumption of wheat by 5 kg3 and increase that of bajra by the same quantity in order to meet its minimum monthly consumption requirement within Rs. 200 per month. Obviously, household's demand for bajra increases from 20 to 25 kg per month despite increase in its price and that of wheat falls to 5 kg.

Changes in demand

In economics the terms change in quantity demanded and change in demand are two different concepts. Change in quantity demanded refers to change in the quantity purchased due to increase or decrease in the price of a product. In such a case, it is incorrect to say increase or decrease in demand rather it is increase or decrease in the quantity demanded.

On the other hand, change in demand refers to increase or decrease in demand of a product due to various determinants of demand, while keeping price at constant. Changes in quantity demanded can be measured by the movement of demand curve, while changes in demand are measured by shifts in demand curve. The terms, change in quantity demanded refers to expansion or contraction of demand, while change in demand means increase or decrease in demand.

Expansion and Contraction of Demand:

The variations in the quantities demanded of a product with change in its price, while other factors are at constant, are termed as expansion or contraction of demand. Expansion of demand refers to the period when quantity demanded is more because of the fall in prices of a product. However, contraction of demand takes place when the quantity demanded is less due to rise in the price o a product. For example, consumers would reduce the consumption of milk in case the

prices of milk increases and vice versa. Expansion and contraction are represented by the movement along the same demand curve. Movement from one point to another in a downward direction shows the expansion of demand, while an upward movement demonstrates the contraction of demand.

Figure-11 demonstrates the expansion and contraction of demand:

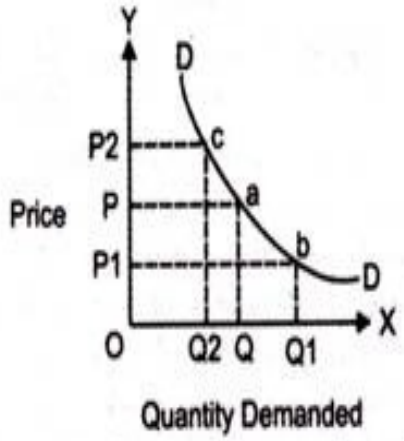


Figure-11: Expansion and Contraction of Demand

In Figure-12, the movement from DD to D1D1 shows the increase in demand with price at constant (OP). However, the quantity has also increased from OQ to OQ1.

Increase and Decrease in Demand:

Increase and decrease in demand are referred to change in demand due to changes in various other factors such as change in income, distribution of income, change in consumer's tastes and preferences, change in the price of related goods, while Price factor is kept constant Increase in demand refers to the rise in demand of a product at a given price.

On the other hand, decrease in demand refers to the fall in demand of a product at a given price. For example, essential goods, such as salt would be consumed in equal quantity, irrespective of increase or decrease in its price. Therefore, increase in demand implies that there is an increase in demand for a product at any price. Similarly, decrease in demand can also be referred as same quantity demanded at lower price, as the quantity demanded at higher price.

Increase and decrease in demand is represented as the shift in demand curve. In the graphical representation of demand curve, the shifting of demand is demonstrated as the movement from one demand curve to another demand curve. In case of increase in demand, the demand curve shifts to right, while in case of decrease in demand, it shifts to left of the original demand curve.

Figure-13 shows the decrease in demand:

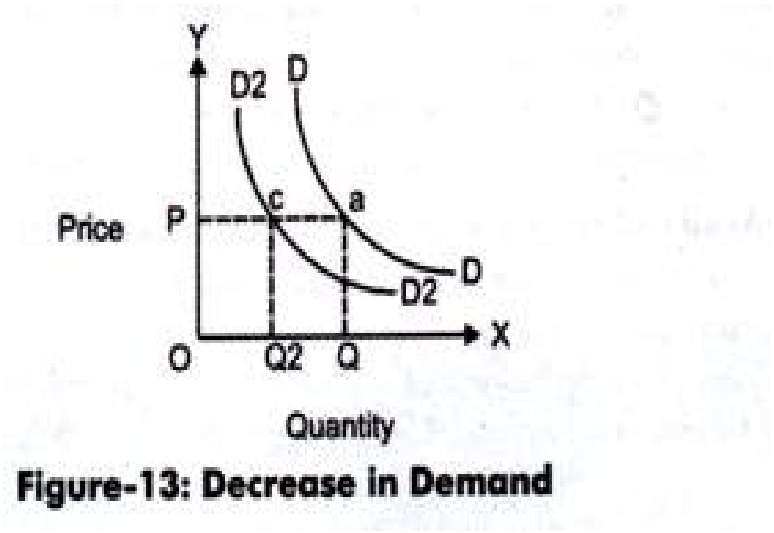


Figure-13: Decrease in Demand

In Figure-13, the movement from DD to D2D2 shows the decrease in demand with price at constant (OP). However, the quantity has also decreased from OQ to OQ2.

Concept of elasticity

The concept of elasticity simply involves the percentage change in one variable associated with a given percentage change in another variable. The elasticity of demand means responsiveness of demand due to change in price of the commodity, income of the consumer and price of related good. The law of demand state only the direction of change in the quantity demanded with the change in price. But what is more important for analyzing the effect of change in price on demand is the elasticity of demand. The law of demand does not give the degree of response of demand to a given change in price but elasticity of demand.

The extent of relationship between the price and the demand is measured by measuring the degree of responsiveness of demand for a product to the change in its price. This is called the elasticity of demand and supply.

Kinds of elasticity of demand:

i. Price elasticity of demand;

ii. Income elasticity of demand; and

iii. Cross-elasticity of demand, *i.e.,* demand elasticity with reference to price of substitutes and complementary goods.

Price elasticity of demand

The price elasticity of demand is defined as the degree of responsiveness or sensitiveness of demand for a commodity to the change in its price. The price elasticity of demand, i.e., the responsiveness of demand for a commodity to change in its price, is measured as the percentage change in the quantity demanded divided by the percentage change in the price. Prof. Alfred Marshall has developed the concept of elasticity of demand in his book 'Principles of Economics' (1890).

$$e_p = \frac{\text{Percentage change in the quantity demanded}}{\text{Percentage change in the price}}$$

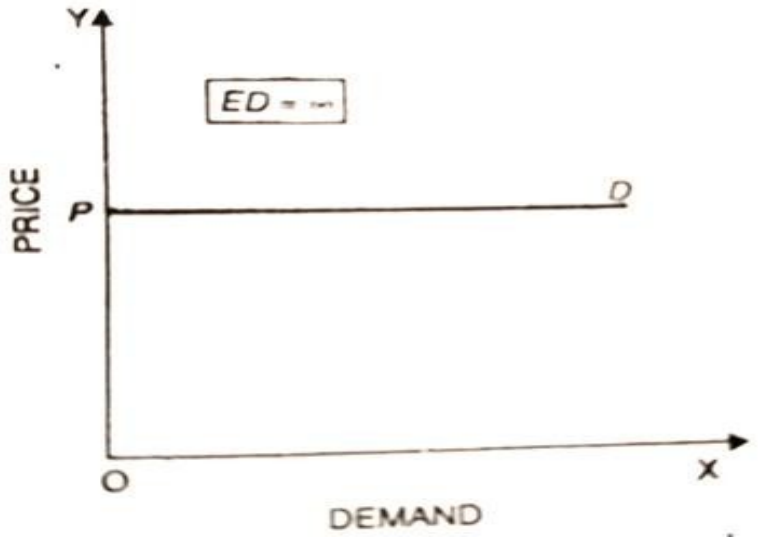
Degrees (or kinds) of price elasticity

The extent of responsiveness of demand with change in the price is not always the same. The demand for a product can be elastic or inelastic, depending on the rate of change in the demand with respect to change in price of a product. Elastic demand is the one when the response of demand is greater with a small proportionate change in the price. On the other hand, inelastic demand is the one when there is relatively a less change in the demand with a greater change in the price.

I. *Perfectly Elastic Demand:*

When a small change in price of a product causes a major change in its demand, it is said to be perfectly elastic demand. In perfectly elastic demand, a small rise in price results in fall in demand to zero, while a small fall in price causes increase in demand to infinity. In such a case, the demand is perfectly elastic or ep = 00.

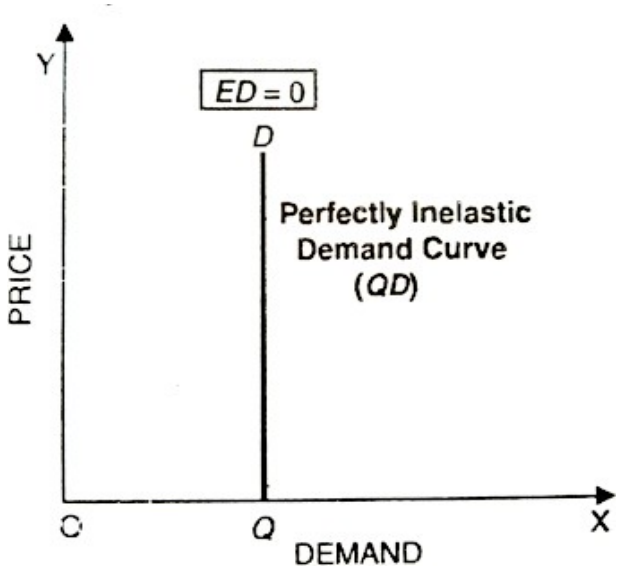
The degree of elasticity of demand helps in defining the shape and slope of a demand curve. Therefore, the elasticity of demand can be determined by the slope of the demand curve. Flatter the slope of the demand curve, higher the elasticity of demand. In perfectly elastic demand, the demand curve is represented as a horizontal straight line, which is shown in Figure.



From Figure, it can be interpreted that at price OP, demand is infinite; however, a slight rise in price would result in fall in demand to zero. It can also be interpreted from Figure-2 that at price *P* consumers are ready to buy as much quantity of the product as they want. However, a small rise in price would resist consumers to buy the product. Though, perfectly elastic demand is a theoretical concept and cannot be applied in the real situation. However, it can be applied in cases, such as perfectly competitive market and homogeneity products. In such cases, the demand for a product of an organization is assumed to be perfectly elastic. From an organization's point of view, in a perfectly elastic demand situation, the organization can sell as much as much as it wants as consumers are ready to purchase a large quantity of product. However, a slight increase in price would stop the demand.

II. *Perfectly Inelastic Demand:*

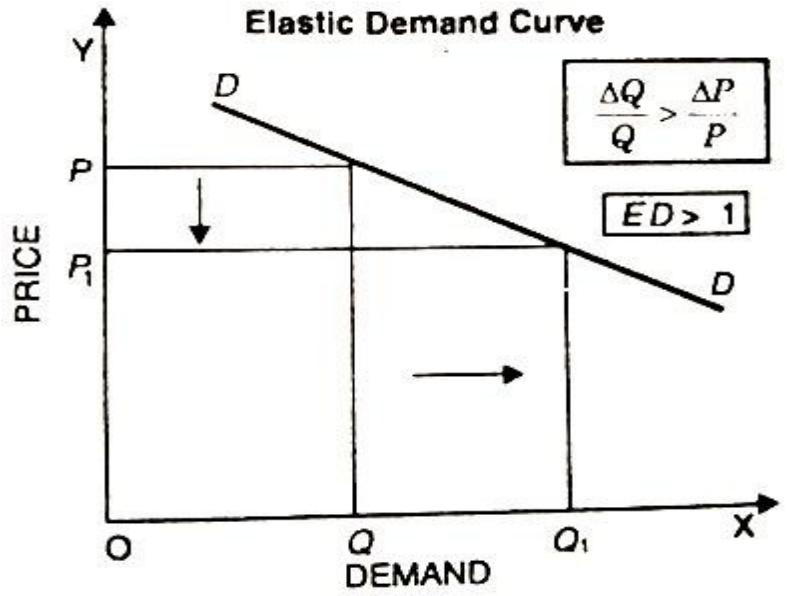
A perfectly inelastic demand is one when there is no change produced in the demand of a product with change in its price. The numerical value for perfectly inelastic demand is zero (ep=0). In case of perfectly inelastic demand, demand curve is represented as a straight vertical line, which is shown in Figure.



The demand remains constant for any value of price. Perfectly inelastic demand is a theoretical concept and cannot be applied in a practical situation. However, in case of essential goods, such as salt, the demand does not change with change in price. Therefore, the demand for essential goods is perfectly inelastic.

III. *Elastic Demand:*

Relatively elastic demand refers to the demand when the proportionate change produced in demand is greater than the proportionate change in price of a product. The numerical value of relatively elastic demand ranges between one to infinity. Mathematically, relatively elastic demand is known as more than unit elastic demand (ep>1). For example, if the price of a product increases by 20% and the demand of the product decreases by 25%, then the demand would be relatively elastic.

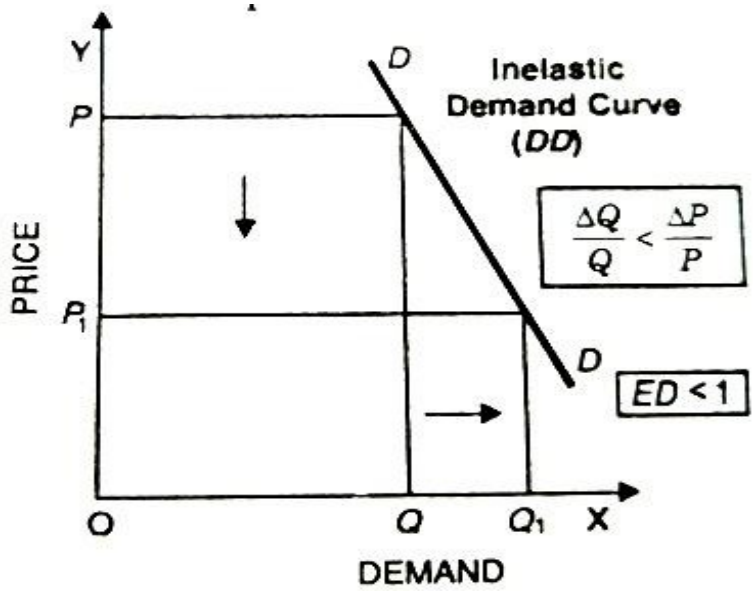


It can be interpreted from Figure that the proportionate change in demand from OQ to OQ1 is relatively larger than the proportionate change in price from OP to OP1. Relatively elastic demand has a practical application as demand for many of products respond in the same manner with respect to change in their prices.

For example, the price of a particular brand of cold drink increases from Rs. 15 to Rs. 20. In such a case, consumers may switch to another brand of cold drink. However, some of the consumers still consume the same brand. Therefore, a small change in price produces a larger change in demand of the product.

IV. *Inelastic Demand:*

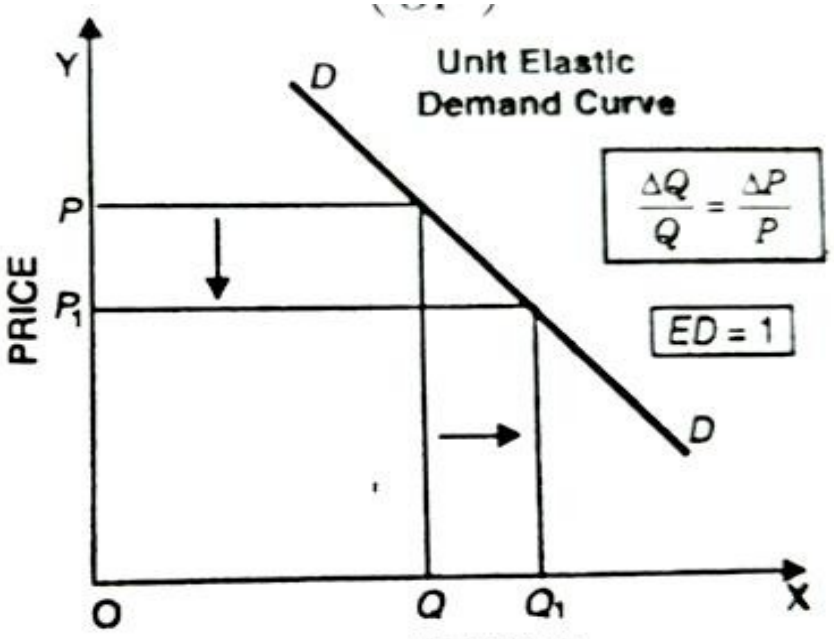
Relatively inelastic demand is one when the percentage change produced in demand is less than the percentage change in the price of a product. For example, if the price of a product increases by 30% and the demand for the product decreases only by 10%, then the demand would be called relatively inelastic. The numerical value of relatively elastic demand ranges between zero to one (ep<1). Marshall has termed relatively inelastic demand as elasticity being less than unity.



It can be interpreted from Figure that the proportionate change in demand from OQ to OQ1 is relatively smaller than the proportionate change in price from OP to OP1. Relatively inelastic demand has a practical application as demand for many of products respond in the same manner with respect to change in their prices.

V. *Unitary Elastic Demand:*

When the proportionate change in demand produces the same change in the price of the product, the demand is referred as unitary elastic demand. The numerical value for unitary elastic demand is equal to one (ep=1).



From Figure, it can be interpreted that change in price OP to OP1 produces the same change in demand from OQ to OQ1. Therefore, the demand is unitary elastic.

Factors determining price elasticity

Now an important question is what are the factors which determine whether the demand for a goods is elastic or inelastic. The following are the main factors which determine price elasticity of demand for a commodity.

1. *The availability of substitutes.*

Of all the factors determining price elasticity of demand the availability of the number and kinds of substiutes for a commodity is the most important factor. If for a commodity close substitutes are available, its demand tends to be elastic. If price of such a commodity goes up, the people will shift to its close substitutes and as a result the demand for that commodity will greatly decline. The greater the possibility of substitution, the greater the price elasticity of demand for it. If for a commodity good, substitutes are not available, people will have to buy it even when its price rises, and therefore its demand would tend to be inelastic.

For instance, if price of Coca Cola were to increase sharply, many consumers would turn to other kind of cold drinks, and as a result, the quantity demanded of Coca Cola will decline very much. On the other hand, if price of Coca Cola falls, many consumers will change from other cold

drinks to Coca Cola. Thus, the demand for Coca Cola is elastic. It is the availability of close substitutes that makes the consumers senstive to the changes in price of Coca Cola and this makes the demand for Coca Cola elastic. Likewise, demand for common salt is inelastic because good substitutes for common salt are not available. If the price of common salt rises slightly, people would consume almost the same quantity of common salt as before since good substitutes are not available. The demand for common salt is inelastic also because people spend a very little part of their income on it and even if its price rises, it makes only negligible difference in their budget allocation for common salt.

2. *The proportion of consumer's income spent.*

Another important determinant of the elasticity of demand is how much it accounts for in consumer's budget. In other words, the proportion of consumer's income spent on a particular commodity also influences price elasticity of demand for it. The greater the proportion of income spent on a commodity, the greater will generally be its price elasticity of demand, and vice versa. The demand for common salt, soap, matches and such other goods tends to be highly inelastic because the households spend only a fraction of their income on each of them. When price of such a commodity rises, it will not make much difference in consumers' budget and therefore they will continue to buy almost the same quantity of that commodity and, therefore, demand for them will be inelastic. On the other hand, demand for cloth in a country like India tends to be elastic since households spends a good part of their income on clothing. If price of cloth falls, it will mean great saving in the budget of many households and therefore this will tend to increase the quantity demanded of the cloth. On the other hand, if price of cloth rises, many households will not afford to buy as much quantity of cloth as before, and therefore, the quantity demanded of cloth will fall.

3. *The number of uses of a commodity.*

The greater the number of uses to which a commodity can be put, the greater will be its price elasticity of demand. If price of a commodity having several uses is very high, its demand will be small and it will be put to the most important uses and if price of such a commodity falls it wil be put to less important uses also and consequently its quantity demanded will rise significantly. To illustrate, milk has several uses. If its price rises to a very high level, it will be used only for

essential purposes such as feeding the children and sick persons. If price of milk falls, it would be devoted to other uses such as preparation of curd, cream, ghee and sweets. Therefore, the demand for milk tends to be elastic.

4. *Complementarity between goods.*

Complementarity between goods or joint demand for goods also affects the price elasticity of demand. Households are generally less sensitive to the changes in prices of goods that are complementary with each other or which are jointly used as compared to those goods which have independent demand or used alone. For example, for the running of automobiles, besides petrol, lubricating oil is also used. Now, if price of lubricating oil goes up, it will mean a very small increase in the total cost of running the automobile, since the use of oil is much less as compared to other things such as petrol. Thus, the demand for lubricating oil tends to be inelastic. Similarly, the demand for common salt is inelastic, partly because consumers do not use it alone but along with other things.

It is worth mentioning here that for assessing elasticty of demand for a commodity all the above three factors must be taken into account. The three factors mentioned above may reinforce each other in determining the elasticity of demand for a commodity or they may operate against each other. The elasticity of demand for a commodity will be the net result of all the forces working on it.

5. *Time and elasticity.*

The element of time also influences the elasticity of demand for a commodity. Demand tends to be more elastic if the time involved is long. This is because consumers can substitute goods in the long run. In the short run, substitution of one commodity by another is not so easy. The longer the period of time, the greater is the ease with which both consumers and businessmen can substitute one commodity for another. For instance, if price of fuel oil rises, it may be difficult to substitute fuel oil by other types of fuels such as coal or cooking gas. But, given sufficient time, people will make adjustments and use coal or cooking gas instead of the fuel oil whose price has risen. Likewise, when the business firms find that the price of a certain material has risen, then it may not be possible for them to substitute that material by some other relatively cheaper one. But with the passage of time they can undertake research to find substitute material

and can redesign the product or modify the machinery employed in the production of a commodity so as to economize in the use of the dearer material. Therefore, given the time, they can substitute the material whose price has risen. We thus see that demand is generally more elastic in the long run than in the short run.

Methods of measurement:

Different methods have been devised by the economists to measure price elasticity. They are:

a) The percentage or proportional method,
b) Arc method,
c) Total outlay or expenditure method

a) *The percentage or proportional method*

The price elasticity of demand is measured by comparing the percentage change in price with the percentage change in demand. The elasticity is the ratio of the percentage change in the quantity demanded to the percentage change in price charged.

$$E_p = \frac{\%\ change\ in\ q}{\%\ change\ in\ p}$$

$$= \frac{\frac{\Delta q}{q}}{\frac{\Delta p}{p}}$$

$$= \frac{\Delta q}{\Delta p} \text{ X } \frac{p}{q}$$

Where q refers to quantity demanded p to price and Δ to change.

If *Ep > 1*, demand is elastic,

If *Ep < 1*, demand is inelastic, and

If *Ep = 1*, demand is unitary elastic.

Suppose the price of a particular commodity declines from Rs. 500 to Rs. 400 each, *i.e.,* by 20%, whereas its demand goes up from 400 to 600, *i.e.,* by 50%, elasticity of demand will be 50/20 or 2.5 percent, which is a case of elastic demand. It has been generally observed that the demand for the *luxuries* is elastic and for *necessaries* like food grains it is inelastic. It may be *zero* for essential commodities.

The elasticity of demand is always negative, although by convention it is taken to be positive.

It is negative because change in quantity demanded is in opposite direction to the change in price. A fall in price is followed by rise in demand and vice-versa, unless the demand curve is abnormal, i.e., sloping upward from right to left.

Strictly speaking, in mathematical terms, there should be minus sign (-) before the figure indicating price elasticity. But by convention, the minus sign is dropped for the sake of simplicity.

b) *Arc method*

The arc method uses the mid-points between the old and new data in the case of both price and quantity demanded. The arc method studies a portion or a segment of the demand curve between the two points. An arc is a portion of a curved line, hence, a portion or segment of a demand curve. Arc elasticity is the elasticity at the mid-point of an arc of a demand curve.

The formula for measuring arc elasticity is

$$E_d = \frac{Change\ in\ quantity\ demanded}{Original\ quality + new\ quantity\ after\ change} \div \frac{Change\ in\ price}{Original\ price + new\ price\ after\ change}$$

Symbolically, the formula may be expressed thus:

$$E_d = \frac{q - q_1}{q + q_1} \div \frac{p - p_1}{p + p_1}$$

- Here, q = original quantity demanded;

  $q1$ = *new* quantity after the change in price;

  p = original price; and

  $p1$ = new price after change.

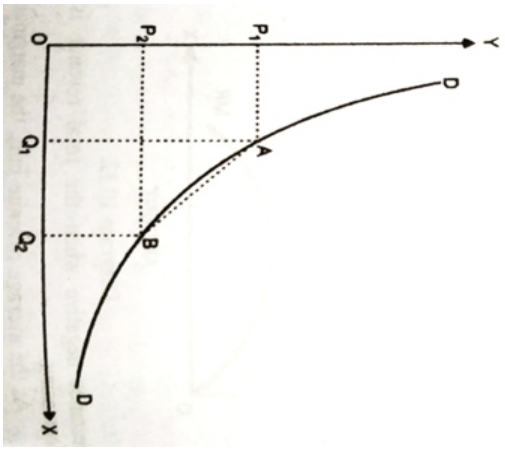
Illustration

Suppose that the price of a commodity is Rs. 5/-and the quantity demanded at that price is 100 units. Now assume that the price of the commodity falls to Rs. 4/-and the quantity demanded rises to 110 units. In terms of the above formula, arc elasticity then will be:

$$E_d = \frac{100 - 110}{100 + 110} \div \frac{5 - 4}{5 + 4}$$

$$= \frac{-10}{210} \div \frac{1}{9} = \frac{-10}{210} \times \frac{9}{1}$$

$$= -\frac{9}{21} = \frac{9}{21}$$



c) Total outlay or expenditure method

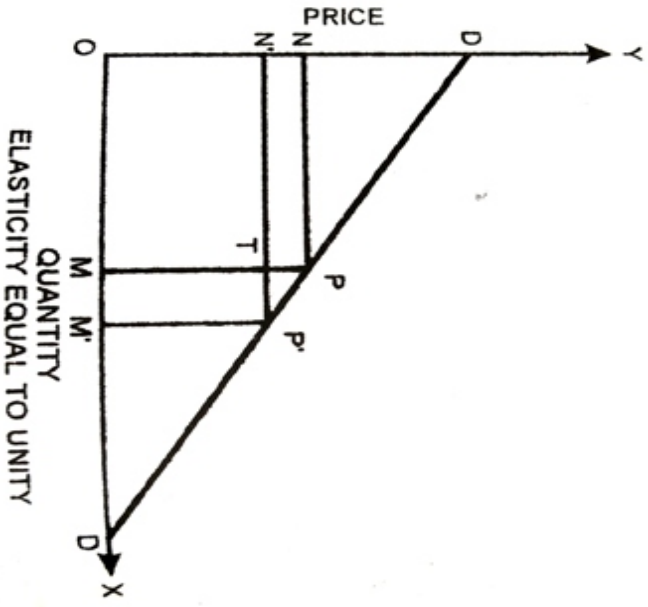
This method is associated with the name of Alfred Marshall. According to this method, if a small fall in price will cause an equally proportionate increase in the quantity demanded. In this method, one would consider the change in price and the subsequent change in the outlay on the purchase of the commodity.

If, for instance, a given change in price does not cause any change in the total amount of money spent on the commodity (i.e., if the total outlay on the commodity remains constant), elasticity of demand is said to be *equal to unity.*

If, as, a result of a given change in price, the total outlay is increased, elasticity of demand is said to be *greater than unity.*

If, on the contrary, as a result of a given change in price, the total outlay is diminished, the elasticity of demand is said to be *less than unity.*

i. *Elasticity of demand is equal to unity*

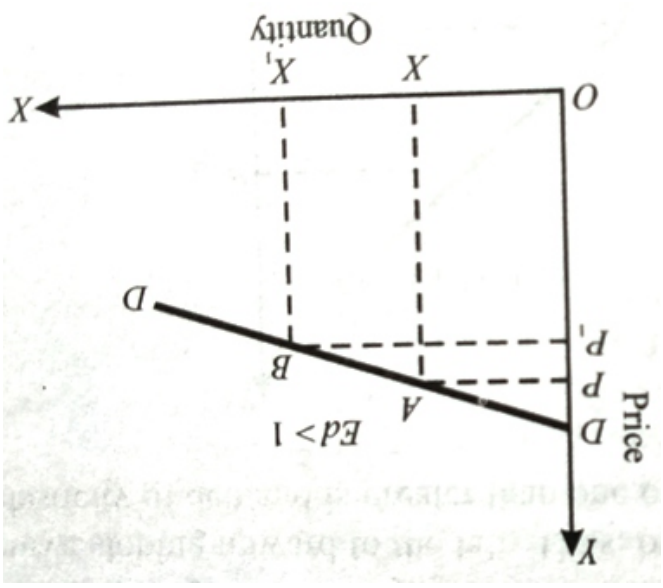


PM is the original price and *OM* is the original quantity demanded. The original outlay, therefore, is *OM X PM = OMPN*. *P'M'* is the new price and *M'* is the new quantity demanded.

The new outlay is *OM'P'N'*. The quantity demanded has, no doubt, increased as a result of a fall in price. But the new total outlay (represented by *OM'P'N'*) is the same as the original outlay (represented by *OMPN*) – the rectangle *NPTN'* showing the decrease in outlay due to fall in price is made good by the increase in outlay represented by the rectangle *TMM'P*, caused by the increase in the quantity demanded at the lower price.

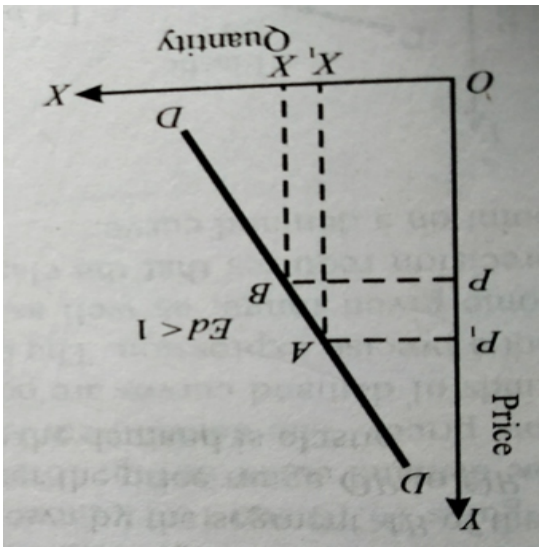
The price falls and the quantity demanded increases by the same percentage with the net result that there is no change in total outlay. Elasticity of demand therefore, equal to unity.

ii. *Elasticity of demand is greater than unity*



In the diagram, it shows a situation where demand is elastic in the price range *OP-OP1*, because here total quantity demanded is greater at lower price (*i.e.,* the area OP1BX1>OPAX).

iii. *Elasticity of demand is less than unity*



In the diagram, the area *OP1AX1* exceeds the area *OPBX* and illustrates a situation where demand in this price range *(OP to OP1)* is decreasing (inelastic).

LAW OF SUPPLY

The Meaning of Supply

As demand is defined as a schedule of the quantities of good that will be purchased at various prices, similarly the supply refers to the schedule of the quantities of a good that the firms are able and willing to offer for sale at various prices. How much of a commodity the firms are able to produce depends on the resources available to them and the technology they employ for producing a commodity. How much of a commodity the firms will be willing to offer for sale depends on the profits they expect to make on producing and selling the commodity. Profits in turn depend on the price of the commodity on the one hand and unit cost of production on the other.

Supply should be distinguished from the quantity supplied. Whereas supply of a commodity is the entire schedule of the quantities of a commodity that would be offered for sale at all possible prices during a period of time, for example, a day, a week, a month and so on, the quantity supplied refers to the quantity of a commodity which the firms are able and willing to sell at a particular price of the commodity. Thus the term ‘supply’ refers to the entire relationship between the price of a commodity and the quantity supplied at various possible prices and is illustrated by the entire supply curve or supply schedule as given in Figure 23.1 and Table

23.1, where the term 'quantity supplied' refers to a point on a given supply curve, that is, quantity supplied at a particular price.

Two things are worth mentioning about the concept of supply. First, supply is a flow concept, that is, it refers to the amount of a commodity that the firms produce and offer for sale in the market per period of time, say a week, a month or a year. Without specifying the time period, supply of a commodity has a little meaning. Second, the quantity supplied at a commodity which the producers plan to produce and sell at a price is not necessarily the same as the quantity actually sold. Sometimes, the quantity which the firms are willing to produce and sell at a price is greater than the quantity demanded, so the quantity actually bought and sold is less than the quantity supplied.

Supply Function

The quantity of a commodity that firms will be able and willing to offer for sale in the market depends on several factors. The important factors determining supply of a commodity are:

1. The price of the commodity
2. The prices of inputs (i.e., resources) used for the production of the commodity
3. The state of technology
4. The number of firms producing and selling the commodity
5. The prices of related goods produced.
6. Future expectations regarding prices.

We will explain these factors determining supply of a commodity in detail in a later section. However, it may be noted that out of the above determinants of supply the own price of the commodity, the prices of inputs (i.e., resources) used to produce the commodity, and the technology used in production are three important factors and therefore the supply function of a commodity is written taking these factors as independent variables. Thus, supply function of a commodity is written as

$$Q_x^s = S(P_x, F_1, F_2 \ldots\ldots\ldots\ldots F_m)$$

where G^{s}_{x} is the quantity supplied of the commodity *X*, *Px* is it own price, *F*1, *F*2,... *Fm* are the prices of inputs used to produce the commodity *X*. The state of technology determines the form of supply function *S*. It must be noted that the form of the function refers to the precise quantitative relation between the independent variables such as the own price of the commodity *Px* and prices of factors such as, *F*1, *F*2 etc.

The Relation between Price and Quantity Supplied: Law of Supply

Supply of a commodity is functionally related to its price. The law of supply relates to this functional relationship between price of a commodity and its quantity supplied. In contrast to the inverse relationship between the quantity demanded and the changes in price, the quantity supplied of a commodity generally varies directly with price. That is, the higher the price, the larger is the quantity supplied of a commodity. The supply schedule and the upward-sloping supply curve reflect the law of supply. According to the law of supply, when the price of a commodity rises, the quantity supplied of it in the market increases, and when the price of the commodity falls, its quantity supplied decreases, other factors determining supply remaining the same. Thus, according to the law of supply, the quantity supplied of a commodity is directly or positively related to price. It is due to this positive relationship between price of a commodity and its quantity supplied that the supply curve of a commodity slopes upward to right as seen from supply curve SS in Figure. 23.1. When price of wheat rises from ` 520 to ` 530 per quintal, the quantity supplied of wheat in the market increases from 200 quintals to 225 quintals per period.

Table 23.1. Supply Schedule of Wheat

Price Per Quantity (₹)	*Quantity Supplied (in quintals)*
500	100
510	150
520	200
530	225
540	250
550	275

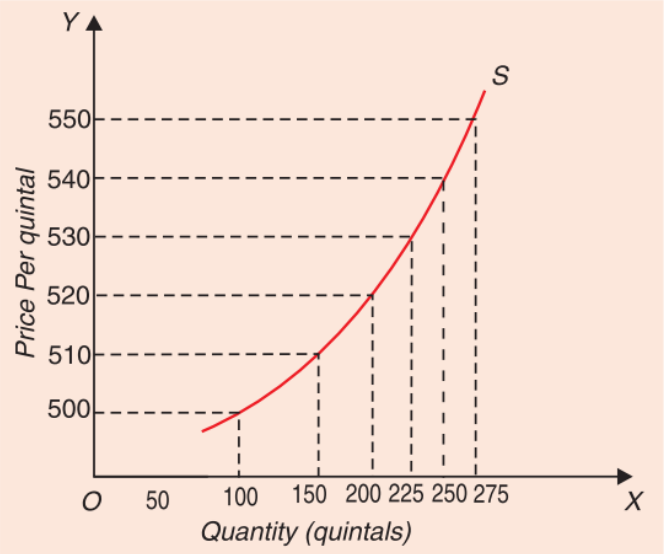


Fig. 23.1. *Supply Curve Showing Direct Relationship Between Price and Quantity Supplied*

Similar to the demand schedule we can construct an individual's supply schedule. Also by totalling up the quantity supplied at various prices by all the sellers in a market, we can obtainthe supply schedule of the market. Supply schedule represents the relation between prices and the quantities that the firms are able and willing to produce and sell at various prices. We have given in Table 23.1 a supply schedule of wheat per day in a market.

It will be seen from the above table that when price of wheat is ` 500 per quintal, the 100 quintals of wheat are supplied in the market. When price of wheat rises to ` 510 per quintal, 150 quintals of wheat are supplied. When price of wheat goes up to ` 550 per quintals, its quantity supplied in the market rises to 275 quintal. By plotting the above supply schedule of wheat on a graph paper we have obtained supply curve SS in Fig. 23.1. In Fig. 23.1, the quantity supplied has been measured along the X-axis and price of wheat has been measured along the Y-axis. It will be seen from this figure that supply curve slopes upward from left to right, which indicates that as price of wheat rises, quantity supplied increases and vice versa. This is in a sharp contrast to the demand curve of a commodity which slopes downward from left to right.

Explanation of the Law of Supply: Why does Supply Curve Generally Slope Upward?

It has been observed that price of a product and quantity supplied of it by firms producing it are positively related to each other, that is, at a higher price more is supplied and vice versa, other

things remaining the same. In analyzing the relation between price of a commodity and the quantity supplied, given *ceteris parilbus* assumption, we are in fact dealing with the supply function, $Qsx = f(Px)$. This positive relationship between price and quantity supplied is an important law of economics. How do we explain it? It should be remembered that firms are driven by profit motive. The higher price of a product, given the cost per unit of output, makes it profitable to expand output and offer more quantity of the product for sale. Thus, higher price serves as an incentive for the producer to produce more of it. The higher the price, the greater the incentive for the firm to produce and supply more of a commodity in the market, other things remaining the same. The basic reason behind the law of supply (*i.e.,* positive relationship between price and quantity supplied) is the way cost changes as output is expanded to offer more for sale. To produce more of a product, firms have to devote more resources to its production. When production of a product is expanded by using more resources, diminishing returns to variable factors occur. Due to the diminishing returns, average and marginal costs of production increase. Therefore, at higher additional cost of producing more units of output; it is profitable to produce and supply more units of output *only at a higher price* so as to cover the rise in additional cost per unit.

Further, the changes in quantity supplied of a product following the changes in its price also depends on the *possibilities of substitution of one product for an other.* For example, if price of wheat in the market rises, the farmers will alter the cropping pattern so as to produce more of wheat by withdrawing land and other resources from the cultivation of gram and devoting them to the production of wheat. This is because high market price for wheat relative to gram induces farmers, who aim at maximizing profits, to use more resources for production of wheat and fewer resources for the production of gram.

Shifts in Supply: Increase and Decrease in Supply

As stated above, the supply of a commodity in economics means the entire schedule or curve depicting the relationship between price and quantity supplied of the commodity, given the other factors influencing supply. These other factors are the state of technology, prices of inputs (resources), prices of other related commodities, etc., which are assumed constant when the relationship between price and quantity supplied of a commodity is examined. It is the change in

these factors other than price that cause a shift in the supply curve. For example, when prices of inputs such as labour and raw materials used for the production of a commodity decline, this will result in lowering the cost of production which will induce the producers to produce and make available a greater quantity of the commodity in the market at each price. This increase in supply of a commodity due to the reduction in prices of inputs will cause the entire supply curve to shift to the right as shown in Figure 23.2 where the supply curve shifts from SS to S' S'. As shown by arrow marks, at price P1, P2 and P3 quantity supplied increases when supply increases causing a rightward shift in the supply curve. Similarly, progress in technology used for production of a commodity which increases productivity and reduces cost per unit will also cause the supply curve to shift to the right.

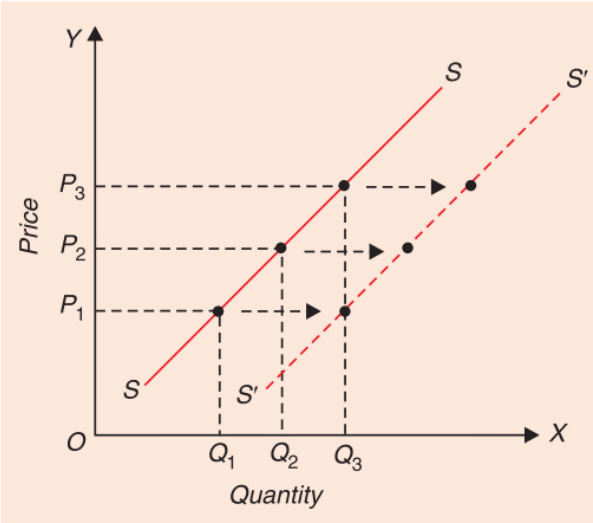


Fig. 23.2. *Increase in Supply Causing a Rightward Shift in the Supply Curve*

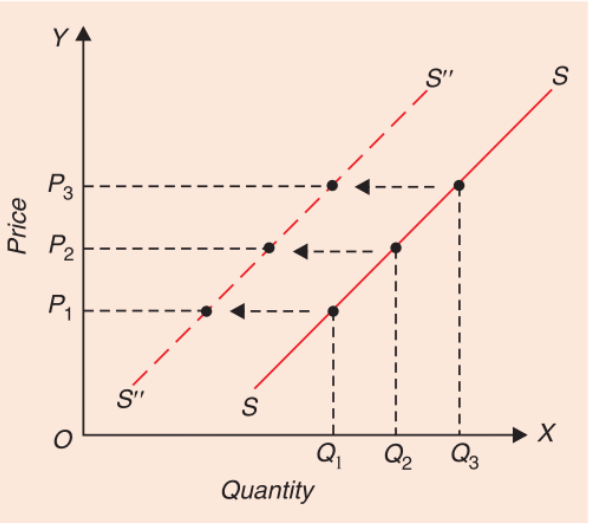


Fig. 23.3. *Decrease in Supply Causing a Leftward Shift in the Supply Curve*

On the other hand, decrease in supply means the reduction in quantity supplied at each price of the commodity as shown in Figure 23.3 where as a result of decrease in supply the supply curve shifts to the left from SS to S″S″. As shown by the arrow marks, at each price such as P1, P2, P3, the quantity supplied on the supply curve S″S″ has declined as compared to the supply curve SS. The decrease in supply occurs when the rise in prices of factors (inputs) used for the production of a commodity produced leads to higher cost per unit of output which causes a reduction in quantity supplied at each price. Similarly, the imposition of an excise duty or sales tax on a commodity means that each quantity will now be supplied at a higher price than before so as to

cover the excise duty or sales tax per unit. This implies that quantity supplied of the commodity at each price will decrease as shown by the shift of the supply curve to the left.

Another important factor causing a decrease in supply of a commodity is the rise in prices of other commodities using the same factors. For example, if the price of wheat rises sharply, it will become more profitable for the farmers to grow it. This will induce the farmers to reduce the cultivated area under other crops, say pulses, and devote it to the production of wheat. This will lead to the decrease in supply of pulses whose supply curve will shift to the left. Further, agricultural production in India greatly depends on the rainfall due to monsoons. If monsoon come in time and rainfall is adequate, there are bumper crops, the supply of agricultural products increases. However, in a year when monsoons are untimely or highly inadequate, there is a sharp drop in agricultural output which causes a shift in the supply curve of agricultural output to the left.

The supply of a commodity in the market at any time is also determined by sellers' expectations of future prices. If, as happens during inflationary periods, sellers expect the prices to rise in future, they would reduce current supply of a product in the market and would instead hoard the commodity. The hoarding of huge quantities of goods by traders is an important factor in reducing their supplies in the market and thus causing further rise in their prices.

Taxes and subsidies also influence the supply of a product. If an excise duty or sales tax is levied on a product, the firms will supply the same amount of it at a higher price or less quantity of it at the same price. This is because excise duty on a commodity is included in price by the sellers to pass it on to the buyers. This implies that imposition of a sales tax or excise duty will cause a leftward shift in the supply curve. On the other hand, when government provides subsidy on a commodity, it will reduce the supply price of a commodity. This implies that grant of subsidy on a commodity will cause the supply curve to shift to the right. It follows from above that technology, prices of factors and products, expectations regarding future prices and taxes and subsidies are the important determinants of supply which cause rightward or leftward shift in the whole supply curve. We thus see that there are several factors other than price which determine the supply of a commodity and any change in these other factors will cause a shift in the entire supply curve.

UNIT -III: THEORY OF PRODUCTION

Introduction

The act of production involves the transformation of inputs into outputs. *The word production in economics is not merely confined to effecting physical transformation in the matter, it is creation or addition of value*. Therefore, production in economics also covers the rendering of services such as transporting, financing, and marketing. Laws of production, or in other words, the generalizations regarding relations between inputs and outputs developed in this chapter will apply to all these types of production. *The theory of production provides a formal framework to help the managers of firms in deciding how to combine various factors or inputs most efficiently to produce the desired output of a product or service*.

The relation between inputs and output of a firm has been called the '*Production Function'*. Thus, the theory of production is the study of production functions. The production function of a firm can be studied by holding the quantities of some factors fixed, while varying the amount of other factors. This is called *short-run production function.* The time period in which at least one factor or input is fixed and production is increased by varying other factors is called the *short run*. The study of short-run production function when at least one factor is kept fixed forms the subject of *law of diminishing returns* which is also known as *law of Variable Proportions.* On the other hand, the time period when *all factors* are variable is called the *long run.* The length of the long run, that is, the time period required for changes in all inputs depends on the industry. For some industries such as making of wooden chairs or tables the long run may be few weeks or months but for production of steel, it may be many years as it takes several yours to expend the capacity of steel production. The behaviour of production when all factors are varied proportionately is the subject-matter of *returns to scale.* Thus, in the theory of production, the study of (a) the returns to a variable factor and (b) the returns to scale is included. Besides this, the theory of production is also concerned with explaining which combination of inputs(or factors of production) a firm will choose so as to minimize its costs of production for producing a given level of output or to maximize output for a given level of cost.

Production Function

Production, as said above, is transformation of physical inputs into physical outputs. The output is thus a function of inputs. The functional relationship between physical inputs and physical output of a firm is known as production function. Algebraically, production function can be written as:

$$q = f(a, b, c, d.....)$$

where q stands for the quantity of output, a, b, c, d etc., stand for the quantities of factors A, B, C, D respectively. This function shows that the quantity (q) of output produced depends upon the quantities, a, b, c, d of the factors A, B, C and D respectively. The above function shows that there exists *some relationship* between output q and the quantities of inputs a, b, c, d etc., but it does not tell us the specific form of this relationship. This unspecified relationship is denoted here by the letter f. If the form of the function f is given, that is, if right-hand side of the equation (i) is given in a *specific* mathematical form, we can then fully find out precisely the quantity of output that the firm would produce with each set of inputs such as labour and capital.

"*The production function is the name given to the relationship between the rates of input of productive services and the rate of output of product. It is the economist's summary of technological knowledge."* Thus, the production function expresses the relationship between the quantity of output and the quantity of various inputs used for the production. More precisely, the production function states the *maximum* quantity of output that can be produced from any given quantities of various inputs per period of time or, in other words, it states the *minimum* quantities of various inputs that are required to yield a given quantity of output per period of time, the technology being assumed to remain constant. It is important to note that when a change in technology occurs such as introduction of a new automated machine or the substitution of skilled labour for unskilled labour we will get a new production function. If a small firm produces wooden tables in a day, its production function will consist of the *maximum* number of tables that can be produced from *given quantities* of various inputs such as wood, varnish, labour time, machine time, floor space. Or, the production function of that firm may also be defined as the *minimum* quantities of wood, varnish, labour time, machine time, floor space, etc., that are required to produce a *given number* of tables

per day.

Knowledge of the production function is a technological or engineering knowledge and is provided to the firm by its engineers or production managers. Two things must be noted in respect of production function. First, production function, like the demand function, must be considered with reference to a *particular period of time, Production function expresses a flow of inputs resulting in a flow of output in a specific period of time.* Secondly, production function of a firm is determined by the state of technology. When there is advancement in technology, the production function changes with the result that the new production can yield greater flow of output from the given inputs, or smaller quantities of inputs can be used for producing a given quantity of output.

In economic theory we, are interested in two types of input-output relations or production functions. First, we study the production function when the quantities of some inputs are kept constant and the quantity of one input (or quantities of few inputs) is varied. This kind of input-output relation forms the subject-matter of the law of variable proportion. Since only in the short run, some factors are required to be held constant, the law of variable proportions relates to the short-run production function. Secondly, we study the input-output relation by varying all inputs proportionally. This forms the subject-matter of returns to scale. Since in the long run all factors can be varied, the question of returns to scale relates to *long-run production function.*

Production function can be represented in various forms; it can be represented by tables, graphs, mathematical equations, showing the maximum quantity of output that a firm can produce per period of time with various combinations of factors (*i.e*., inputs). When two factors have to be explicitly shown, production function can be represented by iso-quants (*i.e.*, equal product curves). Production function can also be represented by *input-output tables.* However, it is worth mentioning that although production function provides quite a useful information about the production possibilities open to a firm, it does not give all the information required for efficient combination of inputs to produce a given level of output or to determine the profit-maximizing rate of output. As mentioned above, production function describes a physical relationship which must be combined with prices of inputs to determine the efficient resource combination of producing a specific level of output.

LAW OF VARIABLE PROPORTIONS

Introduction

Law of variable proportions occupies an important place in economic theory. This law examines the production function with one factor variable, keeping the quantities of other factors fixed. In other words, it refers to the input-output relation when output is increased by varying the quantity of one input. When the quantity of one factor is increased keeping the quantity of the other factors constant, the proportion between the variable factor and the fixed factor is altered; the ratio of employment of the variable factor to that of the fixed goes on increasing as the quantity of the variable factor is increased also. Since under this law we study the *effects on output of variations in factor proportions,* this is known as the *law of variable proportions*. Since according to this law when one factor increases, other factors held constant, after a point, marginal returns to the variable factor diminishes, this is also called law of diminishing returns. This law has played a vital role in the history of economic thought and occupies an equally important place in modern economic theory and has been supported by the empirical evidence about the real world. The law of diminishing returns has been stated by various economists in the following manner: G. J. Stigler, a Nobel Prize winner in economics writes, "*As equal increments of one input are added; the inputs of other productive services being held constant, beyond a certain point the resulting increments of product will decrease, i.e., the marginal products will diminish.*

Similarly, Samuelson, another Nobel Laureate, writes "An increase in some inputs relative to other fixed inputs will, in a given state of technology, causes output to increase; but after a point the extra output resulting from the same additions of extra inputs will become less and less".

Assumptions of the Law

The law of variable proportions or diminishing returns as stated above holds good under the following conditions:

First, the state of technology is assumed to be given and unchanged. If there is improvement in technology, then marginal and average product may rise instead of diminishing.

Secondly, there must be some other inputs such as capital must be kept fixed. It is only in this way then that we are able to measure the changes in output caused by increase in a variable factor that we can alter the factor proportions and know its effects on output. This law does not apply in case all factors are proportionately varied. Behaviour of output as a result of the variations in all inputs is discussed under "returns to scale".

Thirdly, the law is based upon the possibility of varying the proportions in which the various factors can be combined to produce a product. The law does not apply to those cases where the factors must be used in fixed proportions to yield a product. When the various factors are required to be used in rigidly fixed proportions, then the increase in one factor would not lead to any increase in output, that is, the marginal product of the factor will then be zero and not diminishing. It may however be pointed out that products requiring fixed proportions of factors are quite uncommon. Thus the law of diminishing marginal returns also known as law of variable proportions applies to most of the cases of production.

Table 18.1 Total Product, Marginal Product, Average Product of Labour

Labour (*i.e.* No. of workers) L	Total Product (TP) Q	Marginal Product (MP) $\left(\frac{\Delta Q}{\Delta L}\right)$	Average Product (AP) $\left(\frac{Q}{L}\right)$	Output Elasticity of Labour (E_L)
1	80	80	80	1
2	170	90	85	1.06
3	270	100	90	1.11
4	368	98	92	1.06
5	430	62	86	0.72
6	480	50	80	0.62
7	504	24	72	0.33
8	504	0	0	0
9	495	–9	55	–0.16
10	480	–15	48	–0.31

Returns to a variable factor, say labour, keeping the quantity of a fixed factor land as constant is illustrated in *Table 18.1* and *Figure18.5*. We shall first explain it by considering *Table 18.1*.

Assume that there is a given fixed amount of land with which more variable factor, labour, is used to produce wheat. With a given fixed quantity of land, as a farmer raises employment of labour from one unit to 7 units, total product increases from 80 quintals to 504 quintals of wheat. Beyond the employment of 8 units of labour, total product diminishes. It is worth noting that up to the use of 3 units of labour, total product increases at an increasing rate and afterwards it increases at a diminishing rate. This fact is clearly revealed from column 3 which shows successive marginal products of labour as extra units of labour are used. Marginal product of labour, it may be recalled, is the increment in total output due to the use of an extra unit of labour.

It will be seen from col. 3 of *Table 18.1* that the marginal product of labour initially raises and beyond the use of three units of labour, it starts diminishing. Thus when three units of labour are employed, marginal product of labour is 100 and with the use of 4th and 5th units of labour marginal product falls to 98 and 62 respectively. Beyond the use of eight units of labour, total product diminishes and therefore marginal product of labour becomes negative. As regards average product of labour, it rises up to the use of fourth unit of labour and beyond that it is falling throughout.

Three Stages of Production

The behaviour of output when the varying quantity of one factor is combined with a fixed quantity of the other can be divided into three distinct stages. In order to understand these three stages it is better to graphically illustrate the production function with one factor variable. This has been done in *Fig. 18.5.* In this figure, on the *X*-axis we measure the quantity of the variable factor and on the *Y*-axis we measure the total product, average product and the marginal product. How the total product, average product and marginal product of the variable factor change as a result of the increase in the quantity of one factor to a fixed quantity of the others will be seen from *Fig. 18.5.* The total product curve *TP* goes on increasing to a point and after that it starts declining. Average and marginal product curves also rise in the beginning and then decline; marginal product curve starts declining earlier than the average product curve. The behaviour of these total, average and marginal products of the variable factor consequent to the increase in its amount is generally divided into three stages which are explained below:

Stage 1: In this stage, total product increases at an increasing rate to a point. In Fig. 18.5 from the origin to the point *F*, slope of the total product curve *TP* is increasing, that is, up to the point *F*, *the total product increases at an increasing rate* (the total product curve *TP* is concave upwards to the point *F*), which means that the marginal product *MP* rises. From the point *F* onwards during the stage 1, the total product goes on rising but its slope is declining which means that from point F to the point H the total product increases at a diminishing rate (total product curve is concave downwards), i.e., marginal product falls but is positive. The point F where the total product stops increasing at an increasing rate and starts increasing at the diminishing rate is called the point of inflection. Corresponding vertically to this point of inflection marginal product is maximum, after which it starts diminishing.

The stage 1 ends where the average product curve reaches its highest point. During the stage 1, when marginal product of the variable factor is falling, it still exceeds its average product and so continues to cause the average product curve to rise. Thus, during the stage 1, whereas marginal product curve rises in a part and then falls, the average product curve rises throughout. In the first stage, the quantity of the fixed factor is too much relative to the quantity of the variable factor so that if some quantity of the *fixed factor* is withdrawn, the total product would increase. Thus, in the first stage *marginal product of the fixed factor is negative.* Stage 1 is called by some economists as the *stage of increasing returns* because average product of the variable factor increases throughout this stage. It is notable that the marginal product in this stage initially increases and in a later part it starts declining but remains greater than the average product throughout in stage 1 so that the average product continues to rise.

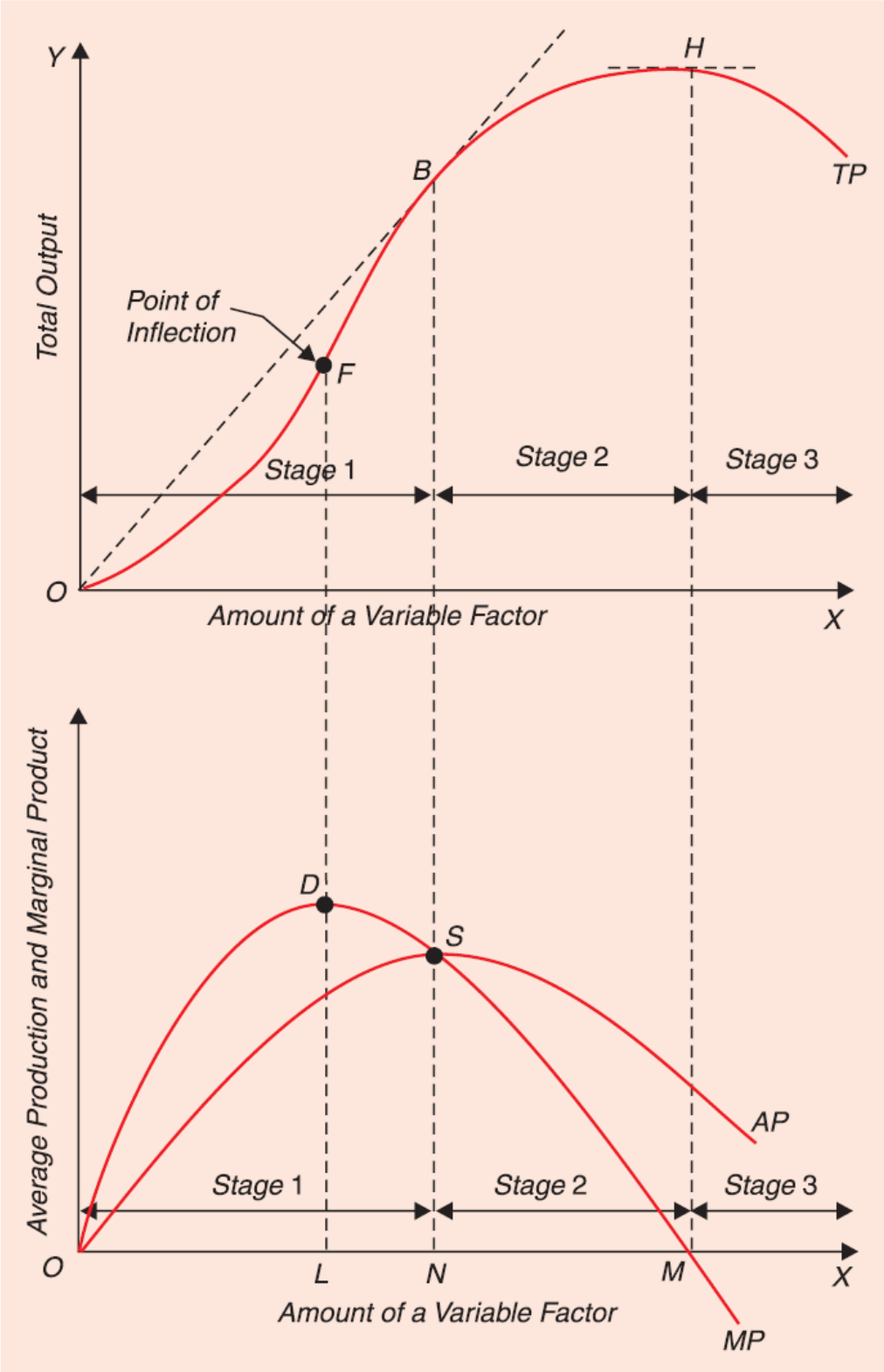


Fig. 18.5. *Three Stages of Production Function with one Variable Factor*

Stage 2: In stage 2, the total product continues to increase at a diminishing rate until it reaches its maximum point *H* where the second stage ends. In this stage, both the marginal product and average product of the variable factor are diminishing but are positive. At the end of the second stage, that is, at point *M* marginal product of the variable factor is zero (corresponding to the highest point *H* of the total product curve *TP*). Stage 2 is very crucial and important because the firm will seek to produce in its range. This stage is known as *the stage of diminishing returns* as both the average and marginal products of the variable factor continuously fall during this stage.

Stage 3: In stage 3, total product declines and therefore the total product curve *TP* slopes downward. As a result, marginal product of the variable factor is negative and the marginal product curve *MP* goes below the *X* axis. In this stage, variable factor is too much relative to the fixed factor. This stage is called the stage of negative returns, since the marginal product of the variable factor is negative during this stage. It may be noted that stage 1 and stage 3 are completely symmetrical. In stage 1 the fixed factor is too much relative to the variable factor. Therefore, in stage 1, marginal product of the fixed factor is negative. On the other hand, in stage 3, variable factor is too much relative to the fixed factors. Therefore, in stage 3, the marginal product of the variable factor is negative.

The Stage of Operation

Now an important question is in which stage a rational producer will seek to produce. A rational producer will never choose to produce in stage 3 where marginal product of the variable factor is negative. Marginal product of the variable factor being negative in stage 3, a producer can always increase his output by reducing the amount of variable factor. It is thus clear that a rational producer will never be producing in stage 3. Even if the variable factor is free, the rational producer will stop at the end of stage 2 where the marginal product of the variable factor is zero. At the end point *M* of the second stage where the marginal product of the variable factor is zero, the producer will be maximizing the total product and will thus be making maximum use of the variable factor.

A rational producer will also not choose to produce in stage 1 where the marginal product of the fixed factor is negative. A producer producing in stage 1 means that he will not be making the best use of the fixed factor and further that he will not be utilizing fully the opportunities of

increasing production by increasing quantity of the variable factor whose average product continues to rise throughout the stage 1. Thus a rational entrepreneur will not stop in stage 1 but will expand further. Even if the fixed factor is free (i.e., costs nothing), the rational entrepreneur will stop only at the end of stage 1 (i.e., at point *N*) where the average product of the variable factor is maximum. At this end point *N* of stage 1 he will be making maximum use of fixed factor.

It is thus clear from above that the rational producer will never be found producing in stage 1 and stage 3. Stage 1 and stage 3 may, therefore, be called stages of economic absurdity or economic nonsense. Thus stages 1 and 3 represent *non-economic region in production function*. A rational producer will always seek to produce in stage 2 where both the marginal product and average product of the variable factor are diminishing. At which particular point in this stage, the producer will decide to produce depends upon the price of variable factor. The stage 2 represents the range of rational production decisions.

Causes of Initial Increasing Marginal Returns to a Variable Factor

In the beginning, the quantity of the fixed factor is abundant relative to the quantity of the variable factor. Therefore, when more and more units of the variable factor are added to the constant quantity of the fixed factor, then the fixed factor is more intensively and effectively utilized, that is, the efficiency of the fixed factor increases as additional units of the variable factors are added to it. This causes the production to increase at a rapid rate. When, in the beginning, the variable factor is relatively smaller in quantity, some amount of the fixed factor may remain unutilized and, therefore, when the variable factor is increased fuller utilization of the fixed factor becomes possible with the result that increasing returns to a variable factor are obtained. The question arises as to why the fixed factor is not initially taken in a quantity which suits the available quantity of the variable factor. Answer to this question is provided by the fact that generally those factors are taken as fixed which are indivisible.

Indivisibility of a factor means that due to technological requirements, a minimum amount of it must be employed whatever the level of output. Thus, as more units of the variable factor are employed to work with an indivisible fixed factor, output greatly increases due to fuller and more effective utilization of the latter. Thus we see that it is the indivisibility of some factors which

causes increasing returns to a variable factor in stage 1. The second reason why we get increasing returns to a factor at the initial stage is that as more units of the variable factor are employed, the efficiency of the variable factor itself increases. This is because when there is a sufficient quantity of the variable factor, it becomes possible to introduce specialization or division of labour which results in higher productivity. The greater the quantity of the variable input, the greater the scope for specialization and hence greater will be the level of productivity and efficiency.

Causes of Diminishing Marginal Returns

The stage of diminishing returns in the production function with one factor variable is the most important. The question arises as to why we get diminishing marginal returns after a certain amount of the variable factor has been added to the fixed quantity of the other factor. It is important to note that the famous law of diminishing returns is in fact concerned with *diminishing marginal returns* and not diminishing total returns *i.e*. total product of a factor.

Scarcity of the fixed factor

As explained above, increasing returns to a factor occur in the first stage primarily because of the more effective and efficient use of the fixed factor as more units of the variable factor are combined to work with it. Once the point is reached at which the amount of the variable factor is sufficient to ensure the efficient utilization of the fixed factor, further increases in the variable factor will cause *marginal and average products to decline because the fixed factor then becomes inadequate relative to the quantity of the variable factor.* In other words, the *contributions to the production made by the variable factor after a point becomes less and less because the additional units of the variable factor have less and less of the fixed factors to work with. The production is the result of the co-operation of various factors aiding each other*. In the beginning in stage 1, the fixed factor is abundant relative to the number of the variable factor and the former provides much aid to the latter. On the other hand, when with the increase in the variable factor the fixed factor becomes more and more scarce in relation to the variable factor so that as the units of the variable factor are increased they receive less and less aid from the fixed factor. As a result, the marginal and average products of the variable factor decline.

Indivisibility of the Fixed Factor

The phenomenon of diminishing marginal returns, like that of increasing marginal returns, rests upon the indivisibility of the fixed factor. As explained above, the important reason for increasing returns in stage 1 is the fact that the fixed factor is indivisible which has to be employed whether the output to be produced is small or large. In stage 1 when the indivisible fixed factor is not being fully used, successive increases in the variable factor add more and more to output since fuller and more efficient use is made of the indivisible fixed factor. But there is generally a limit to the range of the employment of the variable factor over which its average product will increase. There will usually be a level of employment of the variable factor at which indivisible fixed factor is being as fully and efficiently used as possible and therefore the marginal product of the variable factor is maximum. It will happen when the variable factor has increased to such an amount that the fixed indivisible factor is being used in the 'best or optimum proportion' with the variable factor. Once the optimum proportion is disturbed by further increases in the variable factor, marginal returns to the variable factor (*i.e.,* marginal product) will diminish primarily because the indivisible factor is being used in wrong proportion with the variable factor. Just as the marginal product of the variable factor increases in the beginning when better and fuller use of the fixed indivisible factor is being made, so the marginal product of the variable factor diminishes when the fixed indivisible factor is being worked too hard.

If the fixed factor was finely divisible, neither the increasing nor the diminishing marginal returns would have occurred. If the factors were perfectly divisible, then there would not have been the necessity of taking a large quantity of the fixed factor in the beginning to combine with the varying quantities of the other factor. *In the presence of perfect divisibility of the factors, the optimum proportion between the factors could be achieved in every case*. Perfect divisibility of the factors implies that a small farm with a miniature combine and one worker would be as efficient as a large farm with a large combine and many workers. The productivity would be the same in the two cases. Thus we see that if the factors were perfectly divisible the question of varying factor proportions would not have arisen and hence the phenomena of increasing and diminishing marginal returns to a factor would not have occurred. It has been rightly said, "Let divisibility enter through the door, law of variable proportions rushes out through the window."

Imperfect Substitutability of the Factors

Joan Robinson goes deeper into the causes of diminishing returns. She holds that *the diminishing returns occur because the factors of production are imperfect substitutes for one another.* As seen above, diminishing marginal returns occur after a point since the fixed factor becomes inadequate relatively to the variable factor. Now, a factor which is scarce in supply is taken as fixed. When there is a scarce factor, quantity of that factor cannot be increased in accordance with the varying quantities of the other factors which beyond the optimum proportion of factors will result in diminishing marginal returns. If now some variable factor was perfect substitute of the scarce fixed factor, then the paucity of the scarce fixed factor after a stage would have been made up by increasing the supply of this perfect substitute with the result that output could be expanded without diminishing marginal returns. Thus, even if the variable factor which we add to the fixed factor were perfect substitute of the fixed factor, then when the fixed factor becomes relatively deficient, its deficiency would be made up on account of the increase in the variable factor which is its perfect substitute.

Thus Joan Robinson says, "What the Law of Diminishing Returns really states is that there is limit to the extent to which one factor of production can be substituted for another, or, in other words, that the elasticity of substitution between factors is not infinite. If this were not true, it would be possible when one factor of production is fixed in amount and the rest are in perfectly elastic supply, to produce part of the output with the aid of the fixed factor, and then, when the optimum proportion between this and other factors was attained, to substitute some other factor for it and to increase output at constant cost." We therefore see that diminishing marginal returns operate because the elasticity of substitution between factors is not infinite.

Causes of Negative Marginal Returns

As the amount of the variable factor continues to be increased to the constant quantity of the other, a stage is reached when the total product declines and the marginal product becomes negative. This phenomenon of negative marginal returns to the variable factor is due to the fact that the amount of the variable factor becomes excessive relative to the fixed factor so that they get in each other's way with the result that the total output falls instead of rising. Besides, too large a number of the variable factor also impairs the efficiency of the fixed factor. The proverb

"too many cooks spoil the broth" aptly applies to this situation. In such a situation, a reduction in the units of the variable factor will increase the total output. Just as in the first stage, marginal product of the fixed factor was negative due to its abundance, in the third stage the marginal product of the variable factor is negative due to its excessiveness.

LAWS OF RETURNS

The term returns to scale refers to the degree by which output changes as a result of a given proportionate change in the amounts of all factors (inputs) used in production.

When both the inputs (labour and capital) are changed proportionately, the scale of production, *i.e.,* the size of the firm, changes. The laws of production, *i.e.* the input–output relationships under the condition of changing scale of production, are called the laws of returns to scale. The laws of returns to scale are a long-term phenomenon. In the long run, supply of both labour and capital is supposed to be elastic. The firms can therefore employ more of both labour and capital to increase their production.

Three Laws of Return to Scale

When both labour and capital are increased proportionately and simultaneously, there are technically three possible ways in which total output may increase:

1. Output may increase more than proportionately to increase in input,
2. Output may increase proportionately to increase in input and
3. Output may increase less than proportionately to increase in input.

For example, if both the inputs (labour and capital) are doubled, the resulting output may be more than double, equal to double or less than double. This kind of input–output relationship gives three kinds of laws of returns to scale:

1. The law of increasing returns to scale,
2. The law of constant returns to scale, and
3. The law of decreasing returns to scale.

1. The Law of Increasing Returns to Scale

When both the inputs—labour and capital—are increased proportionately and simultaneously and output increases more than proportionately, it gives the law of increasing returns to scale. The law of increasing returns to scale implies that output increases more than proportionately to the increase in inputs and the rate of increase in output goes on increasing with each subsequent increase in inputs. For example, suppose inputs are increased by 50 per cent and output increases by more than 50 per cent, say by 75 per cent, and when inputs are again increased again by 50 per cent and output increases by 100 per cent and so on. This kind of input–output relationship shows that the law of increasing returns to scale is in operation. This kind of returns to change in scale is illustrated in *Figure 12.9.* The three isoquants—Q_1, Q_2 and Q_3—represent three different levels of production—10, 25 and 50 units, respectively. Product lines *OA* and *OB* show the relationship between input and output. For instance, movement from point *a* to *b* denotes doubling the inputs, labour and capital. As *Figure 12.9* shows, input combination increases from 1K + 1L to 2K + 2L. The movement from *a* to *b* also indicates increase in output from 10 to 25 units. This means that when inputs are doubled, output is more than doubled.

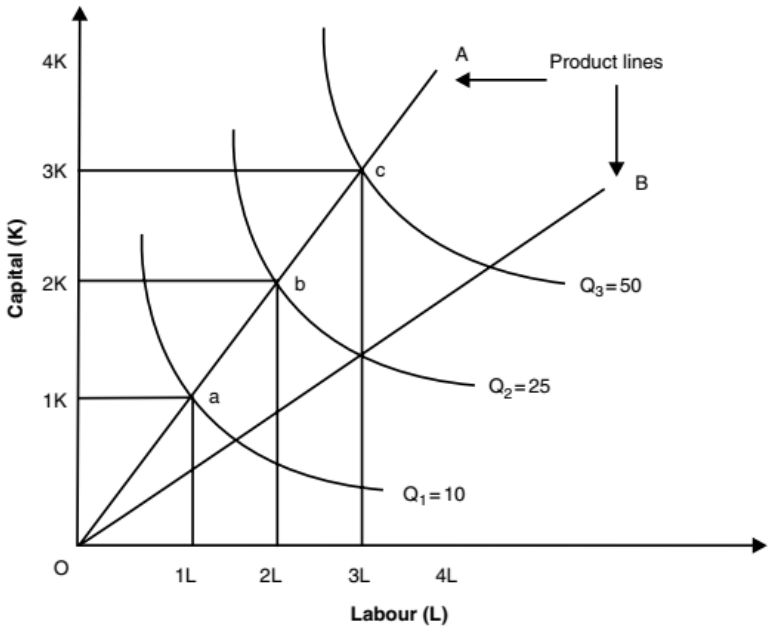


Figure 12.9 *Increasing Returns to Scale*

Similarly, movement from point *b* to *c* shows increase in inputs from $2K + 2L$ to $3K + 3L$, *i.e.,* a 50 per cent increase in inputs, and a rise in output from 25 to 50 units, i.e. a 100 per cent rise in output. This also gives a more than proportionate increase in the output in response to rise in inputs. This reveals the *law of increasing returns to scale*.

Factors Causing Increasing Returns to Scale:

The Economies of Scale The law of increasing returns to scale comes into operation because of economies of scale. There are at least three kinds of economies of scale that make plausible reasons for increasing returns to scale.

i. *Technical and Managerial Indivisibilities.*

Certain inputs, particularly machinery and managerial skills, used in the process of production are available in a given size. Such inputs are indivisible. That is, capital and managers cannot be divided into parts to suit the small scale of production. For example, half a turbine cannot be used; a part of a locomotive engine cannot be used; one third or a part of a composite harvester or earthmover cannot be used. Similarly, half of a manager cannot be employed, if part-time employment is not acceptable to him, and so on. Because of their indivisibility, such factors have to be employed in a minimum quantity even if scale of production is much less than their production capacity. Therefore, when scale of production is increased by increasing all inputs, the productivity of indivisible factors increases exponentially. This results in increasing returns to scale.

ii. *Higher Degree of Specialization.*

Another factor causing increasing returns to scale is higher degree of specialization of both labour and managerial manpower, which becomes possible with increase in the scale of production. The use of specialized labour and management increases productivity per unit of inputs. Their cumulative effects contribute to the increasing returns to scale. Managerial specialization contributes a great deal too increasing production.

iii. *Dimensional Relations.*

Increasing returns to scale is also a matter of dimensional relations. For example, when the size of a room (15' × 10' = 150 sq. ft.) is doubled to 30' × 20', the area of the room is more than doubled, i.e., 30' × 20' = 600 sq. ft. When diameter of a pipe is doubled, the flow of water is more than doubled. Following this dimensional relationship, when the labour and capital are doubled, the output is more than doubled over some level of output.

2. The Law of Constant Returns to Scale

When change in output is proportional to the change in inputs, it shows *constant returns to scale*. In other words, if quantities of both the inputs, *K* and *L*, are doubled and output is also doubled, then the returns to scale are constant. The constant returns to scale is illustrated in Figure 12.10. The lines *OA* and *OB* are product lines indicating two hypothetical techniques of production. The iso-quants, $Q_1 = 10$, $Q_2 = 20$ and $Q_3 = 30$ indicate three different levels of output. In the figure, the movement from point *a* to *b* indicates doubling both the inputs—capital increases from 1*K* to 2*K* and labour increases from 1*L* to 2*L*. When inputs are doubled, output is also doubled, *i.e.,* output increases from 10 to 20. The movement from point *b* to *c* indicates 50 per cent increase in the inputs, as *K* increases from 2*K* to 3*K* and *L* from *2L* to *3L*.

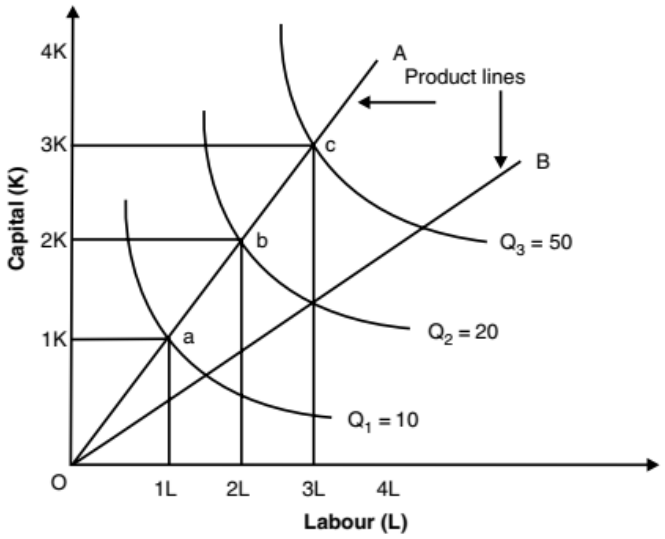


Figure 12.10 *Constant Returns to Scale*

As a result, output increases from 20 to 30, i.e., by 50 per cent. This relationship between the change in inputs and the proportionate change in output may be summed up of as follows.

$$1K+1L=Q=10$$

$$2K+1L=2Q=20$$

$$3K+2L=3Q=30$$

This kind of input–output relationship exhibits the constant returns to scale.

Reason for Constant Returns to Scale:

The constant returns to scale are attributed to the limits of the economies of scale. With the expansion in the scale of production, economies arise from such factors as indivisibility of certain inputs, greater possibility of specialization of capital and labour, use of labour-saving techniques of production and so on. But, there is a limit to the economies of scale. When economies of scale disappear and diseconomies are yet to begin, the returns to scale become constant. The diseconomies arise mainly because of decreasing efficiency of management and scarcity of certain inputs.

The constant returns to scale are said to occur also in productive activities in which factors of production are perfectly divisible. When the factors of production are perfectly divisible, the production function is homogenous of degree 1 like Cobb–Douglas production function.

3. **The Law of Decreasing Returns to Scale**

When output increases less than proportionately to increase in inputs, *K* and *L*, and the rate of rise in output goes on decreasing it is called decreasing returns to scale. Decreasing returns to scale are illustrated in *Figure 12.11.* As the figure shows, when inputs, *K* and *L*, are doubled, *i.e.,* inputs are increased from *1K+1L* to *2K+2L*, the output increases from 10 to 18 units, *i.e.,* 80 per cent increase, which is less than the proportionate increase in inputs. The movement from point b to c indicates a 50 per cent increase in the inputs. But, the output increases only by 33.3 per cent. This shows decreasing returns to scale.

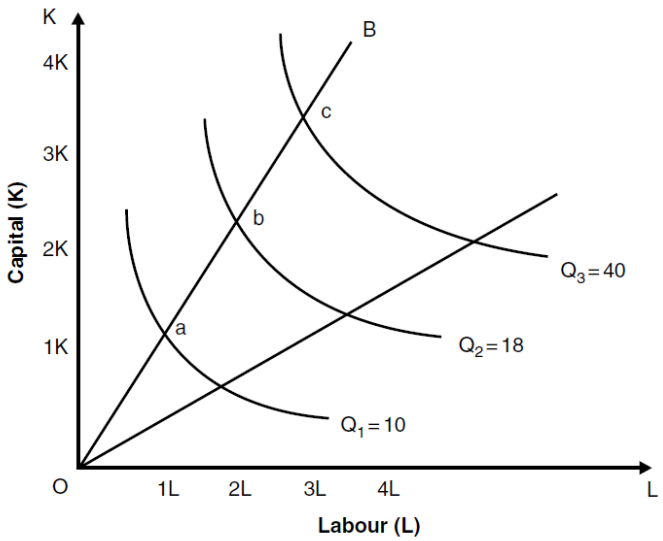


Figure 12.11 *Decreasing Returns to Scale*

Causes of Diminishing Returns to Scale

Decreasing returns to scale are caused by the diseconomies of scale. The most important factor causing diminishing returns to scale is 'the diminishing return to management', *i.e.,* due to managerial diseconomies. As the size of the firm expands, managerial efficiency decreases causing decrease in the rate of increase in output.

Another factor responsible for diminishing returns to scale is the limitedness or exhaustibility of the natural resources. For example, doubling the size of coal-mining plant may not double the coal output because of limitedness of coal deposits or difficult accessibility to coal deposits. Similarly, doubling the fishing fleet may not double the fish output because the availability of fish may decrease when fishing is carried out on an increasing scale.

ECONOMIES OF SCALE

The economies of scale are the cost reducing factors that arise due to the increase in the scale of production. Cost reducing factors arises both inside the firm and outside the firm. Accordingly, the economies of scale are classified under:

1. Internal Economies and
2. External Economies

Internal Economies

Internal economies are the economies that arise within the firm due to the expansion of its scale of production. In other words, internal economies are available exclusively to an expanding firm. Increasing scale of production may be in the form of expansion of the existing plant or adding more plants to the existing ones. In case production scale is expanded, there may be one product or production may be diversified. Internal economies are also called 'real economies'. The economies are 'real' in the sense that they arise out of increasing in productivity per unit of cost or decrease in cost of production caused by increase in the scale of production. Internal economies are classified under the following categories.

1. Economies in production,
2. Economies in marketing—buying inputs and selling outputs,
3. Managerial economies, and
4. Economies in transportation and storage cost.

Economies in Production

Economies of scale in production arise mainly from the increasing returns to scale resulting from the expansion of the production scale. The expansion of the production scale may be in the form of a proportionate or un-proportionate increase in the inputs. Production economies are of two kinds, *viz.,* technical economies and labour economies. Let us now look at technical and labour economies in some detail.

Technical Economies

Technical economies include the economies that arise due to the advantage of (i) opportunity for using specialized kind of machinery, (ii) indivisibility of specialized machinery forcing its optimum utilization, (iii) once-for-all cost of large-scale set-up, (iv) scope of building reserve capacity and (v) advantage of large-scale inventories. An expanding firm is in a position to enjoy these technical advantages by increasing its scale of production. Modern technology provides a specialized-capital equipment combining the entire process of producing a commodity. For example, given the modern technology, a large-scale cotton textile mill can use one composite and indivisible plant combining such units as: (i) spinning, (ii) weaving, (iii) printing and pressing and (iv) packaging. Likewise, a modern milk dairy plant combines: (i) milk processing, (ii) skimming and toning, (iii) chilling and (iv) bottling units. It gives increasing returns to scale as it saves both time and cost. Production by this kind of technology gives a higher productivity of capital per unit of time. A higher productivity of capital reduces unit cost of production compared to production in small scale. A small size firm cannot afford this kind of technology, nor can it have the technical economies of scale. A large-scale expanding firm can afford technically advanced plant and enjoy technical economies.

Labour Economies

Labour economies arise from the increase in labour productivity. Productivity of labour increases due to: (i) advantages of division of labour and (ii) specialization of labour and improved skill. Advantages of division of labour and specialization are described here briefly. When firm's scale of production expands, more and more workers of specialized skills and qualifications are employed. With the employment of larger number of workers, it becomes increasingly possible to divide the labour according to their qualification, skill and to place them in the process of production where they are best suited. This is known as division of labour. Division of labour leads to specialization. It increases efficiency and labour productivity which, in turn, reduces cost of production. Besides, specialized workers develop more efficient tools and techniques and gain speed of work. These advantages of division of labour improve productivity of labour per unit of cost and time.

Economies in Marketing—Buying Inputs and Selling Outputs

Economies in marketing arise from the large-scale purchase of raw materials and other material inputs and large-scale selling of firm's own products. As regards to economies in purchase of inputs, the large-size firms normally make bulk purchases of their inputs. The large sale purchase entitles the firm for certain discounts which are not available on the small purchases. As such, the growing firms gain economies on the cost of their material inputs.

The economies in marketing firm's own product arise due to: (i) economies in advertisement cost, (ii) economies in large-scale distribution through wholesalers, and so on and (iii) low managerial cost due to large-scale marketing. With the expansion of the firm, the total production increases. But the expenditure on advertising the product does not increase proportionately. Similarly, selling through the wholesale dealers reduces the cost on distribution of the firm's production. The large-scale firms also gain on large-scale distribution through better utilization of 'sales force, distribution of sample and so on.' This kind of economy, however, does not directly affect the production conditions.

Managerial Economies

Managerial economies arise from (i) specialization in different areas of management and (ii) mechanization of managerial functions. For a large-size firm, it becomes possible to divide its management into specialized departments under specialized personnel, such as, production manager, sales manager, personal manager, human resource managers and so on. This increases efficiency of management at all the levels because of decentralization of decision making. Large-size firms have the opportunity to use advanced techniques of communication, telephones and telex machines, computers, and their own means of transport. These factors lead to quick decision making, help in saving the valuable time of the management, and thereby improve the managerial efficiency. For these reasons, although managerial cost increases but less than proportionately to the increase in production scale, up to a certain level.

Economies in Transportation and Storage Cost

Economies in transportation and storage costs arise from full utilization of transport and storage facilities of the firm. Transportation costs are incurred on both production and sales sides. Similarly, storage costs are incurred on both raw materials and finished products. The large-size firms may acquire their own means of transportation and they can thereby reduce the unit cost of transportation compared to the market rate, at least to the extent of profit margin of the transport companies. Besides, own transport facility prevents the delays in transporting goods. Some large-scale firms have their own railway tracks from the nearest railway point to the factory, and thereby they reduce the cost of transporting goods in and out. For example, Bombay Port Trust has its own railway tracks and oil companies have their own fleet of tankers. Similarly, large-scale firms can build their own godowns in the various centres of product distribution and can save on storage cost.

External Economies

External economies are also called 'pecuniary economies'. External or pecuniary economies accrue to the expanding firms from the advantages due to conditions changing outside the firm. Pecuniary economies accrue to the large-size firms in the form of discounts and concessions on: (i) large scale purchase of raw material, (ii) large scale acquisition of external finance, particularly from the commercial bank, (iii) massive advertisement campaigns and (iv) large scale hiring of means of transport and warehouses and so on. These benefits are available to all the firms of an industry—they are not specific to any one particular firm. Besides, expansion of an industry invites and encourages the growth of ancillary industries which supply inputs and complementary parts. In the initial stages, such industries also enjoy the increasing returns to scale. In a competitive market, therefore, input prices go down. This benefit accrues to the expanding firms in addition to discounts and concessions. For example, growth of automobile industry helps the development of tyre industry and other motor parts industries. If Maruti Udyog Limited starts producing tyres for its cars and ancillaries, cost of Maruti cars may go up. Consider another example, growth of fishing industry encourages growth of firms that manufacture and supply fishing nets and boats. Competition between such firms and laws of increasing returns in the initial stages, reduce cost of inputs for the expanding firms. This is an important aspect of external economies.

ISO-QUANTS

The term 'iso-quants' has been derived from a Greek word 'iso' meaning *equal* and a Latin word 'quantus' meaning *quantity*. The 'iso-quants curve' is, therefore, also known as *equal product curve* and *production indifference curve*. By definition, *an iso-quants is locus of points representing different combinations of two inputs (labour and capital) yielding the same output*.

An iso-quants curve is analogous to consumer indifference curve with two differences:

1. while an indifference curve represents different combinations of two consumer goods yielding the same level of satisfaction, an *iso-quants curve* represents different combination of two producer goods (labour and capital) producing the same quantity of a commodity; and
2. while an indifference curve represents immeasurable 'utility', i.e., the level of satisfaction, an iso-quants represents a measurable quantity of output of a product.

The possibility that a given quantity of a commodity can be produced by different combinations of two inputs (labour and capital) is based on the assumption that a large variety of techniques of production is available. For example, a certain acre of wheat crop can be harvested per unit of time by 20 labour with 20 sickles (i.e., little of capital) or two labour and a harvesting machine (i.e., a large capital). Consider another example. A certain length of road can be constructed per unit of time by using 50 labours and 20 spades and levelling instrument or by using only five labour and a road roller, and so on. These technical possibilities are shown by an iso-quants curve.

Derivation of Iso-quants Curve

Assumptions

Iso-quants curves are drawn on the basis of the following assumptions:

1. There are only two inputs, labour (L) and capital (K), to produce a commodity, say X;
2. The two inputs (L and K) can be substituted for one another at a diminishing rate, up to a certain limit and

3. Production function is continuous, implying that labour and capital are perfectly divisible and can be substituted in any small quantity.

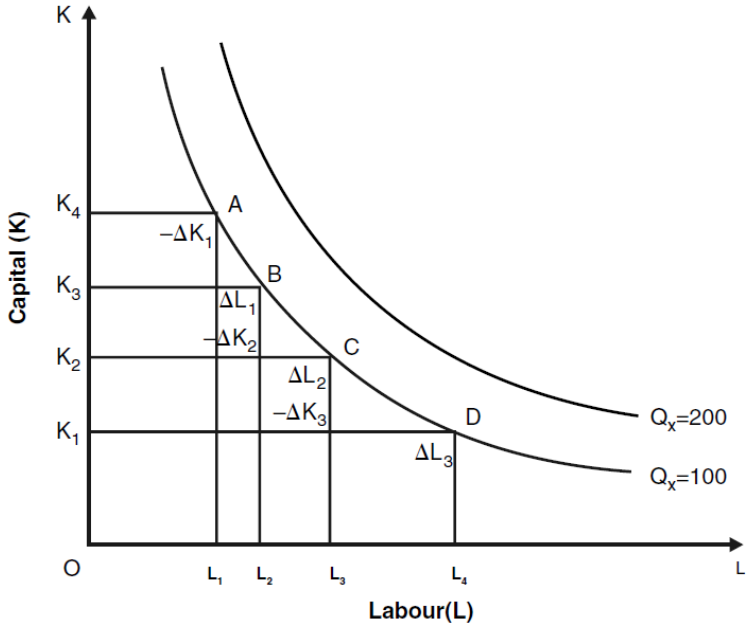


Figure 12.1 *Isoquant Curves*

Table 12.1 *Capital–Labour Combinations and Output*

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

Given these assumptions, it is always possible to produce a given quantity of commodity *X* with various combinations of capital and labour. The factor combinations are so formed that the substitution of one factor for the other leaves the output unaffected. This technological fact is presented through an iso-quants curve (Qx= 100) in Figure 12.1. The curve IQ_1 all along its length represents a fixed quantity, 100 units of product *X*. This quantity of output can be

produced with a number of labour–capital combinations. For example, points A, *B*, *C* and *D* on the iso-quants *Qx* represent four different combinations of inputs, *K* and *L*, as given in Table 12.1, all yielding the same output—100 units. Note that movement from *A* to *D* indicates decreasing quantity of *K* and increasing number of *L*. This implies substitution of labour for capital such that all the input combinations yield the same quantity of commodity *X*, i.e., *Qx* = 100, whatever the combination of labour and capital.

Properties of Iso-quants Curves

Iso-quants have the same *properties* as *indifference curves*. They are explained here in terms of input and output, under the condition that *two inputs are not perfect substitutes*.

1. *Iso-quants have a Negative Slope*

Iso-quants has a negative slope in the *economic region*1 or in the relevant range. The economic region is the region on the iso-quants plane in which substitution between inputs is technically efficient. It is also known as the *product maximizing region*. The negative slope of the iso-quants implies substitution of one input for another so that output remains the same. It means that if one of the inputs is reduced, the other input has to be increased that the total output remains unaffected. For example, movement from *A* to *B* on *Qx* (Figure 12.1) means that if K_4K_3 units of capital are removed from the production process, L_1L_2 units of labour have to be brought into maintain the same level of output. The substitution of one input for another gives iso-quants a negative slope.

2. *Iso-quants are Convex to the Origin*

Convexity of iso-quants means that it has a bend towards the point of origin. Iso-quants are convex to origin because the rate at which one input is substituted for the other goes on diminishing along their length. The rate at which inputs are substituted one for another at different levels is called the *marginal rate of technical substitutions*2 (*MRTS*). The *MRTS* is defined as:

$$MRTS = \frac{-\Delta K}{\Delta L} = Slope\ of\ Isoquants$$

In plain words, *MRTS* is the rate at which labour can substitute capital at margin, and *vice versa*, without affecting the total output. This rate is indicated by the slope of the iso-quants. The *MRTS* decreases for two reasons:

i. no factor is a perfect substitute for another and
ii. inputs are subject to diminishing marginal return.

It is for these reasons that, more and more units of an input are needed to replace each successive unit of the other input. This means diminishing marginal rate of substitution. That *MRTS* goes on diminishing along the iso-quants can be proved by deriving the *MRTS* from $IQx = 100$ in Figure 12.1. Suppose that in Figure 12.1, $X_4X_3 = X_3X_2 = X_2X_1$, it means that $\Delta K_1 = \Delta K_2 = \Delta K_3$.

But as the figure shows, the subsequent units of L substituting K go on increasing, i.e.,

$$L_1 L_2 < L_2 L_3 < L_3 L_4 \text{ or } \Delta L_1 < \Delta L_2 < \Delta L3$$

Let us workout the $MRTS = \Delta K/\Delta L$.

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

This shows that *MRTS* goes on decreasing.

Iso-quants do not intersect or Are Tangent to Each Other

The intersection or tangency between any two iso-quants implies two inconsistent production possibilities that:

i. the same combination of inputs can produce two different quantities of the same commodity and
ii. a given quantity of a commodity can be produced with a smaller as well as a larger input combination.

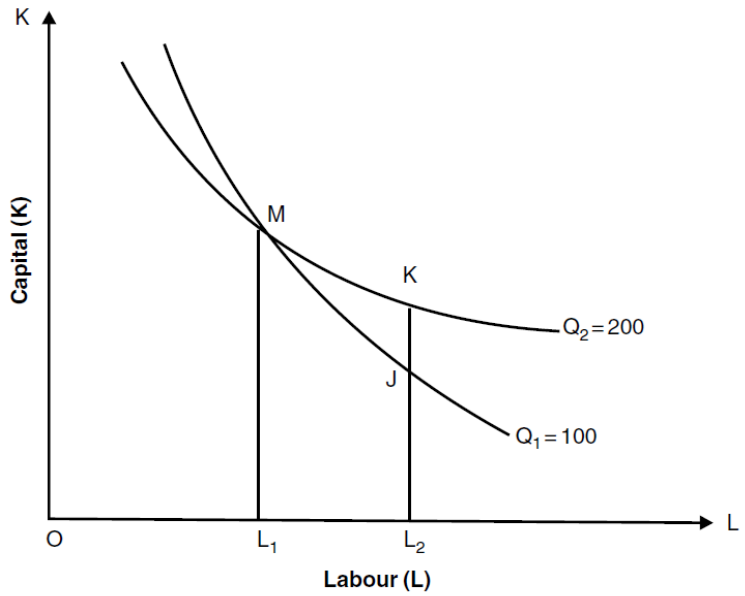


Figure 12.2 *Intersecting Isoquants*

These conditions contradict the laws of production unless marginal productivity of inputs is zero or less than zero. In Figure 12.2, two iso-quants intersect at point *M*. At point *M*, input combination is given as ML_1 of capital and OL_1 of labour. Since point *M* falls on both the iso-quants (Q_1 = 100 and Q_2 = 200), it means that the same combination of inputs can produce 100 and 200 units of the commodity. This is practically impossible, unless productivity of some input is equal to zero. To prove inconsistency, consider two other points—point *J* on iso-quants marked *Q*1 = 100 and point *K* on iso-quants marked *Q*2 = 200. One can easily infer that a quantity that can be produced with the combination of *K* and *L* at point *M* can be produced also with factor combination at points *J* and *K*. On the iso-quants *Q*1 = 100, factor combinations at point *M* and *J* are equal in terms of their output. On the iso-quants *Q*2 = 200, factor combinations at *M* and *K* are equal in terms of their output. Since point *M* is common to both the iso-quants, it follows that input combinations at *J* and *K* are equal in terms of output.

This implies that, in terms of output, $OL_2 + JL_2 = OL_2 + KL_2$. Since OL_2 is common to both the sides, it means that, in terms of output, JL_2 of $K = KL_2$ of *K*. But this cannot be possible because, as can be seen in Figure 12.2, $JL_2 < KL_2$. But the intersection of the iso-quants means that output from JL_2 and KL_2 units of capital are equal. This cannot happen as long as *MP* of capital is greater than zero. That is why iso-quants will not intersect or be tangent to one another. If they do, it violates the law of production.

Upper Iso-quants represent a Higher Level of Output

Between any two iso-quants, the upper one represents a higher level of output than the lower one. The reason is that any point on upper iso-quants implies a larger input combination, which, in general, produces a larger output. Therefore, upper iso-quants indicate a higher level of output. For instance, iso-quants Q_2 in Figure 12.3 will always represent a higher level of output than iso-quants Q_1. For, any point at iso-quants Q_2 consists of more of either capital or labour or both.

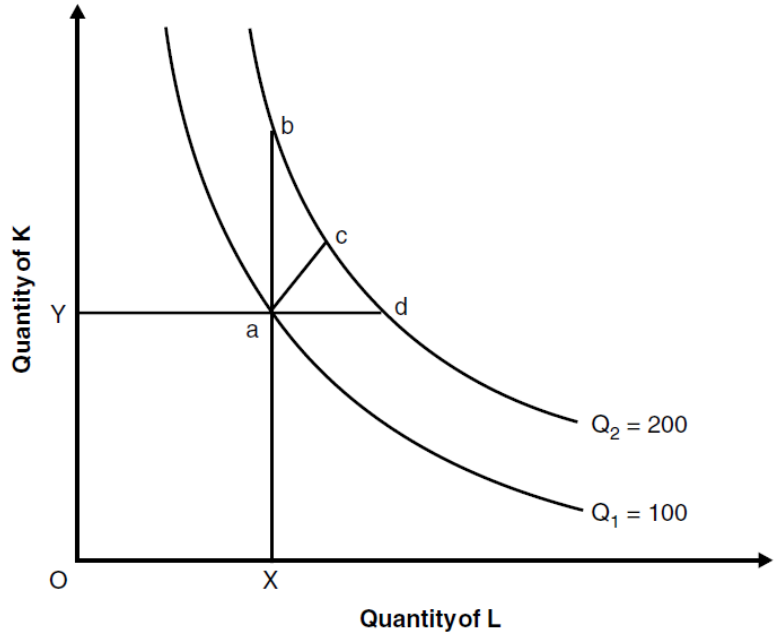


Figure 12.3 *Comparison of Out at Two Isoquants*

For example, consider point a on iso-quants Q_1 and compare it with any point at iso-quants Q_2. Point *b* on iso-quants Q_2 indicates more of capital *(ab),* point d more of labour *(ad)* and point *c* more of both. Therefore, iso-quants Q_2 represents a higher level of output (200 units).

Marginal Rate of Technical Substitution (MRTS)

A reference has already made to the marginal rate of technical substitution (MRTS) in the previous section. Before we proceed to discuss the laws of returns to scale, let us explain MRTS in some detail and look at the reason behind the decline in MRTS along iso-quants. As mentioned earlier, the marginal rate of technical substitution (MRTS) is the rate at which one input can be substituted for another, output remaining constant. The rate at which one input can

be substituted for another at the margin, holding the output constant, is given by the slope of the iso-quants. The slope of the iso-quants (moving downward) on iso-quants is given as follows:

$$MRTS = \frac{-\Delta K}{\Delta L} = Slope\ of\ Isoquants$$

The condition that the total output should remain constant implies that

$$(K\ .\ MP_K) = (L\ .MP_L)$$

where MP_K = marginal product of capital and MP_L = marginal product of labour.

By rearranging the terms in above Eq. we get

$$\frac{-\Delta K}{\Delta L} = \frac{MP_L}{MP_K}$$

Table 12.2 *Alternative Methods of Producing 100 Units of a Commodity*

Input Combination			Output	Changes in K and L		MRTS = –ΔK/ΔL
K	+	L		ΔK	ΔL	
10		2	100	–1	1	–1.0
9		3	100	–1	5	–0.2
8		8	100	–1	10	–0.1
7		18	100			

Since

$$\frac{-\Delta K}{\Delta L} = MRTS$$

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

Thus, *MRTS* of *L* for *K* is determined as the ratio of the marginal product of labour *(MPL)* to the marginal product of capital *(MP_K)*. To illustrate *MRTS* numerically, let us suppose that a given

production function may be presented in a tabular form as given in Table 12.2. The table presents five alternative combinations of K and L that can be used to produce a given quantity, say, 100 units of a commodity. The downward movement in the table shows substitution of labour for capital. As a result, the amount of capital decreases while the number of workers increases, output remaining constant. As the table shows, the units of labour which can substitute one unit of capital (or the quantity of capital that can substitute one unit of labour) goes on increasing. As a result, the $MRTS = -\Delta K/\Delta L$ goes on decreasing. The reason is that both the factors are subject to the laws of diminishing marginal return. As the number of labour increases, its marginal productivity decreases. On the other hand, with the decrease in the quantity of capital, its marginal productivity increases. Therefore, to substitute each subsequent unit of capital, more and more units of labour are required to maintain the production at the same level. That is why the *MRTS* decreases.

Iso-quants Map and Economic Region of Production

Iso-quants Map

An iso-quants map is a set of iso-quants presented on a two-dimensional plane as shown by iso-quants Q_1, Q_2, Q_3 and Q_4 in *Figure 12.4.* Each iso-quant shows various combinations of labour and capital that can be used to produce a given level of output. As shown in *Figure 12.3*, upper iso-quants is formed by a greater quantity of one or both the inputs than the input combination indicated by the lower iso-quants. For example, iso-quants *Q2* indicates a greater input combination than that shown by iso-quants Q_1, and so on.

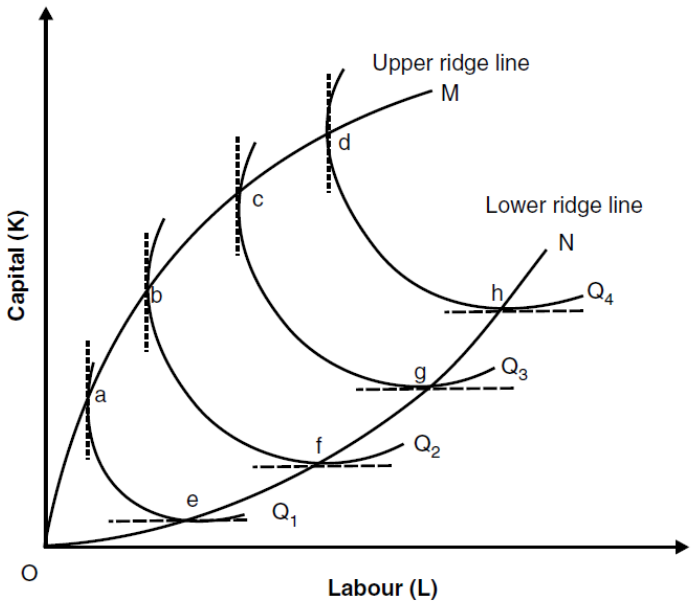


Figure 12.4 *Isoquant Map and Economic Region*

Since upper iso-quants indicate a larger input combination than the lower ones, each successive upper iso-quants indicates a higher level of output than the lower ones. For example, in Figure 12.4, if iso-quants Q_1 represents an output equal to 100 units, iso-quants Q_2 represents an output greater than 100 units, i.e., 200 units. As one of the properties of iso-quants, no two iso-quants can intersect or be tangent to one another.

Economic Region of Production

It is noteworthy that the whole iso-quants map or production plane is neither technically efficient nor every point on an iso-quants technically efficient. The reason is that, on convex iso-quants, the *MRTS* decreases along the iso-quants and zero is the limit to which the *MRTS* can decrease. The point at which *MRTS* equals zero marks the limit to which one input can substitute another. It also determines the minimum quantity of an input which must be used to produce a given output. Beyond this point, an additional employment of one input will necessitate employing additional units of the other input. Such a point on an is quant may be obtained by drawing a tangent to the iso-quants and parallel to the vertical and horizontal axes, as shown by dashed lines in Figure 12.4. By joining the resulting points *a, b, c* and *d*, we get a line called the *upper ridge line*, *OM*. Similarly, by joining the points *e, f, g* and *h*, we get the *lower ridge line, ON*. The

ridge lines are locus of points on the iso-quants where the marginal products of the inputs are equal to zero. The upper ridge line implies that marginal productivity of capital is zero along the line, *OM*. The lower ridge line implies that marginal productivity of labour is zero along the line, *ON*. The area between the two ridge lines, *OM* and *ON*, is called '*economic region*' or '*technically efficient region*' of production. Any production technique, i.e., capital–labour combination, within the economic region is technically efficient to produce a given output. And, any production technique outside this region is technically inefficient, since it requires more of both the inputs to produce the same quantity.

For example, suppose that the quantity represented by iso-quants Q_2 is to be produced, we have two points, *b* and *f*, on the iso-quants Q_2, which fall on the ridge lines. Consider first point *b*, i.e., the point of intersection between the iso-quants Q_2 and the upper ridge line. Point *b* indicates minimum of labour and maximum of capital required to produce Q_2. A smaller amount of capital, given the labour input at point *b*, would be insufficient to produce Q_2. Beyond point *b*, producing Q_2 would require more of capital and labour, which is technically inefficient. It would mean producing the same quantity with a larger input combination. It may be inferred from the above that: (i) at point *b*, marginal productivity of capital is zero and (ii) further substitution of capital for labour is technically inefficient. Similarly, point *f* indicates minimum of capital and maximum of labour required to produce Q_2. Any smaller input of labour would be insufficient to produce Q_2. Any further substitution of labour for capital is technically inefficient, since *MP* of labour at point *f* is zero. Any addition to the quantity of capital would not yield any additional output. Capital is therefore redundant. Even if both labour and capital is used, the level of output will not increase. These limits determine the *economic religion*.

PRODUCTION EQUILIBRIUM

Producer's equilibrium or optimization occurs when producer earns maximum profit with optimal combination of factor inputs. A profit maximization firm faces two choices of optimal combination of factors inputs.

i. To minimize its cost for a given output; and
ii. To maximize its output for a given cost.

Cost-minimization for a given output:

In the theory of production, the profit maximization firm is in equilibrium when, given the cost-price function, it maximizes its profits on the basis of the least cost combination of factors. For this, it will choose that combination which minimizes its cost of production for a given output. This will be the optimal combination for it.

Assumptions

i. There are two factors, *labour* and *capital*.
ii. All units of labour and capital are homogeneous.
iii. The prices of units of labour *(w)* and that of capital *(r)* are given and constant.
iv. The cost outlay is given.
v. The firm produces a single product.
vi. The price of the product is given and constant.
vii. The firm aims at profit maximization.
viii. There is perfect competition in the factor market.

Explanation

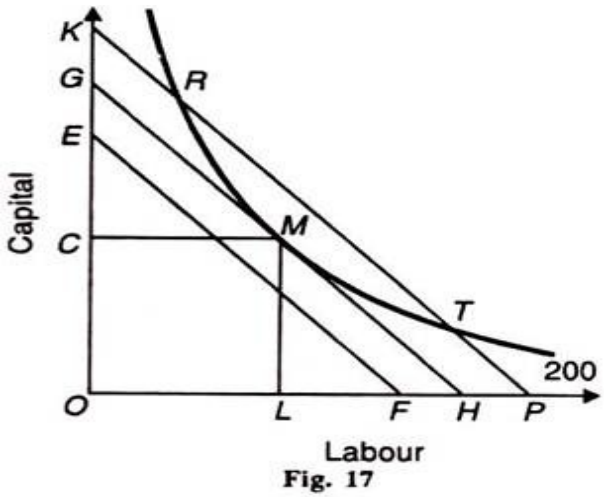


Fig. 17

Given these assumptions, the point of least-cost combination of factors for a given level of output is where the Iso-quant curve is tangent to an Iso-cost line. In *Figure 17*, the Iso-cost line *GH* is tangent to the Iso-quant 200 at point *M*. The firm employs the combination of *OC* of

capital and *OL* of labour to produce 200 units of output at point M with the given cost-outlay *GH*. At this point, the firm is minimizing its cost for producing 200 units.

Any other combination on the Iso-quant 200, such as *R* or *T*, is on the higher Iso-cost line *KP* which shows higher cost of production. The Iso-cost line *EF* shows lower cost but output 200 cannot be attained with it. Therefore, the firm will choose the minimum cost point *M* which is the least-cost factor combination for producing 200 units of output. *M* is thus the optimal combination for the firm. The point of tangency between the Iso-cost line and the Iso-quant is an important first order condition but not a necessary condition for the producer's equilibrium.

There are two essential or second order conditions for the equilibrium of the firm:

The first condition is that the slope of the Iso-cost line must equal the slope of the Iso-quant curve. The slope of the Iso-cost line is equal to the ratio of the price of labour *(w)* to the price of capital *(r) i.e... W/r.* The slope of the Iso-quant curve is equal to the marginal rate of technical substitution of labour and capital *($MRTS_{LC}$)* which is, in turn, equal to the ratio of the marginal product of labour to the marginal product of capital *(MPL/MPC).* Thus the equilibrium condition for optimality can be written as: $W/r = MP_L/MP_C = MRTS_{LC}$.

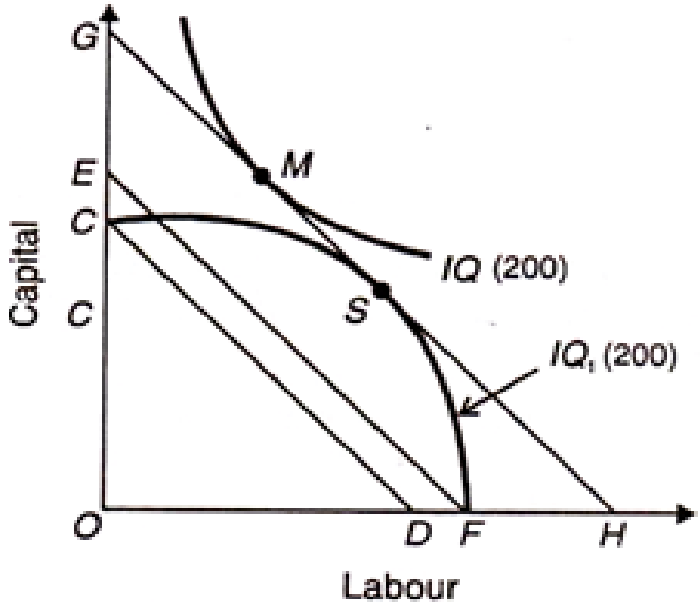


Fig. 18

The second condition is that at the point of tangency, the Iso-quant curve must he convex to the origin. In other words, the marginal rate of technical substitution of labour for capital *(MRTS$_{LC}$)* must be diminishing at the point of tangency for equilibrium to be stable. In Figure 18, *S* cannot be the point of equilibrium, for the Iso-quant *IQ$_1$* is concave where it is tangent to the Iso-cost line *GH*.

At point *S*, the marginal rate of technical. substitution between the two factors increases if move to the right or left on the curve *lQ$_1$* .Moreover, the same output level can be produced at a lower cost *CD* or *EF* and there will be a corner solution either at *C* or *F*. If it decides to produce at *EF* cost, it can produce the entire output with only *OF* labour. If, on the other hand, it decides to produce at a still lower cost *CD*, the entire output can be produced with only *OC* capital.

Both the situations are impossibilities because nothing can be produced either with only labour or only capital. Therefore, the firm can produce the same level of output at point *M* where the Iso-quant curve *IQ* is convex to the origin and is tangent to the Iso-cost line *GH*. The analysis assumes that both the Iso-quants represent equal level of output *IQ = IQ$_1$ = 200.*

Output-maximization for a given cost:

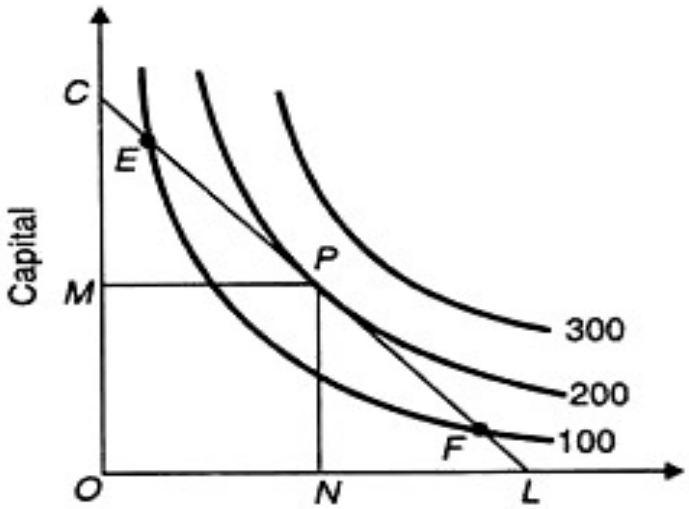


Fig. 19

The firm also maximizes its profits by maximizing its output, given its cost outlay and the prices of the two factors. This analysis is based on the same assumptions, as given. The conditions for the equilibrium of the firm are the same, as discussed.

i. The firm is in equilibrium at point *P* where the Iso-quant curve 200 is tangent to the Iso-cost line *CL* in Figure 19.

At this point, the firm is maximizing its output level of 200 units by employing the optimal combination of *OM* of capital and *ON* of labour, given its cost outlay *CL*. But it cannot be at points *E* or *F* on the Iso-cost line *CL*, since both points give a smaller quantity of output, being on the Iso-quant 100, than on the Iso-quant 200.

The firm can reach the optimal factor combination level of maximum output by moving along the Iso-cost line CL from either point E or F to point P. This movement involves no extra cost because the firm remains on the same Iso-cost line.

The firm cannot attain a higher level of output such as Iso-quant 300 because of the cost constraint. Thus the equilibrium point has to be P with optimal factor combination *OM+ON*. At point P, the slope of the Iso-quant curve 200 is equal to the slope of the Iso-cost line CL. It implies that $w/r = MP_L/MPC = MRTS_{LC}$

ii. The second condition is that the Iso-quant curve must be convex to the origin at the point of tangency with the Iso-cost line, as explained above in terms of Figure 18.

COST

The term 'cost' has different meaning. Accountants view of cost is different from that of economists. The accountants tend to focus on the explicit and historical costs. On the other hand, economists emphasize that for efficient decision making by the firm it is the opportunity cost rather than explicit and historical cost that must be considered. And, as will be explained below, the opportunity cost concept includes both the explicit and implicit costs. It is therefore necessary that we should explain the meaning of cost as used by economists and which is relevant for decision-making by a producer.

Opportunity Cost

The concept of opportunity cost is a basic concept of cost in economics. As the resources of the society are scarce, we cannot produce all goods that the people may desire. Therefore, when more resources are allocated to one product, some other product has to be foregone or sacrificed.

The opportunity cost of a product is therefore the value of the next best alternative product that is forgone so as to release resources for greater production of the former.

The knowledge of opportunity cost is essential for decision making by a firm as it is also faced with the problem of constrained optimization. It works with a budget constraint. Whenever more resources are allocated by it to one division or department they have to be withdrawn from another and therefore the opportunity cost is involved. For example, if some resources are allocated to Research and Development (R & D) Division, it may result in withdrawal of resources from Production Department which means reduction in the current output of the product, though allocation of more resources to R & D may result in more and better products in future yielding more profits to the firm in future. Of course, such a decision should be made if the manager of the firm is quite sure that greater profits in future will outweigh the loss of output and profits in the current period.

However, in the theory of costs we are more concerned with opportunity cost of a factor of production used in the production of a commodity. In this sense, the opportunity cost of an input or factor of production is its value (i.e. earnings) in the next best alternative use or employment. The firm must consider the opportunity cost of all inputs and factors of production. If a firm pays lower price for a factor than it can get in any other firm or use, it will not be able to retain it. That is, actual payment to an input or a factor of production which is called explicit cost must not be less than its opportunity cost, otherwise it will not be able to retain it.

Similarly, it will not be worthwhile for the firm to use a self-owned factor in its business if it can get a higher price by selling it to others, or hiring it out to others. For example, if an entrepreneur has set up his firm and manages his own business, he foregoes the salary that he could have earned for working as a manager in another's firm. Similarly, if an entrepreneur invests his own financial capital in his firm, he foregoes interest or return that he could earn by investing that in fixed deposits in banks or purchasing shares of companies. The interest which he could earn on fixed deposits or return which he could get by making investment of his financial capital in shares of other companies is the opportunity cost of financial capital which he invests in his own business. The salary which an entrepreneur gets elsewhere by working as a manager and return on his own financial capital which he can get by investing elsewhere are the implicit costs which are not generally considered by accountants in their financial statement of the firm must be taken

into account in rational decision making by management of the firm. It follows from above that economic opportunity costs include both explicit and implicit costs and therefore both should be taken into account in rational decision making by management. Thus, both the explicit and implicit costs are relevant and therefore must be considered while making decision regarding whether to produce a product or use inputs for its production. The failure to consider the opportunity cost and rely on historical cost will lead to inefficient decision. An example will bring out the importance of the opportunity cost in making optimal managerial decisions. Let us consider the problem of inventory evaluation. Suppose a firm purchased a raw material at price of ` 1000 per quintal. The accountant would continue to consider the value of the inventory of raw material at its historical cost, namely, `1000 per quintal. However, the economists would consider the value of inventory of raw material at its current price at which it can be presently replaced. If a firm considers the historical cost of inventories of raw materials rather than their current replacement value, it may take a wrong decision. For example, on the basis of higher historical cost of the raw material it may decide not to produce the commodity in whose production it is used because it would involve losses. But if the raw material is valued at its current replacement price of Rs. 600 per quintal it may be profitable to produce the commodity.

The resources which are used for the manufacture of armaments may also be used for the production of cars or other automobiles. Therefore, the opportunity cost of production of an armament is the output of cars and other automobiles foregone or sacrificed, which could have been produced with the same amount of factors that have gone into the making of an armament. To take another example, a farmer who is producing wheat, can also produce potatoes with the same factors. Therefore, the opportunity cost of a quintal of wheat is the amount of output of potatoes given up. Thus, the opportunity-cost of anything is the next best alternative that could be produced instead by the same resources having the same monetary value.

The alternative or opportunity cost of a good can be given a money value. In order to produce a good the producer has to employ various factors of production and has to pay them sufficient prices to get their services. These factors have alternative uses. The factors must be paid at least the price they are able to obtain in the alternative uses. The opportunity cost of a factor to a firm is its earnings in the next best alternative use. For instance, consider the case of a farmer who works on his own farm. While calculating cost he must take into

account the wages of his own labour rendered on his farm. These wages must be equal to the maximum wages that he would earn if he worked as hired labour on other farms. The maximum wages earned elsewhere by a farmer is the opportunity cost of his labour rendered on his own farm. The total alternative earnings in the next best use of the various factors, both supplied by others to the entrepreneur and those supplied by the entrepreneur himself employed in the production of a good will constitute the opportunity cost of the good.

Real Costs

It is a philosophical concept which refers to all those efforts and sacrifices undergone by various members of the society to produce a commodity. Like monetary costs, real costs do not tell us anything what lies behind these costs. *Prof. Marshall* has called these costs as the "Social Costs of Production."

According to Marshall, "Real costs are the exertion of all the different kinds of labour that are directly or indirectly involved in making it together with the abstinence rather than the waiting required for saving the capital used in making it, all these efforts and sacrifices together will be called the real cost of production of the commodity."

In this way, real cost means the trouble, sacrifice of factors in producing a commodity. Though, this concept gained momentum for sometime it has been relegated to the background in modern times due to its impracticability.

Short-run and Long-run Costs

Two other important cost concepts associated with variable and fixed cost concepts that often figure in economic analysis of cost behaviour are short-run and long-run costs. Short-run costs include fixed cost and the variable cost, i.e., the costs which vary with the variation in output, the size of the firm remaining the same. Long-run costs are the costs incurred in the long run. In the long run, there is no fixed cost. All the costs are variable cost. It implies that even the costs incurred on fixed assets, like plant, building, machinery and so on become the variable costs.

In other words, in the long run, even the fixed costs become variable costs. Firms can hire more of all the inputs if they decide to increase the size of the firm or scale of production. Broadly speaking, 'the short-run cost are those associated with variable costs in the utilization of fixed

plant or other facilities, whereas long-run cost-behaviour encompasses changes in the size and kind of plant'.

Total, Average, and Marginal Costs

Total cost (TC) represents the cost of the total resources used in the production of goods and services. If refers to the total outlays of money expenditure, both explicit and implicit, on the resources used to produce a given output. The total cost for a given output is obtained from the cost function. Average cost (AC) is of statistical nature, rather than being an actual cost. It is obtained simply by dividing the total cost (TC) by the total output (Q), i.e., TC/Q = average cost. Marginal cost (MC) is the addition to the total cost on account of producing one additional unit of product. Or, marginal cost is the cost of marginal unit produced. Total, average and marginal cost concepts, used in the economic analysis of the firm's productive activities, are discussed in detail in the following section.

Fixed and Variable Costs

Fixed costs are the costs which are fixed in amount for a certain level of output. Fixed costs do not vary with the variation in the output between zero and a certain level or output. The costs that do not vary over a certain level of output are known as fixed cost. Fixed cost includes cost of (i) managerial and administrative staff; (ii) depreciation of machinery, building and other fixed assets and (iii) maintenance of land. The concept of fixed cost is associated with short run. Variable costs are those which vary with the variation in the total output. Variable costs are functions of the output. Variable costs include direct labour cost, cost of raw materials, and running cost of fixed capital, such as fuel, ordinary repairs, routine maintenance expenditure and the costs of all other inputs that vary with output.

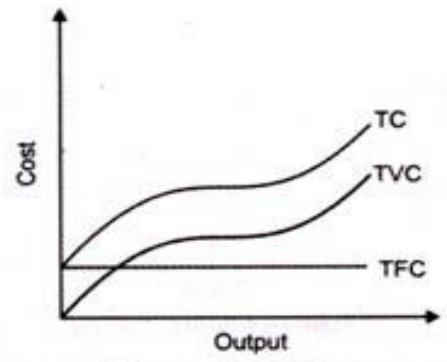


Figure-5: TC Curve

It should be noted that both *TVC* and *TC* increase initially at decreasing rate and then they increase at increasing rate Here, decreasing rate implies that the rate at which cost increases with respect to output is less, whereas increasing rate implies the rate at which cost increases with respect to output is more.

LONG-RUN COST CURVE

In the long run, all the factors of production used by an organization vary. The existing size of the plant or building can be increased in case of long run.

There are no fixed inputs or costs in the long run. Long run is a period in which all the costs change as all the factors of production are variable.

There is no distinction between the *Long-run Total Costs (LTC)* and *Long-run Variable Cost (LVC)* as there are no fixed costs. It should be noted that the ability of an organization of changing inputs enables it to produce at lower cost in the long run.

Cost concepts that are taken into consideration in the long-run:

i. Long-run Total Cost *(LTC)*
ii. Long-run Average Cost *(LAC)*
iii. Long-run Marginal Cost *(LMC)*

i. *Long-run Total Cost (LTC)*

Long-run Total Cost *(LTC)* refers to the minimum cost at which given level of output can be produced. According to *Leibhafasky*, "the long run total cost of production is the least possible cost of producing any given level of output when all inputs are variable." *LTC* represents the least cost of different quantities of output, it is always less than or equal to short run total cost, but it is never more than short run cost.

The LTC curve

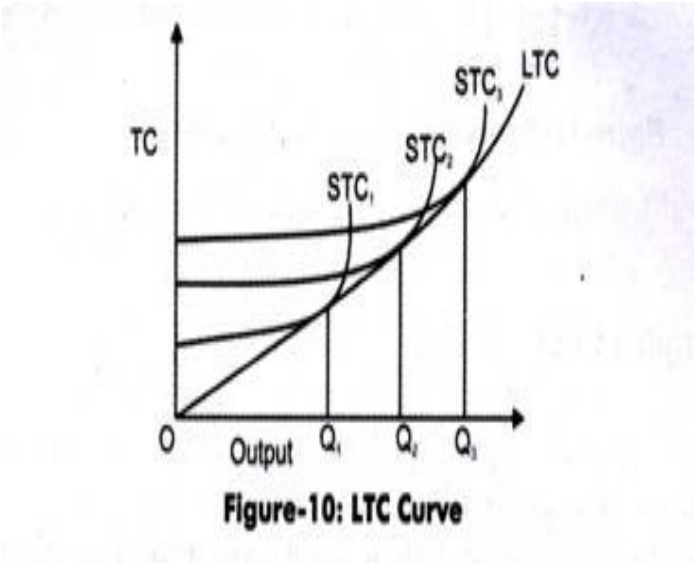


Figure-10: LTC Curve

As shown in *Figure-10*, short run total costs curves; STC_1, STC_2, and STC_3 are shown depicting different plant sizes. The *LTC* curve is made by joining the minimum points of short-run total cost curves. Therefore, *LTC* covers the *STC* curves.

ii. *Long-run Average Cost (LAC)*

Long-run Average Cost *(LAC)* is equal to long-run total costs divided by the level of output. The derivation of long-run average costs is done from the short-run average cost curves. In the short-run, plant is fixed and each short-run cost curve corresponds to a particular plant. The long-run average costs curve is also called *planning curve* or *envelope curve* as it helps in making organizational plans for expanding production and achieving minimum cost.

Derivation of *LAC* curve

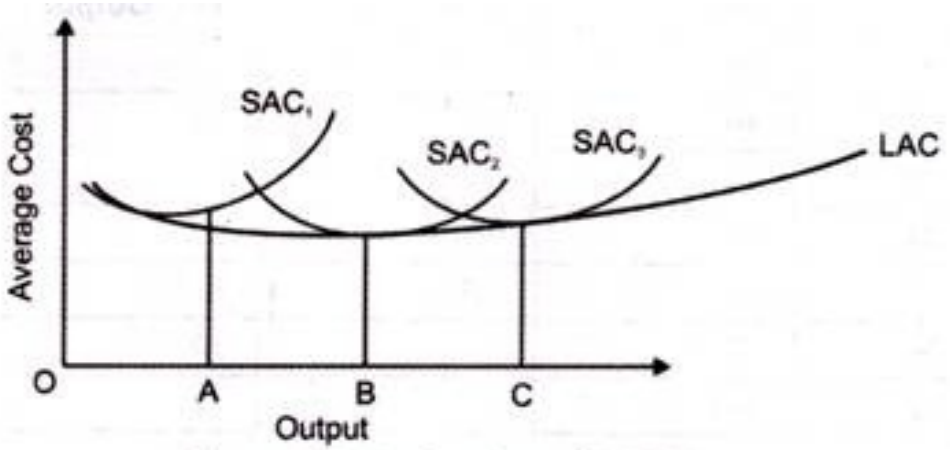


Figure-11: Derivation of LAC Curve

Suppose there are three sizes of the plant and no other size of the plant can be built. In short run, the plant sizes are fixed thus, organization increase or decrease the variable factors. However, in the long run, the organization can select among the plants which help in achieving minimum possible cost at a given level of output.

From Figure-11, it can be noted that till OB amount of production, it is beneficial for the organization to operate on the plant SAC^2 as it entails lower costs than SAC^1. If the plant SAC^2 is used for producing OA, then cost incurred would be more. Thus, in the long run, it is clear that the producer would produce till OB on plant SAC^2. On SAC^2, the producer would produce till OC amount of output. If an organization wants to exceed output from OC, it will be beneficial to produce at SAC^3 than SAC^2. Thus, in the long run, an organization has a choice to use the plant incurring minimum costs at a given output. *LAC* depicts the lowest possible average cost for producing different levels of output. The *LAC* curve is derived from joining the lowest minimum costs of the short run average cost curves.

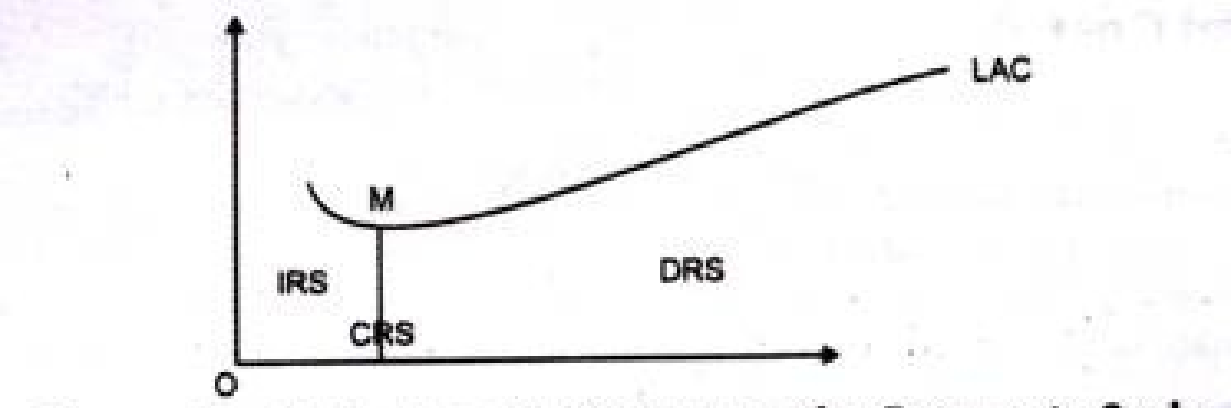


Figure-12: Derivation of LAC curve under Returns to Scale

It first falls and then rises, thus it is *U-shaped* curve. The returns to scale also affect the *LTC* and *LAC*. Returns to scale implies a change in output of an organization with a change in inputs. In the long run, the output changes with respect to change in all inputs of production.

In case of increasing returns to scale *(IRS),* organizations can double the output by using less than twice of inputs. *LTC* increases less than the increase in the output, thus, *LAC* falls. In case of constant returns to scale *(CRS),* organizations can double the output by using inputs twice.

LTC increases proportionately to the output; therefore, *LAC* becomes constant. On the other hand, in case of decreasing returns to scale *(DRS),* organizations can double the output by using inputs more than twice. Thus, *LTC* increases more than the increase in output. As a result, *LAC* increases.

As shown in Figure-12, up to *M*, *LAC* slopes downward. This is because at this stage *IRS* is applied. On the other hand, at *M*, *LAC* becomes constant. After *M*, *LAC* slopes upwards implying *DRS*.

Long-run Marginal Cost (LMC)

Long run Marginal Cost *(LMC)* is defined as added cost of producing an additional unit of a commodity when all inputs are variable. This cost is derived from short-run marginal cost. On the graph, the *LMC* is derived from the points of tangency between *LAC* and *SAC*.

LMC curve

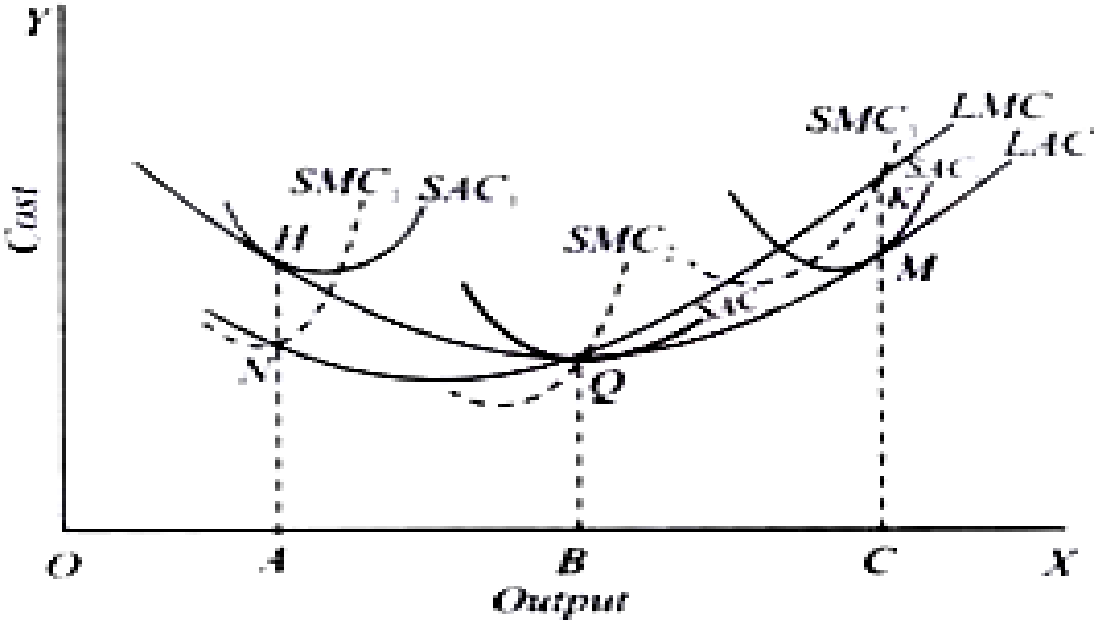
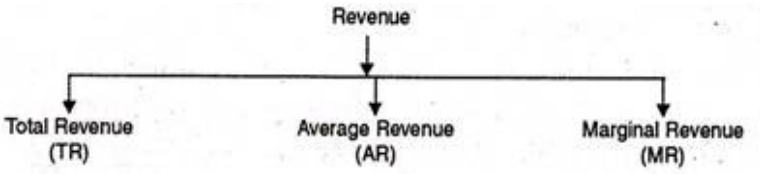


Fig.19.12 . *Derivation of Long-Run Marginal Cost Curve*

If perpendiculars are drawn from point *A, B,* and *C,* respectively; then they would intersect *SMC* curves at *H, Q,* and *M* respectively. By joining *N, Q,* and *K,* the *LMC* curve would be drawn. It should be noted that *LMC* equals to *SMC,* when *LMC* is tangent to the *LAC.*In Figure-13, OB is the output at which: *SAC2* = *SMC2* = *LAC* = *LMC.* The relation between *LMC* and *LAC* as follows: When *LMC* < *LAC, LAC* falls, When *LMC* = *LAC, LAC* is constant, When LMC > LAC, LAC rises.

REVENUE

The term revenue refers to the income obtained by a firm through the sale of goods at different prices. In the words of Dooley, *'the revenue of a firm is its sales, receipts or income'*. The revenue concepts are concerned with Total Revenue (TR), Average Revenue (AR) and Marginal Revenue (MR).



Total Revenue (TR)

The income earned by a seller or producer after selling the output is called the total revenue. In fact, total revenue is the multiple of price and output. The behavior of total revenue depends on the market where the firm produces or sells.

"Total revenue is the sum of all sales, receipts or income of a firm." —*Dooley*

Total revenue may be defined as the "product of planned sales (output) and expected selling price." — *Clower and Due*

"Total revenue at any output is equal to price per unit multiplied by quantity sold." — *Stonier and Hague*

Thus,

$$TR = AP. \times Q$$

where

TR = Total Revenue

AR = Average Revenue or Price per Unit

Q = Output

For example if the price of a commodity is Rs. 100 and total units sold are 20 in that case total revenue will be

$$TR = 100 \times 20 = 200$$

$$TR = 2000$$

Average Revenue (AR)

Average revenue refers to the revenue obtained by the seller by selling the per unit commodity. It is obtained by dividing the total revenue by total output.

"The average revenue curve shows that the price of the firm's product is the same at each level of output." — *Stonier and Hague*

Thus :

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

where AR = Average Revenue

TR = Total Revenue

Q = Output

According to **McDonnell**, "Average Revenue is the per unit revenue received from the sale of one unit of a commodity."

$$TR = \text{Price} \times \text{Output}$$

$$TR = Pq$$

$$AR = \frac{Pq}{q} = P$$

and $P = f(Q)$ is an average curve which shows that price is a function of quantity demanded. It is also a demand curve.

Marginal Revenue (MR)

Marginal revenue is the net revenue obtained by selling an additional unit of the commodity.

"Marginal revenue is the change in total revenue which results from the sale of one more or one less unit of output." — *Ferguson*. Thus, marginal revenue is the addition made to the total revenue by selling one more unit of the good.

In algebraic terms, marginal revenue is the net addition to the total revenue by selling *n* units of a commodity instead of *n – 1*.

Therefore,

$$MR = \frac{\Delta TR}{\Delta Q}$$

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

Whereas

TR_n = Total Revenue of 'n' units

TR_{n-1} = Total Revenue from $(n - 1)$ units

$MR_{(nth)}$ = Marginal revenue from nth unit

n = Any given number

"The marginal revenue is the change in total revenue resulting from selling an additional unit of the commodity." — *A. Koutsoyiannis*

If total revenue from *(n)* units is 110 and from *(n–1)* units is 100, in that case

$$MR_{nth} = TR_n - TR_{n\_1} = 100 - 110$$

$$MR_{nth} = 10$$

MR in mathematical terms is the ratio of change in total revenue to change in output

$MR = \Delta TR/\Delta q$ *or* $dR/dq = MR$

UNIT –IV: PRODUCT PRICING

Concepts of firms, industry, equilibrium

In economics, a firm holds important position as at the firm level managerial decisions are taken. In common language a firm is considered as a manufacturing unit involved in production of goods. The scope of the term firm in economics is broad. It represents any business organization inhering service & agriculture organization also.

Definition of Firm

Some definitions of firm given by renowned economists are given below.

i. Firm is a unit of production that employs factors of production (or inputs) to produce goods & services under given state of technology.
ii. It is an independently administered business unit – Hanson.
iii. It is a center of control where the decisions about what to produce & how to produce are taken.
iv. It is a business unit which hires productive resources for the purpose of producing goods & services.
v. A firm is an independent organization whose destiny is determined by the magnitude of the aggregate pay off & in which the aggregate pay off production & sale of goods or services (Harvey Leibenstein).

All these definitions have evolved during different time periods. They try to emphasis the different economic problems faced by firms. From various definitions different characteristic features of firm emerge which are listed below.

1. It is a place where all decision making related to production are taken viz what, where & how much to produce.
2. It is a place where manpower is hired for production.
3. It is a place where all the resources of production are brought together, production is done as well as sale & distribution of the manufactured product is carried out.
4. The state of technology is defined by the firms production function.

A firm is required to carry out all diverse functions related to production & marketing at the same time. While earning profit, firm as a production unit tries to manufacture the goods or provide service as per the consumer demand. The main objective of any firm is to maximize profit. Traditionally it was assumed that firm tries to maximize profit in each time period. But now it is realized that firm's objective should be to maximize profit in long run irrespective of profit or loss in short run.

Industry

Industry is a group of related firms. The relationship between the firms may be either based upon product or process criterion, e.g. dairy industry or food processing industry etc. The concept of industry is helpful to government and businessmen to formulate their policies.

Types of industry

The activities which are undertaken to produce, convert, extract and fabricate raw materials into finished goods are termed as industries. It is the process where goods are made usable and consumable. There are four different types of industries. These are:

a) **Genetic Industry:** It involves activities in reproducing and multiplying certain species of plants and animals for the sake of earning profit from their sale. Fish culture, cattle breeding, goatery and piggery are included in genetic industries.

b) **Extractive Industry:** The industries engaged with the discovery or extracting natural resources like minerals soil, water and forests are called extractive industries. Mining, agriculture and fishing are best examples of extractive industries.

c) **Manufacturing Industries:** The industries engaged in the conversion of raw material into finished products are called manufacturing industries. Cotton textile, sugar, iron and steel are the best examples of manufacturing industries.

d) **Construction Industry:** The industries in the construction of infrastructure like building, dams, roads, bridges and canals are called construction industries

Equilibrium of Firm and Industry

The word equilibrium has been taken from science. It is a state of no change where opposite forces become equal. The consumer is in equilibrium when he is getting maximum satisfaction from his income.

Similarly, for an industry, equilibrium refers to a situation when there is no tendency for new firms to enter or exit. Now, the question arises, under what conditions such equilibrium situations will be achieved.

Equilibrium of Firm:

"A firm is a unit engaged in the production for sale at a profit and with the objective of maximizing profit." -Watson

A firm is in equilibrium when it is satisfied with its existing level of output. The firm wills, in this situation produce the level of output which brings in greatest profit or smallest loss. When this situation is reached, the firm is said to be in equilibrium.

"Where profits are maximized, we say the firm is in equilibrium". -Prof. RA. Bilas

"The individual firm will be in equilibrium with respect to output at the point of maximum net returns." -Prof. Meyers

Conditions of the Equilibrium of Firm:

A firm is said to be in equilibrium when it satisfies the following conditions:

i. The first condition for the equilibrium of the firm is that its profit should be maximum.
ii. Marginal cost should be equal to marginal revenue.
iii. MC must cut MR from below.

The above conditions of the equilibrium of the firm can be examined in two ways:

a) Total Revenue and Total Cost Approach

b) Marginal Revenue and Marginal Cost Approach.

Total Revenue and Total Cost Approach:

A firm is said to be in equilibrium when it maximizes its profit. It is the point when it has no tendency either to increase or contract its output. Now, profits are the difference between total revenue and total cost. So in order to be in equilibrium, the firm will attempt to maximize the difference between total revenue and total costs. It is clear from the figure that the largest profits which the firm could make will be earned when the vertical distance between the total cost and total revenue is greatest.

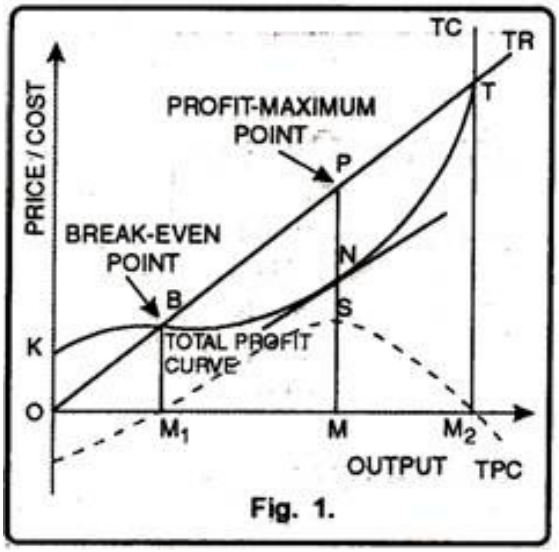


Fig. 1.

In fig. 1 output has been measured on *X*-axis while price/cost on *Y*-axis. *TR* is the total revenue curve. It is a straight line bisecting the origin at 45°. It signifies that price of the commodity is fixed. Such a situation exists only under perfect competition. *TC* is the total cost curve. *TPC* is the total profit curve. Up to OM_1 level of output, *TC* curve lies above *TR* curve. It is the loss

zone. At OM_1 output, the firm just covers costs $TR=TC$. Point B indicates zero profit. It is called the break-even point. Beyond OM_1 output, the difference between *TR* and *TC* is positive up to OM_2 level of output. The firm makes maximum profits at *OM* output because the vertical distance between *TR* and *TC* curves *(PN)* is maximum. The tangent at point *N* on *TC* curve is parallel to the *TR* curve. The behaviour of total profits is shown by the dotted curve. Total profits are maximum at *OM* output. At OM_2 output *TC* is again equal to *TR*. Profits fall to zero. Losses are minimum at *OM*) output. The firm has crossed the loss zone and is about to enter the profit zone. It is signified by the break-even point-B.

Marginal Revenue and Marginal Cost Approach:

Joan Robinson used the tools of marginal revenue and marginal cost to demonstrate the equilibrium of the firm. According to this method, the profits of a firm can be estimated by calculating the marginal revenue and marginal cost at different levels of output. Marginal revenue is the difference made to total revenue by selling one unit of output. Similarly, marginal cost is the difference made to total cost by producing one unit of output. The profits of a firm will be maximum at that level of output whose marginal cost is equal to marginal revenue.

Thus, every firm will increase output till marginal revenue is greater than marginal cost. On the other hand, if marginal cost happens to be greater than marginal revenue the firm will sustain losses. Thus, it will be in the interest of the firm to contract the output. It can be shown with the help of a figure. In fig. 2 *MC* is the upward sloping marginal cost curve and MR is the downward sloping marginal revenue curve. Both these curves intersect each other at point *E* which determines the *OX* level of output. At *OX* level of output marginal revenue is just equal to marginal cost.

It means, firm will be maximizing its profits by producing *OX* output. Now, if the firm produces output less or more than *OX*, its profits will be less. For instance, at OX1 its profits will be less because here $MR = JX_1$, while $MC = KX_1$ So, $MR > MC$. In the same fashion at OX_2 level of output marginal revenue is less than marginal cost. Therefore, beyond *OX* level of output extra units will add more to cost than to revenue and, thus, the firm will be incurring a loss on these extra units. Besides first condition, the second order condition must also be satisfied, if we want

to be in a stable equilibrium position. The second order condition requires that for a firm to be in equilibrium marginal cost curve must cut marginal revenue curve from below. If, at the point of equality, *MC* curve cuts the *MR* curve from above, then beyond the point of equality *MC* would be lower than *MR* and, therefore, it will be in the interest of the producer to expand output beyond this equality point. This can be made clear with the help of the figure.

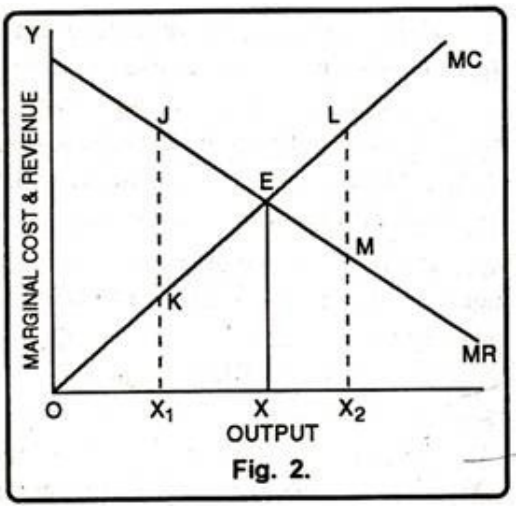


Fig. 2.

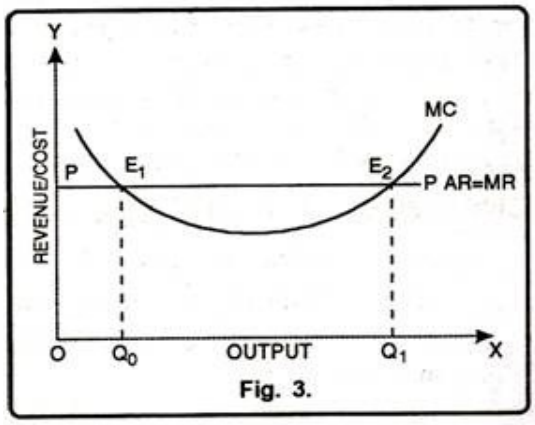


Fig. 3.

In figure 3 output has been measured on X-axis while revenue on Y-axis. MC is the marginal cost curve. PP curve represents the average revenue as well as marginal revenue curve. It is clear from the figure that initially MC curve cuts the MR curve at point E1. Point E1 is called the 'Break Even Point' as MC curve intersects the MR curve from above. The profit maximizing

output is OQ1 because with this output marginal cost is equal to marginal revenue (E2) and MC curve intersects the MR curve from below.

Determination of Short Run Equilibrium of Firm:

Short-run refers to that period in which fixed factors remaining unchanged the firms in order to incur maximum profits can vary their output by changing the variable factors like labour, raw material etc. In the short period, it is not necessary that the firms must earn super-normal or normal profits but even the firms may have to sustain the losses. A firm may earn supernormal profits because in the short run, firms cannot enter the industry. Moreover, a firm may suffer losses, because in the short run, may not step up production even when price of the product falls. In case, it stops production temporarily, it will have to bear the loss of fixed cost which will constitute the minimum losses of the firm.

However, all the above stated possibilities have been explained as under:

i. Supernormal Profits:

A firm is said to be in equilibrium when its marginal cost is equal to marginal revenue and marginal cost curve cuts the marginal revenue curve from below. A firm in equilibrium enjoys supernormal profits if average revenue exceeds marginal cost. This fact has been shown in fig 4. In figure 4 outputs has been shown on horizontal axis and revenue on vertical axis. MC and AC are the marginal cost and average cost curves respectively. PP is the average revenue curve. It is clear from the figure that MC curve intersects the MR curve from below at point N which shows output OX. At this level of output price is NX and average cost is MX. Since average revenue is greater than average cost, the firm is earning super-normal profits MN per unit of output. Thus, the total super-normal profits of a firm will be equal to PLMN.

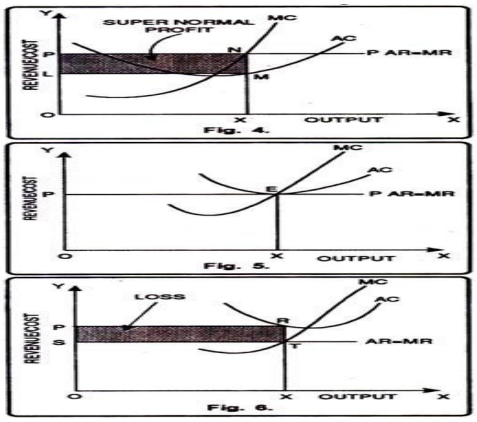
ii. Normal Profit:

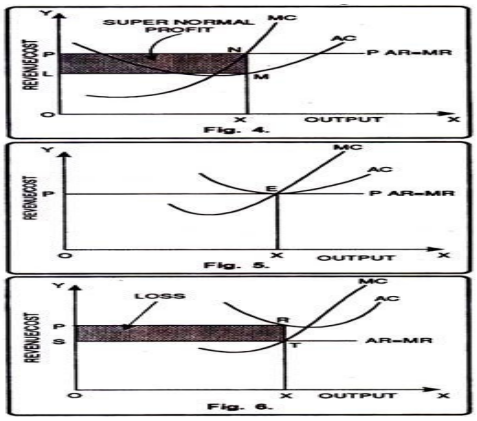
Normal profits refer to those profits where the average cost of the firm equals the average revenue. These profits cover just the reward for entrepreneurial services and are included in the cost of production. It can be shown with the help of a figure. In figure 5 the equilibrium has been depicted at point E. At point E marginal revenue is equal to marginal cost and marginal cost

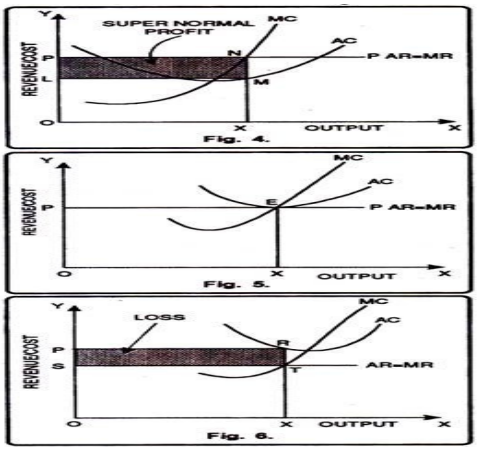
intersects the marginal revenue curve from below. The firm earns normal profits at OX output because at this output both the conditions of equilibrium are fulfilled.

iii. Minimum Losses:

A firm in equilibrium incurs losses when it does not cover the average cost. In other words, when average revenue falls short of average cost, the firm has to sustain losses. In figure 6 the firm is said to be in equilibrium at point T. At this level of output both the conditions of equilibrium are satisfied i.e., marginal revenue is equal to marginal cost and marginal cost curve intersects the marginal revenue curve from below. Thus, it determines the OX level of output correspondingly price is OP. It means loss per unit of output is RT. Therefore, losses will be PSTR.



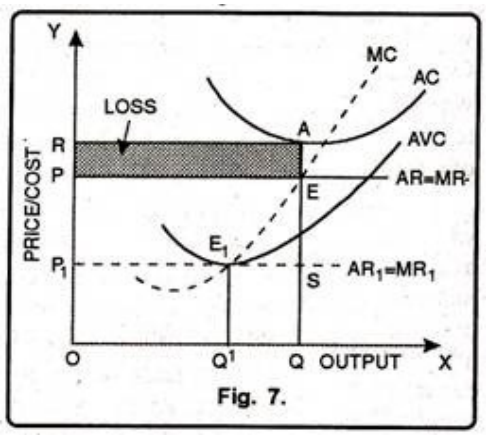




iv. Shut Down Point:

Simple question is why firms continue producing the product if they are making losses. In the short run, the firms cannot go out of the industry by disposing off the plant. Why do they not shut down? It is because they cannot change the fixed factors and they have to face fixed costs even if the firm is shut down.

The firm can avoid only variable costs but it has to bear the fixed costs whether to produce or not. The firm will continue producing till the price covers the average variable cost. If the price covers some part of the average fixed costs besides the variable costs, the producer will continue producing. Thus the firm will continue producing so long as price exceeds average variable cost. The shut down point can be shown with the help of a diagram.



In diagram 7 equilibrium is at E where MR = MC and MC cuts MR from below. The price is EQ and OQ is the output. This price covers the average variable cost. Average cost corresponding to this output is AQ. In that way loss per unit is AE which is equal to average fixed cost. The total losses are equal to total fixed costs. If price is slightly below OP, level, the firm will not produce at all. The firm will simply shut down production and wait for some good days to come.

Shut Down Point (Losses=Total Fixed Costs):

However, the firm may continue to operate even under such a situation because of the following reasons:

i. The firm may continue to operate because a higher valuation (value) is given to an on-going concerns rather than a closed down firm.

ii. More prestige is attached to the owner or manager of a on-going concern than to that of a firm that has closed down or ceased to operate.

iii. By keeping the operation going, the firm will not lose competent personnel.

iv. The firm may continue to operate in the hope of earning profits in future.

Determination of Long Run Equilibrium of the Firm:

Long run refers to that period in which the producer can change its supply by changing all the factors of production. In other words, the producer has the sufficient time to adjust their supplies according to the changed demand conditions. Moreover, new firms can also enter and existing firms can leave the industry. In the long-run, the firm is said to be in equilibrium when marginal cost is equal to price. Besides it, the firm under perfect competition to be in equilibrium-price must be equal to average cost. Generally, in the long run, firm in equilibrium earns normal profits. If the firms happen to earn the super normal profits in the long period, the existing firms will increase their production.

Lured by super normal profits some new firms will enter into the industry. The total supply of the product will increase and the price falls down. Thus, due to fall in price the firms will get normal profits. In case price of the product is less than the average cost, the firms would make losses. These losses would induce some firms to leave the industry. Consequently the output of

the industry will fall which will raise the price, hence, the firms will begin to earn normal profits. It can be shown with the help of a figure 8. In figure 8 output has been depicted on X-axis while revenue on Y-axis. SAC is the short run average cost curve and LAC is the long run average cost curve. Similarly, SMC and LMC are the short run marginal cost and long run marginal cost curves respectively. Let us suppose that the industry determines OP price. At this price firms are producing with SAC1 and is earning super normal profits equal to the shaded area PLNM. Lured by these super normal profits, the existing firms will increase their production capacity, thus, the new firms will enter the industry. As a result of the entry of the new firms supply of the product will increase which will lead to a fall in price.

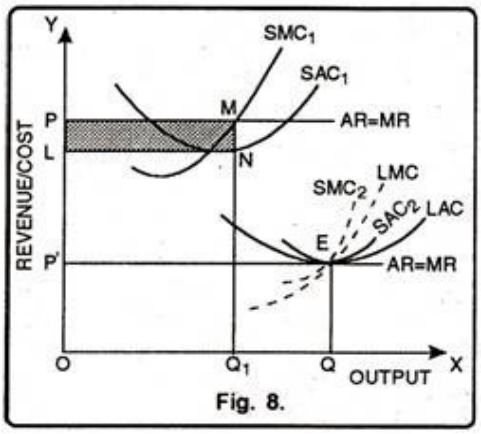


Fig. 8.

Thus, the price will fall to OP'. At this price, the firm will be in equilibrium at point E and will produce OQ level of output. It is due to the reason that at point E, marginal revenue, long run marginal cost, average revenue and long run average cost are all equal and the firm earns normal profits.

Symbolically:

$$MR = LMC = AR = LAC = SAC = SMC = Price$$

Difficulties of TR-TC Approach:

The main difficulties of TR and TC approach are as under:

i. It is very difficult to analyze at what level of output profits are maximum.
ii. It is difficult to see at a glance the maximum vertical distance between TR and TC approach.
iii. It is very difficult to discover the price per unit of output.

Equilibrium of Industry:

The group of firms producing homogeneous product is called industry. In fact the concept of industry exists only under perfect competition. An industry is said to be in equilibrium when it has no tendency to increase or decrease its level of output. According to Prof. Hansen, "An industry will be in equilibrium when there is no tendency for the size of the industry to change i.e., when no firms wish to leave it and no new firms are being attracted to it." New firms will have no tendency to enter the industry when existing firms are enjoying normal profits. The normal profits earned by a firm are included in total cost. In this way equilibrium for the industry means that firms are neither moving in or nor moving out. It means that the level of profits in it is neither above nor below the normal level and hence is equal to it.

Conditions of Equilibrium of an Industry:

i. Constant Number of Firms:

An industry will be in equilibrium when the number of firms remains constant. In this situation, no new firms will enter and no old firms will leave the industry.

ii. Equilibrium of Firms:

An industry will be in equilibrium when all firms operating in it are in equilibrium and have no tendency to increase or decrease the level of output.

Short Run Equilibrium of Industry:

In the short run, the industry is said to be in Equilibrium when all the firms operating under it are in equilibrium. But for the industry to be in full equilibrium in the short run is very rare. Full equilibrium position is possible when firms earn normal profits. In the short run firms can also earn supernormal profits or incur losses. It can be shown with the help of fig. 9.

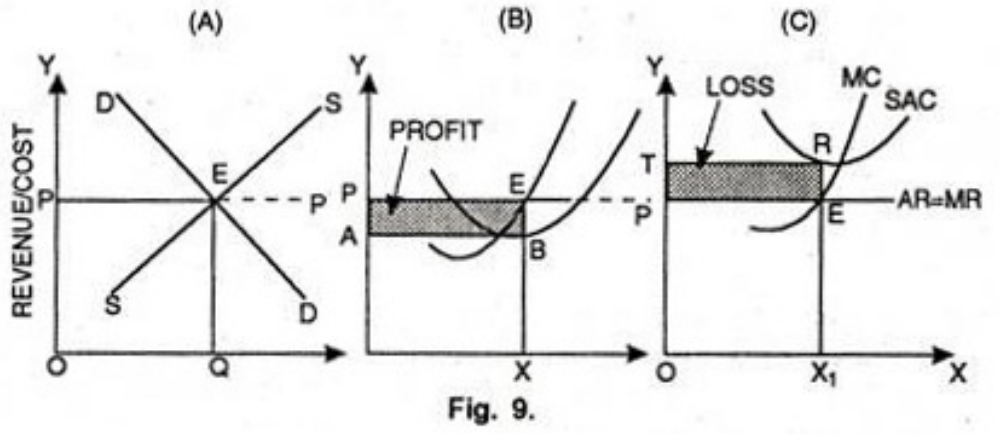


Fig. 9.

In fig. 9 (A) *DD* is the industry's demand curve and *SS* represents the supply curve. Both these curves intersect each other at point *E* which establishes equilibrium of the industry. At this equilibrium point, industry sets price *OP* and produces *OQ* level of output. But, it will not be the full equilibrium of the industry. In fig. B the firms are enjoying supernormal profits as indicated by *ABED*. In fig. 9 *C*, the firms are incurring losses equal to the shaded area *PERT*. In the long run, firms incurring losses will leave the industry. On the other hand, firms getting supernormal profits will expand their production capacity. Lured by supernormal profits new firms will enter the industry. Consequently, industry will be in equilibrium in the short run only if all firms are enjoying normal profits.

Long Run Equilibrium of the Industry:

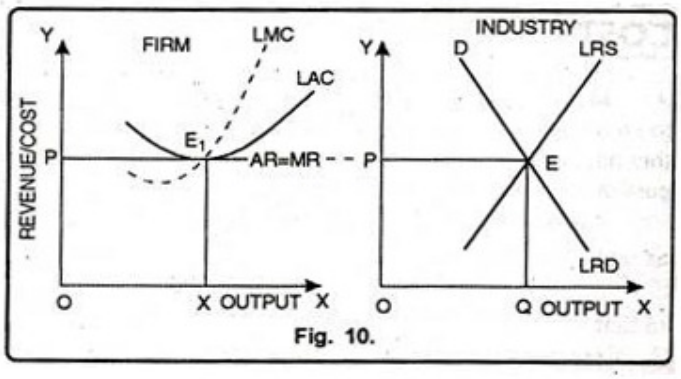


Fig. 10.

The long run equilibrium of the industry can be shown with the help of a figure 10. In the long run the industry will be in equilibrium at a point where long run supply (LRS) is equal to long run demand (LRD). This determination of price is OP and output OQ. The firm will follow this price and will be in equilibrium at E1. Here, the firms will earn just normal profits. Thus, according to Left-witch, "The existence of long run industry equilibrium requires long run individual equilibrium at no profit no loss level of operation".

PERFECT COMPETITION

Introduction

A perfectly competitive market is one in which there is *large number of buyers and sellers of a homogeneous product* and *neither a seller nor a buyer has any control on the price of the product.*

Perfect competition is perceived by the economists as *a rare phenomenon*. Nevertheless, analysis of price and output determination under perfect competition *'lays the foundation' of pricing theory.* This kind of *a notional market* is therefore created by assumption for theoretical purpose.

Characteristics

1) Large number of sellers and buyers.
2) Homogeneous product.
3) Perfect mobility of factors of production.
4) Free entry and free exit.
5) Perfect knowledge about the market conditions.

6) No government interference.

7) Absence of collusion and independent decision making by firms.

1) *Large number of sellers and buyers.*

Under perfect competition, the number of sellers (the firms) is assumed to be so large that the share of each seller in the total supply of a product is very small. 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.

2) *Homogeneous product.*

The commodities 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 the 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, the supplier, to charge a price higher than the market price.

3) *Perfect mobility of factors of production*

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. There is no barrier to labour mobility—*legal*, *linguistic*, *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 implies that factors of production—*land*, l*abour*, *capital* and *entrepreneurship*—can enter or exit a firm or the industry at will.

4) *Free entry and free exit.*

In a perfectly competitive market, 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 exit it at its will.

Therefore, when firms in the industry make supernormal profit for some reason, new firms enters the industry and supernormal profits are eliminated. Similarly, when profits decrease or more profitable opportunities are available elsewhere, firms exit the industry.

5) *Perfect knowledge about the market conditions*

Both buyers and sellers have perfect knowledge about the market conditions. This means that all the buyers and sellers have *full information* regarding the prevailing and future prices and availability of the commodity. Information regarding market conditions is available free of cost. There is no uncertainty in the market.

6) *No government interference.*

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

7) *Absence of collusion and independent decision making by firms.*

Perfect competition assumes that there is no collusion between the firms, i.e., the firms 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, *i.e.,* there are no consumers' associations. Buyers and sellers take their decisions independently and they act independently.

Perfect versus Pure Competition

The difference between the two kinds of competition is a matter of degree. While *'perfect competition'* has all the features mentioned, under *'pure competition',* there are no perfect mobility of factors and no perfect knowledge about market conditions. 'Pure competition' is 'pure' in the sense that it has absolutely no element of monopoly. The perfect competition, with

its all characteristics 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.

Role of a firm in a perfectly competitive market

In a perfectly competitive market, the role of a firm is limited to producing a commodity or service and selling it at the market determined price.

In fact, an individual firm is one among a very large number of firms producing an almost identical commodity. The share of a firm in the total supply of the commodity is, therefore, very small.

A firm's status in a perfectly competitive market can be described as follows.

a) *A firm has no control over price.*

The market share of an individual firm is so small, rather insignificant, that a firm cannot determine the price of its own product, nor it can influence the prevailing market price by changing its supply. In other words, *an individual firm has no control over the market price.*

b) *A firm is a price-taker.*

Under perfect competition, an individual firm does not determine the price of its own product. Price for its product is determined by the *market demand* and *market supply* for the industry as a whole.

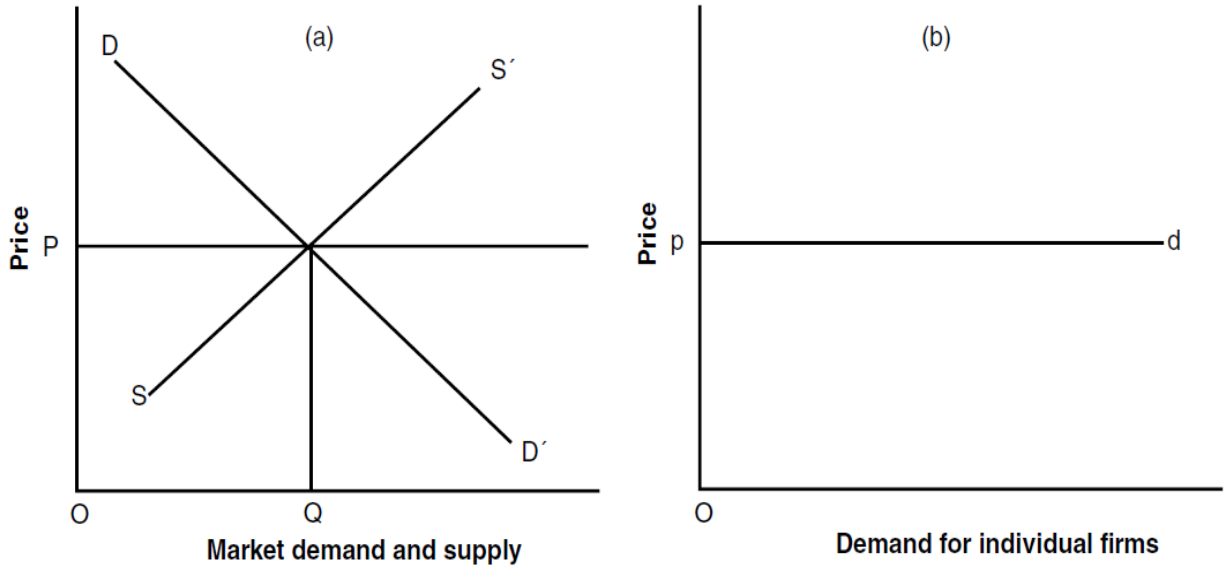


Figure 2.1 *Determination of Market Price and Demand for Individual Firms*

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.

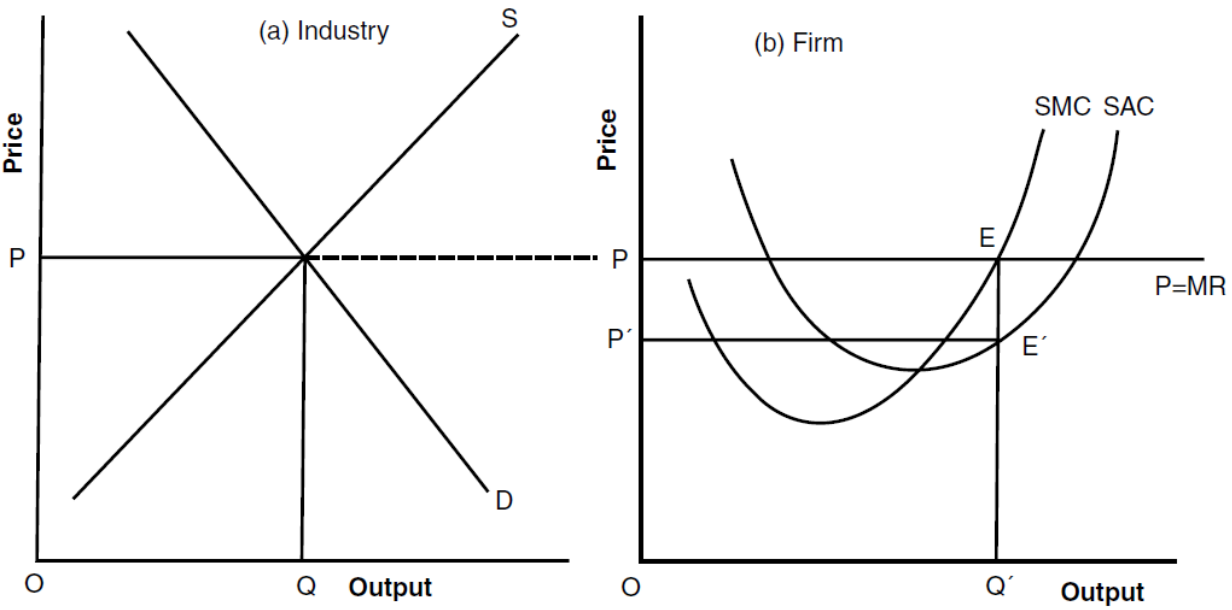
As *Figure 2.1(a)* shows, market price for the industry as a whole is determined at *OP* and equilibrium output for the industry is determined at output *OQ*. Equilibrium price *OP* is given for all the firms of the industry. No firm has power to change this price. At this price, a firm can sell any quantity. It implies that the demand curve for an individual firm is a straight horizontal line, as shown by the line, *pd*, in *Figure 2.1(b),* with infinite elasticity.

c) *No control over cost.*

Because of its small purchase of inputs *(labour and capital)*, under perfect competition 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* ce with a given iizes its profits duce a quantity



the following

and,

MC.

level of output

Figure 2.2 *Short-run Equilibrium of the Firm*

at which its *MC* = *MR*. This condition applies in both short run and long run, even though *MR* and *MC* conditions are somewhat different in the long run.

Assumptions

The short-run equilibrium of a firm is analyzed under the following assumptions:

a) capital cost is fixed but labour cost is variable;
b) prices of inputs are given;
c) price of the commodity is fixed; and the firm is faced with short-run U-shaped cost curves.

The determination of market price is shown in panel *Figure 2.2(a).* As shown in *Figure 2.2(a),* the market price of a commodity is determined at *OP* by the market forces—demand and supply—in a perfectly competitive market.

The price *OP* is fixed for all the firms of the industry. Therefore, a firm faces a straight line or 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 of the firm are shown by *SAC* and *SMC*, respectively.

Firm's short-run equilibrium is illustrated in panel *Figure 2.2(b),* the *SMC* curve intersects the *P = MR* line at point *E*, from below. At point *E*, *SMC = MR*. Point *E* determines, therefore, the point of firm's equilibrium. 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 maximum pure profit is shown by the area *PEE′P′* which equals *PP′* x *OQ* (=*PE*) where *PP′* is the per unit super normal profit at output *OQ*.

Short-run equilibrium of industry/ price and output determination of industry:

An industry is in equilibrium in the short-run when market is cleared at a given price, *i.e.,* when the total supply of the industry equals the total demand for its product. The price at which the market is cleared is the equilibrium price. When an industry reaches its equilibrium, there is no tendency to expand or to contract the output.

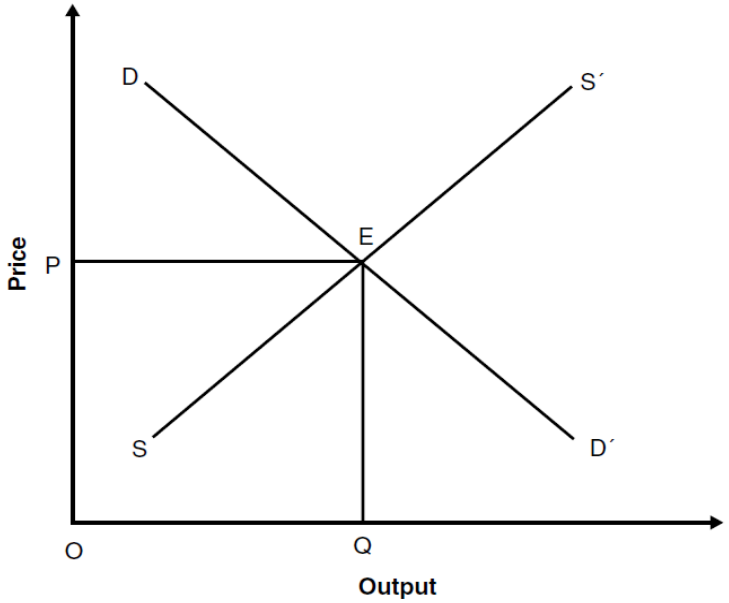


Figure 2.7 *Equilibrium of the Industry*

The industry demand curve *DD′* and supply curve *SS′* intersect at point *E*, determining equilibrium price *OP*. At price *OP*, *D*=*S*. The industry is supplying as much as consumers demand.

In the short-run equilibrium of the industry, some individual firms may make pure profits, some normal profits and some may make even losses, depending on their cost and revenue conditions, this situation will, however, not continue in the long run.

Link between short-run equilibrium of the industry and the firm

The short-run equilibrium of the *firm* and *industry* have been analyzed separately there exists, however, a link between a firm's and industry's equilibrium.

In a perfectly competitive market, change in the *equilibrium* of an individual firm does not affect the *industry's equilibrium*, for the simple reason that the total output of a single firm constitutes a small fraction of the industry's output. But, a change in the industry's equilibrium does alter the equilibrium of an individual firm.

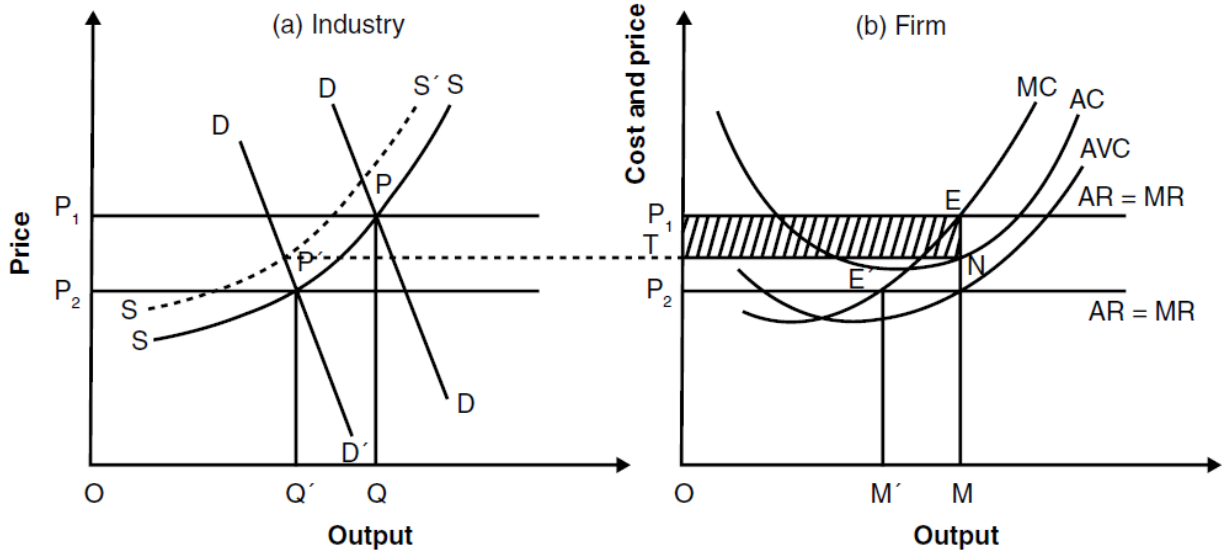


Figure 2.8 *Industry's Vs. Firm's Equilibrium in the Short Run*

The link between industry's and firm's equilibrium is illustrated in Figure 2.8. Suppose industry's initial demand and supply curves are given as *DD* and *SS* respectively (Figure 2.8(a)). As shown in Figure 2.8(a), industry's demand and supply curves intersect each other at point *P*,

determining the market price at *PQ*=*OP*1 and industry's equilibrium output is *OQ*. Thus, the price *PQ* is given to all the firms of the industry. Given the price *PQ* and firm's cost curves, an individual firm finds its equilibrium at point *E* in Figure 2.8(b), where its *MC*=*MR*.

Firm's equilibrium output is *OM* in Figure 2.8(b). At price *EM* = *PQ*, the firm is making an abnormal profit in the short run to the extent of *EN* per unit of output. The firm's total pure profit is shown by the shaded area, *P*1*ENT*. If industry demand curve *DD* now shift downward for some reason to *DD′*, supply curve remaining unchanged. As a result, market price falls to *P′Q′* and industry's equilibrium output falls to *QQ′*. With the fall in price, firm's equilibrium shifts from point *E* to *E′* where its *MC*=*MR*. At this point, the firm is making a loss because its *AR* which equals *E′M′* is lower than its *AC*. Thus, change in industry's equilibrium changes firm's equilibrium.

Thus, change in industry's equilibrium changes firm's equilibrium. Firms making loss is, however, a short-run situation. Losses will disappear in the long run through a process of market adjustment. The process of market adjustment begins with loss-making firms exiting the industry. When loss-making firms quit the industry, supply declined and the supply curve shifts left side as shown by the dotted supply curve *SS′*. Price goes up and loss disappears and firm reaches another equilibrium point.

Long-run equilibrium of the firm

A comparative look at the short- and long-run market conditions:

The short run is, by definition, a period in which

i. firm's cost and revenue curves are given,

ii. firms cannot change their size—their capital is fixed,

iii. existing firms do not have the opportunity to leave the industry and

iv. new firms do not have the opportunity to enter the industry.

In contrast, *long run* is a period in which these constraints disappear. Long run permits improvement in production technology and a larger employment of both, labour and capital, *i.e.,*

firms can change their size. Some of the existing firms may leave and new firms may enter the industry. In the long run, supply curve not only shifts downward but also becomes more elastic. In this section, we will analyze the equilibrium of the firm and industry in the long run.

Equilibrium of the Firm in the Long-run

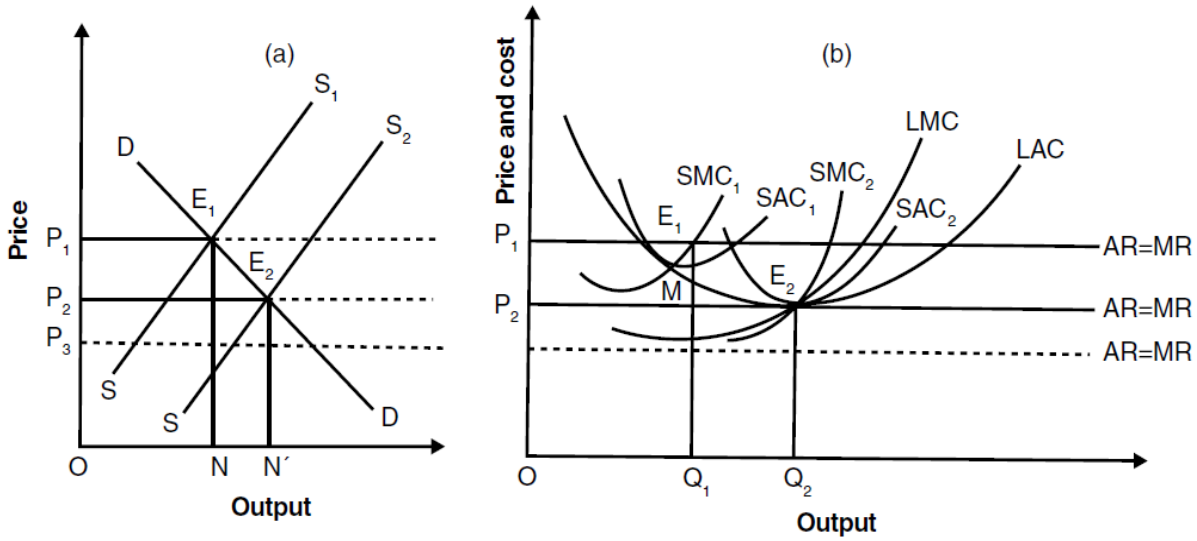


Figure 2.9 *Long-run Equilibrium of the Firm and Industry*

Suppose (i) short-run price is given at *OP*1 *(Figure 2.9(a))* and (ii) that firm's short-run cost curves are given by *SAC*1 and *SMC*1, as shown *Figure 2.9(b)*. Given the price *OP*1, firms are in equilibrium at point *E*1. It can be seen in *Figure 2.9(b)* that the firms are making an abnormal profit of $E1M = E1Q1 - MQ1$ per unit of output.

Abnormal profit brings about two major changes in the industry. *First*, existing firms get incentive to increase the scale of their production. Their average and marginal costs go down caused by the economies of scale. This phenomenon is shown by *SAC*2 and *SMC*2. When we draw the *LAC* and *LMC* curves, these curves show decreasing costs in the long run. *Secondly*, attracted by the abnormal profit, new firms enter the industry increasing the total supply.

Given the new market price, *OP*2, firms attain their equilibrium in the long run at point *E*2 where *AR = MR = LMC = LAC = SMC = SAC* as shown in *Figure 2.9(b).* As the figure shows, the firms of industry reach their equilibrium in the long run where both short- and long-run equilibrium conditions are satisfied simultaneously. In a perfectly competitive market, the cost and revenue conditions are given for the firms. Therefore, when price goes down to *OP*2, what firms are required to do is to adjust their output to the given revenue and cost conditions in order to maximize their profit. Through this process of adjustment for output, the firms reach the equilibrium in the long run at point *E*2. Point *E*2 is the point of equilibrium for all the firms in the long run.

In case market price falls below *OP*2, say, to *OP*3, all the firms make losses. This brings in a reverse process of adjustment. While some firms quit the industry, some firms cut down the size of the firm. As a result, total supply decreases, demand remaining the same. Consequently, price tends to rise. This process of output adjustment continues until industry reaches back to its equilibrium at point *E*2, where *LAC* is tangent to *P = AR = MR* for each firm in the industry. At point *E*2, the point of equilibrium, *P = MR = LMC = LAC = SMC = SAC*. Since *P = LAC*, the firms make only normal profits in the long run. If firms deviate from point *E*2, due to some short-run disturbances, the market forces will restore the equilibrium.

Equilibrium of industry in long-run

An industry is in equilibrium at a price and output at which its market demand equals its market supply. The equilibrium of the industry is illustrated in *Figure 2.9(a).* When an industry is in equilibrium, all its firms are supposed to be in equilibrium *(as shown in Figure 2.9(b))* and earn only normal profits. This is so because under the conditions of perfect competition, all the firms are assumed to achieve the same level of efficiency in the long run. Since industry yields only normal profits, there is no incentive for new firms to enter the industry.

These conditions are fulfilled at price *OP*2 in *Figure 2.9(a)* and *(b).* At price *OP*2 , all the firms are in equilibrium, as for each firm, *LMC = LMR = SMC = SAC = P = LAC.*

Since *P = LAC*, all the firms are earning only normal profit. At industry's equilibrium output OM, market demand equals market supply *(Figure 2.9(a)).* At price *OP2*, therefore, market is cleared. The output *OM* may remain stable in the long run. For, there is no incentive for new

firms to enter the industry and no reason for the existing ones to leave the industry. The industry is, therefore, in equilibrium.

Role of time element in the theory of price determination

The price of a commodity is determined under perfect competition by its demand-supply. But, demand and supply do not exercise the same degree of influence on the determination of the price of a commodity in all circumstances.

In order to explain the relative role of demand and supply in price determination Alfred Marshall has stressed the role of time in the theory of value.

Alfred Marshall has mentioned the four different time periods and has shown the relative influence of demand and supply in each of these periods:

1) The very short period or momentary equilibrium

In the very short-period market, the supply of a commodity remains absolutely fixed, *i.e.,* the supply curve becomes a vertical straight line. In such a period demand or marginal utility is more active than supply in determining the price of a commodity. As the supply remains absolutely fixed *(e.g., the supply of fish or of vegetables, etc., in a local market),* the prices rise or fall with the rise or fall in the demand.

When the demand increases, the new demand curve will cut the vertically straight line supply curve at a higher period and the new equilibrium will be attained at a higher price. The opposite happens when the demand falls in the very short-period market. The equilibrium in such a period is called *momentary* equilibrium, and the price prevailing in such a market is called the market price.

2) *The short-period equilibrium*

In the short period, a firm does not get sufficient time to alter its fixed equipment and plant size. In such a market, the supply can be partly adjusted to the change in demand. The price will rise when demand is higher, but the rise will not be of such an extent as is observed in the very short-period. Such a price is called the *short-run* normal price.

Supply or cost of production has some influence over the short-run normal price in a perfectly competitive market. It is found that a competitive firm produces a commodity up to that level at which *P=MC* (rising). As in the short-run, neither the free entry of new firms nor the free exit of the old firm is possible, the short-run competitive price may be greater than equal to, or less than *ATC* provided it is greater than, or equal to *AVC*. The short-run competitive price must cover the full amount of variable cost and at least a portion of the fixed cost. Or, alternatively the average variable cost sets the louver limit below which the short-run price must not fall (the shut-down point).

3) *The long-period equilibrium*

The long run is a period which is long enough to bring about a change in the existing scale of operations and equipment. Now, supply can be fully adjusted to changes in demand. The equilibrium price in the long-run is called the long-run normal price. The supply or cost of production is more important than the demand or marginal utility in the determination of the long-run normal price.

Moreover, in the long run, a competitive firm produces an output up to that level at which *P=MC*. But, due to free entry of new firms and free exit of the existing firms, the long-run competitive price cannot be greater or less than the average cost. It should be exactly equal to average cost. So, the long-run competitive price is equal to both marginal and average costs.

4) *The very long-period equilibrium*

The very long-period is one in which increased or decreased demand for capital equipment leads to changes in their supply. In such a situation the supply of a commodity can be fully adjusted to demand, and the supply or cost of production is more important than the demand in the determination of price. During this period cost may fell due to technological progress or innovation.

Thus, in the final analysis, both demand and supply forces play a greater role in the price determination in different time periods. Which forces are stronger than the other depends on the length of time. Just like the pair of blades of a scissors, both demand blade and supply blade are necessary in market price determination.

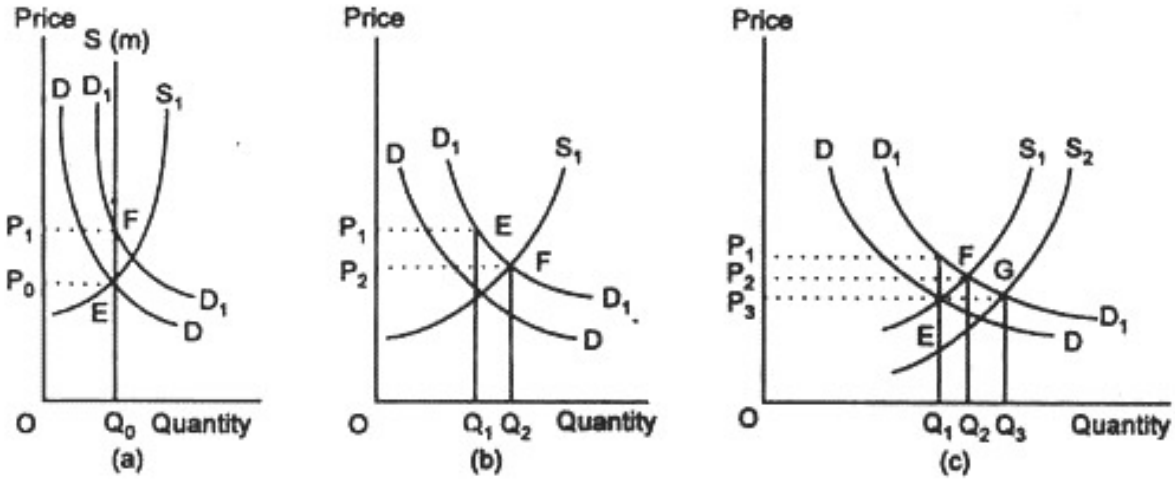


Fig. 10 : Periods of supply and equilibrium price

In Fig. 10(a), we see the effect of an increase in demand in the momentary period. Supply is perfectly inelastic and price rises from OP_0 to OP_1 that is, by the full extent of the change in demand.

Fig. 10(b) shows the situation after sufficient time has elapsed for the industry to react to the increase in demand. The higher price has stimulated an increase in output and this is depicted by a movement along the existing short-period supply curve. Quantity supplied has increased and price has fallen from OP_1 to OP_2. This higher (short period) price, if sustained, will lead to changes in the productive capacity of the industry. Existing firms will be eager to expand and new firms will join the industry.

In Fig. 10(c) we see the effects of this increase in the scale of production. The new supply situation is shown by an entirely new short-term supply curve S_2. The increase in supply has caused the price to fall from OP_2 to OP_3 because the quantity demanded has increased. The price OP_3 may be higher or lower than the original price OP_0, depending upon the extent to which the expansion of the industry has yielded economies of scale.

The shorter the time period, the greater becomes the influence of demand and marginal utility and the longer the time period, the greater becomes the influence of supply and cost of production on value.

MONOPOLY

Meaning and definitions

The word monopoly has been derived from Greek word *monos*, meaning 'alone' and *polein* meaning 'seller'. The term "monopoly" is used to describe the market structure in which there is only one producer of a good or service for which there are no close substitutes and entry into and exit from the industry is impossible. A monopoly firm enjoys an absolute power to produce and sell a commodity.

The single producer may be in the form of *individual owner* or *a single partnership* or *a joint stock company.* In other words, under monopoly there is no difference between firm and industry. Monopolist has full control over the supply of commodity. Having control over the supply of the commodity he possesses the market power to set the price.

"Pure monopoly is represented by a market situation in which there is a single seller of a product for which there are no substitutes; this single seller is unaffected by and does not affect the prices and outputs of other products sold in the economy." *Bilas*

"Monopoly is a market situation in which there is a single seller. There are no close substitutes of the commodity it produces, there are barriers to entry". *Koutsoyiannis*

"Under pure monopoly there is a single seller in the market. The monopolist demand is market demand. The monopolist is a price-maker. Pure monopoly suggests no substitute situation".
A. J. Braff

"A pure monopoly exists when there is only one producer in the market. There are no dire competitions." *Ferguson*

"Pure or absolute monopoly exists when a single firm is the sole producer for a product for which there are no close substitutes." *McConnel*

Characteristics

1) Single Seller of a Product

Under monopoly, there is a single seller selling the product. As a result, the monopoly firm and industry is one and the same thing. Monopolist has full control over the supply and price of the product. However, there are large numbers of buyers of monopoly product and no single buyer can influence the market price.

2) Barriers to Entry and Exit

There exist strong barriers to entry of new firms and exit of existing firms. As a result, a monopoly firm can earn abnormal profits and losses in the long run. These barriers may be due to legal restrictions like licensing or patent rights or due to restrictions created by firms in the form of cartel.

3) No Close Substitutes

The product produced by a monopolist has no close substitutes. So, the monopoly firm has no fear of competition from new or existing products. For example, there is no close substitute of electricity services provided by NDPL. However, the product may have distant substitutes like inverter and generator.

4) Imperfect Knowledge

Imperfect knowledge about the product and market between buyers and seller prevails only in monopoly market it is not possible in other type of markets.

5) Price Discrimination

A monopolist may charge different prices for his product from different sets of consumers at the same time. It is known as ‘price discrimination’.

6) No Supply Curve

Since a monopolized industry is a single firm industry therefore there is no distinction between a firm and an industry in a monopolistic market structure. Hence the demand curve of a monopolistic firm is same as the market demand curve.

7) Price Maker

In case of monopoly, firm and industry is one and the same thing. So, firm has complete control over the industry output. As a result, monopolist is a price-maker and fixes its own price. It can influence the market price by changing the supply of the product.

Sources and kinds of monopolies

The emergence and survival of monopoly are attributed to the factors which prevent the entry of other firms into the industry. The barriers to entry are, therefore, the sources of monopoly power.

The major sources of barriers to entry are

a) *Legal restrictions,*

Some monopolies are created by law in public interest. Such monopolies may be created in both public and private sectors. Most of the state monopolies in the public utility sector, including *postal*, *telegraph and telephone services*, *generation and distribution of electricity*, *railways*, *airlines* and *state roadways*, etc., are public monopolies. Such monopolies are created by the government in the public sector.

The government may create monopolies in the private sector also through license or patent. Such monopolies are intended to reduce cost of production by the economies of scale and investment in technical innovations. Such monopolies are also known as *franchise monopolies.*

b) *Sole control over the supply of certain scarce and key raw materials,*

Some firms acquire monopoly power from their legally granted control over certain scarce and key raw materials that are essential for the production of certain other goods, e.g. *bauxite*, *graphite*, *diamond*, etc. Such monopolies are often called *'raw material monopolies'*. The

monopolies of this kind emerge also because of monopoly over certain specific *technical knowledge* or *techniques of production.*

c) *Efficiency and*

A primary and technical reason for growth of monopolies is the economies of scale. In some industries, long-run minimum cost of production, i.e., the most efficient scale of production coincides almost with the size of the market. In such industries or products, a large size firm finds it profitable, in the long run, to eliminate competition by cutting down its price for a short period.

Once a monopoly is established, it becomes almost impossible for the *new firms* to enter the industry and survive. Monopolies born out of efficiency are known as natural monopolies. A natural monopoly may emerge out of the *technical conditions of efficiency* or may be created by the *law on efficiency grounds* in public interest.

d) *Patent rights*

Another source of monopoly is the patent rights of the firm *for a product* or *for a production process.* Patent rights are granted by the government to a firm to produce a commodity of *specified quality and character* or to use *a specified technique of production*. Patent rights give firm *exclusive rights* to produce the *specified commodity* or to use the *specified technique of production.* Such monopolies are called patent monopolies.

Cost and revenue curves under monopoly

The cost curves—the *AC* and *MC* curves—faced by the monopoly firm are *U*-shaped. The monopoly firms face *U*-shaped *AC* and *MC* cost curves because the laws of production apply to monopoly firms. The demand of *AR* and *MR* curves that a monopoly firm faces are different from those faced by the firms under perfect competition.

In a perfectly competitive market, there is dichotomy between the firm and the industry. While firms face a horizontal, straight-line demand curve, industry faces a downward sloping demand curve. Under monopoly, however, there is no distinction between the firm and the industry. The monopoly industry is a single-firm industry and industry demand curve has a negative slope. A

monopoly firm faces, therefore, a downward sloping demand curve—it may be a linear or a non-linear demand curve. Given the demand curve, a monopoly firm has the option to choose between price to be charged or output to be sold. Once it chooses price, the demand for its output if fixed. Similarly, given its demand curve, if the firm decides to sell a certain quantity of output, then its price is fixed—it cannot charge any other price inconsistent with the demand curve.

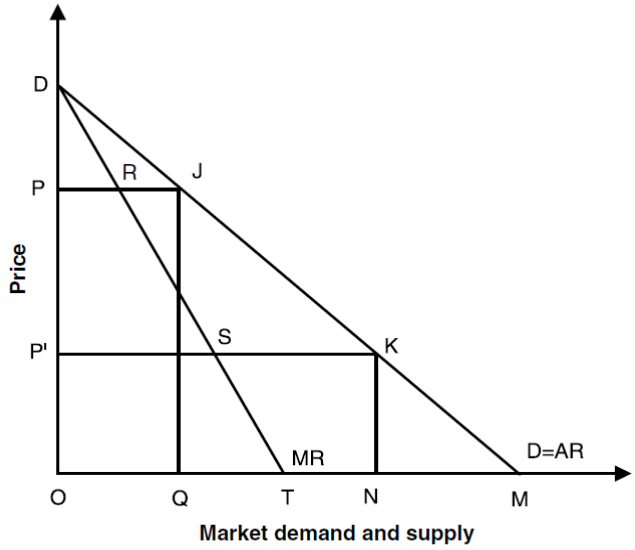


Figure 3.1 *AR and MR Curves for Monopoly*

The price–quantity constraint is demonstrated in Figure 3.1. Suppose that the demand curve for a monopolized industry is given as *DM* in Figure 3.1. Demand curve, *DM*, shows the quantities that can be sold at different prices.

For instance, if monopoly firm chooses price *OP*, the quantity that it can sell at this price is fixed at *OQ*—no other quantity can be sold at this price. Similarly, if it decides to sell quantity *ON*, its price is fixed at *OP'*—it cannot sell *ON* output at a higher price. This means that if demand curve is given, the options of monopoly firm become limited—it can choose either price or quantity at a time, not a price and a quantity inconsistent with the demand curve.

AR and *MR* Curves under Monopoly

The *AR* curve for a monopoly firm is the same as its demand curve. Since a monopoly firm faces a downward sloping demand curve, its *AR* also slopes downwards to the right. For example, the demand curve *DM* in Figure 3.1 is the same as the firm's *AR* curve.

What is much more *important* in the analysis of equilibrium of a monopoly firm is the *relationship between the AR and MR* curves. When price is fixed, as in case of perfect competition, firm's demand curve takes the form of a horizontal line. In that case, *AR = MR* and *MR* is a straight line too. But, in case of a monopoly firm, demand curve has a negative slope. Therefore, its *MR* curve too has a negative slope. There is, however, a *specific relationship* between *AR* and *MR*, *i.e.*, *the slope of MR curve is twice that of that AR curve*. That is, given the linear demand function, marginal revenue curve is twice as steep as the average revenue curve.

Short-run equilibrium of the monopoly: price and output determination

According to the traditional theory of firm, a firm is said to be in equilibrium where it maximizes its profit. As in case of perfect competition, equilibrium of a monopoly is studied under both *short-run* and *long-run* conditions.

The short-run equilibrium of monopoly can be explained by two approaches:

a) Total revenue–total cost *(TR–TC)* approach and
b) Marginal revenue–marginal cost *(MR–MC)* approach

a) *Monopoly Equilibrium by Total Revenue–Total Cost Approach*

According to the *total revenue–total cost* (*TR–TC*) approach, a profit maximizing monopoly firm is in equilibrium at the level of output and price at which it's *TR–TC = Total Profit* is maximum.

The equilibrium of monopoly by *TR–TC* approach is illustrated graphically in *Figure 3.2* under the following assumptions:

i. The monopoly firm faces a *cubic TC function* of the form $TC = F + bQ - cQ2 + dQ3$ (where *F = fixed cost*) and

ii. Its demand curve is given by a demand function of the form $Q = a - bP$.

When *TC function* is graphed, it produces a *TC* curve, from the demand function, derived a price function as $P = a/b - Q/b$. Using this price function, derived monopoly's *TR* function as $TR = (aQ – Q2)/b$.

The *TR function* when graphed produces a *TR* curve as shown in *Figure 3.2*. The *TC* curve shows monopoly's *total cost* at different level of output and *TR* curve shows its total revenue at different level of output and price. As *Figure 3.2* shows, the monopoly firm faces a loss till output *OQ*1 and beyond output *OQ*3 That is, monopoly's profitable range of output lies between *OQ*1 and *OQ*3 because it is only in this range of output that monopoly's *TR*>*TC*.

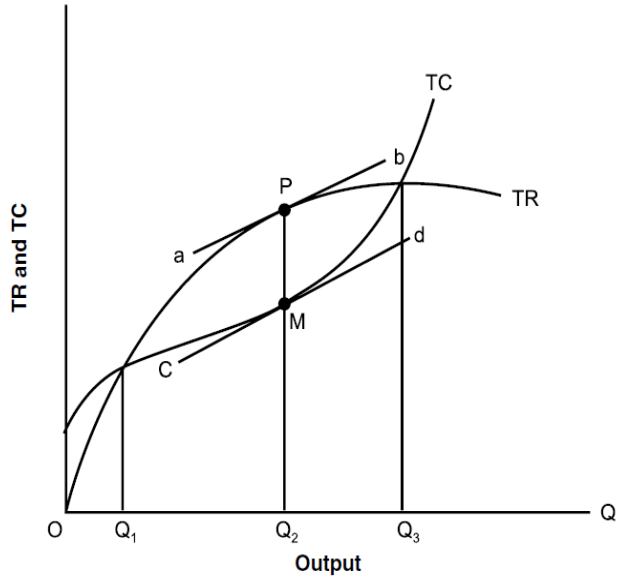


Figure 3.2 *Short-run Equilibrium of Monopoly: TR–TC Approach*

It is obvious that total profit is maximum where the vertical difference between *TR* and *TC* curves is maximum. The maximum difference between the *TR* and *TC* curves can be obtained by a simple technique, *i.e.,* by drawing *parallel tangents* to *TR* and *TC* curves as shown by the tangent *ab* and *cd*. Note that the line *ab* is tangent to the *TR* curve at point *P* and line *cd* is tangent to the *TC* curve at point *M* and line *ab* and *cd* are parallel. As a matter of rule, the vertical gap between tangential points *P* and *M* is maximum.

That is, given the revenue and cost conditions, the monopoly firm can make a maximum profit of *PM*. A line drawn from point *P*, through point *M* to *X*-axis determines profit maximizing output at OQ_2. It means that a profit maximizing monopoly reaches its equilibrium at output OQ_2. This equilibrium solution satisfies the necessary condition of profit maximization that profit is maximum where *MR* = *MC*. Recall that the slope of the *TR* curve gives $\partial TR/\partial Q = MR$ and the slope of the *TC* curve gives $\partial TR/\partial Q = MR$ at their respective points of tangency. Since tangents

ab and *cd* are parallel, their slopes are equal. It means that at the tangential points, *MR* = *MC*. This satisfies the necessary condition of profit maximization.

b) *Monopoly Equilibrium by MR–MC Approach*

The short-run revenue curves of the monopoly firm are shown by the *AR* and *MR* curves and its short-run cost curves are given by the *SAC* and *SMC* curves. The *AR* and *MR* curves can be derived from the *TR* function used in *TR–TC* approach. Similarly, *SAC* and *SMC* curves can be derived from the *TC* function.

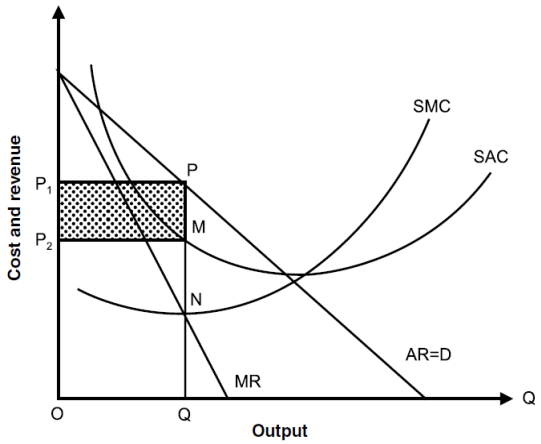


Figure 3.3 *Monopoly Equilibrium: MR–MC Approach*

Given the revenue and cost curves and the profit maximization rule, the equilibrium of the monopoly firm can easily be traced. Recall once again the profit maximization rule, i.e., profit is maximum where *MR* = *MC*. It can be seen in the figure that *MR* and *MC* curves intersect at point *N*. Note that point *N* satisfies both the conditions of profit maximization: (i) *MR* = *MC* and (ii) *MC* curve intersects *MR* curve from below. Point *N*, therefore, determines the equilibrium output and price.

An ordinate drawn from point *N* to *X*-axis determines the profit maximizing output at *OQ*. The ordinate *NQ* extended upwards to the *AR* curve gives the price *PQ* at which output *OQ* can be disposed of, given the demand function. Thus, the *MR–MC* approach to monopoly equilibrium determines both equilibrium output and price simultaneously. No other output and price can increase the monopoly's profit.

Once equilibrium price and output are determined, given the revenue and cost curve, the maximum monopoly profit can be easily determined as follows. Per unit monopoly profit = *AR-SAC.* In Figure 3.3, *AR* = *PQ* and *SAC* = *MQ.* By substitution, we get per unit monopoly profit = *PQ - MQ* = *PM.* Given the equilibrium output *OQ,* total monopoly profit = $OQ \cdot PM$

Since *OQ* = *P2M,* total monopoly profit at equilibrium can be worked out as $P_2M\ PM = P_1PMP_2$. *The total monopoly profit is shown by the shaded area in the Figure 3.3. Since cost and revenue conditions of the monopoly firm are supposed to be given, the monopoly equilibrium is supposed to be stable.*

Monopoly equilibrium in the long-run

The long-run equilibrium conditions of a monopolist are different from those faced by the competitive firms in another important respect. The main differentiating factor is the possibility of the entry of new firms into the industry. While in a competitive market, there is free entry to the industry, a monopoly firm is protected by the *barriers to entry*.

The barriers to entry may be in the form of *patent rights, legal protection, economies of scale* and *the well established long standing of the monopolist and its powers to eliminate the potential competitors by waging a price war.*

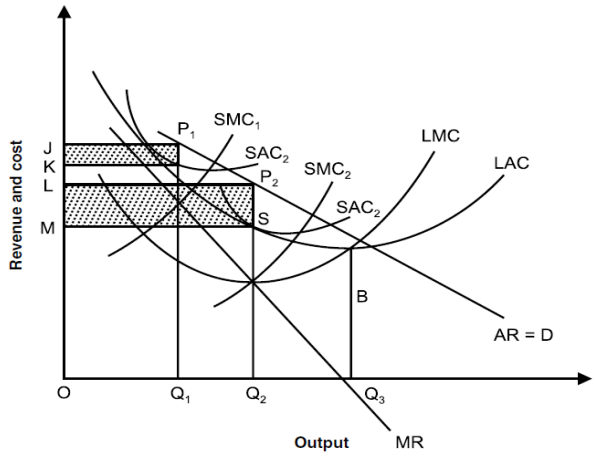


Figure 3.7 *Long-Run Equilibrium of the Monopoly Firm*

More importantly, a monopoly firm is free to choose between the alternatives available to it in the long run. The alternatives are whether to close down in case of losses or to continue in the business. If *SAC>AR*, the monopolist incurs losses in the short run. If market size is so small that

no plant-size can ensure pure profit in the long run (Figure 3.4), the monopolist goes out of business. If its *AR>SMC*1 it earns a short-run profit as shown at output *OQ*1 in *Figure 3.3*. The firm would, therefore, not only continue in the business but would also expand its business to the size that yields maximum profit in the long run.

A monopolist gets an opportunity to expand the size of its plant with a view to maximizing its long run profits. The expansion of the plant size may, however, be subject to such conditions as: (a) size of the markets; (b) expected economic profits and (c) risk of inviting legal restrictions.

A general case of monopoly equilibrium in the long run is presented in *Figure 3.7*, assuming none of the above conditions limits the expansion of monopoly firm. The *AR* and *MR* curves show the market demand and marginal revenue conditions faced by the monopoly. The *LAC* and *LMC* curves show the long-term cost conditions. As shown in the figure, the point of intersection between *LMC* and *MR* curves determine the equilibrium output at OQ_2. Given the *AR* curve, price is determined at *P*2Q_1 Thus the long-run equilibrium output is OQ_2 and equilibrium price is P_2Q_2 This price–output combination maximizes the monopolist's long-run profits. The total long-run profit has been shown by the area $LMSP_2$. Here the monopoly firm is in the long-run equilibrium. Price P_1Q_1 and output OQ_1 present the short-run equilibrium. Its total short-run profit is shown by the smaller shaded area.

PRICE DISCRIMINATION UNDER MONOPOLY

The theory of pricing under monopoly gives the impression that once a monopolist fixes the price of its product, the same price will be charged from all the consumers. This is, however, not the case generally. A monopolist, simply by virtue of its monopoly power, is capable of charging different prices from different consumers or groups of consumers.

When the same (or somewhat differentiated) product is sold at different prices to different consumers, it is called *price discrimination*. When a monopolist sells an identical product at different prices to different buyers, it is called a *discriminatory monopoly*. Consumers are discriminated in respect of price on the basis of their *income or purchasing power, geographical location, age, sex, quantity they purchase, their association with the seller, frequency of purchases, purpose of the use of the commodity or service*, and also *on other grounds* which the monopolist may find suitable.

Some common examples of consumers being discriminated on the basis of their incomes are found in the following cases:

i. Consulting physicians charge different fees from different clients on the basis of their paying capacity even if quantity and quality of service rendered is the same.
ii. Price discrimination on the basis of age is found in railways, roadways and airways: children between 3 and 12 years are charged only half the adult rates.
iii. Price discrimination on the basis of quantity purchased is very common. It is generally found that private businessmen charge lower price (or give discount) when bulk purchase is made.
iv. In case of public utility services, however, lower rates are charged when commodity or service is consumed in smaller quantity. *For example,* MNTL (New Delhi) charges lower rates on the night telephone calls, and DESU charges lower tariff rates on lower slabs of electricity consumption.
v. The most common practice of price discrimination is found in entertainment business, *e.g.* cinema shows, musical concerts, game shows, *etc.* Different rates are charged from different class of audience.

The product or service in question may be identical or slightly modified. For example, services of consulting *physicians and lawyers* are identical. The services of *railways*, *roadways* and *entertainment* shows may be slightly modified by providing more comfortable seats, sleepers, security and air-conditioning, *etc.* for the purpose of price discrimination.

The modification in service may involve some additional cost. But price differentials are much more than what is justified by cost differentials. Although price discrimination is a common practice under monopoly, it should not mean that this practice exists only under monopoly. Price discrimination is quite common also in other kinds of market structures, particularly when market imperfection exists. Most business firms discriminate between their customers on the basis of personal relationship, quantity purchased, duration of customer ship and so on.

Necessary conditions for price discrimination

Although price discrimination is a general practice adopted by the monopoly firms, it can be adopted successfully under the following conditions.

i. *Markets are so separated that resale is not profitable.*

The markets for different classes of consumers are so separated that buyers of low-price market do not find it profitable to resell the commodity in the high-price market. The factors that separate markets include are: (i) geographical distance involving high cost of transportation, *e.g.* domestic versus foreign markets; (ii) exclusive use of the commodity, *e.g.* doctor's services, entertainment shows, *etc.* and (iii) lack of distribution channels, *e.g.* transfer of electricity and gas.

ii. *Price elasticity of demand is different in different markets.*

If market is divided into sub-market, the elasticity of demand at a given price must be different in each sub-market. It is the difference in price elasticities that provides opportunity for price discrimination. If price elasticities of demand in different markets are the same, price discrimination would not be gainful.

iii. *The firm must have some monopoly power to control production and price. This condition applies also to discrimination under imperfect competition in the market.*

The monopoly firm must possess some monopoly power over the supply of the product to be able to distinguish between different classes of consumers and to charge different prices.

Degrees of price discrimination

The degree of price discrimination refers to the extent to which a seller can divide the market and can take advantage of market division in extracting the consumer's surplus. According to *Pigou*, there are three degrees of price discrimination practiced by the monopolists:

First-degree price discrimination;

Second-degree price discrimination and

Third-degree price discrimination.

First-degree price discrimination

The discriminatory pricing that attempts to take away the entire consumer surplus is called first-degree price discrimination. First-degree discrimination is possible only when a seller is in a position to know the price each buyer is willing to pay. That is, the monopolist knows buyer's demand curve for the product. What the seller does is that he first sets price at the highest possible level at which all those who are willing to buy purchase at least one unit each of the commodity.

When the consumer surplus of this section of consumers is exhausted, he gradually lowers down the price, so that the consumer surplus of the lower income groups can be extracted. This procedure is continued until the whole consumer surplus available at the price where *MR=MC* is extracted. In the case of services of exclusive use, *e.g.* medical services. A doctor, who knows or can guess the paying capacity of his patients, can charge the highest possible fee from visibly the richest patient and the lowest fee from the poorest one. The first degree of price discrimination is the limit of discriminatory pricing.

Second-degree price discrimination

The second-degree price discrimination is adopted by the monopolist when marked fairly large and divided according to different kinds of buyers but the ability-to-pay of different groups is unknown. The second degree of discriminatory pricing is adopted also to charge different prices for the different quantities of purchase. The second-degree price discrimination is also called *'block pricing system'*. A different price is charged from different category of consumers.

A monopolist adopting the second-degree price discrimination intends to take only the major part of the consumer surplus, rather than the entire of it. The second-degree price discrimination is feasible where

a) The number of consumers is large and price rationing can be effective, as in case of utilities like telephones, natural gas and also consumer durables;
b) Demand curves of all the consumers are identical and
c) A single rate is applicable only for a group of large number of buyers.

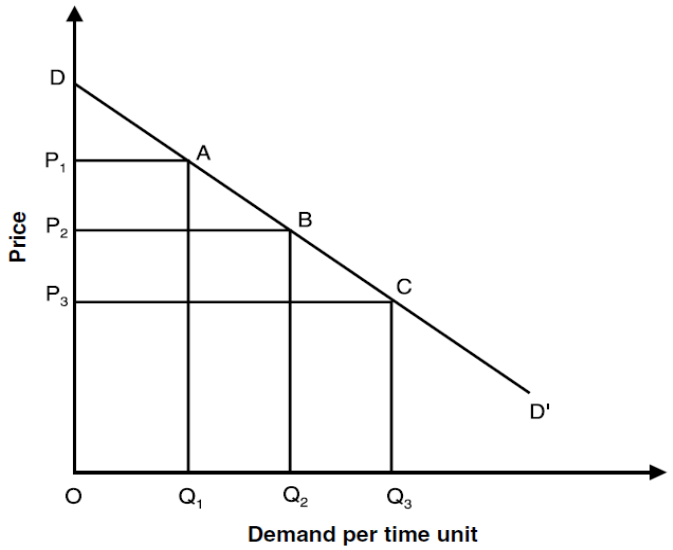


Figure 3.14 *Second-Degree Price Discrimination*

A monopolist using second-degree price discrimination sets the price first at OP_1 and sells OQ_1 over a period of time. After this market segment is exhausted, the monopolist sets a lower price, OP_2, and sells $Q1O_2$ units over a period of time to the second market segment. After the sale of $Q1Q2$ additional units, the monopolist sets still a lower price, OP_3, for the third lowest segment of the consumer group to sell an additional quantity of Q_2Q_3 units. Thus, by adopting a *block-pricing system*, the monopolist maximizes his total revenues (TR) as

$$TR = (QO_1 \times OP_1) + (Q_1O_2 \times OP_2) + (Q_2O_3 \times OP_3)$$

If monopolist is restrained from price discrimination and is forced to choose any one of the three prices—OP_1, OP_2, or OP_3 —his total revenue will be much less.

Third-degree price discrimination

When a profit maximizing monopoly firm sets different prices in different markets having demand curves with different elasticities, it is using third-degree price discrimination.

A monopolist is often faced with two or more markets, completely separated from each other—each having a demand curve with different elasticity. Therefore, a uniform price cannot be set for all the markets without loosing the possible profits. The monopolist, therefore, allocates total output between the different markets and fixes different prices, so that profit is maximized in each market. Profit in each market would be maximum only when $MR=MC$ in

each market. The monopolist, therefore, allocates its total output between the markets in such proportions that in each market *MR = MC*.

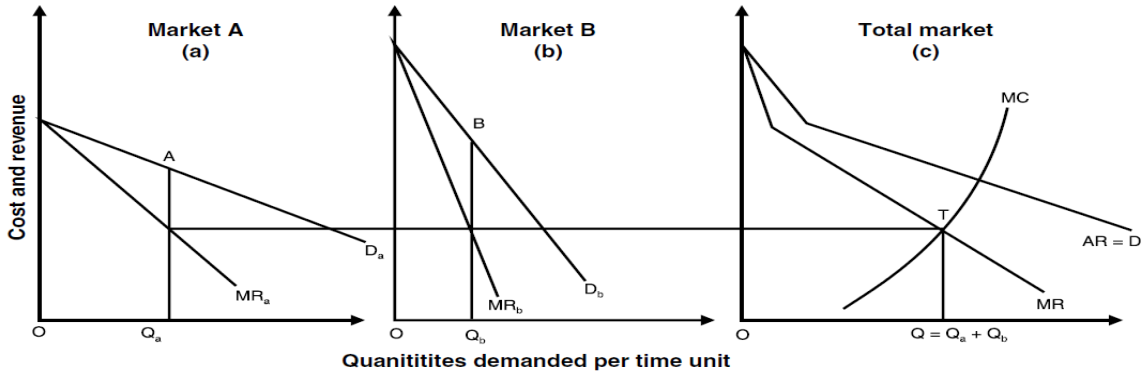


Figure 3.15 *Third-Degree Price Discrimination*

The process of output allocation and determination of price for different markets is illustrated in *Figure 3.15.* Suppose that a monopolist has to sell its goods in only two markets, *A* and *B*. The two markets are so separated that resale of commodity is not feasible.

The demand curve (*Da*) and marginal revenue curve (*MRa*) given in *Figure 3.15(a)* represent the *AR* and *MR* curves in market *A* and the curves *Db* and *MRb* in *Figure 3.15(b)* represent the *AR* and *MR* curves, respectively, in market *B*. The horizontal summation of demand curves *Da* and *Db* gives the total demand curve for the two markets, as shown by the curve *AR=D*, and the horizontal summation of *MR*d and *MRb* as given by the curve *MR* in *Figure 3.15(c).* Firm's marginal cost is shown by *MC*. Given the *MR* and *MC* curve, the firm's profit maximization output is determined by the intersection of the two curves. *Panel (c) of Figure 3.15* shows *MR* and *MC* curves interested at point *T* determining the equilibrium output at *OQ*. Thus, the optimum level of output for the firm is determined at *OQ*. The whole of *OQ* cannot be profitably sold in any one market because of their limited size.

Therefore, the monopolist has to allocate output *OQ* between the two markets in such proportions that the necessary condition of profit maximization is satisfied in both the markets, *i.e.,* on both the markets *MC=MR*. The profit maximizing output for each market can be obtained by drawing a line from point *T*, parallel to *X*-axis, through *MRb* and *MRa*. The points of intersection on curved *MRa* and *MRb* determine the optimum share for each market. As shown in

Figure 3.15, the monopolist maximizes profit in market *A* by selling *OQa* units at price *AQa* and by selling *OQb* units in market *B* at price *BQb*.

The firm's total equilibrium output is $OQ = OQa + OQb$. Since at *OQa*, $MRa = MC$ in market *A*, and at *OQb*, $MRb = MC$ in market *B*,

$$MC = TQ = MRa = MRb$$

Thus, the equilibrium condition is satisfied in both the sub-markets and the monopoly firm adopting the third degree method of price discrimination maximizes its profits.

The third degree method of price discrimination is most suitable where the total market is divided between the home and the foreign markets. However, it may be suitably practiced between any two or more markets separated from each other by any two or more of such factors as *geographical distance*, *transport barriers*, *cost of transportation*, *legal restrictions on the inter-regional or inter-state transfer of commodities by individuals*, etc.

MONOPOLISTIC COMPETITION

Introduction

Monopolistic competition refers to a market structure in which a large number of sellers sell differentiated products, which are close substitutes for one another. Monopolistic competition combines the basic elements of both perfect competition and monopoly.

The element of monopoly in monopolistic competition arises from the fact that each firm has an absolute right to produce and sell a *branded* or *patented* product. Other firms are prevented by laws from producing and selling a branded product of other firms.

The element of competition comes from the fact that each branded product has several close substitutes and firms selling branded products of the same generic category have to compete for the market share.

One index of the competition between them is the amount that they spend advertising their product. Some of the industries looking monopolistically competitive may be *oligopolistic* in which there are only a few sellers selling differentiated or homogenous products.

Characteristics

Monopolistic competition combines the elements of both *perfect competition* and *monopoly power*, therefore, the main characteristics of monopolistic competition are the blend of perfect competition and monopoly.

The main features of monopolistic competition *vis-à-vis* perfect competition and monopoly are described as:

i. Product differentiation
ii. Large number of sellers
iii. Free entry and free exit
iv. Selling costs
v. Downward sloping demand curve.

i. *Product differentiation*

Product differentiation is the basis of and the main distinctive characteristic of monopolistic competition that distinguishes it from monopoly and perfect competition. Under monopolistic competition, the firms differentiate their products from one another in respect of their *shape, size, colour, design, minor qualitative differences, efficiency in use, some extra facility, packaging, after-sale-service, guarantee and warrantee and so on.*

Product differentiation may be real or fanciful and spurious. The basic purpose of product differentiation is to make the consumers believe that a product is different from others and, thereby, to create brand loyalty of the consumers. Product differentiation affects firm's demand curve in a significant way.

ii. *Large number of sellers*

Under monopolistic competition, the number of sellers is *large*. How large? It is difficult to specify number of firms: it may be 10, 20 or more depending on the size of the market. However, the question 'how large' can be answered in conceptual terms with reference to perfect competition. Under perfect competition, the number of sellers is so large that a firm becomes a *price taker*.

In contrast, under monopolistic competition, the number of firms is only so large that a firm retains its power to be a *price maker*. The monopolistically competitive firms have the power to set the price of their product depending on the objective of the firm.

iii. *Free entry and free exit*

As in case of perfect competition, there is no barrier on the entry of new firms and exit of old ones from the industry. New firms are free to enter the monopolistically competitive industry and to quit at will.

Entry of new firms reduces the market share of the existing ones and exit of firms does the opposite. These consequences of free entry and free exit lead to intensive competition among the firms for retaining as well as increasing their market share.

iv. *Selling costs*

Unlike firms under *perfect competition* and *monopolies*, firms under monopolistic competition make heavy expenditure on advertisement and other sales promotion schemes for their product. This is an important feature that distinguishes monopolistic competition from *perfect competition* and *monopoly*. Selling costs include all the expenditure on advertisement, sales promotion schemes and salaries of sales personnel.

v. *Downward sloping demand curve*

As in case of *monopoly*, a *monopolistically* competitive firm faces a downward sloping demand curve. The reason is that a monopolistically competitive firm can, by exercising its monopoly power, increase its price and still retain some buyers with brand loyalty and can increase the demand for its product by decreasing the price because of a relatively higher cross-elasticity of the competitive product.

Basic elements

There are three basic elements of monopolistic competition such as:

i. *Product differentiation and firm's perceived demand curve;*

Product differentiation is the basis of competition among the monopolistically competitive firms. A general class of product is differentiated if any significant basis exists for distinguishing the goods (or services) of one seller from those of others. Such a basis may be real or fancied, as long as it is of any importance to buyers, and leads to a preference for one variety of product over another.

The basic purpose of product differentiation is to make customers distinguish the product of a firm from those of the others in the industry and to develop a preference or band loyalty. Once brand loyalty is developed, it alters the course of the demand curve for the product.

Product differentiation aims at changing the slope and position of the demand curve for the product and converting it from a horizontal demand line (as under perfect competition) to a downward sloping demand curve. The downward sloping demand curve gives the firm power to use its discretion in fixing the price of its product.

The preview demand curve

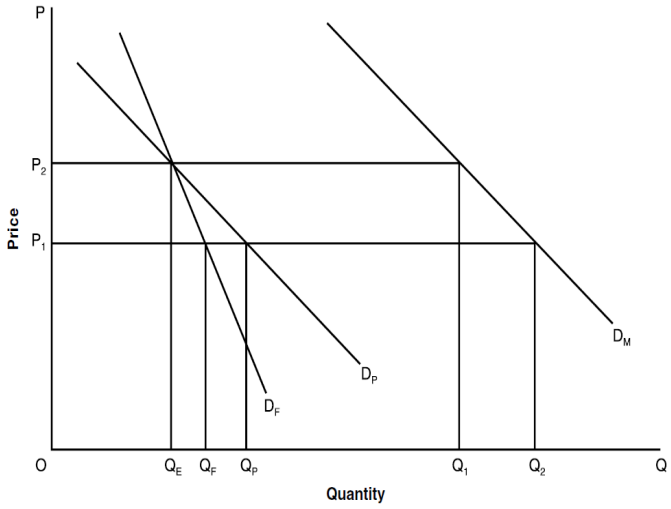


Figure 4.1 *Demand Curves for the Industry and the Firms*

Under monopolistic competition with product differentiation, each firm perceives that the demand curve for its own product is more elastic than that of the rival firms. This aspect is illustrated in *Figure 4.1.* Suppose industry's demand curve (*i.e.* market demand curve) is given by the curve D_M. If industry demand is proportionately divided between the firms, each firm is supposed to have a demand curve shown by D_F.

However, firms under monopolistic competition do not take D_F to be the demand curve for their individual product. Each firm perceives that the demand curve for its own product is *more elastic* than that of the other firms. Its *perceived demand curve* is shown by the demand curve D_P. Note that the *perceived demand curve*, D_P, is more elastic than the demand curve D_F. The basis of the perceived demand curve is the firm's belief that if it changes the price of its own product, it will go unnoticed by the other firms and they will not react to change the price of their products. The reason for this assumption is the firm's belief that the number of firms is so large that price changes made by a single firm is very much likely to go unnoticed by the rival firms.

Given the industry demand curve *DM* in *Figure 4.1,* if price for the industry (the 'product group') is given at OP_2, the demand for the industry as a whole will be $OQ1$ and each firm will be selling an equal quantity, OQ_E. Given this price–quantity combination, an individual firm perceives that if it cuts down the price of its own product to OP_1, the other firms will not change their price, and then the demand for its product will increase by Q_EQ_P. The additional increase in

the demand for its product is the result of cross-elasticity, *i.e.,* when one firm decreases its substitute price and other firms do not, its substitute product becomes relatively cheaper.

Since products of all the firms are close substitutes for one another, some customers will switch over from the constant-price products to the product which has a lower price. Similarly, if only one firm increases its price, it looses its customers to other firms. As a result, demand for its product decreases more than indicated by the demand curve D_F. This holds for all price changes and gives rise to a perceived demand curve D_P. The perceived demand curve plays a significant role in price and output determination in monopolistic competition with price competition.

ii. *Selling costs and firm's cost structure and*

The selling cost incurred by the firms under monopolistic competition and its impact on their cost structure is another important aspect of *Chamberlin's* theory of monopolistic competition. Introduction of selling costs in the theory of price and output determination under monopolistic competition is another innovative contribution made by *Chamberlin*. *Chamberlin* defines *selling costs* as 'costs incurred in order to alter the position or the slope of the demand curve for a product'. *Chamberlin's* concept of selling costs is not exactly the same as advertisement cost: it is advertisement cost plus. By *Chamberlin's* definition, selling costs include:

i. Cost of advertisement;
ii. Expenditure on sales promotion schemes (including gifts and discounts to buyers);
iii. Salary and commission paid to sales personnel;
iv. Allowance to retailers for displays and
v. Cost of after-sale-services.

Also, *Chamberlin* distinguishes *selling costs* and *production cost* on the basis of their basic purpose and functions. According to Chamberlin, costs that are incurred to create a product or service of utility and making it available to the consumers are *production costs*. In Chamberlin's perception, production cost includes also the cost of transportation. The basic function of the production cost is to create a commodity and to make it available to the consumers. The selling costs, on the other hand, perform the following functions:

i. Informing potential buyers about the availability of the product;
ii. Increasing demand for the product by attracting customers of the rival products and
iii. Making the demand curve shift upward.

What is more important in price and output determination is the effect of selling costs on the total cost that figure in pricing decisions. In his model of monopolistic competition, *Chamberlin* assumes the traditional *U-shaped* cost curves—*AC*, *AVC* and *MC*—and also a *U-shaped average selling cost* (*ASC*) curve. *ASC* is defined as *SC/S* (where *SC* = *selling costs* and *S* = *sales*). The *U-shaped ASC* curve is illustrated in *Figure 4.2.*

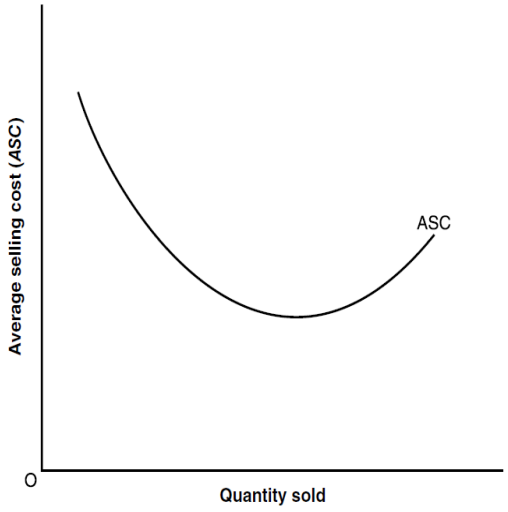


Figure 4.2 *The Average Selling Cost Curve*

As *Figure 4.2* shows, the *ASC* first decreases until it reaches its minimum and then begins to increase. In the beginning, it is very high because a little selling cost (or advertisement cost) is not effective enough to promote sales. With increase in selling costs, however, sales increase at a rate higher than the rate of increase in selling costs. As a result, *ASC* decreases. The decrease in *ASC* is attributed to 'increasing returns' to advertising and economies of scale in advertisement cost.

It must, however, be noted that returns to selling cost are determined by the following factors:

Price of the product. A high price makes selling cost less effective;

Price of the substitute. A lower price of the substitutes makes selling cost less productive;

Buyers' income. Advertising a costly product in low-income society has no pay-off and

Buyers' loyalty to rival brands. The stronger the loyalty to rival brands, the lower the cross elasticity and the less effective the selling cost.

However, even if all these factors are favorable, a stage is finally reached when returns to selling costs, especially to advertisement cost, tends to become constant. This is the stage of saturation. The stage of saturation may be marked by the size of the market and/or competitive advertisement by the rival firms.

Increasing selling costs or advertisement cost at the stage of saturation tends to become less and less effective in attracting more buyers. Therefore, sales increase at a much lower rate than the increase in the selling costs. Consequently, *ASC* begins to increase and goes on increasing. This is how *ASC* gets its *U-shape*, as shown in *Figure 4.2.* The *ASC* curve is added to the *AC* curve in determining the profit maximizing level of output and price.

Optimum level of advertising cost

The optimum level of advertising cost is determined by the objective of the firm—whether it is *profit maximization*, *retaining market share* or *countervailing the advertisement by the rival firms*.

Optimization of advertising expenditure is under the following assumptions.

a) Objective of the firm is to maximize its profit;
b) Price of the product is given;
c) Average production cost (*APC*) and *MPC* curves are given and
d) Average selling cost (*ASC*) curve is also known

Under these conditions, the optimum level of selling cost is determined where the firm's overall marginal cost (*MC*) including 'marginal cost of production' (*MCP*) and marginal cost of advertising (*MCA*) equals the price. Since price is given by assumptions, $price = AR = MR$. Thus, the condition for the optimum selling cost can be expressed as $MC = MPC + MAC = AR = MR$.

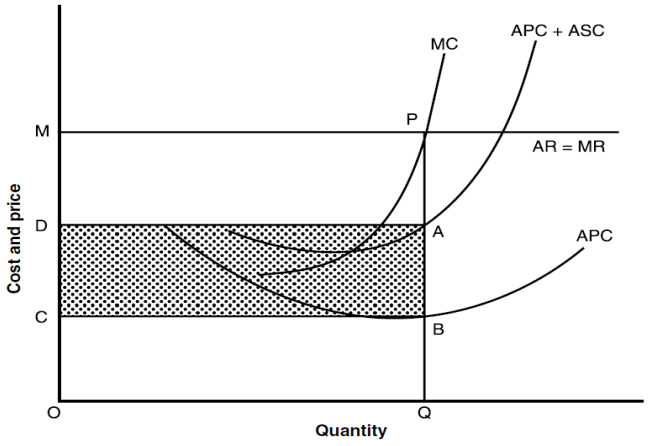


Figure 4.3 *Optimization of Selling Cost*

The determination of the optimum selling cost is illustrated in *Figure 4.3*. The price of the product is given at *OM* and firm's *APC* is shown by the curve *APC*. The firm's *APC* curve added vertically with *ASC* curve (not given in the figure) is shown by the curve labeled *APC + ASC*. The vertical distance between the *APC* and *APC + ASC* gives the measure of the *average selling cost* (*ASC*). For example, at output *OQ*, the *ASC* equals *AB*. Finally, firm's overall *MC* (associated with *APC + ASC)* is shown by the *MC* curve.

As *Figure 4.3* shows, price line, *AR=MR*, and *MC* curve intersect at point *P* determining the equilibrium price at *OM*. An ordinate drawn from point *P* to the quantity axis determines the profit-maximizing output at *OQ*. Once profit-maximizing output is determined, the optimum level of all costs (given the cost curves) is automatically determined. It can be seen in *Figure 4.3* that at profit-maximizing output *OQ*, price equals *OM=PQ=BQ+AB+AP*. Note that at output *OQ*, average selling cost (*ASC*) equals *AB*. That is, at the profit-maximizing level of output, average advertising cost equals *AB*. Therefore, *AB* is the *optimum average selling cost* (*ASC*). The total optimum selling cost or advertisement expenditure can be obtained by multiplying *ASC* with output *OQ*. That is, *total optimum advertisement expenditure* = *OQ*×*AB*. Since *OQ* = *CB*, the total optimum selling cost can be written as *CB*×*AB* = *DABC*, as shown by the shaded area.

iii. Product differentiation and the concept of industry/ Concept of industry and product groups

An industry under perfectly competitive conditions is defined as a group of firms *producing a homogeneous product*. But, this concept of industry cannot be applied to the cases where *products are differentiated*. Where products are differentiated—slightly or substantially—each firm is, in a sense, an industry in itself, exactly as a monopoly firm is an industry in itself.

Since under perfect competition, product is homogeneous, demand curve for an industry can be obtained by adding individual demand curve of individuals firms. But, under monopolistic competition, product is made heterogeneous through product differentiation techniques and in case of heterogeneous products, the demand for individual products cannot be added to obtain market demand and supply curves.

The monopolistically competitive industry defined as a *'group'* of firms producing a 'closely related' commodity, called *product group*. The products of the 'group' must be *close, technological and economic substitutes*. The two products are *technological substitutes* for each other if they technically satisfy the same want.

Firm's equilibrium under monopolistic competition

Under monopolistic competition, however, a firm can alter its sales prospects by the following *three methods*:

i. *By changing the price of its product;*

As regards to changing price, since a firm under monopolistic competition faces a downward sloping demand curve with elasticity less than infinity, it has the option to change the price.

ii. *By changing 'the nature of the product' and*

In regard to changing the nature of the product, a firm can do it by changing the quality of its product by making technical changes, by introducing a new design, by using superior material, by making a new style of packaging, by establishing a close link with buyers and so on.

iii. *By incurring the advertisement outlays.*

As regards the methods of sales promotion, a firm can increase its sales by prompt and courteous service, credit facilities and by enhancing expenditure on advertisement.

While making changes in price and output is a short-run phenomenon, changing the quality of the product and attracting larger number of buyers are long-run phenomena. Therefore, Chamberlin's theory (monopolistic competition) of price and output determination is discussed under *short-* and *long-run* conditions.

Assumptions

i. There are a large number of firms selling slightly differentiated products, which are close substitutes for one another.
ii. The number of firms in a product group is so large that their activities, especially strategies of price and output, go unnoticed by the rival firms.
iii. Demand and cost curves for all the products and for all the firms of the group are uniform, *i.e.,* firms face identical demand (including *perceived* one) and cost curves.
iv. Consumer's preferences are evenly distributed among the different products and product differentiations are not such that they make a difference in cost.

Short-run equilibrium of the firm

The *short-run equilibrium* of the firm under monopolistic competition is illustrated in *Figures 4.4* and *4.5*. *Figure 4.4* illustrates how firms in the state of disequilibrium adjust their price and output to move to the state of equilibrium. *Figure 4.5* presents the final position of the firm's equilibrium and also the determination of equilibrium prices and output.

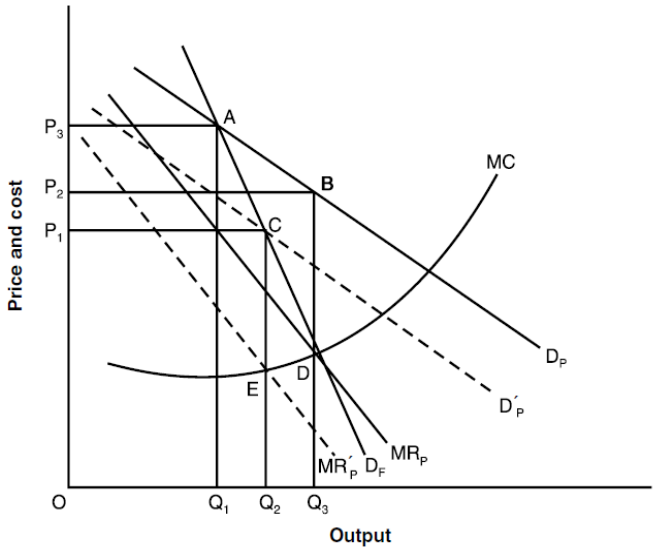


Figure 4.4 *Price and Output Adjustments in the Short-run*

In *Figure 4.4*, the firm's perceived and proportional market demand curves are given by the curves *DP* and *DH*, respectively. The *MRP* curve shows their *marginal revenue* curve corresponding to their perceived demand curve (*DP*) and their *marginal cost curve* is shown by the curve *MC*. Now suppose that in the short-run market price is given at *OP*3 determined by the firms themselves or by 'custom', and all the firms are at point *A*, each one selling output *OQ*1. Incidentally, *OQ*1 equals *DM*/*n* (where *DM* is the total market demand for the industry as a whole and *n* is the number of firms.

Under monopolist competition, however, this position is *not* stable because each firm, given its perceived demand curve (*DP*), expects to sell more by cutting down the price of its product. Therefore, they cut down their price in anticipation of increase in their sales. They continue to cut down their price until price and output reach the level at which they maximize their profits. The profit-maximizing quantity and price is determined in *Figure 4.4* at point *D*, the point of intersection between firms' *MC* and *MR* (perceived) curves. Point *D* determines the profit-maximizing output at OQ_3 and price at OP_2. That is, each firm produces OQ_3, and charges price at OP_2.

It can be seen in *Figure 4.4* that total production based on perceived demand has increased very heavily—by ($Q_1Q_3 \times n$). The firms find that there is excess production even at a lower price. In

order to get rid of their surplus production, the firms begin to cut down their price. As a result, their perceived demand curve (D_P) shifts downward as shown by the dashed perceived demand curve (D'_P). This shift continues until the firms reach the level of final demand curve (D'_P) and they reach their final equilibrium through point *E* in *Figure 4.4.*

The final short-run equilibrium position of the firm under monopolistic competition is illustrated in *Figure 4.5.* In this figure, the firm's perceived demand curve and the corresponding marginal revenue curve are shown by *DP* and *MRP* curves, respectively, its short-run average and marginal cost curves are shown by the *SAC* and *SMC* curves, respectively.

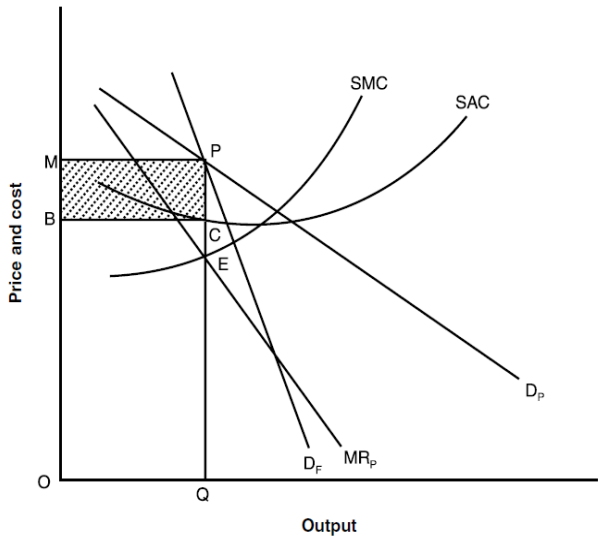


Figure 4.5 *Short-run Equilibrium of the Firm Under Monopolistic Competition*

As *Figure 4.5* shows, the *MR* and *MC* curves intersect at point *E* determining the profit-maximizing output at *OQ* and price at *M*. Each firm produces *OQ* and sells at price *PQ*. At this price and output, the firms maximize their short-run profit. The firms are, therefore, in short-run equilibrium at point *E*. Their maximum pure profit equals *BC×PC=MPCB* as shown by the shaded area. Any other price and output will reduce their total profit. Therefore, firms have no incentive to change their price and output. Note that the final price is determined at the point of intersection between D_P and D_F curves. This, however, should not mean that all the firms in monopolistic competition make pure profits in the short run. Chamberlin does not rule out the possibility of some firms making losses.

Long-run equilibrium of the firm

The long-run conditions differ from the short-run conditions because in the long run:

i. new firms enter the industry,
ii. firms indulge in price competition,
iii. changes *(i)* and *(ii)* take place simultaneously and
iv. firms advertise their product more vigorously.

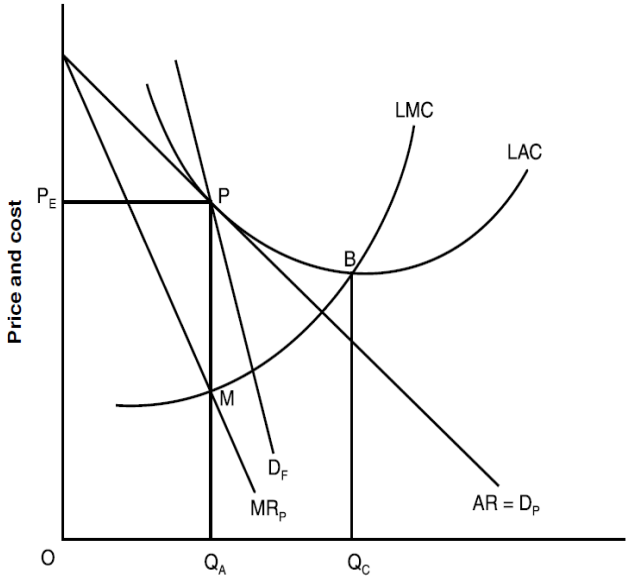


Figure 4.6 *Long-run Equilibrium of Firms Under Monopolistic Competition*

The long-run equilibrium of the firms under monopolistic competition is illustrated in *Figure 4.6.* The revenue and cost curves in Figure 4.6 are similar to those given in *Figure 4.5.* As *Figure 4.5* shows, each firm makes a *pure* or *supernormal profit* of *PC* per unit of output, *i.e.,* to the extent of the difference between the price and *SAC*. Supernormal profit attracts new firms to the industry as there is no barrier to entry. With the entry of new firms, the existing firms loose a part of their market share to the new entrants. As a result, the proportional demand curve (*DP*) of the firms shifts leftward and perceived demand curve (*DP*) shifts downward. This shift continues until firms reach a new equilibrium point with their perceived *MRP=MC*. This stage is shown by point is *M* when *MC* intersects with *MRP*. Point *M* determines the equilibrium output at *OQA* and price at *PQA*. Note that at point *P*, demand curve *DP* is tangent to *LAC* curve and demand curve *DP* intersects with them. Since price equals *LAC* at point *P*, no firm is making pure profit.

There is, therefore, no incentive for new firms to enter the industry nor is there any disincentive for the existing ones to quit the industry. The firms are, therefore, in the state of their long-run equilibrium at point *P*. It is important to note here that point *P* is not imaginary or coincidental. It is determined by the intersection of the *MR* and *MC* curve determining profit-maximizing output at *OQA* and price at *PQA*. At their equilibrium, firms produce and sell quantity *OQA* at price *PQA* = *OPE*. An important outcome of this analysis is that, in the long run all firms in monopolistic competition make only normal profit.

Wastes in monopolistic competition

The following ones are the six major wastes of monopolistic competition. They are:

i. *Competitive advertisement*

One of the important wastes of monopolistic competition is the incurring of expenditure on competitive advertisement by firms. Excess advertisement adds to costs and prices. Expenditure on packing, *colour*, *flavour*, *etc*. and on media like *TV*, *radio*, *cinema*, *newspapers*, *etc.* create unnecessary product differentiation.

As a result, irrational preference for certain brands of products are created in the minds of consumers which tend to push the sales of one firm at the cost of others. Expenditure of competitive advertisement is also resorted to by all firms at least to keep their respective customers attached to their brand of the product. But all such expenditure is socially wasteful.

ii. *Product differentiation*

Another waste of competition is the production of varieties of a product which each firm produces. This is done by creating artificial or imaginary product differentiation so as to distinguish the product of one seller from those of another. This is done by changing the *colour, design, fragrance, packing, etc.* of the same product by the same producer. For instance, The Brooke Bond Tea Company sells such brands of tea as *Green Label*, *Red Label*, *Yellow Label*, *etc.*

Due to product differentiation each firm produces varied collection of types and qualities for its own customers and often confusing them. Rather than producing only one type of product and

charging uniform price, they charge different prices for each brand of the same product. Thus a large number of brands, styles, etc. confuses the consumer and adds to costs and prices, thereby making the products costly. This leads to wastage of resources and to loss of economic efficiency.

iii. *Expenditure on cross transportation*

The expenditure on cross transport is another waste of monopolistic competition. Each producer tries to sell his products in the far-off markets rather than in the markets near its place of manufacture. This involves huge transport costs and also expenses on advertisement and propaganda. Rather than save these expenses and reduce prices, firms under monopolistic competition prefer to incur expenses on transportation and advertisement. This is apparently wastage of resources.

iv. *Inefficient firms*

Under monopolistic competition, there is a large number of inefficient firms. The price charged by each firm exceeds the long-run marginal cost because both the *AR* and *MR* curves are downward sloping under monopolistic competition. The firm's equilibrium condition is *Price=LAC>LMC=MR*. Therefore, resources are under allocated to firms in the market and misallocated in the economy.

Under perfect competition all firms are of the most efficient size in the long-run because *P=LAC=LMC=MR*. Moreover, under monopolistic competition, an inefficient firm will have to lower its price in order of sell more and to expand. For this, it will have to lower its average costs per unit. But an inefficient firm may not be in a position to lower its average costs per unit and to lower its price. Thus such firms may continue to exist on the strength of their customers but without attracting the customers of their rivals.

v. *Excess capacity*

All firms under monopolistic competition possess excess capacity. Since the demand curve *(AR)* of a monopolistic competitive firm is downward sloping, its tangency point with the *LAC* curve will always occur to the left of its minimum point. Thus when the firm is in long-run

equilibrium, it underutilizes its optimum scale plant. This leads to the existence of more firms in the industry than required.

All firms work under less than the optimum capacity, and all charge higher than the competitive price. The failure of the firms to produce less than the optimum output due to a downward sloping demand curve is a clear wastage of resources.

vi. *Unemployment.*

As a corollary to the above, unutilized resources lead to unemployment when firms under monopolistic competition try to maintain the price of their product instead of maintaining production.

OLIGOPOLY

Definition

Oligopoly is the market structure in which there are a few sellers selling homogeneous or differentiated products. However, economists do not specify what number of sellers make the market oligopolistic. In fact, it depends on the market size. However, two sellers is the *limiting case* of oligopoly. When there are only two sellers, the market is called *duopoly*.

In any case, if 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, if 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, cigarettes, etc., are some examples of industries characterized by *differentiated* or *heterogeneous oligopoly*.

Factors Causing Oligopoly

The main factors that lead to the growth of oligopoly are described here briefly.

i. *Huge Capital Investment.*

Some industries are by nature capital intensive, e.g., firms manufacturing automobiles, aircraft, ships, TV sets, refrigerators, steel and aluminium goods, etc., and hence require huge investment. Therefore, only a few firms can enter these kind of industries. In fact, a huge investment requirement works as a natural barrier to entry to the oligopolistic industries.

ii. *Economies of Scale.*

By virtue of huge investment and large scale production, large units enjoy absolute cost advantage due to economies of scale in their purchase of industrial inputs, acquiring external finance, and in sales organization. This gives the existing firms a comparative advantage over new firms, especially in price competition. This works not only as a deterrent for the entry of new firms, but also causes exit of high cost firms.

i. *Patent Rights.*

In case of differentiated oligopoly, firms get their differentiated product patented which gives them monopoly power, i.e., 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.

ii. *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.

iii. *Merger and Acquisition.*

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 an important factor creating oligopoly. In fact, in modern times, this is the most important

factor that gives rise to oligopolies and strengthens the oligopolistic tendency in modern industries.

Features of Oligopoly

i. *Small Number of Sellers.*

As already mentioned, there are a small number of sellers under oligopoly. How small the number of sellers is not given precisely: 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 so large that a single firm can influence the market price and the business strategy of its rival firms. The number may vary from industry to industry.

ii. *Interdependence of Decision Making.*

The most striking feature of an oligopolistic market structure is the interdependence of business decision of oligopoly firms. 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 counteraction in the absence of collusion between the firms. Since the number of firms in the industry is small, the business decision and strategy of each firm in respect of pricing, advertising, product modification is closely watched by the rival firms and it evokes retaliatory actions. What is equally important in strategic business decisions is that firms initiating a new business strategy anticipate and take into account the counteraction by the rival firms. This is called interdependence of oligopoly firms.

iii. *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) resistance by the established firms by 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 franchise branches of MNCs.

iv. *Indeterminate Price and Output.*

Another important feature, though controversial, 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, collusion may last long or it may breakdown. An opposite view is that price under oligopoly is sticky, i.e., if price is once determined, it tends to stabilize.

The Oligopoly Models: An Overview

Under oligopolistic conditions, rival firms adopt an intricate pattern of actions, reactions and counteractions showing a variety of behavioral patterns. The uncertainty arising out of unpredictable behaviour, actions and reactions of oligopoly firms makes systematic analysis of oligopoly an extremely difficult task. As Baumol puts it,

Under (these) circumstances, a very wide variety of behaviour pattern becomes possible. Rivals may decide to get together and cooperate in the pursuit of their objectives … or, at the other extreme, may try to fight each other to death. Even if they enter an agreement, it may last or it may breakdown.

Economists have, therefore, found it extremely difficult to make a systematic analysis of price and output determination under oligopoly. The complexity of the problem, however, could never deter the economists from their efforts to find a reasonable solution to the problem.

In accordance with the wide variety of behaviour patterns, economists have developed a variety of analytical models based on different behavioural assumptions. The widely quoted models of oligopoly include Cournot's duopoly model (1838), Bertrand's duopoly model (1880), Edgeworth's duopoly model (1897), Stackelberg's model (1933),

Sweezy's kinked demand curve model (1939), Neumann and Margenstern Game Theory model (1944), and Baumol's sales maximization model (1959). None of these models, however, provides a universally acceptable analysis of oligopoly, though these models do provide an insight into oligopolistic behaviour.

DUOPOLY

The uncertainty is respect of behaviour pattern of a firm under oligopoly arising out of their unpredictable action and reaction makes a systematic analysis of oligopoly difficult. However, classical and modern economists have developed a variety of models based on different behavior assumptions. These models can broadly be classified into two categories classical duopoly models and modern oligopoly duopoly models, when there are only two sellers a product, there, exists duopoly.

Duopoly is a special case of oligopoly. Duopoly is a special case in the sense that it is limiting case of oligopoly as there must be at least two sellers to make the market oligopolistic in nature.

a) The Cournot's Duopoly Model
b) The Chamberlin Duopoly Model
c) The Bertrand's Duopoly Model
d) The Edgeworth Duopoly Model

MONOPSONY

A monopsony occurs when a firm has market power in employing factors of production (e.g. labour). A monopsony means there is one buyer and many sellers. It often refers to a monopsony employer – who has market power in hiring workers. This is a similar concept to monopoly where there is one seller and many buyers.

Monopsony in Labour Markets

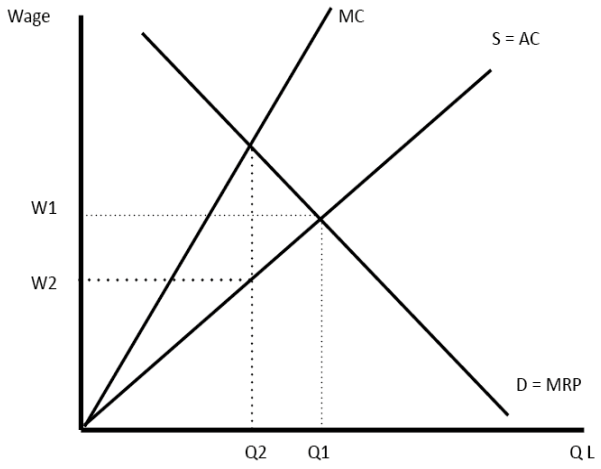
An example of a monopsony occurs when there is one major employer and many workers seeking to gain employment. If there is only one main employer of labour, then they have market power in setting wages and choosing how many workers to employ.

Examples of monopsony in labour markets

Coal mine owner in town where coal mining is the primary source of employment.

The government in the employment of civil servants, nurses, police and army officers.

Diagram of monopsony



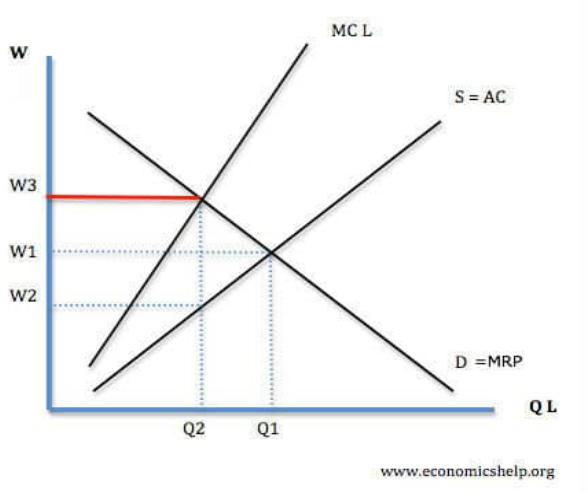
In a competitive labour market, the equilibrium will be where $D=S$ at Q_1, W_1. However, a monopsony can pay lower wages (W_2) and employ fewer workers (Q_2).

Profit Maximization for a Monopsony

The marginal cost of employing one more worker will be higher than the average cost – because to employ one extra worker the firm has to increase the wages of all workers. To maximize the level of profit, the firm employs Q_2 of workers where the marginal cost of labour equals the marginal revenue product $MRP = D$. In a competitive labour market, the firm would be a wage taker. If they tried to pay only W_2, workers would go to other firms willing to pay a higher wage.

Minimum wage in a monopsony

In a monopsony, a minimum wage can increase wages without causing unemployment.



A monopsony pays a wage of *W2* and employs *Q2*. If a minimum wage was placed equal to *W1*, it would increase employment to *Q1*. A minimum wage of *W3* would keep employment at *Q2*.

Monopsony in the real world

Even if a firm is not a pure monopsony, it may have a degree of monopsony power, due to geographical and occupational immobility, which make it difficult for workers to switch jobs and find alternative employment.

For example, there are several employers who might employ supermarket checkout workers. However, in practice, it is difficult for workers to switch jobs to take advantage of slightly higher wages in other supermarkets. There is a lack of information and barriers to moving jobs. Therefore, although there are several buyers of labour, in practice the big supermarkets have a degree of monopsony power in employing workers.

Monopsony and the gig economy

The gig economy refers to recent trends towards self-employment and very flexible labour practices. In practice, workers in the gig economy can easily face a monopsony employer. For example, *Uber* drivers have little control over rates of pay and have to meet strict criteria from *Uber*. In theory, they could work elsewhere but in practice it is difficult to replicate that job.

Problems of monopsony in labour markets

Monopsony can lead to lower wages for workers. This increases inequality in society. Workers are paid less than their marginal revenue product. Firms with monopsony power often have a degree of monopoly selling power. This enables them to make high profits at the expense of consumers and workers. Firms with monopsony power may also care less about working conditions because workers don't have many alternatives to the main firm.

Monopsony in product markets

In several industries, there is one buyer and several sellers. Supermarkets have monopsony power in buying food from farmers. If farmers don't sell to the big supermarkets, there are few alternatives. This has led to farmer protests about the price of milk. Amazon.com is one of the biggest purchases of books. If publishers don't sell to Amazon at a discounted price, they will miss out on selling to the biggest distributor of books.

UNIT –V: FACTOR PRICING

Nature of factor markets

A factor market is a market in which companies buy the factors of production or the resources they need to produce their goods and services. Companies buy these productive resources in return for making payments at factor prices. This market is also referred to as the input market.

A factor market is different from the product, or output, market—the market for finished products or services. In the latter, households are buyers and businesses are sellers. But in a factor market, the reverse is true: households are sellers and businesses are buyers. The primary difference between product markets and factor markets is that factors of production like labor and capital are part of factor markets and product markets are markets for goods. The relationship between the factor market at the product market is determined by derived demand, or the demand for productive resources, as determined by the demand for goods and services output, or products. When consumers demand more goods and services, producers increase their demands for the productive resources used to make those goods and services.

Every individual takes part in the factor market. People who are looking for jobs take part in the factor market. Employee wages that are paid by firms are part of the factor market. Investors who receive any form of compensation like a dividend or rental payments also take part in this market. Households thus become sellers because they are selling their services for money paid for by the buyers, who are the businesses. The combination of the factor markets along with the goods and services market forms a closed loop for the flow of money. Households supply labor

to firms, which pay them wages that are then used to buy goods and services from the same firms. This is a symbiotic relationship that benefits the economy.

The price for each factor is based on supply and demand. But that demand is derived because it's based on the demand for output. So the amount of input depends on how much a company will produce. In a booming economy with a tight labor market, wages will rise because the demand for workers is high. So when there's a high demand for a product, a company will increase its workforce. Conversely, in recessionary conditions where unemployment is high and demand for goods is low, wages will remain stagnant or even fall. Companies may cut back on hiring and may even lay off workers to deal with the drop in demand.

Marginal productivity theory of distribution

What determines the prices of factors of production?

A theory which tries to answer this question and which has been fairly widely held by professional economists is known as marginal productivity theory of distribution. The essence of this theory is that the price of a factor of production depends upon its marginal productivity. It also seems to be very fair that a factor of production should get its reward according to the contribution it makes to the total output, i.e. , its marginal productivity. Marginal productivity theory was first put forward to explain the determination of wage i.e, reward for labour but subsequently prices of other factors of production such as land, capital etc. also were explained with doctrine of marginal productivity.

The origin of the concept of marginal productivity can be traced to Ricardo and West. But both Ricardo and West applied the marginal productivity doctrine only td land. The concept of marginal productivity is implicit in the *Ricardian Theory of Rent*. But the idea of marginal productivity did not gain much popularity till the last quarter of 19th century, when it was rediscovered by economists like *J.B. Clark*, *Jevons*, *Wicksteed*, *Walaras* and later *Marshall* and *J.R. Hicks* popularized the doctrine of marginal productivity. It may however be pointed out that there are many versions of marginal productivity theory. Theories of marginal productivity propounded by various economists differ from each other in some respects. There are in fact so many versions of marginal productivity theory that Joseph Schumpeter has gone so far as to remark that there are almost as many marginal productivity theories as there are economists.

Marginal Productivity Theory: *Clark's Version*

J. B. Clark, an American economist, developed marginal productivity theory of distribution in his book *"The Distribution of Wealth"*. In order to bring out the fundamental factors at work in the mechanics of income distribution, Clark assumed a *completely static society,* free from the disturbances caused by economic growth or change. In other words, he assumed constant population, a constant amount of capital and unchanging techniques of production.

Besides the assumption of static economy, he has also assumed perfect competition in the factor market and perfect mobility on the part of both labour and capital. Besides assuming that total stock of capital remains constant, Clark also supposes that the *form of capital* can be varied at will. In other words, physical instruments of production can be adapted to varying quantities and abilities of available labour. Further, *he treats labour as a homogeneous factor by taking identical labour units* and discusses how the wage rate of labour is determined.

Optimal use or employment of a factor

For an individual firm or industry, marginal productivity of labour will decline as more and more workers are added to the fixed quantity of capital. He will go on hiring more and more labour units as long as the addition made to the total product by a marginal labour unit is greater than the wage rate he has to pay for it. The employer will reach equilibrium position when the wage rate is just equal to the marginal product of labour.

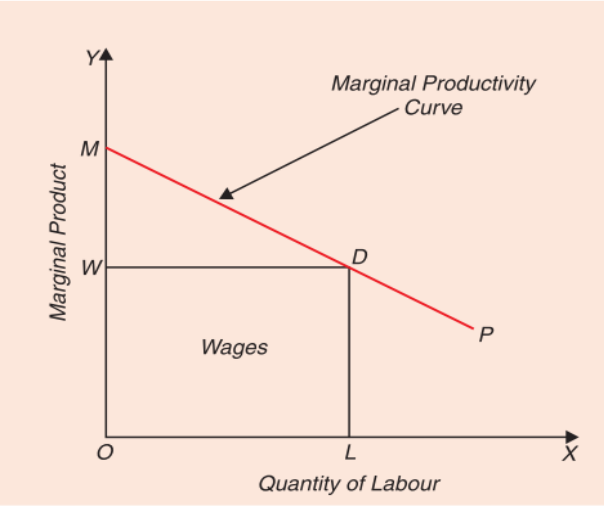


Fig. 50.1. *Wage rate is equalised with marginal productivity of labour*

Just consider the adjoining Fig. 50.1 where units of labour are represented on the X-axis and the marginal product of labour on the *Y*- axis. The *MP* curve shows the diminishing marginal product of labour as more units of it are employed. If the prevailing wage rate which an employer must pay is equal to *OW,* then it will be profitable for the employer to go on employing additional workers until the marginal product of labour becomes equal to the prevailing wage rate *OW*. It will be evident from Fig. 50.1 that if the prevailing wage rate is *OW,* then the employer will employ *OL* units of labour, since the marginal productivity of labour is equal to *OW* at *OL* employment of labour.

He would not employ more than *OL* amount of labour as the marginal product of labour will fall below the wage rate *OW* and he would therefore be incurring losses on the employment of marginal workers beyond *OL.* Thus the employer would be maximizing his profits by equalizing the marginal product with the wage rate *OW*. Since perfect competition is assumed to be prevailing in the labour market, an individual firm or industry will have got no control over the wage rate. An individual firm or industry has therefore to determine only the number of factor units (labour in the present case) to which it has to give employment at the prevailing (existing) wage rate. Thus, at micro level *(i.e.,* for individual firm or industry) marginal productivity theory is the theory of employment.

Determination of a factor price

A marginal-product schedule or curve shows a particular wage-employment relationship. Since *Clark* has assumed a stationary state, *he takes the total supply of labour available for employment in the whole economy as given and constant.*

In other words, in *Clarkian* analysis, aggregate supply curve of labour has been assumed to be *perfectly inelastic*, that is, it is of vertical shape. Given the total supply of labour in the economy, the wage rate will be determined by the marginal product of the available amount of labour assuming that all labour get employment. Given the aggregate amount of labour that is seeking employment, the wage rate that the laborers will secure will be equal to the *addition made to the total product* by the employment of the marginal unit of labour.

In other words, if the total quantity of labour seeking employments is '*n*' units, then each unit of labour will get wage which will be equal to the difference between total production when *n*

labour units were employed and that when n -1 labour units were employed. In other words, in the competitive labour market the wage rate will be *determined by the marginal product of a given quantity of labour force.*

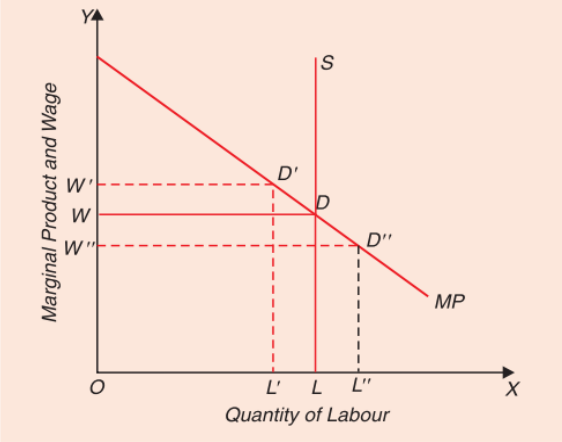


Fig. 50.2. *Wage rate is determined by marginal productivity*

If the labourers compete with each other for obtaining jobs, they will bid the wage rate down if some of them find themselves unemployed. The employers will bid the wage rate up if the prevailing wage rate is smaller than the marginal product of available labour force. This is so because at the wage rate lower than marginal product the employers' demand for labour force will be more than the available number of labourers.

Consider Fig. 50.2. In this figure, *MP* curve shows diminishing marginal product of labour as more units of labour are employed in the economy, assuming the quantities of other factors used as unchanged. Now, if the available quantity of labour force is *OL* in the whole economy, *LS* is the supply curve of labour which is perfectly inelastic. The marginal product of *OL* quantity of labour is *LD*. The wage rate will be determined by this marginal product *LD* and, therefore, equilibrium wage rate which will be determined by demand for and supply of labour in the market will be equal to *LD* or *OW*. At a higher wage rate *OW'* , the employers will employ *OL'* amount of labour leaving *LL'* amount of labour unemployed. Unemployed workers in their attempt to get employment will bring the 'wage rate down to the level *OW* at which all are employed.

On the other hand, at a lower wage rate than *OW,* say *OW''* , the employers will demand *OL''* amount of labour since their profits will be maximum if they are employing *OL''* amount of

labour at the wage rate *OW″*. But the available amount of labour is *OL*. Thus, at a lower wage rate than *OW″* the demand for labour by the employers will be greater than the available supply of labour. In a bid to get more labour, competition among employers will push the wage rate up to *OW* at which wage rate the employers demand just the same amount
of labour which is actually available.

If the actually available labour force in the whole economy is *OL″* , that is, labour supply curve is vertical line *LD″*, then the marginal productivity of labour force in equilibrium will be *L″D″* and therefore the wage rate will be equal to *OW″*. If the actual quantity of labour force is *OL′*, then the wage rate will be equal to *OW′*. Thus, given the quantity of labour in the country, wage rate is *determined* by marginal productivity of labour. One assumption which is implicit in the Clarkian marginal productivity theory as applied to the economy as a whole is that of *full employment*. In other words, it is assumed that the existing amount of labour in the economy is fully employed. To sum up 'in Clark's presentation, the marginal productivity of a given quantity of available labour determines. Its wage level when we consider the market as a whole. In the disaggregated picture, however, where a single employer finds the wage level determined by the forces beyond his control, the marginal product of labour determines the level of employment.

Clark's marginal productivity theory of distribution may be divided for analytical purposes into the following *three* component parts:

Firstly, there is the premise that a rational employer will be guided by the marginal productivity of the factor in determining the number of units of that factor he has to employ. This premise has been called by Prof. Cartter as *the marginal productivity principle.* (This is essentially based on the twin assumptions that law of diminishing returns is working and that the employers are rational).

Secondly, there is the assumption of perfect competition so that market forces tend to
equalize rates of return for all units of a factor.

Thirdly, there is the premise of long-run general equilibrium in all market.

Given the above three assumptions, it can be said that with a given fixed supply of labour in the market, the level of wages will be *determined* by the marginal product of labour.

RENT

Introduction

Rent is the payment made for the use of land and other kinds of rentable assets. The basic theory of factor pricing, i.e., the *marginal productivity theory* applies to land rent also. However, much prior to the formulation of modern theory of factor pricing, many classical economists have used different concepts and have formulated different theories for different factor prices. So is the case with the theory of rent. Ricardo had formulated a theory of rent, which is widely known as the *classical theory of rent*. After Racardo, Marshall made some significant contributions to the concept of rent, e.g., the concept of *quasi-rent* and *economic rent*. Modern economists have made some important contributions to the concept and applicability of the rent theory. The basic conceptual difference between the Ricardian and the modern theories of rent is that Ricardo considered rent as a '*surplus produce of land*', modern economists interpret rent as '*economic surplus*', the concept which applies to other factors of production with fixed supply in the short run also. This chapter deals with different aspects of the rent theory in detail.

Land as a Factor of Production

The classical economists treated land as a 'free gift of nature'. Land was treated as a special factor of production distinct from man-made means of production including capital, technology and desert reproducible human labour. Land as a means of production has been used in three different senses: a) the dry and desert area of land within the territory of a country; (b) the area of fertile land available for cultivation and (c) natural resources underground and on the surface of land.

When the term 'land' is used in sense (a), the supply of land is traditionally assumed to be absolutely inelastic. Ricardo considered land as a non-reproducible means of production. But, his view applies to land only in its physical sense and not in economic sense. In economic sense of the term, 'land' is defined as 'cultivable fertile area' and as a 'basket of natural resources', which can be exhausted and replenished to some

extent. Land is thus not much different from man-made resources and its supply is subject to variable1 Maintaining the supply of land, defined as fertile cultivable area, requires a considerable amount of effort and cost on the improvement of land, maintenance of the fertility of soil (e.g., cost of manuring and soil conservation) and creation of irrigation facilities and so on. Otherwise, the supply of cultivable land will decrease. The supply of cultivable land can be increased by means of land reclamation, soil conservation, fertilization and expansion of irrigation facilities. On the other hand, supply of cultivable land tends to decrease when cost of soil conservation and so on. goes up and vice versa. Similarly, if land is defined to include all the natural resources underground and on the surface of land, the exploitation of resources and converting them into usable intermediate goods, requires a heavy cost. Since the magnitude of natural resources, available in or on the land of a country's territory is limited, their supply can be increased by increasing cost of exploiting natural resources. For example, water resources, forest resources, minerals, petroleum and natural gas and so on, are available in a limited supply, scattered over the territory of a country. Besides, exploitation of natural resources is subject to diminishing returns, i.e., their supply can be increased but at an increasing cost. Thus, the supply of land, in the economic sense, is not perfectly inelastic. Land supply is, however, less elastic than the supply of man-made means of production.

Historical Background of Rent Theory

The classical theory of rent, as formulated by David Ricardo, has an interesting antecedent. In the early 19th century, food prices in Britain had increased heavily partly due to Napoleanic War and partly due to increase in population and the consequent increase in demand for food. The rise in food prices caused a great deal of anxiety to the British government. Therefore, both the House of Lords and the House of Commons appointed a committee to investigate the causes of rise in food prices. The committee reported that '*food prices were high because rents were high*'. The contemporary economists, namely, West, Torrens, Malthus and Ricardo disagreed to this suggestion. They offered, separately, alternative explanations to the problem. In their opinion, *food prices were high not because rents were high, instead rents were high because food prices were high*.2 According to them, food prices had gone up due to Napoleanic War and increase in

population causing increase in demand for food. Scarcity of food led to increase in food prices which, in turn, increased profitability of cultivation. This resulted in increase in demand for cultivable land, which caused rise in rents. Ricardo added that the landed aristocracy (the landlords) was thriving on the misfortune of the rest of the society and causing misery to the tenant farmers. For defending his view, Ricardo was criticized as being anti-landed aristocracy. Ricardo expounded a theory of rent in his efforts to establish his argument.

The Ricardian Theory of Rent

Ricardian Definition of Rent

Ricardo had defined rent as 'that portion of the produce of the earth which is paid to the landlord for the use of the *original* and indestructible powers of the soil'.3 Ricardo considered payment of rent as an indication of *niggardliness* of nature. This was contrary to the opinion of French economists, known as 'Physiocrats' who considered rent as the result of *bounty* of nature. By niggardliness of nature, Ricardo meant 'fixed supply' of land with different levels of productivity. Given the fixed supply of land, it proves to be scarce with the growth of population. Growth of population forces extension of cultivation to inferior lands. According to Ricardo, the need for the payment of rent arises due to differences in surplus accruing to the cultivators of different grades of land. The reason for the difference in surplus product is the difference in the fertility of soil of different grades of land.

The Ricardian Theory of Rent

Ricardian theory of rent is based on the principle of demand and supply of land. If supply of land in a country exceeds the total demand for land, no rent will be paid, like nothing is paid for the use of air. In Ricardo's words, 'If all lands had the same properties, if it were unlimited in quantity, and uniform in quality, no charge could be made for its use, unless where it possessed peculiar advantages of situation.' Rent is chargeable '…because land is not unlimited in quantity and uniform in quality and because (due to increase in population), land of an inferior quality, or less advantageously situated, is called into cultivation…'. Ricardo has observed that, with increase in demand for food, land may be used extensively and intensively and that *rent arises in both extensive and intensive cultivation of land*.

Extensive cultivation means extending cultivation to inferior grades of land with the same amount of capital and labour applied to all grades of land. When cultivation is extended from superior to inferior lands, then superior land bears a rent. The rent on superior land equals the difference between the produce of the superior land and that of the most inferior land. Suppose there are three grades of land—*A*, *B* and *C*—suitable for wheat cultivation but land *A* is of the best quality and land *C* of the worst quality. If an equal amount of capital and labour is used to cultivate an equal area of each grade of land, then the respective yields are 100, 80 and 70 quintals of wheat. If, in a country, the supply of *A*-grade land is greater than what must be cultivated to meet the food requirement of the entire population, no rent is payable on the land of *A* grade. When population increases, demand for food increases and, therefore, demand for land increases. And, if cultivation has to be extended beyond grade-*A* land, then the land of grade *B* will be brought under cultivation. But, compared to yield from land *A* (i.e., 100 quintals), land *B* yields only 80 quintals of wheat, with the same quantities of capital and labour. According to the Ricardian theory of rent, the difference in the yields from lands of category *A* and *B* gives rise to rent on land of grade *A*. The rent on land *A* equals 100 - 80 = 20 quintals of wheat. Similarly, when population increases further, land of grade *C* is also brought under cultivation. The *C*-grade land yields only 70 quintals of wheat. This gives rise to a rent on land *B* and raises rent on land *A*. The rent on the three grades of land can be computed as follows:

Rent on land *A* = 100-70 = 30 quintals of wheat

Rent on land *B* = 80-70 = 10 quintals of wheat

Rent on land *C* = 70-70 = 0 quintals of wheat

If the value of capital and labour used in cultivation equals the value of 70 quintals of wheat, the land of grade *C* will not bear any rent. Land *C* is, therefore, called '*marginal land*' or '*no-rent land*'. According to Ricardo, the land of grade *C* will remain a no-rent land until cultivation is extended to land of grade *D*, if any.

In case of *intensive cultivation*, as Ricardo observed, it is often found that before land *B* is brought under cultivation, additional capital is employed to increase productivity of land *A*. But, it is quite likely that when capital expenditure on land *A* is doubled, output is not doubled. It may yield only 95 quintals (instead of 100 quintals) of wheat. Note that the yield from *A*-grade land is

still greater than the yield expected from land *B*. The cultivators would, therefore, intensify cultivation of land *A*, instead of employing their capital on land *B* or any inferior land. In this case, the rent on land *A* would be 5 quintals (= 100 - 95 quintals) of wheat. Thus, in case of intensive cultivation, additional capital will be employed on land *A* till the yields from subsequent units of capital used on land *A* are greater than that of land *B* without using extra capital and labour. As more and more units of capital and labour are employed on land *A*, the yield from the successive units of capital and labour decreases. When yield from additional units of capital equals the yield from land *B* (without extra capital), then the cultivators may continue to cultivate land *A* or move on to land *B*, depending on their choice on non-economic considerations. This has two repercussions: (1) rent on land *A* increases and (2) the inferior land, i.e., land *B*, is brought under cultivation. It shows that Ricardian concept of rent is based on the *law of diminishing returns*. This is, in brief, the Ricardian theory of rent determination.

Criticism

Ricardian theory of rent has been criticized on the following grounds.

First, Ricardo's concept of rent is based on the assumption that powers of soil are 'original and indestructible'. This view can hardly be accepted. 'Powers of soil', i.e., productivity of land, can be created through techniques of soil conservation and land reclamation, and can be destroyed through continuous and intensive use of land, soil erosion, spread of desert, flood and so on. Destruction of 'power of soil' by flood and spread of desert is widely known.

Secondly, Ricardo's idea that rent is peculiar to land as a factor of production has been questioned by the modern economists. The differential surplus as rent accrues also to other factors—labour, capital and entrepreneurship—as well as to land.

Thirdly, Ricardo assumed only one use of land, i.e., growing corn, and hence, there is no transfer earning. So, all that is paid in the name of rent becomes *economic rent*. There are, however, alternative uses of land. There are, therefore, *transfer earnings*, and the total rent cannot be economic rent.

Finally, Ricardo considered land supply to be fixed because he considered land of the economy as a whole. For an individual cultivator, however, the supply of land has an elasticity greater than zero. This alters the concept of rent as envisaged by Ricardo.

In spite of these points of criticism, the Ricardian theory of rent holds the ground. The reason is that criticism of the theory is not strong enough to refute it altogether. For instance, whatever one says, the productivity of land is original and natural, though it can be increased or decreased. Similarly, one can apply the Ricardian concept of rent to other factors also but it does not challenge the original concept. So is the case with the supply of land. Although land supply can be increased, it is limited to the boundary walls of a nation.

Transfer Earning and Economic Rent: The Modern Approach

The two other important concepts that are used in modern economic analysis of factor prices are the concepts of 'economic rent' and what Marshall called 'transfer earning'. In modern economist's perception, these two concepts constitute the actual factor earning. Thus, the factor price at equilibrium consists of two components:

1. Transfer earning and
2. Economic rent.

The *transfer earning* is also known as *opportunity cost* and 'reservation price'. Suppose a factor has alternative uses and it is put to its best use. It will not be transferred till its earning is equal to the earning expected from the second best use. Thus, transfer earning is defined as the *amount that a factor must earn to remain in its present occupation*. Alternatively, the transfer earning can be defined as the amount that a factor expects to earn if transferred to its *second best use*. For example, if wheat cultivation on an area of land yields Rs 60,000 annually and when used for running a school it yields Rs 80,000 annually, the land will be used for running a school. In that case, the transfer earning of land is Rs 60,000—the earning expected from wheat cultivation.

Consider the case of a labour. Suppose a doctor earns Rs 50,000 per month from his private practice, the alternative available to him is to serve in a hospital as an employee where he expects to earn Rs 30,000 per month. Thus, doctor's transfer earning is Rs 30,000 per month. He must earn a minimum of Rs 30,000 per month to remain in his private practice. So long as he

earns Rs 30,000 per month from his private practice, he has no incentive to join a hospital as an employee.

Economic rent (*ER*) is the excess of *actual earning* (*AE*) of a factor over its transfer earning (*TE*). Economic rent may thus be defined as *ER* = *AE* - *TE*. In our example of land, the use of land for running a school yields an economic rent of Rs 20,000 = Rs 80,000 - Rs 60,000. Similarly, doctor's monthly economic rent is the difference between his actual earning and his opportunity cost, i.e., doctor's economic rent equals Rs 50,000 - 30,000 = Rs 20,000.

Note that the term 'economic rent' has a different meaning compared to the term 'rent' used in common parlance. In its common usage, the term 'rent' means the *actual* payment to the landlord, i.e., contractual rent, much of which is transfer earning. But, when an economist uses the term 'rent', he means 'economic rent', i.e., the difference between the actual earning and the transfer earning.

Quasi-Rent: The Short-Term Earning of Fixed Factors

The concept of *quasi-rent* was introduced by Marshall6 to analyse the short-term earnings of man-made factors including machinery and other capital equipments. An important feature of capital equipments is that their *supply is inelastic in the short-run* and *elastic in the long run*. The short-term earning of manmade factors may be termed, to begin with, as the *surplus of total revenue after the payment to variable factors*. In Marshall's view, short-run earning of man-made factors cannot be called 'rent' (in usual sense of the term) because, unlike rent, it is subject to fluctuation in the short run. Short-run variation in factor payment depends on the change in the price of the goods they produce. If demand for the product increases in the short run, its price goes up. But more of capital equipments cannot be hired to increase production because their supply in the short run is fixed. Therefore, price of the goods remains high. This increases the total revenue. Consequently, surplus of total revenue over the cost of variable factors increases. As a result, the surplus accruing to man-made factors exceeds their *normal earnings*, i.e., their rentals. Similarly, when demand for the product decreases, its price goes down. As a result, total revenue goes down. Therefore, the surplus over the variable cost goes down, reducing the earning of the man-made factors. It may go below the normal rate of rentals. That is, short-term

earnings sometimes exceed the normal rentals and sometimes go much below it. That is why, perhaps, Marshall did not find it appropriate to use the term 'rent' for the short-run earnings of the man-made factors. Marshall coined a new term '*quasi-rent*' for the short-term earnings of the man-made factors in inelastic supply, in the short run.

WAGES

Introduction

Wage-rate is the sum of money which an employer contracts to pay a worker in return for services rendered. This definition includes salaries as well as wages, and makes no distinction between time- and piece-rates. Earnings are what the worker actually receives in his pay-packet (his 'take-home' pay) plus deductions which have been made for insurance, income tax, superannuation, etc.

Wage Differentials – Meaning

Wage differentials bear a direct relationship to the diversity in occupation and industries that exist in the economic sphere of activity in a country. A certain job requiring a certain skill is paid more or less than another job requiring a different skill either in the same or some other industry. There are a variety of contributory factors.

When there is a variation in workers' skills, *i.e.,* highly skilled, skilled, semi-skilled and unskilled, their wage rates will differ. The variance is due to the complexity of the skill acquired, its scarcity, and the time taken in training to acquire it. The unskilled category which requires none of these is relatively less well paid.

Factors Responsible for the Differences in Wages

1. Difference in Efficiency:

All persons are not equally efficient. They differ in abilities. Some are more efficient and some are less efficient. Some others are not efficient at all. An efficient worker gives better output. Hence, he is paid higher wages than others. Moreover, the efficiency requirement in different jobs varies.

2. Presence of Non-Competing Groups:

Society is divided into a number of working groups, which are noncompeting. Caste system creates such groups in India. As a result, a child born to a sweeper will most likely be a sweeper just as a black smith's son will be a black smith. Besides, the chances of receiving training for better-paid occupations depend on the resources of the family. Thus, inheritance, environment, training and gender are some factors, which create noncompeting groups in the society. Hence, workers belonging to different groups are paid at different wage rates.

3. Immobility of Labour:

Labour is not perfectly mobile. They are normally shy to move. It has inertia to stick to one job. Sometimes, people are not prepared to accept higher wages if it necessitates a change of place. This accounts for difference in wage in different places. The presence of noncompeting groups in society makes labour more immobile. Political barriers against the free movement of labour from one country to another country result in the difference in wages in different countries.

4. Nature of Employment:

The nature of work also influences wage rates. Dangerous and disagreeable work brings higher money wages to attract larger supply of labour. For example, a coal miner gets higher wages than a clerk in the office. High money wages act as compensation.

5. Training and Qualification:

Jobs requiring special qualification and apprenticeship generally command higher wages than jobs learnt easily and for which no special training is required.

6. Productivity

Productivity differs in different occupations. The Cobbler's job is not as productive as that of a skilled motor mechanic or of clerk as that of a principal of a college.

7. Regularity of Employment:

If there is regular employment in a job, one may demand lower wages. If the job is irregular or seasonal, wage has to be higher. In case of India, young men prefer low paid jobs under government due to security and regularity of employment to irregular and insecure private jobs with more remuneration.

8. Future Prospects:

There are some jobs where promotion prospects are better than other jobs. Even if initial salary is low, if promotion prospects are there people prefer these jobs to others jobs.

9. Scope for Extra Earning:

If a job has scope for extra earnings, the regular wage may be lower. A doctor may start with a lower salary than a lecturer but the former can make up the deficiency by private practice.

Collective Bargaining

Collective bargaining is a process of negotiation between employers and a group of employees aimed at agreements to regulate working salaries, working conditions, benefits, and other aspects of workers' compensation and rights for workers.

Industrial disputes between the employee and employer can also be settled by discussion and negotiation between these two parties in order to arrive at a decision. This is also commonly known as collective bargaining as both the parties eventually agree to follow a decision that they arrive at after a lot of negotiation and discussion.

According to *Beach*, "Collective Bargaining is concerned with the relations between unions reporting employees and employers (or their representatives). It involves the process of union organization of employees, negotiations administration and interpretation of collective

agreements concerning wages, hours of work and other conditions of employees arguing in concerted economic actions dispute settlement procedures".

According to *Flippo*, "Collective Bargaining is a process in which the representatives of a labor organization and the representatives of business organization meet and attempt to negotiate a contract or agreement, which specifies the nature of employee-employer union relationship".

"Collective Bargaining is a mode of fixing the terms of employment by means of bargaining between organized body of employees and an employer or association of employees acting usually through authorized agents. The essence of Collective Bargaining is bargaining between interested parties and not from outside parties".

According to an ILO Manual in 1960, the Collective Bargaining is defined as:

"Negotiations about working conditions and terms of employment between an employer, a group of employees or one or more employers organization on the other, with a view to reaching an agreement."

It is also asserted that "the terms of agreement serve as a code defining the rights and obligations of each party in their employment relations with one another, if fixes large number of detailed conditions of employees and during its validity none of the matters it deals with, internal circumstances give grounds for a dispute counseling and individual workers".

Collective Bargaining Involves:

i. Negotiations.
ii. Drafting.
iii. Administration.
iv. Interpretation of documents written by employers, employees and the union representatives.
v. Organizational Trade Unions with open mind.

Forms of Collective Bargaining:

i. bargaining may be between the single employer and the single union.

ii. the bargaining may be between a single firm having several plants and workers employed in all those plants.
iii. instead of a separate union bargaining with separate employer, all the unions belonging to the same industry bargain through their federation with the employer's federation of that industry.

Main Features of Collective Bargaining:

i. It is a group action
ii. It is a continuous process
iii. It is a bipartite process
iv. It is a process
v. It is flexible and mobile and not fixed or static
vi. It is industrial democracy at work
vii. It is dynamic
viii. It is a complementary and not a competitive process
ix. It is an art

INTEREST

Introduction

The basic activity in the financial market of a country is borrowing and lending of money. The lender charges and the borrower agree to pay an amount in addition to the amount lent and borrowed. This excess amount is called, in common usage, interest. In economic terminology, like rent and wages, interest is a factor payment. Interest is paid for the services of capital as a factor of production. In economic sense of the term, interest is the cost of capital per unit of time.

The term 'capital' is used in two senses: (i) money or financial capital, i.e., loanable or investible stock of money and (ii) physical assets, e.g., land, building, plant and machinery and so on. Money capital in the form of bank term deposits, shares and debentures and so on. yield different forms of incomes—interest and dividend. Investment in physical capital yields income called 'return on capital'. Money capital finally takes the form of physical capital and interest paid on money capital takes the form of 'cost of capital'. Based on the two concepts of capital, there are two types of interest theories: (i) monetary theories of interest and (ii) real theories of interest, entwined with the Theory of Capital. A comprehensive treatment of theories of category (ii) falls outside the purview of this book.

Capital as a Factor of Production

It may be noted at the outset that the theoretical problems associated with capital and interest are much more complicated than those pertaining to other factors and factor prices. The complications arise for at least three reasons.

First, unlike land and labour, physical capital is a man-made factor—a produced means of production. The supply of capital is, therefore, under human control. It is produced only when it is needed. Production of each capital asset is a matter of economic decision. Besides, there are various types of capital assets with varying productivity.

Secondly, another peculiarity of capital is that producing capital goods requires investment. Availability of funds for investment depends on savings. Given the income, additional saving requires cutting the present consumption. Cutting present consumption with the purpose of having a higher level of income and consumption in future is the matter of making choice between the present and future consumption. Making choice between the present and future consumption brings in the question of *time preference*.

Thirdly, production of capital is a 'round about process' which is time consuming. For example, to produce tractors requires steel, which requires mining of iron ore. To mine iron requires mining machinery and equipment, and the process goes on. Thus, tractor production takes a long time. Besides, capital goods generally have a long productive life but are subject to wear and tear and obsolescence. This necessitates maintaining capital intact and replacement of obsolete capital. It, therefore, requires a near-accurate prediction of timings of expenditure on capital goods and returns from them. All these considerations are taken into account in the theory of interest and investment decisions.

The Classical Theory of Interest

The classical theory of interest refers, according to Keynes, to the theories (or views) of Marshall, Cassel, Tausig and Walras and so on. In fact, *none of these economists, whom Keynes groups at modern classical school, has given a precise or an explicit account of the interest theory*.

Keynes has, however, reconstructed the classical theory of interest himself which is presented here. According to the classical theory, the rate of interest is determined by the *demand for investment* and *willingness to save*. In the words of Keynes,

> Investment represents the demand for investible resources; saving represents the supply, and the rate of interest is the 'price' of investible resources at which the two are equated. Just as the price of a commodity is necessarily fixed at that point where the demand for it is equal to the supply, so the rate of interest necessarily comes to rest under the play of market forces at the point where the amount of investment at the rate of interest is equal to the amount of saving at the rate.

In simple words, rate of interest is determined by demand for investible funds and the supply of investible funds, i.e., supply of savings. The rate of interest is determined where demand for investible funds equals the supply of savings. The classical theory of interest is presented in Figure 10.2. The investment demand schedule is inversely related to interest as shown by I_1 and I_2—the investment demand schedules. In contrast, the supply of savings schedule, *S*, is positively related to interest. The investment demand schedule, I_1, intersects with the supply of savings schedule at point *E*. The rate of interest is thus determined at *OR* by the point of intersection of the two schedules. At this rate of interest, the demand for investible funds, *OQ*, equals the supply of funds. The interest rate *OR* is therefore the equilibrium rate of interest. The equilibrium rate of interest will be disturbed only when there is a change in the demand for investible funds and in the supply of savings. But, investment and savings will automatically adjust themselves to a new equilibrium rate of interest. For instance, given the saving supply schedule, S', if investment demand schedule shifts downward to I_2, intersecting S-schedule at E', a new (lower) rate of interest will be determined at OR'.

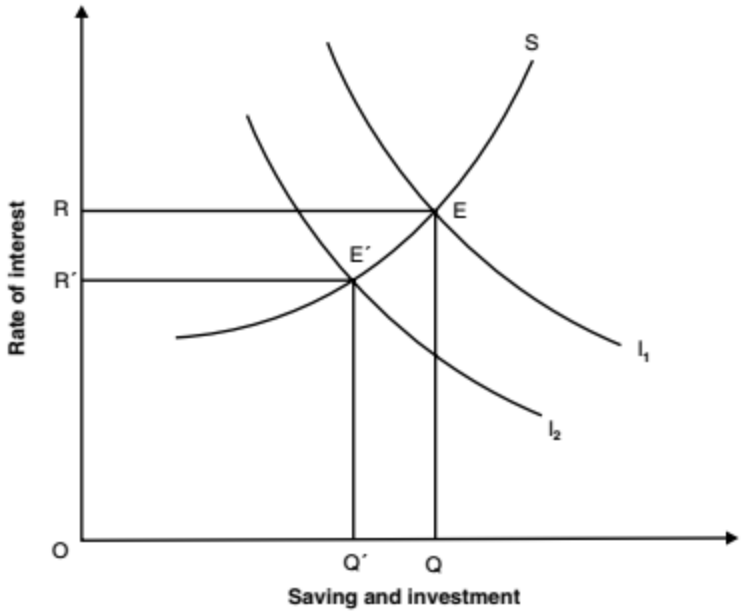


Figure 10.2 *The Classical Theory of Rate of Interest*

Keynes' Criticism of Classical Theory

Keynes himself criticized the classical theory of interest on the following grounds.

First, classical theory implicitly assumes *income to be given* and saving to be a unique function of interest, i.e., $S = f(i)$. Keynes, however, argues that the classical assumption of constant *income* implies that there exists an important relationship between saving and income, i.e., $S = f(Y)$ also. Savings are, thus, not the function of interest alone but also of income, i.e., not only $S = f(i)$, but also $S = f(Y)$. But the classical school neglects this important relationship between income and savings, which leads to a theoretical error in the classical theory of interest.

Secondly, in spite of its theoretical error, classical theory of interest concludes that, given the investment and saving schedule, interest will be determined by the investment demand and saving schedules. It implies that interest rate will be determined even if there is a shift in the investment demand schedule. But interest rate is indeterminate. This point is illustrated in Figure 10.3. Suppose that rate of interest was initially determined at Or_3 by intersection of I_1 and S_1 schedules at point E_1. If investment demand schedule shifts leftwards to I_2, due to say, a fall in the marginal efficiency of capital, according to the classical theory, interest rate will be determined at Or_1 by intersection of S_1 and I_2 curves at point E_2. Similarly, if both investment

demand and savings schedules shift leftward, interest will be determined at Or_2 by point E_3. As Keynes puts it, 'the classical theory of rate of interest seems to suppose that, if the demand curve for capital shifts or if the curve relating the rate of interest to the amount saved out of a

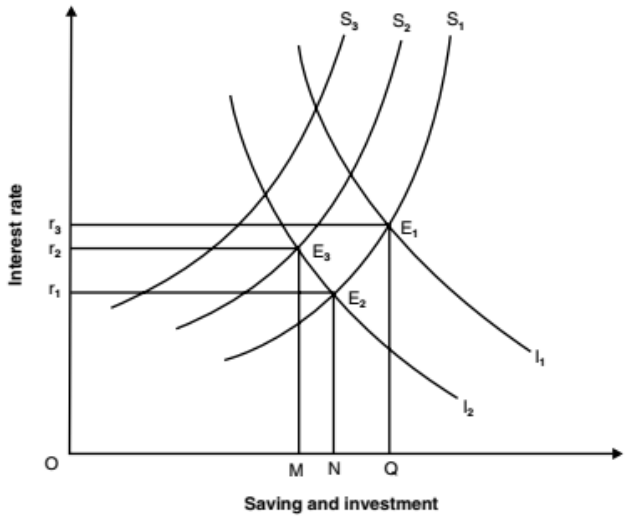


Figure 10.3 *Change in Interest Rate and Levels of Income*

given income shifts or if both these curves shift, the new rate of interest will be given by the point of intersection of the new position of the two curves' (Keynes, p. 179). 'But this,' according to Keynes, 'is nonsense theory.' The error in classical theory lies in its assumption that investment demand schedule (I) can shift without causing a shift in the saving schedule (S). In fact, when investment schedule shifts, it means a change in investment. Change in investment causes a change in income (Y) because $Y = f(I)$. When income changes, savings change too because $S = f(Y)$. For instance, if investment schedule I_1 shifts downward to I_2 ($S1$ remaining the same), investment falls by QN. Since $Y = f(I)$, incomes will fall due to fall in investment. Since $S = f(Y)$, savings will also fall due to fall in income and saving schedule S, shifts backward to S_2. As a result, interest rises from Or_1 to Or_2 and investment falls further by NM (Figure 10.3). This fall in investment will again generate a chain of adjustments.

It is, therefore, inconsistent to assume that investment demand and saving schedules can shift independent of each other. A shift in investment demand schedule does cause a shift in saving schedule. If investment demand and saving schedules keep shifting from one position to another, the whole classical scheme of interest determination breaks down. Interest rate cannot find its equilibrium. Thus, according to Keynes *classical theory of interest is indeterminate*.

Neo-classical theory of interest

A variant of classical theory is the loanable fund theory of interest, also called the neo-classical theory of interest. The economists who have contributed to the growth of this theory include Wicksell, Ohlin, Robertson, Pigou and Viner. According to the loanable fund theory, rate of interest is determined by the intersection of demand and supply schedules of loanable funds.

According to the loanable fund theory of interest, the demand for loanable funds consists of:

i. Investors' demand for funds (I_D);
ii. Consumers' demand for funds (C_D); and
iii. Demand for funds for hoarding (H_D).

All the three kinds of demand for funds are inversely related to the interest as shown in Figure 10.4. The horizontal summation of the three kinds of demand curves gives the aggregate demand for loanable funds (DL) as shown by the schedule DL in Figure 10.4.

According to Robertson, the supply of loanable funds consists of:

i. Voluntary savings *(V_s),* i.e., savings out of disposable income;
ii. Bank credits *(B_C),* i.e., the supply of credits by banks; and
iii. Activated idle balances or dishoardings *(D_h)*.

All the three components of the supply of loanable funds are positively related to interest. A horizontal summation of the schedules V_s,B_c and D_h gives the aggregate supply schedule of loanable funds (S_L) as shown in Figure 10.4. The demand schedule for loanable funds (D_L) intersects the supply schedule of loanable funds (S_L) at point P, determining interest rate at *OR*.

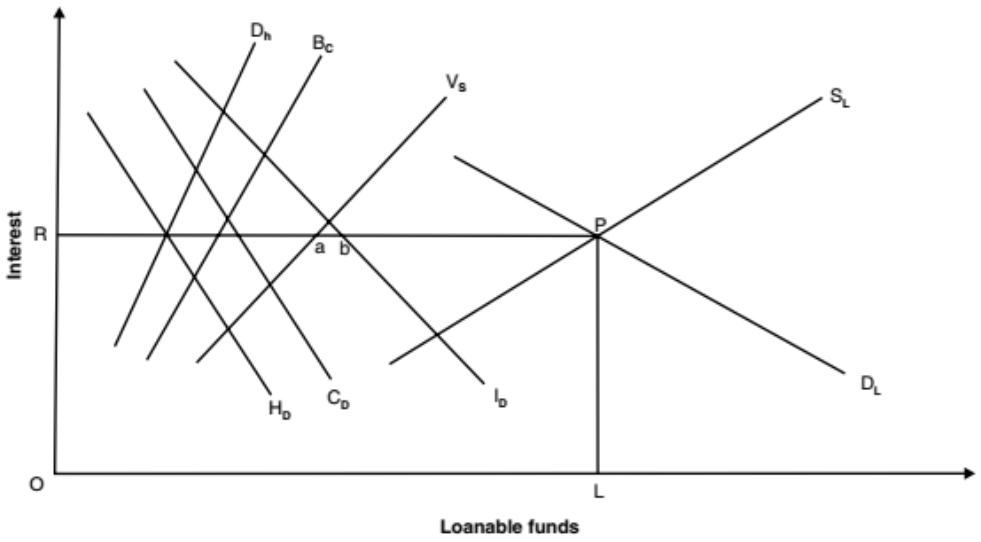


Figure 10.4 *Loanable Fund Theory of Interest*

Classical vs Neo-Classical Theory of Interest

The neo-classical theory of interest is superior to the classical theory in so far as it considers the demand for funds other than the demand for investment, and also the other sources of fund supply than voluntary savings. Besides, the loanable fund theory, as presented in Figure 10.4, reveals important information that planned savings may not be equal to the planned investment at equilibrium rate of interest. The planned savings are represented by the schedule V_s and planned investment by the schedule I_D. As Figure 10.4 shows, at equilibrium rate of interest, *OR*, planned savings, *Ra*, is less than planned investment, *Rb*.

Criticism

Keynes's criticism of classical theory of interest applies to the loanable fund theory also. According to Keynes, like classical theory of interest the loanable funds theory is also indeterminate. In the loanable fund theory, supply of investible funds include voluntary savings, bank money and the dishoarding of idle balances. Of the three components of the total supply of inventible fund, 'savings' accounts for the largest proportion in the total and savings are functions of the 'disposable incomes'. In fact, almost entire supply of investible funds depends on the disposable incomes. Therefore, interest cannot be known unless income is known; income cannot be known unless investment is known and investment cannot be known unless interest is

known. As in classical theory, the error in the neo-classical theory lies in the implicit assumption that demand for and supply of investible funds can vary independent of each other.

It is, however, suggested that Keynes's criticism of classical theory does not apply to the neo-classical theory. For, unlike classical theory, the neo-classical theory considers savings to be the function of the preceding year's income which is known and cannot be influenced by the current investment.

Having criticized the classical theories of interest, Keynes propounded his own Liquidity Preference Theory of Interest. Keynesian theory of interest is a purely monetary theory of interest. Also, it considers aggregate demand for and aggregate supply of money in the determination of interest rate. Therefore, Keynesian theory of interest is discussed in a macro-monetary framework. Since this book is concerned with only micro analysis, Keynesian theories of interest and later development in interest theory have been omitted.

Keynes's Liquidity Preference Theory of Interest

In his epoch-making book, "The General Theory of Employment, Interest and Money" the late Lord Keynes gave a new view of interest. According to him, "Interest is the reward for parting with liquidity for a specified period." A man with a given income has to decide first how much he is to consume and how much to save. The former will depend on, what Keynes calls, the propensity to consume. Given this propensity to consume, the individual will save a certain proportion of his given income. He now has to make another decision. Should he hold his savings? How much of his resources will he hold in the form of ready money (cash or noninterest-paying bank deposits) and how much will he part with or lend depend upon what Keynes calls his "liquidity preference". Liquidity preference means the demand for money to hold or the desire of the public to hold cash.

Demand for Money or Motives for Liquidity Preference

Liquidity preference of a particular individual depends upon several considerations. The question is : Why should the people hold their resources liquid or in the form of ready money, when they can get interest by lending such resources ? The desire for liquidity arises because of three

motives: (*i*) the transactions motive, (*ii*) the precautionary motive, and (*iii*) the speculative motive.

The Transactions Motive

The transactions motive relates to the demand for money or need for cash for the current transactions of individual and business exchanges. Individuals hold cash in order "to bridge the interval between the receipt of income and its expenditure". This is called the 'Income Motive'. Most of the people receive their incomes by the week or the month, while the expenditure goes on day by day. A certain amount of ready money, therefore, is kept in hand to make current payments. This amount will depend upon the size of the individual's income, the interval at which the income is received and the methods of payments prevailing in the society.

The businessmen and the entrepreneurs also have to keep a proportion of their resources in ready cash in order to meet current needs of various kinds. They need money all the time in order to pay for raw materials and transport, to pay wages and salaries and to meet all other current expenses incurred by business firms. Keynes calls this as 'Business Motive' for keeping money. It is clear that the amount of money held under this business motive will depend to a very large extent on the turnover (i.e., the volume of trade of the firm in question). The larger the turnover, the larger in general will be the amount of money needed to cover current expenses.

Precautionary Motive

Precautionary motive for holding money refers to the desire of the people to hold cash balances for unforeseen contingencies. People hold a certain amount of money to provide for the danger of unemployment, sickness, accidents, and the other uncertain perils. The amount of money held under this motive will depend on the nature of the individual and on the conditions in which he lives.

Speculative Motive

The speculative motive relates to the desire to hold one's resources in liquid form in order to take advantage of market movements regarding the future changes in the rate of interest (or bond

prices). The notion of holding money for speculative motive is a new typically Keynesian idea. Money held under the speculative motive serves as a store of value as money held under the precautionary motive does. But it is a store of money meant for a different purpose. The cash held under this motive is used to make speculative gains by dealing in bonds whose prices fluctuate. If bond prices are expected to rise, which, in other words, means that the rate of interest is expected to fall, businessmen will buy bonds to sell when their prices actually rise. If, however, bond prices are expected to fall, i.e., the rate of interest is expected to rise, businessmen will sell bonds to avoid capital losses. Nothing being certain in this dynamic world, where guesses about the future course of events are made on precarious basis, businessmen keep cash balances to speculate on the probable future changes in bond prices (or the rate of interest) with a view to making profits.

Given the expectation about the changes in the rate of interest in future, less money will be held under the speculative motive at a higher current or prevailing rate of interest arid more money will be held under this motive at a lower current rate of interest. The reason for this inverse correlation between money held for speculative motive and the prevailing rate of interest is that at a lower rate of interest less is lost by not lending money or investing it, that is, by holding on to money, while at a higher rate of interest holders of cash balances would lose more by not lending or investing.

Thus, the demand for money under speculative motive is a function of the current rate of interest, increasing as the interest rate falls and decreasing as the interest rate rises. Thus, demand for money under this motive is a decreasing function of the rate of interest. This is clear from Fig. 56.5. Along the X-axis is represented the speculative demand for money and along the Y-axis the rate of interest. The liquidity preference curve LP is a downward sloping towards the right signifying that the higher the rate of interest, the lower the demand for speculative motive, and vice versa. Thus, at the high current rate of interest Or, a very small amount OM is held for speculative motive. This is because at a high current rate of interest

much money would have been lent out or used for buying bonds and therefore less money willbe kept as inactive balances. If rate of interest falls to Or', then a greater amount OM' is held under speculative motive. With the further fall in the rate of interest to Or", money held under speculative motive increases to OM". It will be seen from Fig. 56.5 that the liquidity

preferencecurve LP becomes quite flat, i.e., perfectly elastic at a very low rate of interest; it is horizontal line beyond point E" towards the right. This perfectly elastic portion of liquidity preference curve indicates the position of absolute liquidity preference of the people. That is, at a very low rate of interest people will hold with them as inactive balances any amount of money they come to have. This portion of liquidity preference curve with absolute liquidity preference has been called liquidity trap by J.M. Keynes.

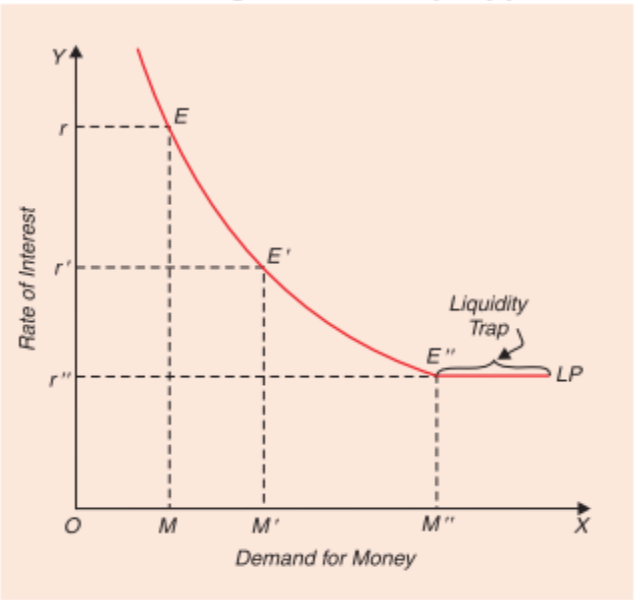


Fig. 56.5. *Liquidity Preference for Speculative Motive*

But demand for money to satisfy the speculative motive does not depend so much upon what the current rate of interest is, as on expectations of changes in the rate of interest. If there is a change in the expectations regarding the future rate of interest, the whole curve or schedule of liquidity preference for speculative motive will change accordingly. Thus, if the public on balance expect the rate of interest to be higher (i.e., bond prices to be lower) in the future than had been previously supposed, the speculative demand for money will increase and the whole liquidity preference curve for speculative motive will shift upward. If the total supply of money is represented by M, we may refer to that part of M held for transactions and precautionary motive as M_1 and to that part held for the speculative motive as M_2. Thus $M = M_1 + M_2$. The money held under the transactions and precautionary motives, i.e. M_1, is completely interest-inelastic unless the interest rate is very high. The amount of money held as M_1, that is, for transactions and precautionary motive, is mainly a function of the size of income and business transactions

together with the contingencies growing out of the conduct of personal and business affairs. We can write this in a functional form as follows:

$$M_1 = L_1(Y) \ldots\ldots (1)$$

where Y stands for income, L_1 for liquidity preference function, and M_1, for money held under the transactions and precautionary motive. The above function implies that money held under the transactions and precautionary motive is a function of income.

On the other hand, money demanded for speculative motive, i.e., M_2, as explained above, is primarily a function of the rate of interest. This can be written as:

$$M_1 = L_1(r) \ldots\ldots (2)$$

where r stands for the rate of interest, L_2 for liquidity preference function for speculative motive. Since total supply of money $M = M_1 + M_2$, we get from (i) and (ii) above

$$M = L_1(Y) + L_2(r) \ldots\ldots (3)$$

It follows from (iii) above that given the supply of money M (and also income) the rate of interest will be determined by the liquidity preference.

Determination of the Rate of Interest: Interaction of Liquidity Preference and Supply of Money

According to Keynes, the demand for money, i.e., the liquidity preference and supply of money determines the rate of interest. It is in fact the liquidity preference for speculative motive which along with the quantity of money determines the rate of interest. We have explained above the speculative demand for money in detail. As for the supply of money, it is determined by the policies of the Government and the Central Bank of the country. The total supply of money consists of coins plus notes plus bank deposits. How the rate of interest is determined by the equilibrium between the liquidity preference for speculative motive and the supply of money is shown in Fig. 56.6.

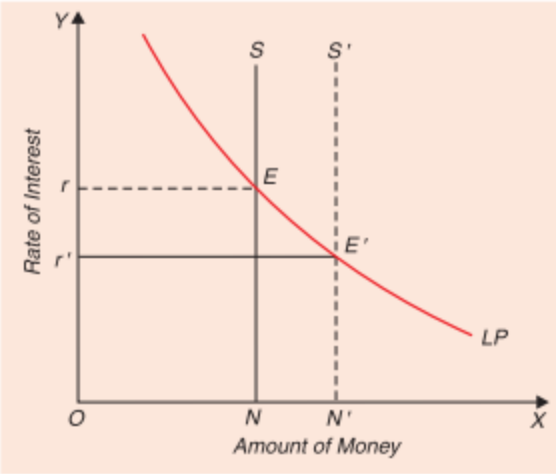


Fig. 56.6. *Equilibrium between Demand for and Supply of Money*

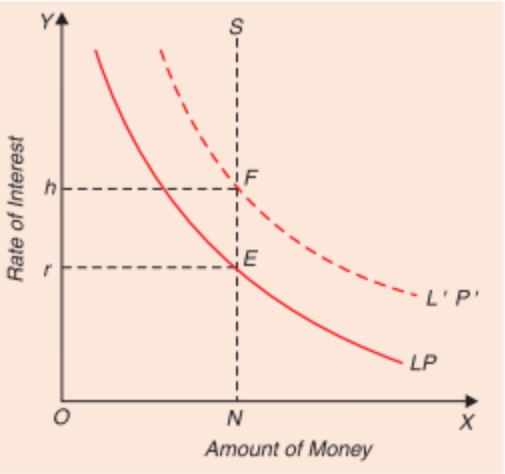


Fig. 56.7. *Effect of Increase in Liquidity Preference on the Rate of Interest*

In Fig. 56.6, *LP* is the curve of liquidity preference for speculative motive. In other words, *LP* curve shows the demand for money for speculative motive. To begin with, *ON* is the quantity of money available for satisfying liquidity preference for speculative motive. Rate of interest will be determined where the speculative demand for money is in balance or equal to the fixed supply of money *ON*. It is clear from the figure that speculative demand for money is equal to *ON* quantity of money at *Or* rate of interest. Hence *Or* is the equilibrium rate of interest. Assuming no change in expectations, an increase in the quantity of money (via open market operations) for the speculative motive will lower the rate of interest. In Fig. 56.6, when the quantity of money increases from *ON* to *ON'*, the rate of interest falls from *Or* to *Or'* because the new quantity of money *ON'* is in balance with the speculative demand for money at *Or"* rate of interest. In this case we move down along the curve. Thus, given the schedule or curve of liquidity preference for speculative motive, an increase in the quantity of money brings down the rate of interest.

But the act of increase in the quantity of money may cause a change in the expectations of the public and thereby cause an upward shift in liquidity preference curve for speculative motive bringing the rate of interest up. But this is not certain. "New developments may only cause wide differences of opinion leading to increased activity in the bond market without necessarily causing any shift in the aggregate speculative demand for money schedule. If the balance of market expectation is changed, there will be *shift* in the schedule. Central Bank policy designed to increase the money supply may therefore be met by an upward shift of speculative demand function leaving the rate of interest virtually unaffected." Thus, a large increase in the quantity of

money may exert only a small influence on the rate of interest in certain circumstances. It is worth mentioning that shift in liquidity preference schedule or curve can be caused by many other factors which affect expectations and might take place independently of changes in the quantity of money by the Central Bank. Shifts in the liquidity function may be either downward or upward depending on the way in which the public interprets a change in events. If some change in events leads the people on balance to expect a higher rate of interest in the future than they had previously supposed, the liquidity preference for speculative motive will increase which will bring about an upward shift in the curve of liquidity preference for speculative motive and will raise the rate of interest.

In Fig. 56.7, assuming that the quantity of money remains unchanged at *ON,* the increase in the liquidity preference curve from *LP* to *L'P '*, the rate of interest rises from *Or* to *Oh* because at *Oh,* the new speculative demand for money is in equilibrium with the supply of money *ON.* It is worth noting that when the liquidity preference for speculative motive increases from *LP* to *L'P',* the amount of money held does not increase ; it remains *ON* as before. Only the rate of interest rises from *Or* to *Oh* to equate the new liquidity preference for speculative motive with the available quantity of money *ON.*

Thus we see that Keynes explained interest in terms of purely monetary forces and not in terms of real forces like productivity of capital and thrift which formed the foundation-stones of both classical and loanable fund theories. According to him, demand for money for speculative motive together with the supply of money determines the rate of interest. He agreed that the marginal revenue product of capital tends to become equal to the rate of interest but the rate cf interest is not determined by marginal revenue productivity of capital. Moreover, according to him, interest is not a reward for saving or thriftiness or waiting but for parting with liquidity. Keynes asserted that it is not the rate of interest which equalizes saving and investment. But this equality is brought about through changes in the level of income.

Critical Appraisal of Keynes's Liquidity Preference Theory of Interest

i. *Keynes ignored real factors in the determination of interest.*

First, it has been pointed out that rate of interest is not a purely monetary phenomenon. Real forces like productivity of capital and thriftiness or saving also play an important role in the determination of the rate of interest. Keynes makes the rate of interest independent of the demand for investment funds. In fact, it is not so independent. The cash-balances of the businessmen are largely influenced by their demand for capital investment. This demand for capital-investment depends upon the marginal revenue productivity of capital. Therefore, the rate of interest is not determined independently of the marginal revenue productivity of capital (marginal efficiency of capital) and investment demand. When investment demand increases due to greater profit prospects or, in other words, when marginal revenue productivity of capital rises, there will be greater demand for investment funds and the rate of interest will go up. But Keynesian theory does not account for this. Similarly, Keynes ignored the effect of the availability of savings on the rate of interest. For instance, if the propensity to consume of the people increases, savings would decline. As a result, supply of funds in the market will decline which will raise the rate of only LM curve which shows various rates of interest at different levels of income interest.

ii. *Keynesian theory is also indeterminate.*

Exactly the same criticism applies to Keynesian theory itself on the basis of which Keynes rejected the classical and loanable funds theories. Keynes's theory of interest, like the classical and loanable funds theories, is indeterminate. According to Keynes, rate of interest is determined by the speculative demand for money and the supply of money available for satisfying speculative demand. Given the total money supply, we cannot know how much money will be available to satisfy the speculative demand for money unless we know how much the transactions demand for money is. And we cannot know the transactions demand for money unless we first know the level of income because money held under transactions motive depends on the level of income. Thus the Keynesian theory, like the classical, is indeterminate. "In the Keynesian case the supply and demand for money schedules cannot give the rate of interest unless we already know the income level; in the classical case the demand and supply schedules for saving offer no solution until the income is known. Precisely the same is true of loanable fund theory. Keynes criticism of the classical and loanable-fund theories applies equally to his own theory, In fact, as we shall see below from Keynes's liquidity preference theory, we can

derive only *L.M*. curve which shows various rates of interest at different levels of income when money market is in equilibrium.

iii. *No liquidity without savings.*

According to Keynes, interest is a reward for parting with liquidity and is in no way a compensation and inducement for saving or waiting. But without saving how can the funds be available to be kept as liquid and how can there be question of surrendering liquidity if one has not already saved money. Jacob Viner rightly maintains, "Without saving there can be no liquidity to surrender." Therefore, the rate of interest is vitally connected with saving which is neglected by Keynes in the determination of interest. It follows from above that Keynesian theory of interest is also not without flaws. But importance Keynes gave to liquidity preference as a determinant of interest is correct. In fact, the exponents of loanable funds theory incorporated the liquidity preference in their theory by laying greater stress on hoarding and dishoarding. We are inclined to agree with Prof. D. Hamberg when he says, "Keynes did not forge nearly as *new* a theory as he and others at first thought. Rather, his great emphasis on the influence of hoarding on the rate of interest constituted an invaluable addition to the theory of interest as it had been developed by the loanable funds theorists who incorporated much of Keynes's ideas into their theory to make it more complete."

PROFIT

Economists' concept of profit is of *pure profit*. It is also called '*economic profit*' or 'just profit'. 'Pure profit' is a return over and above the opportunity cost, *i.e.* the payment that would be 'necessary to draw forth the factors of production from their most remunerative alternative employment'. Pure profit may thus be defined as 'a residual left over after all contractual costs have been met, including the transfer costs of management, insurable risks, depreciation, and payments to shareholders sufficient to maintain investment at its current level'. In other words, *pure profit* equals *net profit* less opportunity costs of management, insurable risk, depreciation of capital, necessary minimum payments to shareholders that can prevent them from withdrawing their capital from its current use.

Profit as a Dynamic Surplus

A popular conception of profits is that they arise in a dynamic economy, that is, in an economy where changes are taking place. In a static economy where nothing changes there can be no profits. It was J.B. Clark who first propounded that profits are a dynamic surplus. He argued that in a stationary state where no changes in conditions of demand and supply are occurring, the prices paid to the factors on the basis of their marginal productivity would exhaust the total value production and no profits would accrue to the entrepreneur. Profits result when selling prices of the goods exceed their cost of production. Now, in a competitive long-run equilibrium, price equals average cost of production (including normal profits which are in fact wages for routine supervision and management) and therefore no pure profits are made. Now, if no changes either in the conditions of demand or in the conditions of supply occur, competitive equilibrium will persist and therefore no profits will be earned by the entrepreneur. On the contrary, if due to the changes in either demand or supply, price exceeds cost of production, profits will emerge. If due to these changes, price falls below the cost of production, negative profits, that is, losses, will accrue to the entrepreneur. It is evident that changes disturb the equilibrium and thereby give rise to profits. In other words, profits arise due to *disequilibrium* caused by the changes in demand and supply conditions. Prof. Stigler rightly says, "Firms in a competitive industry may receive profits.....because of a state of disequilibrium....these profits can arise even if all entrepreneurs are identical, for disequilibrium can characterize a whole industry. If prices are higher, or costs lower than were anticipated, entrepreneurs will receive a return in excess of the alternative product of their resources. If

prices were lower or costs higher than were anticipated, entrepreneurs will receive less than the alternative product of their resources, *i.e.,* negative profits. Positive profits may persist for a long time if firms outside the industry are slow to enter the industry and negative profits can persist as long as specialized equipment yields more when used in the industry than when used elsewhere, say as scrap."

It should be noted that these disequilibrium profits arise from *unanticipated* changes in demand or cost conditions. If the changes could have been foreseen in advance, then suitable adjustments could have been made according to the anticipated changes so that forces of competition would have driven profits to zero.

Now the question is, what changes occur in the economy and give rise to profits? J. B. Clark mentioned five changes that occur in a dynamic economy and which give rise to profits.

These five changes are: changes in the quantity and quality of human wants, changes in methods or techniques of production, changes in the amount of capital, and changes in the forms of business organization. These changes are constantly taking place and bring about the divergence between price and cost and thereby give rise to profits, positive or negative. If the demand for a commodity increases due to the increase in population or increase in the incomes of the people or due to the increase in consumers' preference for the commodity, the price of the commodity will rise, and if cost remains the same, profits would accrue to the entrepreneurs producing the commodity. On the other hand, cost of production may go down as a result of the adoption of a new technique of production, or as a result of cheapening of the raw material, and if price remains constant or does not fall to the same extent, the profits would emerge.

Innovations and Profits: Schumpeter's Theory of Profits

Successful innovations as important dynamic changes and as source of profit has been, in brief, explained above. But since innovation has been signed out as a very important factor responsible for the occurrence of profits to the entrepreneurs, it requires to be dealt with separately. It has been held by Joseph Schumpter that the main function of the entrepreneur is to introduce innovations in the economy and profits are reward for his performing this function. Now, what is innovation? Innovation, as used by Schumpeter, has a very wide connotation. Any *new* measure or policy adopted by an entrepreneur to reduce his cost of

production or to increase the demand for his product is an innovation. Thus, innovations can be divided into two categories.

First type of innovations are those which reduce cost of production, or in other words, which change the production functions. In this first type of innovations are included the introduction of a new machinery, new and cheaper technique or process of production, exploitation of a new source of raw material, a new and better method of organizing the firm, etc. Second types of innovations are those which increase the demand for the product, or in other words, which change the demand or utility function. In this category are included the introduction of a new product, a new variety or design of the product, a new and superior method of advertisement, discovery of new markets etc. If an innovation proves successful, that is, if it achieves its aim of either reducing the cost of production or enhancing the demand for the product, it will give rise to profit. Profits emerge because due to successful innovation either cost falls below the prevailing price of the product or the entrepreneur is able to sell more and at a better price than before. It should be noted that profits accrue not to him who conceives innovation nor to him who finances it but to him who introduces it. Further, whenever any new innovation is contemplated to be introduced, it always calls for a new combination of factors or reallocation of resources.

It is here worth mentioning that profits caused by a particular innovation are only temporary and tend to be competed away as others imitate and also adopt that. An innovation ceases to be new or novel, when others also come to know of it and adopt it. When an entrepreneur introduces a new innovation, he is first in a monopoly position, for the new innovation is confined to him only. He therefore makes large profits. When after some time others also adopt it in order to get a share, profits will disappear. If the law allows and the entrepreneur is able to get his new innovation, *e.g.,* new product patented, and then he will continue to earn profits.

But in a competitive economy and without patent laws, the existing competitors or the new firms will soon adopt any successful innovation and profits would be eliminated. But in a competitive and progressive economy the entrepreneurs always continue to introduce new innovations and thus profits continue emerging out of them. Thus. Prof. Stigler says, “Unless one can construct a permanent monopoly, such profits as are realized by successful innovation

are essentially transitional and will be eliminated by the attempts of other firms to share them. But these profits may exist for a considerable time because of the ignorance of other firms of their existence or because of the time required for the entry of new firms. More important, the successful innovator can continuously seek new disequilibrium profits since the horizon of conceivable innovations is unlimited."

Risk, Uncertainty and Profits: Knight's Theory of Profits

An important theory associates profit with risk and uncertainty. According to F. H. Knight, profit is a reward for uncertainty bearing. Even before Knight, F. B. Hawley and A. C. Pigou had pointed out that entrepreneurs earned profits because they had to bear the risks of undertaking production. But Knight has greatly developed the theory of profits based on uncertainty. He has distinguished between risk and uncertainty on the one hand and predictable and unpredictable changes on the other. According to him, dynamic changes give rise to profits in so far as changes and their consequences are unpredictable in character. Only those changes whose occurrence cannot be known before hand give rise to profit.

Profits arise because of the uncertainty of future. If the future conditions could be completely foreknown in the present, then competition would certainly adjust things to the ideal state where all prices would equal costs and profits would not emerge. Thus, it is our ignorance about the future and uncertainty of it that give rise to profits. In other words, it is the divergence of actual conditions from those which have been expected and on the basis of which business arrangements have been made that give rise to uncertainty and profits. Prof. A. K. Dass Gupta rightly maintains, "Uncertainty is thus a permanent feature of economic system. It is one of the limitations of human ingenuity that it cannot unearth the contents of the future. Trained instincts of businessmen coupled with statistical information may go a long way, but in so far as the course of nature (both physical and human) is anything but rhythmical, the future would always remain more or less of mystery." He further writes: "So long as entrepreneurs start operations with imperfect knowledge about the state of the market and so long as the anticipated marginal product of the hired factors deviates from their actual product, so long a surplus would persist."

We thus see that entrepreneurs have to undertake the work of production under conditions of uncertainty. In advance they have to make estimates of the future conditions regarding demand for the product and other factors which affect price and costs. In view of their estimates and anticipations, they make contract with the suppliers of factors of production in advance at fixed rates of remuneration. They realize the value of the output produced by the hired factors after it has been produced and sold in the market. But a good deal of time is spent in the process of producing and selling the product. It follows, therefore, that a good time gap elapses between the contracts made by the entrepreneur with the factors of production at fixed rates and the realization of sale proceeds from the output made by them. As mentioned before, these contracts are based upon anticipations about the future conditions. But between the time of contracts and the sale of the output many changes may take place which may upset anticipations for good or for worse and thereby give rise to the profits, positive and negative. Now, if the conditions prevailing at the time of the sale of output could be known or predicted when the entrepreneurs enter into contractual relationships with the factors of production about their rates of remuneration, there would have been no uncertainty and, therefore, no profits. Thus uncertainty, that is, ignorance about the future conditions of demand and supply, is the core of profits. It should be noted that positive profits accrue to those entrepreneurs who make correct estimate of the future or whose anticipations prove to be correct. Those whose anticipations prove to be incorrect will have to suffer losses.

We thus see that profit is a residual and non-contractual income which accrues to the entrepreneurs because of the fact of uncertainty. The entrepreneur is unhired factor; he hires others for work of production. It is, therefore, entrepreneur who bears uncertainty and earns profits as a reward for that. J.F. Weston who has been a prominent exponent and supporter of uncertainty theory of profits explains the emergence of profits in the following way: "Under uncertainty total product may not be equal to total costs (explicit and implicit) *because plans are not fulfilled.* How this occurs is briefly indicated. Two classes of owners of productive services are distinguished. First, those with rates of compensation fixed in advance of the determination of the results of operations are called *hired factors* and receive contractual returns. Second, those with rates of compensation dependent upon the results of operation are referred to as *unhired factors* who receive non-contractual or residual returns. Whatever the basis upon which contractual relationships have been entered, actual results will not have

been accurately foreseen because of uncertainty. Hence whatever *the basis upon which contractual commitments have been made events actually do not turn out that way.* This is the significance of economic profit. *It is not possible to plan in advance exactly what total product or total costs will be."*

Now, the question is what changes cause uncertainty? As has been explained earlier, there are *two types of changes* which take place and are responsible for conditions of uncertainty. First types of changes refer to the innovations (for example, introduction of a new product or a new and cheaper method of production etc.) which are introduced by the entrepreneurs themselves. These innovations not only create uncertainty for the rivals or competitors who are affected by them but they also involve uncertainty for the entrepreneur who introduces them, for one cannot be certain whether a particular innovation will be definitely successful. The second type of changes which cause uncertainty are those which are external to the firms and industries. These changes are: changes in tastes and fashions of the people, changes in Government policies and laws especially taxation, wage and labour policies and laws, movements of prices as a result of inflation and depression, changes in the incomes of the people, changes in production technology etc. All these changes cause uncertainty and bring profits, positive or negative, into existence.

We have seen above that entrepreneurs work under conditions of uncertainty and that they bear uncertainty and earn profits as a reward for that. Here a distinction drawn by F. H. Knight between insurable risk and non-insurable risk is worth mentioning. Because of the changes that are continuously occurring in the economy, entrepreneur has to face many risks. But all risks do not cause uncertainty and give rise to profits. It is only non-insurable risks that involve uncertainty and the entrepreneur earns profits for bearing these non-insurable risks. This raises the question as to what kind of risks are insurable and what non-insurable. The entrepreneur faces risks like fire, theft, accident etc. which may cause him huge losses. But these risks of fire, theft, accident etc. can be insured against on the payment of a fixed premium. Insurance premium is included in the cost of production. Thus, no uncertainty arises due to insurable risks as far as individual entrepreneurs are concerned and therefore they cannot give rise to profits.

Only those risks can be insured the probability of whose occurrence can be calculated. Thus, an insurance company knows by its calculation on the basis of past statistics that how

much percentage of the factories will catch fire in a year. On the basis of this information, it will fix the rate of premium and is able to insure the factories against this risk. But there are risks which cannot be insured and therefore they have to be borne by the entrepreneurs. These non-insurable risks relate to the outcomes of the price-output decisions to be taken by the entrepreneurs. Whether it will pay him to increase output, reduce output and what will be the outcome in terms of profits or losses as a result of his particular output decision. Again, whether it will pay him to lower price or to raise it and when he takes a particular price decision whether he would make profits or losses. Similarly, he has to face risks as a result of his decisions regarding mode of advertisement and outlay to be made on it, product variation etc. For taking all these decisions he has to guess about demand and cost conditions and there is always risk of suffering losses as a result of these decisions. No insurance company can insure the entrepreneurs against commercial losses which may emerge out of decisions regarding price, output, product variation and also against the losses which may fall upon the entrepreneurs due to the structural, cyclical and other exogenous changes which take place in the economy. It is, therefore, clear that *it is non-insurable risks that involve uncertainty and give rise to profits.* To quote Knight, "It is 'uncertainty' distinguished from insurable risk that effectively gives rise to the entrepreneurial form of organization and to the much condemned 'profit' as an income form."

Knight's theory rightly emphasized that uncertainty in a dynamic business environment gives rise to profits. The bearing of uncertainty is an important function to be performed by the entrepreneurial class in a society. This makes the profits as functional income that is earned for performing the function of uncertainty bearing. But all profits are not functional nor do they arise due to uncertainty. As we have studied in the theories of monopoly, monopolistic competition, oligopoly in product markets and monopsony in factor markets, that a significant amount of profits arises due to the monopoly power enjoyed by the producers. Knight's theory ignores the fact of monopoly power giving rise to profits. M. Kalecki whose theory of profits we will discuss below propounded that monopoly power was an important source of profits in modern free-market economies. It is also worth mentioning that monopoly profits can arise even in a stationary economy where no changes are taking place.

Further, in our view that while the profits may emerge as a result of dynamic changes, innovations and uncertainty but the *appropriation* of profits by the recipients is *institutionally* determined. The *separate* theories must be developed to explain the *origin* of profits and *sharing out* of profits. For instance, Weintraub points out that "The entrepreneur's function is directed and executed with an eye towards profit making; the sharing of profits is an entirely separate issue." Likewise, R.A. Gordan states: "It is important to bear in mind that accounting for the origin and existence of an income does not in itself explain its allocation to particular persons or classes. The way in which a given income accrues to various individuals will depend largely on how those individuals fit into the institutional setting which prevails. Thus, though 'pure profits' may be explainable in terms of change, uncertainty and friction (essentially a nonfunctional explanation), the fact that these gains go to particular persons is related not to exercise of entrepreneurial function, but rather to the nature and distribution of ownership rights."

Bibliography

Prescribed Books:

M.L. Seth, Micro Economics/ Principles of Economics
M.L. Jhingan, Micro Economics
H.S. Agarwal, Micro Economics

Reference Books:

Henderson J and R.E Quandt (1980), Microeconomic Theory: A mathematical approach, McGraw Hill; New Delhi.

Koutsoyiannis, A. (1990), Modern Microeconomics, Macmillan.

Lipsey, R.G. and K.A. Chrystal (1999), Principles of Economics (9th Ed) Oxford University Press, Oxford.

Samuelson, P.A. and W.D. Nordaus (1998), Economics, Tata McGraw Hill, New Delhi.

Stonier, A.W. and D.C. Hague (1972), A Textbook of Economic Theory, ELBS and Lognman Group, London.

Microeconomics I by D. N. Dwivedi

Microeconomics II by D. N. Dwivedi

Advanced Economic Theory (Microeconomic Analysis), 21st Edition by Ahuja H.L.

https://www.economicsdiscussion.net/