B.A. First Year

Economics, Paper - I

MICRO ECONOMICS

मध्यप्रदेश भोज (मुक्त) विश्वविद्यालय – भोपाल

MADHYA PRADESH BHOJ (OPEN) UNIVERSITY - BHOPAL

Reviewer Committee

1. Dr (Prof) Manish Sharma
   Professor
   IEHE, Bhopal

2. Dr Kalpana Malik
   Associate Professor
   IEHE, Bhopal

3. Dr. Bindu Mahawar
   Assistant Professor
   Navin College, Bhopal

Advisory Committee

1. Dr Jayant Sonwalkar
   Hon'ble Vice Chancellor
   Madhya Pradesh Bhoj (Open) University, Bhopal

2. Dr H.S.Tripathi
   Registrar
   Madhya Pradesh Bhoj (Open) University, Bhopal

3. Dr L.P.Jharia
   Director Student Support
   Madhya Pradesh Bhoj (Open) University, Bhopal

4. Dr (Prof) Manish Sharma
   Professor
   IEHE, Bhopal

5. Dr Kalpana Malik
   Associate Professor
   IEHE, Bhopal

6. Dr. Bindu Mahawar
   Assistant Professor
   Navin College, Bhopal

COURSE WRITERS

Dr. D. N. Dwivedi, *Professor of Economics, Maharaja Agrasen Institute of Management Studies*
(Units: 1-5)

Published by Registrar, MP Bhoj (Open) University, Bhopal in 2020

Vikas® is the registered trademark of Vikas® Publishing House Pvt. Ltd.

VIKAS® PUBLISHING HOUSE PVT. LTD.
E-28, Sector-8, Noida - 201301 (UP)
Phone: 0120-4078900 • Fax: 0120-4078999
Regd. Office: A-27, 2nd Floor, Mohan Co-operative Industrial Estate, New Delhi 1100 44
• Website: www.vikaspublishing.com • Email: helpline@vikaspublishing.com

SYLLABI-BOOK MAPPING TABLE

Micro Economics

Syllabi	Mapping in Book
Unit-1: Definition, Scope and Nature of Economics, Methods of Economic Analysis - Inductive and Deductive, Basic Concepts; Micro Economics, Utility, Demand, Supply, Commodity, Free Goods. Value and Price, Market Administered Price.	**Unit-1:** Introduction to Economics **(Pages 3-26)**
Unit-2: Law of Demand and its Exceptions, Giffin Goods, Elasticity of Demand; Price, Income and Cross. Law of Supply, Law of Diminishing Marginal Utility and Equi-Marginal Utility, Consumer's Surplus, Indifference Curves-Characteristics and Consumers Equilibriums.	**Unit-2:** Demand and Supply **(Pages 27-68)**
Unit-3: Production Law of Production, Law of Variable Proportions, Returns to Scale, Economies of Scale, Iso Product curves, characteristics and Producer's Equilibrium, Concepts of Cost and Revenue, Total Marginal & Average.	**Unit-3:** Production, Cost and Revenue **(Pages 69-100)**
Unit-4: Market Meaning, Type of Markets, Demand and Supply Equilibrium, Price and Output Determination for Firm and Industry under Perfect Competition, Monopoly and Monopolistic Competition.	**Unit-4:** Market and Price Determination **(Pages 101-157)**
Unit-5: Factor Pricing; Marginal Productivity Theory, Adding up Theorem, Modern Theories of Wages, Interest, Profit and Rent.	**Unit-5:** Factor Pricing and Factor Market **(Pages 159-218)**

CONTENTS

INTRODUCTION

NOTES

Micro Economics is that branch of economics that studies the behaviour of individuals and firms in making decisions regarding the allocation of scarce resources and the interactions among these individuals and firms. In micro economics, the following theories are dealt with:

Demand theory deals with consumers' behaviour. It answers such questions as: How do the consumers decide whether or not to buy a commodity? How do they decide on the quantity of a commodity to be purchased? When do they stop consuming a commodity? How do the consumers behave when price of the commodity, their income and tastes and fashions, etc., change? At what level of demand, does changing price become inconsequential in terms of total revenue? The knowledge of demand theory can, therefore, be helpful in making the choice of commodities, finding the optimum level of production and in determining the price of the product.

Production theory explains the relationship between inputs and output. It also explains under what conditions costs increase or decrease; how total output behaves when units of one factor (input) are increased keeping other factors constant, or when all factors are simultaneously increased; how can output be maximized from a given quantity of resources; and how can the optimum size of output be determined? Production theory, thus, helps in determining the size of the firm, size of the total output and the amount of capital and labour to be employed, given the objective of the firm.

Price theory explains how price is determined under different kinds of market conditions; when price discrimination is desirable, feasible and profitable; and to what extent advertising can be helpful in expanding sales in a competitive market.

Thus, price theory can be helpful in determining the price policy of the firm. Price and production theories together, in fact, help in determining the optimum size of the firm. Profit making is the most common objective of all business undertakings. But, making a satisfactory profit is not always guaranteed because a firm has to carry out its activities under conditions of uncertainty with regard to: (*i*) demand for the product, (*ii*) input prices in the factor market, (*iii*) nature and degree of competition in the product market, and (*iv*) price behaviour under changing conditions in the product market, etc. Therefore, an element of risk is always there even if the most efficient techniques are used for predicting the future and even if business activities are meticulously planned. The firms are, therefore, supposed to safeguard their interest and avert or minimize the possibilities of risk.

Profit theory guides firms in the measurement and management of profit, in making allowances for the risk premium, in calculating the pure return on capital and pure profit and also for future profit planning.

NOTES

This book, *Micro Economics* is divided into five units that follow the self-instruction mode with each unit beginning with an Introduction to the unit, followed by an outline of the Objectives. The detailed content is then presented in a simple but structured manner interspersed with Check Your Progress Questions to test the student's understanding of the topic. A Summary along with a list of Key Terms and a set of Self-Assessment Questions and Exercises is also provided at the end of each unit for recapitulation.

UNIT 1 INTRODUCTION TO ECONOMICS

NOTES

Structure

1.0 INTRODUCTION

A general and, perhaps, natural curiosity of a student who begins to study a science or discipline is to know the nature and scope of his subject of study. As such, a student of economic science would like to know 'what is economics'; 'what is its subject matter'; and 'what is the nature of economic science'. It may be noted at the very beginning that there is no precise answer to these questions. Attempts made by economists to answer these questions have not yielded any precise and universally accepted definition of economics. In fact, there has been a great deal of controversy among economists of different generations on the nature and scope of economics. The main reason for inadequate and controversial definitions of economics is that it is still an unfinished science and an attempt to define an unfinished science is bound to yield an inadequate and controversial definition. As J.S. Mill pointed out nearly one-and-a half centuries ago, *definition of a science invariably follows, not precedes the creation of a science,* and economics has not yet matured into a perfect science. Zeuthen's remark that 'economics is an unfinished science' still holds good. It should, therefore, not be surprising if economic science could not be defined precisely.

Economists have defined economics and delimited its scope and subject matter differently at different stages of its growth as a social science. A brief review of some popular definitions of economics will show the evolution of the definition of economics. It will also help in understanding the nature and scope of economics.

NOTES

1.1 OBJECTIVES

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

- Define the term macroeconomics and microeconomics
- State the nature and scope of economics
- Discuss the methods of conducting economic analysis

1.2 DEFINITION, NATURE AND SCOPE OF ECONOMICS

The various definitions and views on the nature and scope of economics may be chronologically grouped under four broad categories, *viz.,* (*i*) Pre-classical Definitions; (*ii*) Classical Definitions; (*iii*) Neo-classical Definitions; and (*iv*) Scarcity or Modern Definitions. In the following sub-sections of this unit, we shall briefly discuss these groups of definitions and also the corresponding scope of economics.

Pre-Classical Definitions

Economics was in its embryonic stage till the middle of the 18th century. It was then a hybrid of politics, ethics, logic and philosophy. During this period, Greek philosophers, *viz*., Aristotle and Xenophon viewed economics as 'an art of household management'. In later years, wealth gained an important place and role in the life of nations. Wealth has been a matter of great importance for all, in all ages. But in the early middle ages, the object which concerned statesmen and merchants most was the supply of precious metals which they thought was the best indication, if not the chief cause, of material prosperity of both individuals and the nation. The economic thinkers of those days, therefore, concerned themselves with the issues pertaining to the acquisition and management of national wealth. As a result, during the period of Mercantilists, economics rose to the status of 'Political Economy', and the scope of economics was widened to include the management of national wealth. But economics remained a part of Social Studies.

Classical Definition: Adam Smith

Attempts to distinguish economics as a separate branch of Social Studies began with Adam Smith's famous treatise, *An Enquiry into the Nature and Causes of Wealth of Nations* (1776), popularly known as *Wealth of Nations*. Adam Smith, sometimes called the father of economics, defined economics as 'a subject concerned with an enquiry into the nature and causes of wealth of nations.' He also called it a 'science of wealth'. Most classical economist followed and supported Smith's definition of economics. Thus, the classical economists limited the scope of economics to the enquiry of material wealth and prosperity of nations. The emphasis on material wealth and prosperity as the subject matter of economic science came under severe criticism by other sections of social thinkers of the Christian Society, who held spirituality higher than other human values and aspirations. Thomas Carlyle, a historian called economics the 'gospel of mammon'.

Ruskin, a social reformer, condemned it as a 'dismal science', as a 'bastard science', and as a 'science of bread and butter'. Obviously, Carlyle, Ruskin and others with a similar inclination towards economics mounted their attack not only on the definition of economics but also on the social relevance of economic laws and their impact on social life.

It may also be argued that the definition of economics given by Adam Smith and his followers took a very narrow view of economic science compared to its modern connotation. It delimited the scope of economic behaviour, the main subject matter of modern economics, relegated economic studies to a position secondary to the acquisition of material wealth and prosperity. Smith's definition however remained in vogue till the end of the 19th century when Alfred Marshall attempted to redefine economics and widened its scope.

Neo-Classical Definition: Alfred Marshall

Alfred Marshall, a pioneer neo-classical economist reoriented economics towards the 'study of mankind' and provided economic science with a more comprehensive definition. In Marshall's own words, 'Political Economy or Economics is a study of mankind in the ordinary business of life; it examines that part of individual and social action which is most closely connected with the attainment and with the use of the material requisites of well being,' He added that economics "is on the one side a study of wealth; and on the other and more important side, a part of study of man.' As is obvious from his definition, Marshall widened the scope of economics to include the study of makind and their activities undertaken to promote their material welfare. He rather emphasized that man and his material welfare are a more important side of economic studies than the 'nature and causes of wealth'. Following Marshall, other economists of the neo-classical tradition defined economics with similar connotations. For example, A.C. Pigou gave a restrictive definition of economics. According to him, enquiry of economics is 'restricted to that part of social welfare which can be brought directly or indirectly into relationship with the measuring rod of money.' In Cannan's view, 'The aim of political economy is the explanation of the general causes on which the material welfare of human beings depends."

Robbins' Criticism of Neo-Classical Definitions

The neo-classical definition of economics, often called the 'welfare definitions' or 'materialist definitions', remained widely accepted and unchallenged until the publication of Lionel Robbins' *An Essay on the Nature and Significance of Economic Science* in 1932. Robbins examined the validity of neo-classical definitions in the light of theories of wages and exchange as developed by them, and pointed out the following *deficiencies* in 'materialist definitions'.

First, Robbins questioned the neo-classical division of human activities between 'economic' and 'non-economic' associated, respectively, with 'material' and 'non-material' welfare. He argued that considering only the economic activities as the subject matter of economics is not plausible with the theory of wages which 'was an integral part of any system of economic analysis'. According to the materialist definition, the service of an orchestra player or a singer is unproductive

NOTES

as it does not yield any material good; it does not have the remotest bearing on material welfare; and hence, it will not form a part of the subject matter of economics. But the service of an orchestra player has a market value and is as much a part of the wage theory as any service in the neo-classical sense, e.g., the service of a textile worker. Similarly, activities of government servants, musicians, Churchmen, lawyers, physicians, buffoons, etc., are 'non-material' in nature and do not contribute to 'material welfare' and yet their activities are a part of the subject matter of economics.

Second, Robbins pointed out that the 'materialist definition of economics contains certain contradictions in itself. For example, on the one hand, neo-classical economists admit that since wars do not contribute to 'material welfare', they are not included in the subject matter of economics. But, on the other hand, they admit the impact of war on the national economy.

Third, Robbins has criticized the 'materialist definition' also for being *classificatory* rather than *analytical*. The Materialist definition classified human activities into such categories as 'economic' and 'non-economic'; 'productive' and 'unproductive'; and activities related to material welfare and otherwise. And, it considers only those human activities that are undertaken to procure material welfare as the subject matter of economics such as it is, other kinds of human activities, whatever their contribution to human welfare, were kept outside the scope of economic studies by the materialist definition.

Fourth, in distinguishing his own definition from the 'materialist definition' he has argued that neo-classical definitions emphasize the *materiality* of human activities, i.e., whether or not they yield material goods. But mere *materiality* of a product does not qualify it for being included in the subject matter of economics. It is in fact the scarcity of the material goods and their exchange value which make them an economic good. 'For', as Robbins puts it, 'it is not the *materiality* of even material means of gratification which gives them their status as economic goods; it is their relation to valuation,' whereas the 'materialist' definition emphasizes only its materialness. The 'materialist' definition of economics therefore misrepresents the science as we know of it.

Finally, Robbins disputed the Marshallian conception of economics as a 'social science'. In Marshall's opinion, economics is a 'social science'. Economics as a 'social science' studies the behaviour of an individual, as Marshall said, 'in his ordinary business of life' as a member of society. Thus, according to Marshall, the activities of an individual living in seclusion fall outside the scope of economics. But according to Robbins, economics is a 'human science' and according to him, must also include the activities of a recluse, like Robinson Crusoe, because he has limited resources (time and other means) to meet his requirements. He has, therefore, to make a choice between his ends and to divide 'his time between the production of real income and the enjoyment of leisure.'

In a nutshell, Robbins has concluded his argument against the 'materialist' definition in the following words, '[It is] misleading to go on describing Economics as the study of the causes of material welfare... . Whatever Economics is concerned with, it is not concerned with the causes of material welfare as such.'

Scarcity Definition: Leonen Robbins

NOTES

Robbins' Definition of Economics: After having argued for the rejection of the neo–classical or what he calls 'materialist' definitions of economics, Robbins provided an alternative 'working' definition of economic science. He defined economics in the following words: 'Economics is a science which studies human behaviour as a relationship between ends and scarce means which have alternative use. 'Robbins's definition implies that economics essentially deals with human behaviour which are related to the allocation of scarce resources between their alternative uses. Briefly speaking, it studies the economizing behaviour of human beings. As Robbins has observed, the problem of allocation of resources—time and other means—arises out of the following 'four fundamental characteristics' of human existence.

First, the 'ends' or human wants to which resources are to be put are 'various', rather unlimited.

Second, 'material means of achieving [these] ends are limited.' That is, means of production available to a society for satisfying the wants of its people during a specific period of time are not unlimited in the economic sense: resources are scarce in relation to human wants.

Third, resources are capable of being put to alternative uses but productivity of resources varies from use to use. This property of resources creates conditions for making a choice between the alternative uses.

Fourth, all ends or human wants are not equally important. Some wants are more urgent and pressing than others. Some wants can be postponed and some cannot be. The variety and high and low urgency of various wants also necessitate the making of choice between wants.

Another fact of human life, which may be added to Robbins' list, is people's desire to maximize their satisfaction or gains—households want to maximize their income out of their limited sources of earning (i.e., land, labour, capital); consumers want to maximize their satisfaction out of limited disposable incomes; firms want to maximize their profits from the limited resources. This is rather a much more important factor in economizing behaviour of the human beings.

In brief, fulfilment of unlimited wants requires the use of limited resources. But, since the resources have alternative uses, again the problem of choice arises. One is supposed to make a choice between the alternative uses of resources so that the selected wants are best fulfilled. This human behaviour is essentially the behaviour of resource allocation, i.e., allocating the limited resources between the competing ends in a manner that best serves human interest. *It is this human behaviour which*, according to Robbins, *is the subject matter of economics*. In his own words, 'Economics...is concerned with that aspect of behaviour which arises from the scarcity of means to achieve given ends.'

Evaluation of Robbins' Definition

Robbins' definition is *superior to classical and neo-classical definitions* in many respects. It has certain merits which other definitions do not possess.

NOTES

First, Robbins's definition, is *analytical* as it focuses attention on a particular *aspect* of human behaviour caused by scarcity of resources. Besides, economics, as conceived by Robbins, seeks to *analyse* and generalize the causes of a particular aspect of human behaviour (i.e., economising behaviour), rather than studying a particular *kind* of behaviour whether it is economic or non-economic.

Second, according to Robbins, 'Economics is entirely neutral between ends...Economics is not concerned with ends as such.' Nor is the economist concerned with ends as such. Economics 'is concerned with the way in which the attainment of ends is limited' irrespective of whether it is noble or base. Thus, *Robbins' conception of economics imparts to it the nature of a pure science*.

Third, Robbins' definition widens the scope of economics to include all those human activities which are performed in relation to various ends and scarce means irrespective of whether they pertain to material welfare or not. Robbins considers producing 'philosophy' as much an economic activity as producing 'potatoes', for if total resources are spent on producing one, the other will have to be foregone. That is, all human activities having any economic aspect lie within the scope of economics.

Finally, Robbins' definition has the quality of being universally applicable. It unifies economic science under all systems. Unlike the 'materialist' definition, Robbins' conception of economics is said to be applicable under barter as well as under money exchange, under individual as well as social conduct, under capitalist as well as under socialist society. For, such basic 'facts' of human life as scarcity of resources and multiplicity of ends exist under all these systems of human existence.

Criticism: Despite the above merits of Robbins' definition of economics, it has been severely criticized. Some of the major criticisms levelled against Robbins' definition are following.

First, Robbins' definition has been criticized for its 'positivism' as it recognizes only the 'positive' aspect of human behaviour. It ignores completely the 'normative' aspect of human life. In other words, it excludes the welfare aspect of human conduct.

Second, Robbins has been criticised also for creating an *artificial* distinction between 'ends' and 'means'. For, what is 'end' at one stage of human activity may take the form of 'means' at another. But Robbins argues that the confusion between 'ends' and 'means' is unjustified. According to him 'ends' are 'tendencies' to which people conduct themselves and 'which can be defined and understood.'

For example, money-making in itself is not an end. 'The money is clearly a means to ultimate purchase. It is sought, not for itself, but for the things on which it may be spent...money-making in this means securing the means for the achievement of all those ends which are capable of achievement by the aid of purchasable commodity.'

Third, it has been alleged that Robbins' conception of economics as a science of choice-making extends the scope of economics to the choices where no cost is

involved. For example, there is an economic problem in making the choice between leisure and worship, but there is no economic problem involved if the choice is between the deities to be worshipped. However, this criticism and alike seem to have arisen only because of an unimaginative extension of scope of economics by critics. It is not difficult to determine what choices involve economic problems and what do not.

Fourth, if one examines Robbins' definition against the present status or scope of economics, one will find the definition very inadequate or even misleading. As Schultz has remarked Robbins' definition is misleading as 'it does not fully reflect two of the major concerns of modern economics, growth and stability.'

Conclusion: Despite the above allegations, Robbins' definition of economics has been widely accepted by economists. The evidence of its wide acceptability is that many modern economists have defined economics in Robbinsonian fashion with, of course, some modifications, without altering the spirit of his definition. For example, Paul A. Samuelson has defined economics as "the study of how people and society end up choosing with or without the use of money, to employ scarce productive resources that could have alternative uses to produce various commodities and distributing them for consumption, now or in the future, among various persons and groups in society. It analyses the costs and benefits of improving patterns of resources allocation.' Samuelson's definition is more comprehensive and self-explanatory. Many other economists, viz., Cairncross, Alfred Stonier and Douglas Hague, Tibour Scitovsky, C.E. Furguson, and others, have also defined economics in terms of 'scarcity of resources' and problem of 'resource allocation' or resource management.

This, however, should not mean that Robbins' definition is the final word on the nature and scope of economics. 'Economics is still a very young science and many problems in it are almost untouched.' Its boundaries continue to be enlarged to cover numerous economic issues never thought of half-a-century ago. As a result, the scope of economics continues to expand. As such, it will be extremely difficult, if not impossible, to determine the scope of economics for all times to come.

Scope

As noted above, the scope of economics is not marked precisely and, as it appears, it cannot be. However, the scope of economics, as it is known today, has expanded vastly in the post-World War II period. *Modern economics* is now divided into two major branches: *Microeconomics* and *Macroeconomics.* A brief description of the subject matter and approach of microeconomics and macroeconomics follows.

Microeconomics

Microeconomics is concerned with microscopic study of the various elements of the economic system and not with the system as a whole. As Lerner has put it, 'Microeconomics consists of looking at the economy through a microscope, as it

NOTES

were, to see how the million of cells in body economic—the individuals or households as consumers and the individuals or firms as producers—play their part in the working of the whole economic organism'. Thus, micro economics is the study of the economic behaviour of the individual consumer and producer and of individual economic variables, i.e., production and pricing of individual goods and services. Microeconomics studies how consumers and producers make their choices; how their decisions and choices affect the market demand and supply conditions; how consumers and producers interact to settle the prices of goods and services in the market; how prices are determined in different market settings; and how total output is distributed among those who contribute to production, i.e., between landlords, labour, capital supplier and the entrepreneurs. Briefly speaking, *theory of consumer behaviour, theories of production and cost of production, theory of commodity and factor pricing, efficient allocation of output and factors of production (called welfare economics)* constitute the main themes of *microeconomics*.

Macroeconomics

Macroeconomics is a relatively new branch of economics. It was only after the publication of Keynes's *The General Theory of Employment, Interest and Money* in 1936, that macroeconomics crystallized as a separate branch of economics. *Macroeconomics* studies the working and performance of the economy as a whole. It analyses behaviour of the national aggregates including national income, aggregate consumption, savings, investment, total employment, the general price level and the country's balance of payments. According to Boulding, 'Macroeconomics is the study of the nature, relationship and behaviour of aggregates and averages of economic quantities'. He contrasts macroeconomics with microeconomics in the following words: 'Macroeconomics ... deals not with individual quantities, as such, but aggregates of these quantities—not with individual incomes, but with the national income, not with individual prices but with price levels, not with individual output but with the national output.' More importantly, macroeconomics analyses the relationship between the national aggregate variables and how aggregate variables interact with one another to determine one another. It studies also the impact of public revenue and public expenditure, government's economic activities and policies on the economy. An important aspect of macroeconomics studies is the consequences of international trade and other economic relations between nations. The study of these aspects of economic phenomena constitutes the major themes of *macroeconomics.*

Specialized Branches of Economic Studies

In addition to microeconomics and macroeconomics, many *specialized branches of economics* have come up over time as a result of the growing need for intensive and extensive study of certain aspects of microeconomics or macroeconomics. Some of the *major specialized fields of economic studies* are listed below with a brief description of their subject matter.

1. ***Economics of Development*** deals with the factors that determine economic development and growth of a country, the causes of under-

development, unemployment and poverty in less developed countries, problems faced in accelerating the pace of development and suggests policy measures to achieve a sustainable high growth rate of the economy and employment.

NOTES

2. ***Public Economics*** examines, the economic role of government, sources of government revenue, government's fiscal policy, effects of taxation and public expenditure, causes and consequences of budgetary and fiscal deficits, if any, rationale for and consequences of public sector economic activities.
3. ***Monetary Economics*** studies the monetary affairs of the country including demand for and supply of money, working of the money market, credit and financial system, and management of the monetary sector.
4. ***International Economics*** studies the causes and consequences of international trade in goods and services, international flow of capital, international monetary and financial institutions, balance of payments and international payment system.
5. ***Industrial Economics*** is concerned with the working, growth and structure of the industrial sector (firms and industries) of the country, management and organization of industries, and problems and prospects of industrial growth.
6. ***Labour Economics*** examines the problems faced by labour as an economic class and problems associated with labour organizations, labour productivity and wages, exploitation of labour, labour welfare schemes and labour laws and their effects.
7. ***Econometrics*** is the study of statistical and mathematical techniques applied to economic data with a view to testing a hypothesis, to quantify the relationship, if any, between dependent and independent economic variables and to measure the effects of economic policies.
8. ***Economic History*** studies past economic record of a country or group of countries and of big historical economic events, e.g., Industrial Revolution and the Great Depression, often with the objective of bringing out unknown facts to light and also to know how past experience can be used to promote economic growth in future.
9. ***History of Economic Thoughts*** is the study of evolution and development of economic thoughts and ideas, their background, their logic and flaws. It contributes to the understanding of economic science.
10. ***Comparative Economic Systems*** is a comparative study of economic systems—capitalist or market economy, socialist or centrally planned and mixed economy systems—to understand their advantages and disadvantages and their strong and weak points and their social desirability.
11. ***Regional Economics*** studies the development of various regions of a country; it looks into the causes of imbalance in regional development, it examines why growth of urban economy is faster that of rural economy.

NOTES

12. ***Industrial Finance*** is concerned with the development and working of the financial sector, especially the financial institutions that cater to the financial requirement of the industries and of the capital market, and it studies how fluctuations in the financial sector affect the working and growth of the industrial sector.
13. ***Environmental Economics*** examines how industrial growth affects, rather destroys, the natural environment of the country and how world industrial growth affects the global environment and causes global warning and affects climatic conditions.
14. ***Managerial Economics*** studies how economic theories, concepts and tools of analysis can be applied to business decision-making and to understand the business environment of the country.

To sum up, the scope of economics is very vast. It may be added here that, in addition to the subject matter mentioned above, economics provides logic and reasoning, tools and technique, and analytical framework to analyse economic phenomena and to predict the consequences of change in economic conditions. It may thus be concluded that economics as a science studies the economic behaviour of people and its consequences at both the micro and macro levels; it brings out the cause-and-effect relationship between economic events; provides the tools and techniques of analying economic phenomenon and the basis for predicting the consequences of economic decisions and economic events. Economics studies economic phenomena systematically and methodically. The scientific method of economic inquiry imparts economics the status of a *science*.

Nature of Economics

The nature of economics can be understood by asking the fundamental question: is economics a Positive or a Normative science?

A positive science studies the phenomena as they actually are or as they actually happen. It does not involve any value judgement on whether what happens is good or bad, desirable or undesirable. A normative science, on the other hand, involves value judgement on whether what happens is socially desirable or undesirable, and if undesirable, how it can be made desirable. As J.N. Keynes puts it, "…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." Friedman has defined 'positive science' more elaborately and clearly. In his own 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." Judged against these definitions of positive and normative science, economics as a social science deals with both positive and normative economic questions: 'what is' and 'what ought to be'. Thus, economics is both a positive and a normative science. Let us look at positive and normative character of economic science in some detail.

NOTES

Economics as a Positive Science

Economics as a positive science seeks to analyse systematically and explain economic phenomena as they actually happen; find common characteristics of economic events; brings out the 'cause and effect' relationship between the economic variables, if any; and generalizes this relationship in the form of a theoretical proposition. One of the main purposes of economic studies is 'to provide a system of generalization' in the form of economic theories that can be used to make predictions about the future course of related events. It means that economics has a positive character. Economics explains the economic behaviour of individual decision-makers under given conditions; their phenomena. This makes economics a positive science. Here, 'positive' does not mean that theoretical statements are positively true: it means that it has a great possibility to occur if conditions are fulfilled.

Economics as a Normative Science

Economics as a normative science is concerned with ideal economic situation, not with what actually happens. Its objective is to examine 'what actually happens' from moral and ethical points of view and to judge whether 'what happens' is socially desirable. It examines also whether economic phenomena like production, consumption, distribution, prices, etc. are socially desirable or undesirable. Desirability and undesirability of economic happenings are determined on the basis of socially determined values. Thus, normative economics involves value judgement and values are drawn from the moral and ethical values and political aspirations of the society. In simple words, normative side of economics deals with such normative questions as 'what ought to be?' and whether 'what happens' is good or bad from society's point of view? It not, then how to correct it.

The need for such studies arises because 'what is' or 'what is being produced and consumed' may not be desirable or it may not be in the interest of the society. For example, production and sale of harmful goods like alcohol, drugs, cigarettes, gutka and pan masala, may be a very profitable business. But, 'Is production and sale of these goods desirable for the society?' is a normative question—a question in public interest. Economics 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 various economics activities and events and prescribes control and regulatory measures. Consider another economic problem—the issue of rent control. Given the growth of population and supply of houses in India, house rents, if not controlled, will increase, and have, in fact, increased exorbitantly. 'Should house rents be allowed to increase depending on the demand and supply conditions or be controlled and regulated to protect the interest of tenants?' is a normative question—a question in public interest.

Economics as a normative science examines the issue from society's angle including interest of both landlords and the tenants,and prescribes the reasonable rate of house rents and measures to implement it. Since economics 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

NOTES

economics, consider the issue of foodgrain prices in India. Recall that in 2001, there was surplus foodgrain production in India, on the one hand, and large-scale starvation and starvation deaths reported from different parts of the country. This was paradoxical situation. Yet, the Food Corporation of India (FCI), responsible for fixing the foodgrain price, did not allow foodgrain prices to go down. This problem can be examined from both positive and normative angles. Examining 'how price of foodgrains is determined?' is a question for positive economics and 'how should the prices of foodgrains be determined?' is a question for normative economics. It may thus be concluded that economics is both a positive and normative science.

However, it is important to note that economics is fundamentally a positive science. It acquires its normative character from the application of economic theories to examine and evaluate 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.

1.2.1 Basic Economic Problem: Choice and Scarcity

The need for making a choice arises because of some *basic facts of economic life.* Let us look at the basic facts of human life in some detail and how they create the problem of choice-making.

1. ***Human Wants, Desires and Aspirations are Limitless:*** The history of human civilization bears evidence to the fact that human desire to consume more and more of better and better goods and services has ever since been increasing. For example, housing need has risen from a hut to luxury palace, and if possible, a house in space; the need for means of transportation has gone up from mules and camels to supersonic jet planes; demand for means of communication has risen from messengers and postal services to cell phones with features of a computer; need for computational facility from manual calculation to superfast computers; and so on. For an individual, only the end of life brings an end to his/her needs. But for homo sapiens, needs and desires continue to grow endlessly.

   Human wants, desires and *needs* are endless in the sense that they go on increasing with increase in people's ability to satisfy them. The *endlessness of human wants* can be attributed to (i) people's insatiable desire to raise their standard of living, comforts and efficiency; (ii) human tendency to accumulate things beyond their present need; (iii) increase in knowledge about inventions and innovations of new goods and services with greater convenience, efficiency and serviceability; (iv) multiplicative nature of some wants (e.g., buying a car creates want for many other things—petrol, driver, cleaning, parking place, safety locks, spare parts, insurance, etc.); (v) biological needs (e.g., food, water, etc.) are repetitive; (vi) imitative and competitive nature of human beings creating needs due to *demonstration* and *bandwagon effects*; and (vii) influence of advertisements in modern times creating new kinds of wants. For these reasons, human wants continue to increase endlessly.

NOTES

Apart from being *unlimited,* another and an equally important feature of human wants is that *they are gradable.* In simple words, all human wants are not equally urgent and pressing at a point of time, or over a period of time. While some wants have to be satisfied as and when they arise (e.g., food, clothes and shelter) some others can be postponed, e.g., purchase of a car. Also, while satisfying some others gives a greater satisfaction than others. Given their intensity and urgency, human wants can be arranged in the order of their priority. The priority of wants, however, varies from person to person, and from time to time for the same person. Therefore the question arises as to 'which want to satisfy first' and 'which the last'. Thus, consumers have to make the choice: 'what to consume' and 'how much to consume'. Economics studies how consumers (individuals and household) make the choice between their wants and how they allocate their expenditure between different kinds of goods and services they choose to consume.

2. ***Resources are Scarce:*** The need for making a choice between the various goods that people want to produce and consume arises mainly because *resources that are available to the people at any point of time for satisfying their wants are scarce and limited.* What are the resources? Conceptually, anything which is available and can be used to satisfy human wants and desire is a *resource.* In economics, however, *resources* that are available to individuals, households, firms, and societies at any point of time are traditionally classified as follows.
   (*i*) *natural resources* (including cultivable land surface, space, lakes, rivers, coastal range, minerals, wildlife, forest, climate, rainfall, etc.);
   (*ii*) *human resources* (including manpower, human energy, talent, professional skill, innovative ability and organizational skill, jointly called *labour*);
   (*iii*) *man-made resources* (including machinery, equipment, tools, technology and building, together called *capital*); and
   (*iv*) *Entrepreneurship (*i.e., the ability, knowledge and talent to put *land, labour* and *capital* in the process of production, and ability and willingness to assume *risk* in business).

   To these basic resources, economists add other categories of resources, viz., *time, technology* and *information.* All these *resources are scarce.* Resource scarcity is a relative term. It implies that resources are scarce in relation to the demand for resources. The *scarcity of resources is the mother of all economic problems.* If resources were unlimited, like human wants, there would be no economic problem and, perhaps, no economics as a subject of study. *It is the scarcity of resources in relation to human wants that forces people to make choices*.

   Furthermore, the problem of making choice arises also because resources have *alternative uses* and alternative uses have *different returns or earnings.* For example, a building can be used to set up a shopping centre, business office, a 'public school', a hospital or for residential purpose. But the return on a building varies from use to use of the building. Therefore, a

NOTES

return maximizing building owner has to make a choice between the alternative uses of the building. If the building is put to a particular use, the landlord has to forego the return expected from its other alternative uses. This is called **opportunity cost**. Economics as a social science analyses how people (individuals and society) make their choices between the economic goals they want to achieve, between the goods and services they want to produce, and between the alternative uses of their resources with the objective of maximizing their gains. The gain maximizers evaluate the costs and benefits of the alternatives while deciding on the final use of resources. Economics studies the process of making choices between the alternative uses. This is what constitutes, according to Robbins, the subject matter of economics.

3. ***People Are Gain Maximizers:*** Yet another important aspect of human nature that leads to the choice-making behaviour is that most people aim at maximizing their gains from the use of their limited resources. 'Why people want to maximize their gains' is no concern of economics. Traditional economics assumes the maximizing behaviour of people as a part of their rational economic behaviour. This assumption is based on observed facts. As *consumers,* they want to maximize their utility or satisfaction; as *producers,* they want to maximize their output or profit; and as *factor owners,* they want to maximize their earnings. *People's desire to maximize their gains* is a very important aspect of economic behaviour of the people giving rise to economics. If people were not to maximize their gains, the problem of choice making would not arise. Consumers would not bother as to 'what to consume' and 'how much to consume'; producers would not bother as to 'what to produce', 'how much to produce' and 'how to produce'; and factor owners would not care as to where and how to use the resources. But, in reality, they do maximize their gains. Economics studies how people maximize their gains.

Check Your Progress

1. How is economics defined by Adam Smith?
2. Define economics in your own words.

1.3 METHODS OF ECONOMIC ANALYSIS

Economic theories which constitute the body of economic science today are the result of scientific investigation into economic facts. The scientific search for economic truths consists of a systematic and logical procedure of arranging and analysing economic facts and establishing the relationship between the facts.

The two kinds of methods which have been adopted at different stages of growth of economic science by different schools of thought are : (*a*) *Deductive Method :* and (*b*) *Inductive Method*. Deduction and induction are, in fact, two different forms of logic which are used to draw inferences. In *the deductive method,*

reasoning proceeds from general to particular or from universal to individuals. In this method, inferences are drawn from general cases to establish the particular case. On the other hand, in the *inductive method*, reasoning proceeds from particular to the general or from individual to the universal, and a general case is made from the individual cases.

Let us now discuss the two methods in detail.

1.3.1 Deductive Method

The deductive method is also known as the *analytical method*. In the deductive or analytical method, initially certain assumptions or postulates are made. On the basis of these assumptions, certain logical conclusions are drawn which become the testable hypotheses. The hypotheses are then tested against observed facts. The hypotheses confirmed by the facts are accepted as tentative theories. If a theory so formulated stands the tests, time and again, it becomes a law, e.g., law of demand, law of diminishing marginal utility, etc.

The deductive approach proeeds by the following major steps : (*a*) selecting the problem for analysis; (*b*) specifying the assumptions or postulates; (*c*) formulating hypotheses on the basis of assumptions; and (*d*) testing the validity of the hypotheses.

The *first* step in any scientific analysis is *to specify the problem of the study*. The problem chosen for the study generally is, and should be, of practical importance to the society. This, however, is not necessary. Economists may, and in fact they do, select a problem of their own interest which may not serve any immediate social purpose. In fact, most early scientific discoveries have been the result of a scientist's own curiosity rather than the result of research undertaken to serve certain predetermined social ends.

The *second* step is to specify the *assumptions*. Assumptions serve several purposes in scientific analysis. They determine the scope and dimensions of the study and also specify the factors to be taken as constant. Assumptions are essentially used as the premise on which hypotheses are built.

Despite its merits, the deductive method has certain *disadvantages. First,* although it is claimed to be a simple method, it is a highly complicated method as it requires great skill and logical acumen to derive conclusions from the postulates. *Second,* since this method often leads to a high degree of abstraction, it involves the risk of yielding theories far from the reality. *Third*, the theories formulated are of limited applicability. That is, they are applicable within the framework of the assumptions which are often unreal. *Finally,* the deductive approach very often turns to be a mere *intellectual exercise* yielding results of little practical use.

1.3.2 Inductive Method

The inductive approach to formulating economic principles is the reverse of the deductive approach. While the deductive method is a descending process which proceeds from general to particular, the inductive method is an ascending process which proceeds from particular to general. Inductive analysis begins with observed facts regarding the recurrence of an economic event or existence of an economic

NOTES

phenomonon and its causes. It then establishes the cause–and–effect relationship between the events, making a general case. The general case is then used to explain individual economic events. For example, people have observed over centuries that when crops are damaged by flood, drought or inclement weather, agricultural production falls and agricultural prices go up. This makes a general case of price behaviour in response to change in supply. This general case may be applied to explain or predict the price behaviour in case of a particular crop, i.e., how price will behave given the supply position.

The inductive method involves the following steps:

The *first* step in formulating inductive economic laws is the same as in the case of the deductive approach, i.e., the selection of the problem for analysis. It may be any economic problem, such as returns to increasing inputs, unemployment, inflation, industrial unrest, etc.

The *second* step is collection, classification and analysis of data by using appropriate statistical techniques in order to find out the relationship between the variables.

The *final* step is to find out the reasons for the relationship established through statistical analysis and to set the rules for the veritication of the principle.

Like the deductive approach, the inductive approach too has its own *merits* and *demerits*. As regards its *merits*, since this approach analyses economic phenomena on the basis of observed facts, it has been claimed to be more close to reality. Besides, inductive method is considered as the most important way of testing or verifying an established economic theory. Further, economic theories based on inductive approach can be a better tool of predicting future course of economic events.

Demerits of the inductive method lie in the problems of data and the statistical tools of analysis which are frequently used in this method. Collecting appropriate, requisite data on a particular economic problem is in itself a difficult task, particularly where conceptual problems are involved. Data–related problems arise mainly because experiments in an economic phenomenon is not possible in the same manner as in natural sciences. Besides, different investigators may arrive at different conclusions, from the same data, if their assumptions differ. This makes the conclusions doubtful. Further, in the inductive approach, 'there can be no absolute assurance that the result of the generalisation will actually be attained in a particular case.'

The inductive generalizations, therefore, turn to be merely statements of tendencies, not even testable hypotheses.

Check Your Progress

3. Name the two major methods of conducting economic analysis.
4. What are the merits of using the inductive approach of conducting economic analysis?

1.4 BASIC CONCEPTS

NOTES

Let us now discuss some of the basic concepts in microeconomics. These are discussed below.

1.4.1 Microeconomics

Microeconomics is the study of how individuals, households and firms find solutions to the problem of maximizing their gains from their limited resources. Microeconomics is essentially the study of *economic behaviour*, i.e., choice-making behaviour, of people. What is economic behaviour? *Economic behaviour* is essentially the process of evaluating the economic opportunities open to an individual or a society and, given the resources, making a choice of the best of the opportunities. The objective behind this economic behaviour is to maximize gains from the available resources and opportunities. In their efforts to maximize their gains from their resources, people have to make a number of choices regarding the use of their resources and spending their earnings. The basic function of economics is to observe, explain and predict how people (individuals, households, firms and governments) as decision-makers make choices about the use of their resources (land, labour, capital, knowledge and skills, technology, time and space, etc.) to maximize their income, and how they as consumers decide how to spend the income to maximize their total utility. Thus, economics is fundamentally the study of choice-making behaviour of people. Studying it in a systematic or scientific manner gives economics the status of a social science.

For the purpose of economic analysis, people are classified according to their decision-making capacity as *individuals, households, firms* and the *society,* and according to the nature of their economic activity as *consumers, producers, factory owners* and *economy managers,* i.e., *the government.* As *consumers,* individuals and households, with their given income, have to decide 'what to consume and how much to consume'. They have to make these decisions because consumers are, by nature, utility maximizers and consuming any commodity in any quantity does not maximize their gains, their satisfaction. As *producers,* firms, farms, factories, shopkeepers, banks, transporters, etc., have to choose 'what to produce, how much to produce and how to produce' because they too are gain maximizers and producing any commodity in any quantity by any technique will not maximize their gains (profits). As *labour*, they have to choose between alternative occupations and places of work because any occupation at any place will not maximize their earnings. Likewise, the *government* has to choose how to tax, whom to tax, how much to spend and how to spend so that social welfare is maximized at a given social cost. Economics as a social science studies how *people* make their choices.

It is this economic behaviour of the individuals, households, firms, government and the society as a whole, which forms the *central theme* of economics as a social science. Thus, *economics is fundamentally the study of how people allocate their limited resources to produce and consume goods and services to satisfy their endless wants with the objective of maximizing their gains.*

NOTES

1.4.2 Utility

The notion of "Utility" was introduced to social thought by the British philosopher, Jeremy Bentham, in the 18th century and to economics by William Stanley Jevons in the 19th century. In its economic meaning, the term "utility" is synonymous with "pleasure", "satisfaction" and a sense of fulfilment by desire. A person consumes a commodity because he or she derives pleasure out it. In other words, he derives utility from the consumption of the goods and services.

In abstract sense, the term *"utility" refers to the power or property of a commodity to satisfy human needs.* For example, bread has the power to satisfy hunger; water quenches our thirst; books fulfill our desire for knowledge; and postal stamps take our letters to their destination, and so on. All the goods that people hold or consume possess utility. Utility can also be defined as the "want-satisfying power" of a commodity. But it is not absolute—it is relative. It is relative to a person's need. In other words, whether a commodity possesses utility depends on whether a person needs that commodity. All the persons need not derive utility from all the commodities. For example, non-smokers do not derive any utility from cigarettes; strict vegetarians do not derive any utility from meat and chicken; a book on economics has no utility for those who are not student of economics, and so on. The utility derived by a person from a commodity depends on his or her intensity of desire for that commodity: *the greater the need, the greater the utility.*

Besides, utility of some commodities depends on the availability of complementary goods. For example, electricity operated gadgets (e.g., TV, VCR, computers, refrigerators, etc.) yield utility only where electricity is available and petrol has utility only for those who possess an automobile.

Furthermore, the concept of utility is **"ethically neutral"**. It is neutral between good and bad and between useful and harmful. For example, some drugs are bad and harmful, for every body but they yield utility to the drug-addicts. Utility is free from moral values. It is not subject to social desirability of consuming a commodity. Eating beef may be immoral or socially undesirable for Hindus, but if a Hindu takes it, it satisfies his hunger.

Measurability of Utility

Measurability of utility has been and remains a debatable issue. Essentially, utility is a psychological phenomenon—it is a feeling of pleasure or a feeling of satisfaction and achievement. Can utility be measured in **absolute terms**?

The classical and neo-classical economists held the view that **utility is quantitatively** or **cardinally measurable.** It can be measured like height, weight, length and temperature. Their method of measuring utility can be described as follows:

(i) Walras, a classical economist, used the term 'util' meaning 'units of utility'. The term was used as an accounting unit like kilogram, meter, etc.

(ii) The classical economists used 'util' as the measure of utility under the assumption that one unit of money equals one 'util'. It implies that price that

a consumer pays for a commodity equals the utility derived from the commodity.

(iii) They assumed that marginal utility of money remains constant, i.e., the utility one derives from each successive unit of money income remains constant whatever the stock of money one holds.

This method of measuring utility has been rejected by the modern economists. For, it was realised over time that absolute or cardinal measurement of utility is not possible. The difficulties in measuring utility proved insurmountable. Money was not found to be a reliable measure of utility because the utility of money itself changes with change in its stock. Neither economists nor psychologists nor other scientists could devise a reliable technique or instrument for measuring the feeling of satisfaction or utility. The modern economists have therefore discarded the concept of cardinal utility.

Notwithstanding the problems in quantitative measurement of utility, the consumption theory based on cardinal utility concept provides deep insight into the consumer psychology and consumer behaviour and remains an indispensable element of consumption theory. In fact, it serves as a starting point in the study of further advances in the theory of consumer behaviour.

1.4.3 Demand and Supply

The term 'demand' refers to the *quantity demanded* of a commodity *per unit of time* at a given price. It implies also a *desire* backed by ability and willingness to pay. A mere desire of a person to purchase a commodity is not his demand. He must possess adequate resources and must be willing to spend his resources to buy the commodity. Besides, the *quantity demanded* has always a reference to 'a price' and 'a unity of time'. The quantity demanded referred to 'per unit of time' makes it a *flow* concept. Apparently there may be some problems in applying this flow concept to the demand for durable consumer goods like house, car, refrigerators, etc. But this apparent difficulty may be resolved by considering the fact that the total service of a durable good is not consumed at one point of time and its utility is not exhausted in a single use. The service of a durable good is consumed over time. At a time, only a part of its service is consumed. Therefore, the demand for the services of durable consumer goods may also be visualised as a demand per unit of time. However, this problem does not arise when the concept of demand is applied to total demand for a consumer durable. Thus, the demand for consumer goods also is a flow concept.

The Law of Supply

Market supply means the quantity of a commodity which all its producers or sellers offer to sell at a given price, per unit of time. Market supply, like market demand, is the sum of supplies of a commodity made by all individual firms.

The law of supply can be stated as the supply of a product increases with the increase in its price and decreases with decrease in its price, other things

NOTES

remaining constant. It implies that the supply of a commodity and its price are positively related. This relationship holds under the assumption that "other things remaining the same". "Other things" include cost of production, change in technology, price of related goods (substitutes and complements), and weather and climate in case of agricultural products.

The law of supply has been discussed in detail in unit 2.

1.4.4 Commodity and Free Goods

We can define a **commodity** as a tangible good that can be bought and sold or exchanged for products of similar value. Examples of commodity include both natural substances such as oil, copper, gold, as well as basic foods such as corn. A commodity basically has two properties. These are:

- Firstly, it is a good that is sold by numerous companies or firms.
- Second, it is uniform in quality between companies that produce and sell it. One cannot tell the difference between one firm's goods and another. To put it another way, a commodity is fungible.

A **free good** is a good that is not scarce, and therefore is available without limit. This means it is a good with zero opportunity cost and can be consumed as much quantity as required without decreasing its availability to others. Some examples of free good are air, sunlight, water, and so on.

1.4.5 Value and Price

Price is basically the monetary value of a good, service or resource established during a transaction. In a monopoly, the price is set by the sellers. In a monopsony, prices are set by the consumer. In a competitive market, the price is set through the market itself. Value refers to the utility of a commodity. However, in economics, the term 'value' has a quite different meaning. Economic value refers to measure of the benefit provided by a good or service to an economic agent. It should be stated that economic value is not the same as market price, nor is economic value the same thing as market value. If a customer is willing to buy a good, it implies that he or she places a higher value on the good than the market price. The difference between the value to the consumer and the market price is called 'consumer surplus.'

1.4.6 Market Administered Price

When prices of goods or services are set by the internal pricing structures of organizations that take into account cost rather than through the market forces of supply and demand, it is known as administered pricing. This type of pricing is common in industries with few competitors and those in which costs tend to be rigid and more or less uniform. They are considered undesirable when they cause prices to be higher than a competitive standard, when they are accompanied by excessive non-price competition, or when they add to inflation.

Check Your Progress

5. What were classical and neoclassical economists, views on measuring utility?
6. Mention the law of supply.

NOTES

1.5 ANSWERS TO 'CHECK YOUR PROGRESS'

1. Adam Smith, sometimes called the father of economics, defined economics as ' a subject concerned with an enquiry into the nature and causes of wealth of nations.'
2. Economics is fundamentally the study of how people allocate their limited resources to produce and consume goods and services to satisfy their endless wants with the objective of maximizing their gains.
3. The two major methods of conducting economic analysis are the Inductive method and Deductive method.
4. The inductive approach has its merits and demerits. As regards its merits, since this approach analyses economic phenomena on the basis of observed facts, it has been claimed to be more close to reality. Besides, inductive method is considered as the most important way of testing or verifying an established economic theory. Further, economic theories based on inductive approach can be a better tool of predicting future course of economic events.
5. The classical and neo-classical economists held the view that utility is quantitatively or cardinally measurable.
6. The law of supply can be stated as the supply of a product increases with the increase in tis price and decreases with decrease in its price, other things remaining constant. It implies that the supply of a commodity and its price are positively related.

1.6 SUMMARY

- Economics was in its embryonic stage till the middle of the 18th century. It was then a hybrid of politics, ethics, logic and philosophy.
- Attempts to distinguish economics as a separate branch of Social Studies began with Adam Smith's famous treatise, *An Enquiry into the Nature and Causes of Wealth of Nations* (1776), popularly known as Wealth of Nations.
- In Marshall's own words, 'Political Economy or Economics is a study of mankind in the ordinary business of life; it examines that part of individual and social action which is most closely connected with the attainment and with the use of the material requisites of well-being,'

NOTES

- Robbins examined the validity of neo-classical definitions in the light of theories of wages and exchange as developed by them, and pointed out the following deficiencies in 'materialist definitions'.
- Economics is a science which studies human behaviour as a relationship between ends and scarce means which have alternative use.
- Robbins' definition is superior to classical and neo-classical definitions in many respects. It has certain merits which other definitions do not possess.
- Microeconomics is concerned with microscopic study of the various elements of the economic system and not with the system as a whole.
- Macroeconomics is a relatively new branch of economics. It was only after the publication of Keynes's The General Theory of Employment, Interest and Money in 1936, that macroeconomics crystallized as a separate branch of economics.
- Microeconomics is the study of how individuals, households and firms find solutions to the problem of maximizing their gains from their limited resources.
- As Marshall defined it, economic laws are statements of economic tendencies. They bring out the relationships between the economic variables. The existence of such relationships is subject to the conditions specified by underlying assumptions.
- Economic laws are not as exact and precise as the laws of natural sciences. Further, economic laws do not operate as steadily as do the laws of natural sciences. For example, any object thrown in the air has always a tendency to fall to the ground.
- Economic theories which constitute the body of economic science today are the result of scientific investigation into economic facts. The scientific search for economic truths consists of a systematic and logical procedure of arranging and analysing economic facts and establishing the relationship between the facts.
- The deductive method is also known as the analytical method. In the deductive or analytical method, initially certain assumptions or postulates are made. On the basis of these assumptions, certain logical conclusions are drawn which become the testable hypotheses.
- The inductive approach to formulating economic principles is the reverse of the deductive approach. While the deductive method is a descending process which proceeds from general to particular, the inductive method is an ascending process which proceeds from particular to general. Inductive analysis begins with observed facts regarding the recurrence of an economic event or existence of an economic phenomenon and its causes.
- Regional Economics studies the development of various regions of a country; it looks into the causes of imbalance in regional development, it examines why growth of urban economy is faster that of rural economy.

NOTES

- In abstract sense, the term "utility" refers to the power or property of a commodity to satisfy human needs. For example, bread has the power to satisfy hunger; water quenches our thirst; books fulfill our desire for knowledge; and postal stamps take our letters to their destination, and so on.
- The classical and neo-classical economists held the view that utility is quantitatively or cardinally measurable. It can be measured like height, weight, length and temperature.
- The term 'demand' refers to the quantity demanded of a commodity per unit of time at a given price. It implies also a desire backed by ability and willingness to pay.
- Market supply means the quantity of a commodity which all its producers or sellers offer to sell at a given price, per unit of time. Market supply, like market demand, is the sum of supplies of a commodity made by all individual firms.
- We can define a commodity as a tangible good that can be bought and sold or exchanged for products of similar value.
- When prices of goods or services are set by the internal pricing structures of organizations that take into account cost rather than through the market forces of supply and demand, it is known as administered pricing.

1.7 KEY TERMS

- **Microeconomics:** It is the study of how individuals, households and firms find solutions to the problem of maximizing their gains from their limited resources.
- **Industrial economics:** It is concerned with the working, growth and structure of the industrial sector (firms and industries) of the country, management and organization of industries, and problems and prospects of industrial growth.
- **Utility:** It refers to the power or property of a commodity to satisfy human needs.
- **Opportunity cost:** It implies the value of the next highly-valued alternative use of that resource.
- **Demand:** It refers to the quantity demanded of a commodity per unit of time at a given price.

1.8 SELF-ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. What is the nature and scope of economics as stated by the classical economists?

2. Write a short note on the nature of economics.
3. State the differences between microeconomics and macroeconomics.
4. Write a short note on the scope of economics.

NOTES

Long-Answer Questions

1. Examine the criticisms raised against the neo-classical definitions of economics by Robbins.
2. Elaborate the basic needs of human life which lead to choice and scarcity.
3. Discuss the inductive and deductive methods of conducting economic analysis.
4. Explain market administered price with the help of an example.

1.9 FURTHER READING

Dwivedi, D. N. 2016. *Microeconomics: Theory and Application*, 3rd Edition. New Delhi: Vikas Publishing House.

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

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

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

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

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

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

UNIT 2 DEMAND AND SUPPLY

NOTES

Structure

2.0 INTRODUCTION

Market mechanism plays a crucial role in solving the basic economic problems of a free market economy and that the entire market system functions in an orderly manner, though some aspects of it may not be desirable. The market system functions in an orderly manner because it works under certain fundamental laws of market known as the laws of demand and supply.

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*.

Supply side of the market refers to the sellers of a product in the market supply is the quantity supplied at a given price per unit of time.

In this unit, you will study about law of demand and its exceptions, Giffen goods, elasticity of demand, law of supply, law of diminishing marginal utility, and equi-marginal utility, consumer's surplus and characteristics of indifference curve.

NOTES

2.1 OBJECTIVES

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

- Explain the law of demand and its exceptions
- Describe elasticity of demand
- Define law of supply
- State law of diminishing marginal utility and equi-marginal utility
- List the characteristics of indifference curve

2.2 LAW OF DEMAND AND ITS EXCEPTIONS

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

1. **Individual demand:** Individual demand refers to 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.
2. **Market demand:** 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.

Individual and market demand curves are discussed ahead in detail. Let us now discuss the law of demand. Since we are concerned in this unit with market demand, the *law of demand* will be discussed in the context of market demand.

The Law of Demand

The law of demand states the relationship between the quantity demanded and the price of a commodity. In general, quantity demanded of a commodity depends on many other factors also, viz., consumer's income, price of the related goods (substitutes and complements), consumer's taste and preferences, advertisement, etc. However, price of a product is the most important and the only determinant of its demand in the *short run* because other factors are taken to remain constant. Therefore, the law of demand is linked to the price of the product.

The law of demand can be stated as 'all other things remaining constant, the quantity demanded of a commodity increases when its price decreases and demand decreases when its price increases'. This law implies that *demand and price are inversely related*. Marshall states the law of demand as 'the amount demanded increases with a fall in price and diminishes with a rise in price'. This law holds under *ceteris paribus* assumption, i.e., *all other determinants of demand remain unchanged*. The law of demand can be illustrated through a *demand schedule* and a *demand curve*.

NOTES

2.2.1 The Demand Schedule

A *demand schedule is* a tabular presentation of quantity demanded of a commodity at different prices per unit of time. A hypothetical market demand schedule is given in Table 2.1. This table presents price of shirts (P_s) and the corresponding number of shirts demanded (Q_s) per month.

Table 2.1 *Demand Schedule for Shirts*

P_s (Price in ₹)	Q_s (Shirts in '000)
800	8
600	15
400	30
300	40
200	55
100	80

Table 2.1 illustrates the law of demand. As data given in the table shows, the demand for shirts (Q_s) increases as its price (P_s) decreases. For instance, at price ₹800 per shirt, only 10,000 shirts are demanded per month. When price decreases to ₹400, the demand for shirts increases to 30,000 and when price falls further to ₹100, demand rises to 80,000. Similarly, one can read the table in reverse order and arrive at the conclusion that as price of shirt increases, its demand decreases. This relationship between the price and the quantity demanded gives the law of demand.

2.2.2 The Demand Curve

A **demand curve** is a graphical presentation of the demand schedule. For example, when the data given in the demand schedule (Table 2.1) are presented graphically as shown in Figure 2.1, the resulting curve *DD'* represents the *demand curve*. The curve *DD'* in Figure 2.1 depicts the law of demand. It slopes downward to the right. That is, it has a negative slope.

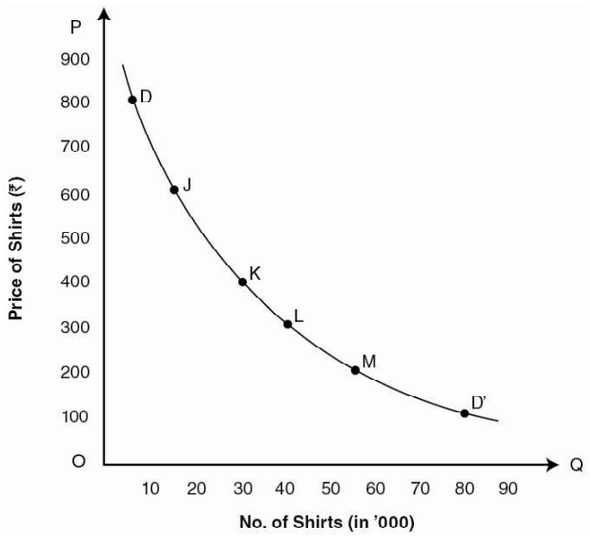

Fig. 2.1 *The Demand Curve*

NOTES

The negative slope of the demand curve *DD′* shows the *inverse relationship* between the price of shirt and its quantity demanded. The inverse relationship means that demand increases with the decrease in price and decreases with the rise in price. As can be seen in Figure 2.1, downward movement on the demand curve *DD′* from point *D* towards *D′* shows fall in price and rise in demand. Similarly, an upward movement from point *D′* towards *D* reads rise in price and fall in demand.

The law of demand is based on an empirical fact, i.e., based on real market data. For example, when prices of cell phones and personal computers (PCs), especially of the latter, were astronomically high, only few rich persons and big firms could afford them. Now with the revolution in computer and cell phone technology and the consequent fall in their prices, demand for these goods has shot up in India though other factors too contributed to rise in demand for these goods.

The Factors Behind the Law of Demand

According to the law of demand, when a price of a product increases, its demand decreases and *vice versa*, all other demand determinants remaining constant. A question arises here: what are the factors behind the law of demand or why does demand decrease when price rises or other way round? The factors *behind the law of demand* are the following.

1. **Income effect:** When price of a commodity falls, purchasing power of the consumers increases since they are required to pay less for the same quantity. According to another economic law, increase in real income (or purchasing power) increases demand for the goods and services in general and for the goods with reduced price in particular. The increase in demand on account of increase in real income is called *income effect*.

   It should, however, be noted that the *income effect* is negative in case of *inferior goods*. In case, price of an inferior good accounting for a considerable proportion of the total consumer expenditure falls substantially, consumers' real income increases. Consequently, they substitute superior goods for inferior ones. Therefore, *income effect* on the demand for the inferior good becomes *negative*.

2. **Substitution effect:** When price of a commodity falls, it becomes cheaper compared to its substitutes, prices of substitutes remaining constant. In other words, when price of a commodity falls, price of its substitutes remaining the same, its substitutes become relatively costlier. Consequently, rational consumers tend to substitute cheaper goods for costlier ones within the range of normal goods—goods whose demand increases with the increase in consumer's income—other things remaining the same. Therefore, demand for the relatively cheaper goods increases. The increase in demand on account of substitution of cheaper good for the relatively costlier one is known as *substitution effect*.

3. **Diminishing marginal utility:** Marginal utility is the utility derived from the marginal unit consumed of a commodity. Diminishing marginal utility is also responsible for the increase in demand for a commodity when its price falls. When a person buys a commodity, he exchanges his money income with the commodity in order to maximize his satisfaction. The utility of money is assumed to remain constant and utility of money is equal to price paid. Under this condition, a consumer continues to buy goods and services so long as marginal utility of his money (MU_m) is less than the marginal utility of the commodity (MU_c). Given the price (P_c) of a commodity, the consumer adjusts his purchases so that $MU_c = MU_m$. When the price of commodity decreases, the utility of money hence to be higher than the product. A utility maximizing consumer spends more money to attain equilibrium at $MU_c = MU_c$. This fact increases the demand with decrease in the price

2.2.3 Exceptions to the Law of Demand (Giffen Goods)

The law of demand is one of the fundamental laws of economics. The law of demand, however, does not apply under 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 just 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 of demand does not apply to the commodities which are used as a 'status symbol'. Prestige goods are the goods 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 ₹5/kg and wheat at ₹10/kg. At these prices, the household spends ₹200 per month on food grains. That is the maximum it can afford. Now, if price of *bajra* increases to ₹6 per kg, the household will be forced to reduce its consumption of wheat by 5 kg and increase that of *bajra* by the same quantity in order to meet its minimum monthly consumption

requirement within ₹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.

NOTES

Check Your Progress

1. Define the term 'demand schedule'.
2. What is a demand curve?

2.3 ELASTICITY OF DEMAND, PRICE, INCOME AND CROSS

Before we proceed to discuss the elasticity of demand, let us have a clear view of the components of demand elasticities. The demand for a product, especially in the long run, depends on several factors as listed below.

1. Price of the product,
2. Consumers' income;
3. Price of substitutes and complements;
4. Advertisement of the product;
5. Future price expectations; and
6. Consumers' taste and fashion.

Of these demand determinants, the effect of change in 'consumers' taste and fashion' is difficult to quantify. In practice, therefore, the overall demand for a product is generally deemed to be determined by the quantifiable demand determinants, viz., *price of the product, consumers' income, price of the substitutes and compliments, ad-spending by the firms and consumers' expectations about the future prices*. Therefore, the overall demand and change in demand for a product depends on the nature and extent of change in these demand determinants. And, the overall elasticity of demand for a commodity depends on the combined effects of changes in these demand determinants. Therefore, the elasticity of demand is measured separately with respect to all its major determinants. Following this practice, we discuss in this unit, the following kinds of demand elasticities.

1. Price elasticity of demand;
2. Income elasticity of demand; and
3. Cross-elasticity of demand, i.e., demand elasticity with reference to price of substitutes and complementary goods.

All these kinds of demand elasticities are discussed in this chapter. However, of these kinds of elasticities of demand, *price elasticity of demand* is of the greatest significance from both theoretical and practical points of view. Therefore, the price elasticity of demand will be discussed here in a greater detail.

NOTES

The Price Elasticity of Demand

The price elasticity of demand is defined as the degree of responsiveness 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. That is,

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

Here, e_p denotes the price elasticity of demand. The numerical value of e_p is called the coefficient of demand elasticity.

A general formula for measuring the price elasticity of demand is derived as follows:

$$e_p = \frac{\frac{Q_2 - Q_1}{Q_1} \times 100}{\frac{P_2 - P_1}{P_1} \times 100} = \frac{\frac{Q_2 - Q_1}{Q_1}}{\frac{P_2 - P_1}{P_1}}$$

Here, Q_1 = original demand, Q_2 = demand after price change, P_1 = original price and P_2 = changed price. By denoting $Q_2 - Q_1$ as Q and $P_2 - P_1$ as P, a general formula for measuring price elasticity coefficient is expressed as follows.

$$e_p = \frac{\Delta Q}{Q_1} \div \frac{\Delta P}{P_1}$$

or

$$e_p = \frac{\Delta Q}{\Delta P} \times \frac{P_1}{Q_1} \qquad (2.1)$$

To measure price elasticity of demand numerically by using the formula given in Eq. (2.1), let us suppose that price of a commodity X and its demand at two different prices are given in Table 2.2. Given the price-quantity data, ΔP and ΔQ can be worked out as

$$\Delta P = ₹8 - ₹10 = -₹2$$

and

$$\Delta Q = 60 - 50 = 10.$$

Table 2.2 *Price of Commodity X and Quantity Demanded*

Price of X (₹)	Quantity Demanded
10	50
8	60

By substituting these values in elasticity formula as given in Eq. (2.1), we get:

$$e_p = -\frac{10}{-2} \times \frac{10}{50} = 1.0$$

Thus, elasticity coefficient (e_p) equals 1.

Note that a minus sign (–) is inserted in the formula for measuring elasticity with a view to making *elasticity coefficient* a *non-negative value*. The coefficient of price elasticity calculated without minus sign in the formula will always be negative,

NOTES

because either P or Q will carry a negative sign depending on whether price increases or decreases. But a negative coefficient of elasticity is rather misleading because elasticity cannot be negative—less than zero. The 'minus' sign is, therefore, inserted in the price elasticity formula as a matter of 'linguistic convenience' to make the coefficient of elasticity a non-negative value. Sometimes, it is also advised to ignore the negative sign of P or Q. The price elasticity measure is, however, always reported with a negative sign just to indicate inverse relationship between the price change and the quantity demanded.

The Arc and Point Elasticity

When price elasticity of demand is measured between any two finite points on a demand curve, it is called **arc elasticity** and elasticity measured at a point on the demand curve is called **point elasticity**. As noted above, price elasticity of demand is measured as the percentage change in the quantity demanded due to a certain percentage change in price. The percentage change in price may be considerably high (e.g., 20 per cent or even higher) or it may be very small—so small that it is not significantly different from zero. When change in price is significantly high, it shows a movement from one point on the demand curve to another point, making an *arc*. Therefore, the price elasticity measured for a considerably high change in price is called *arc elasticity of demand*. And, when price elasticity is measured for very small changes in price—not significantly different from zero—it is called *point elasticity*.

Method of Measuring Arc Elasticity

As noted above, the arc elasticity of demand is the measure of elasticity between any two finite points on the demand curve. Suppose a demand function for commodity X is given as

$$Q_x = 80 - 2P_x$$

The price quantity data generated by this demand function presented graphically, produces the demand curve, PM, as shown in Figure 2.2. The measure of elasticity between any two points on the demand curve, PM, gives the arc elasticity. For example, the measure of the price elasticity of demand between points J and K on the demand curve PM in Figure 2.2 is the measure of *arc elasticity*. The movement from point J to K on the demand curve PM shows a fall in price of commodity X from ₹25 to ₹15 and the consequent increase in demand from 30 to 50 units. Here, $P = 15 - 25 = -10$ and $Q = 50 - 30 = 20$. The arc elasticity between points J and K (moving from J to K) can be measured by using the formula given in Eq. (2.1), as given below.

$$e_p = \frac{\Delta Q}{\Delta P} \times \frac{P_0}{Q_0}$$

where P_0 and Q_0 are the original price and the original quantity demanded, respectively.

NOTES

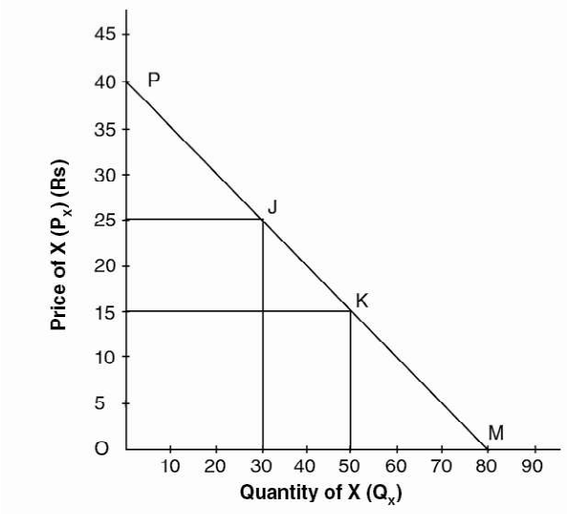

Fig. 2.2 *Change in Price and Arc Elasticity Coefficient*

$$e_p = -\frac{20}{-10} \times \frac{25}{30} = 1.66 \tag{2.2}$$

Point Elasticity *is the measure of price elasticity at a finite point on a demand curve*. However, as 'point' is defined in geometry, it occupies no space and has no dimensions. It implies that there is no change in the price and hence no change in the quantity demanded. Therefore, the concept of 'point elasticity' may not appear to be reasonable. However, from practical point of view, point elasticity concept is applied to an insignificant change in the price and the consequent change in the quantity demanded. Point elasticity is, in fact, the measure of the proportionate change in the quantity demanded in response to a *very* small proportionate change in the price. The concept of point elasticity is useful where change in the price and the consequent change in the quantity demanded are infinitesimally small. Besides, it offers an alternative to the arc elasticity. Point elasticity may be symbolically expressed as

$$e_p = \frac{\partial Q}{\partial P} \cdot \frac{P}{Q} \tag{2.3}$$

Determinants of Price Elasticity of Demand

The price elasticity of demand varies from commodity to commodity depending on the nature of the commodity. While the demand for some commodities is highly elastic, for some it is highly inelastic. Besides, given the nature of a commodity, there are several other factors which determine the price elasticity of demand for a commodity. The effect of the main determinants of the price elasticities of demand is described in this section.

1. **Availability of substitutes***:* One of the most important determinants of price elasticity of demand for a commodity is the availability of its substitutes. *The closer the substitute, the greater the price elasticity of demand for a commodity*. For instance, coffee and tea may be considered as close

NOTES

substitutes for one another. If price of one of these goods (say, coffee) increases, then the demand for coffee decreases more heavily. The reason is that the other commodity (tea) becomes *relatively* cheaper. Therefore, consumers buy more of the relatively cheaper good (tea) and less of the costlier one. Besides, the wider the range of the substitutes, the greater the elasticity. For instance, soaps, toothpastes, cigarettes, etc. are available in different brand names, each brand being a close substitute for the other, all other things remaining the same. Therefore, the price elasticity of demand for each brand will be much greater than the generic commodity. On the other hand, sugar and salt do not have their close substitute and hence their price elasticity is lower.

2. **Nature of commodity***:* Price elasticity of demand depends also on the nature of a commodity. Commodities can be grouped broadly as luxuries, comforts and necessities, on the basis of the degree of intensity of the need they satisfy. Demand for *luxury goods* (e.g., air conditioners, costly TV sets, cars, and decoration items) is more elastic than the demand for other kinds of goods because consumption of luxury goods can be postponed when their price rises. On the other hand, consumption of *necessities* (e.g., sugar, clothes, vegetables, and electricity, medicines) cannot be postponed and hence their demand has lower inelasticity. Demand for *comforts* is generally more elastic than that for necessities and less elastic than the demand for luxuries. Commodities may also be classified as durable goods and non-durable goods. Demand for durable goods is more elastic than that for non-durable goods—mainly necessities because when the price of the former increases, people either get the old one repaired instead of replacing it or buy a 'second-hand'.
3. **Proportion of income spent***:* Another factor that influences the elasticity of demand is the proportion of consumer's income spent on a particular commodity. If proportion of income spent on a commodity is very small, its demand will be inelastic, and *vice versa*. Classic examples of such commodities are salt, sugar, books, toothpastes, which claim a very small proportion of consumers' income. Demand for these goods is generally inelastic because increase in the price of such goods does not substantially affect consumer's budget.
4. **Time factor***:* Price elasticity of demand for high-price goods depends also on the time consumers can take to adjust their consumption expenditure to buy a new commodity—the shorter the time taken, the greater the elasticity. Consumers are able to adjust their expenditure pattern to price changes over a short period of time. For instance, if price of TV sets is decreased, demand will immediately increase if people possess excess purchasing power and require a short time to take decision. But, if not, then people may not be able to adjust their expenditure pattern over a short period of time to buy a TV set at the (new) lower price. If consumption adjustment takes a long period, it creates uncertainty and makes elasticity lower.
5. **Range of alternative uses of a commodity***:* The wider the range of alternative uses of a product, the higher the elasticity of its demand for

decrease in price and the lower elasticity for rise in price. Decrease in the price of a multi-use commodity encourages the extension of their use. Therefore, the demand for such a commodity generally increases more than the proportionate decrease in its price. For instance, milk can be taken as it is, it may be converted into curd, cheese, *ghee* and butter milk. The demand for milk will, therefore, be highly elastic for decrease in price. Similarly, electricity can be used for lighting, cooking, heating and for industrial purposes. Therefore, demand for electricity is highly elastic, especially for decrease in price. Reverse is the case for rise in their price.

6. **The proportion of market supplied***:* Technically, the elasticity of market demand depends also on the proportion of the market supplied at the ruling price. If less than half of the market is supplied, elasticity of demand will be higher and if more than half of the market is supplied elasticity will be lower. That is, towards the upper end, demand curve is more elastic than towards the lower end.
7. **Direction of change in price***:* The direction of change in price, i.e., whether price rises or falls, also determines the elasticity coefficient. Between any two points on the demand curve, price elasticity coefficient is higher for the fall in price and it is lower for the same rise in price.

Other Elasticities of Demand

We have discussed above the various aspects of price elasticity of demand. The price of a product is undoubtedly the most important determinant of its demand, especially in the short run. But price is not the only determinant of demand, especially in the long run. Going by the *dynamic demand function*, there are some other important demand determinants, viz., (i) price of the related good—substitutes and complements, (ii) income of the consumers, (iii) advertisement of the product, and (iv) future price expectation. The elasticity of demand with respect to these demand determinants plays a significant role in determining the future demand prospects and in business planning. In this section, we will discuss elasticities of demand with respect to the price of related goods, i.e., *cross elasticity*, and consumer's income, i.e., *income elasticity*.

Cross-Elasticity of Demand

Cross-Elasticity is the measure of responsiveness of demand for a commodity to the changes in the price of its substitutes and complementary goods. For instance, in case of **substitutes**, the cross-elasticity of demand for tea (T) is the percentage change in quantity demanded of tea due to a change in the price of its substitute, coffee (C). The formula for measuring cross-elasticity of demand is the same as the measure of price elasticity with a minor modification. The cross elasticity of demand for tea ($e_{t,c}$) with respect to price of coffee (P_c) can be measured as follows.

$$e_{t,c} = \frac{\text{Percentage change in demand for tea } (Q_t)}{\text{Percentage change in price of coffee } (P_c)}$$

Going by the price elasticity formula, the cross-elasticity of demand for tea in response to change in price of coffee is given as follows:

$$e_{t,c} = \frac{P_c}{Q_t} \cdot \frac{\Delta Q_t}{\Delta P_c} \tag{2.4}$$

NOTES

For a numerical example, suppose that price of coffee (P_c) increases from ₹10 to ₹15 per cup and as a result, demand for tea increases from 20 to 30 cups per week, price of tea remaining constant. By substituting these values in Eq. (2.4), we get cross-elasticity of demand for tea with respect to the price of coffee, as

$$\begin{aligned} e_{t,c} &= \frac{10}{20} \cdot \frac{20-30}{10-15} \\ &= \frac{10}{20} \cdot \frac{-10}{-5} \\ &= 1.0 \end{aligned}$$

The same formula is used to measure the cross-elasticity of demand for a good in response to change in the price of its **complementary goods**. Electricity to electrical gadgets, petrol to automobile, butter to bread, sugar and milk to tea and coffee, are the examples of complementary goods.

It is **important** to note here that when two goods are substitutes for each other, their demand has a *positive cross-elasticity* because increase in the price of one increases the demand for the other. But, the demand for complementary goods has *negative cross-elasticity*, because increase in the price of a good decreases the demand for its complementary goods.

Another *important* aspect of cross-elasticity is that it provides the basis for treating a commodity as a substitute or a complementary good. If cross-elasticities between any two goods are positive, the two goods can be treated as substitutes for each other. Also, the higher the cross-elasticity, the closer the substitute. Similarly, if cross-elasticity of demand for any two related goods is negative, the two good may be considered as complementary for each other: the higher the negative cross-elasticity, the higher the degree of complementarity.

Income Elasticity of Demand

Apart from price of a product and its substitutes, another important determinant of demand for a product is consumer's income. As noted earlier, the relationship between demand for normal goods and consumer's income is of positive nature. In simple words, the demand for normal goods and services increases with increase in consumer's income and *vice versa*. The responsiveness of demand to the change in consumer's income is known as **income elasticity** of demand.

Income elasticity (e_m) of demand for a product, say X, with respect to change in money income (M) can be defined as:

$$e_m = \frac{\Delta Q_x / Q_x}{\Delta M / M} = \frac{M}{Q_x} \cdot \frac{\Delta Q_x}{\Delta M} \tag{2.5}$$

where Q_x = quantity of X demanded; M = disposable money income; Q_x = change in quantity demanded of X; and M = change in income.

As shown in Eq. (2.5), unlike price elasticity of demand (which is negative except in case of Giffen goods), income elasticity of demand for normal goods has a positive sign because there is a positive relationship between the income and the quantity demanded of the product. There is an exception to this rule. Income elasticity of demand for an *inferior good* is negative, because of negative income effect. The demand for inferior goods decreases with increase in consumer's income and *vice versa*. When income increases, consumers switch over to the consumption of superior commodities. That is, they substitute superior goods for inferior ones. For instance, when income rises, people prefer to buy more of rice and wheat and less of inferior food grains like *bajra*, ragi, etc. and use more of taxi and less of bus service and so on.

Nature of Commodity and Income Elasticity For all normal goods, income elasticity is positive though the degree of elasticity varies depending on the nature of commodities. As noted above, consumer goods are generally grouped under three broad categories, viz., necessities (essential consumer goods), comforts, and luxuries. The general pattern of income elasticities for goods of different categories for increase in income and their impact on sales are given in Table 2.3.

Table 2.3 *Nature of Commodities, Income Elasticity and Expenditure*

Commodities	Coefficient of Income Elasticity	Impact on Expenditure
1. Necessities	Less than unity ($e_y < 1$)	Less than proportionate change in expenditure
2. Comforts	Almost equal to unity ($e_y \cong 1$)	Almost proportionate change in expenditure
3. Luxuries	Greater than unity ($e_y > 1$)	More than proportionate increase in expenditure

Income elasticity of demand for different categories of goods may, however, vary from household to household and from time to time, depending on choice, taste and preference of the consumers; levels of their consumption and income; and their susceptibility to 'demonstration effect'. The other factor which may cause deviation from the general pattern of income elasticities is the frequency of increase in income. If income increases regularly and frequently, income elasticities will conform to the general pattern, otherwise not.

Uses of Income Elasticity: Some *important uses* of income elasticity are following:

First, the concept of income elasticity can be used to estimate the future demand for a product provided the rate of increase in income and income elasticity of demand for the product are known. The knowledge of income elasticity can be used for forecasting demand, when a change in personal income is expected, other things remaining the same.

Secondly, the concept of income elasticity can also be used to define the 'normal' and 'inferior' goods. The goods whose income elasticity is *positive* for all levels of income are termed as 'normal goods'. On the other hand, the goods for which income elasticities are *negative*, beyond a certain level of income, are termed as 'inferior goods'.

NOTES

Application of Demand Elasticity

Although Samuelson condemned the concept of elasticity as an 'essentially arbitrary' and a more or less 'useless concept', it has many important uses in both economic analysis and formulation of economic policies. Some important uses of elasticity of demand are described here briefly.

1. **Application of elasticity in business decisions:** The concept of elasticity of demand plays a crucial role in business decisions regarding manoeuvring of prices with a view to making larger profits. For instance, when cost of production is increasing, the firm would like to raise the price. Firms may decide to change the price even without change in cost of production. But, whether raising price following the rise in cost or otherwise will prove beneficial or not depends on (a) the price elasticity of demand for the products and (b) its cross-elasticity because when the price of a product increases, its substitutes become automatically cheaper even if their prices remain unchanged. Raising price will be beneficial only if (i) demand for a product has an elasticity less than 1 and (ii) demand for its substitute has cross-elasticity less than 1. Although most businessmen, intuitively, are aware of the elasticity of demand of the goods they make, use of precise estimates of elasticity of demand adds precision to the business decisions.
2. **Application of elasticity in formulation of government policies:** The elasticity of demand can be used in formulating government policies, particularly in respect of (a) commodity taxation policy aiming at raising revenue or controlling demand; (b) granting subsidies to the industries; (c) determining prices for public utilities; (d) fixing prices of essential goods; and (e) in determining export and import duties and the rate of devaluation of domestic currency. To consider an example, suppose government wants to impose sales tax on a particular commodity with the sole objective of raising revenue. Whether adequate revenue can be raised or not depends on the price elasticity of that commodity. If demand is highly elastic, the revenue yield will be much less than expected. The tax will instead cause price distortion and affect production adversely. But, if objective is to control demand, then the price elasticity must be greater than 1.
3. **Application of elasticity in economic analysis:** The concept of elasticity is useful in economic analysis, at least for specifying the relationship between the dependent and independent variables. Besides, the elasticity concept is used in specifying and estimating demand functions. The most common form of a dynamic demand function used in empirical research is the 'constant elasticity demand function' of the form given below.

$$Q_X = AP_X^B \Upsilon^C P^D E^{FT}$$

in which P_X, Y, P_Y and E_{ET} represent, respectively, price of commodity X, consumer's income, price of other goods and a trend factor of 'taste', and superscripts B, C, D are the respective elasticity coefficients, and A is a constant.

Check Your Progress

3. State the formula for calculating the price elasticity of demand.
4. What is arc elasticity?

2.4 LAW OF SUPPLY

In a market economy, while buyers of a product constitute the demand side of the market, sellers of that product make the supply side of the market. In this section, we discuss the supply side of the market.

Meaning of Market Supply

Supply means the quantity of a commodity which its producers or sellers offer for sell at a given price, per unit of time. Market supply, like market demand, is the sum of supply of a commodity made by all individual firms or suppliers.

The Law of Supply

In general sense of the term, the supply of a commodity depends on its price. In other words, supply of a product is the function of its price. The law of supply is expressed generally in terms of price–quantity relationship. The **law of supply** can be stated as follows: ***The supply of a product increases with the increase in its price and decreases with decrease in its price, other things remaining constant***. It implies that the supply of a commodity and its price are positively related. This relationship holds under the assumption that "other things remain the same". "Other things" include technology, price of related goods (substitute and complements), consumers' taste and preferences, and weather and climatic conditions in case of agricultural products.

The Supply Schedule and Supply Curve

The law of supply can be depicted by a *supply schedule* and a *supply curve*. A supply schedule is a table showing quantity that suppliers are willing to offer for sale at different prices. Table 2.4 presents a hypothetical supply schedule of shirts, i.e., number of shirts supplied per month at different prices.

Table 2.4 *Supply Schedule of Shirts*

Price (in ₹)	Supply (Shirts in '000)
100	10
200	35
300	50
400	60
600	75
800	80

NOTES

The ***supply curve*** is a graphical presentation of the supply schedule. The supply curve *SS*' given in Figure 2.3 has been drawn by plotting the price and supply data given in Table 2.4. The points *S, P, Q, R, T* and *S*' show the price-quantity combinations on the supply curve *SS*'. The supply curve, *SS*', depicts *the law of supply*. The upward slope of the supply curve indicates the rise in the supply of shirts with the rise in its price and *vice versa*. That is, the supply of shirts increases with the rise in its price and *vice versa*. For example, at price ₹200, only 35,000 shirts are supplied per month. When price rises to ₹400, supply increases to 60,000 shirts.

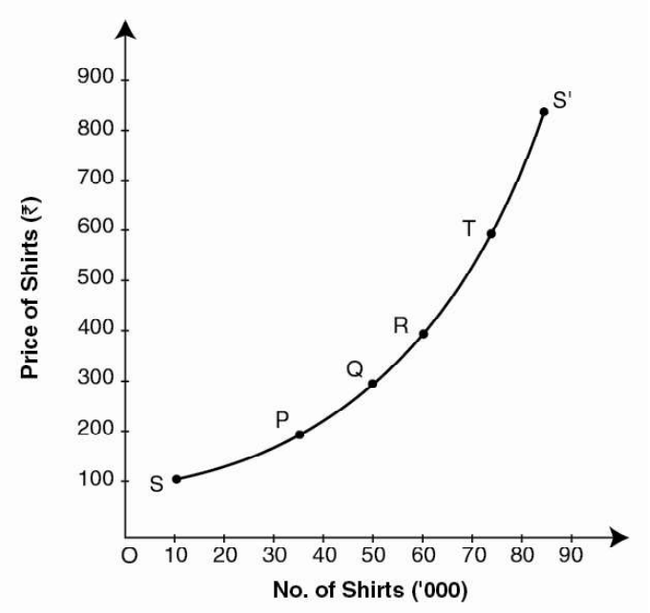

Fig. 2.3 *Supply Curve of Shirts*

As shown in Figure 2.3, *a supply curve has a positive slope*. The positive slope of the supply curve is caused by seller's desire to make larger profit and, more importantly, by the rise in cost of production. In fact, when price of a commodity increases, its suppliers tend to supply more and more. To supply more and more, they need to produce more and more. When they increase production, cost of production increases due to the law of diminishing returns. In fact, supply curve is derived from the marginal cost curve.

2.4.1 Shift in the Supply Curve

We have shown above that a change in the price of a commodity causes a change in its quantity supplied along a given supply curve. Although price of a commodity is the most important determinant of its supply, it is not the only determinant. Several other factors influence the supply of a commodity. Given the supply curve of a commodity, when there is a change in its other determinants, the supply curve shifts rightward or leftward depending on the effect of such changes. Let us now explain how other determinants of supply cause shift in the supply curve.

1. **Change in input prices:** Input prices include the price of labour, raw materials, overheads, etc. Input prices determine the cost of production. When input prices decrease, the use of inputs increases. As a result, product supply increases and the supply curve *SS* shifts to the right to *SS″*, as shown in Figure 2.4. Similarly, when input prices increase, product supply curve shifts leftward from *SS* to *SS′*.

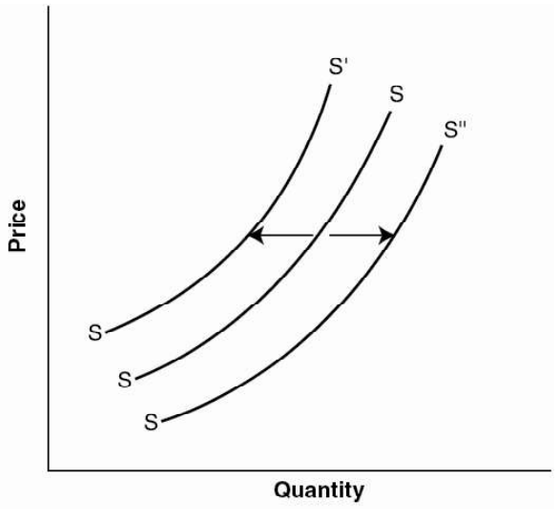

Fig. 2.4 *Shift in the Supply Curve*

2. **Technological progress:** Technological progress reduces cost of production or increases labour productively or does both. Technological progress that reduces cost of production or increases efficiency causes increase in product supply. For instance, introduction of high-yielding variety of paddy and new techniques of cultivation increased per-acre yield of rice in India in the 1970s. Such changes make the supply curve shift to the right.
3. **Product diversification and cost reduction:** In production of many commodities, it is possible to produce some other goods which require a similar technology. For example, a refrigerator company can also produce ACs; Tatas famous for truck production can also produce Nano and other types of cars; Maruti Udyog can produce trucks and so on. Product diversification may cause reduction in the production cost of the main product. This may lead to the rise in the supply of the main product due to capacity utilization for profit maximization.
4. **Nature and size of the industry:** The supply of a commodity depends also on whether an industry is monopolized or competitive. Under monopoly, supply of a product is shorter than it is in a competitive market. When a monopolized industry is made competitive, the total supply increases. Besides, if size of an industry increases due to new firms joining the industry, the total supply increases and supply curve shifts rightward.

NOTES

5. **Government policy:** When government imposes restrictions on production, e.g., import quota on inputs, rationing of or quota imposed on input supply, etc., production tends to fall. Such restrictions make supply curve to shift leftward.
6. **Non-economic factors:** The factors like labour strikes and lock-outs, war, droughts, floods, communal riots, epidemics, etc. also affect adversely the supply of commodities making supply curve shift leftward.

Supply Function

A *supply function* is a mathematical statement which states the relationship between the quantity supplied of a commodity and its determinants. The short-run market supply function is based on the law of supply. The law of supply states the nature of relationship between the price and the quantity supplied, i.e., supply increases with the increase in price. A supply function that specifies the relationship between the price and supply of a product is expressed as

$$Q_x = dP_x \tag{2.6}$$

where Q_x denotes the quantity supplied of commodity *X;* P_x denotes its price; and d gives the measure of relationship between Q_x and P_x.

Once the relationship between Q_x and P_x is measured in numerical terms, i.e., the numerical value of 'd' is known, then the supply function can be expressed numerically. For example, suppose $d = 10$, then the factual supply function can be expressed as

$$Q_x = 10P_x \tag{2.7}$$

Given the supply function (2.7), a supply schedule can be obtained by substituting numerical values for P_x. For example, if $P_x = 2$, $Q_x = 20$ and if $P_x = 5$, $Q_x = 50$. By plotting the supply schedule, a supply curve can be obtained.

Check Your Progress

5. What is market supply?
6. Define supply curve.

2.5 LAW OF DIMINISHING MARGINAL UTILITY AND EQUI MARGINAL UTILITY

The law of diminishing *MU* is the fundamental law on which is based the cardinal utility analysis of the consumer behaviour. This law states that as the quantity consumed of a commodity increases per unit of time, the utility derived by the consumer from the successive units goes on decreasing, provided the consumption of all other goods remains constant. This law is founded on the basis of some basic facts of life: (i) the utility derived from a commodity depends on the intensity or urgency of the need for that commodity, and (ii) as more and more quantities of

a commodity is consumed, the need gets satisfied and therefore the intensity of need decreases. For these reasons, the utility derived from the marginal unit goes on diminishing. For example, suppose you are very hungry and you are offered sandwiches to eat. The utility that you derive from the first piece of sandwich would be the maximum because intensity of your hunger is the highest. When you eat the second piece, you derive a lower satisfaction because intensity of your hunger is reduced. As you go on eating more and more sandwiches, the intensity of your hunger goes on decreasing and therefore the satisfaction which you derive from the successive units goes on decreasing. If you continue to eat sandwiches, a point is reached when your hunger is fully satisfied and therefore the last piece of sandwich gives you zero utility. Eating sandwiches any more will give you a negative utility in the form of discomfort or stomachache. This relationship between quantity consumed and utility derived from each successive unit consumed is called the **law of diminishing MU**.

Numerical Example

Table 2.5 presents a numerical illustration of the law of diminishing *MU*. As the table shows, *TU* increases with increase in consumption of sandwiches, but at a decreasing rate. It means that *MU* decreases with increase in consumption. This is shown in the last column of the table. It can also be seen in the table that the *TU* reaches its maximum level at 100 at four sandwiches consumed. The consumption of the fifth sandwich gives no utility, i.e., its *MU* = 0. Consumption of the sixth sandwich yields a negative utility of 10 and the *TU* declines to 90.

Table 2.5 *Total and Marginal Utility*

Sandwiches	Total Utility	Marginal Utility = $TU_n - TU_{n-1}$
1	40	40 - 0 = 40
2	70	70 - 40 = 30
3	90	90 - 70 = 20
4	100	100 - 90 = 10
5	100	100 - 100 = 00
6	90	90 - 100 = - 10

Graphical Illustration

The law of diminishing *MU* is graphically illustrated in Figure 2.5. The *TU* and *MU* curves have been obtained by plotting the data given in Table 2.5. The *TU* curve is rising till the fourth sandwich is consumed. Note that the *TU* curve is rising but at a diminishing rate. It shows decrease in the *MU*, i.e., the utility added to the total. The diminishing *MU* has been shown by the *MU* curve. Beyond five sandwiches consumed, the *MU* turns negative. It means that additional consumption of sandwiches yields disutility in the form of discomfort.

NOTES

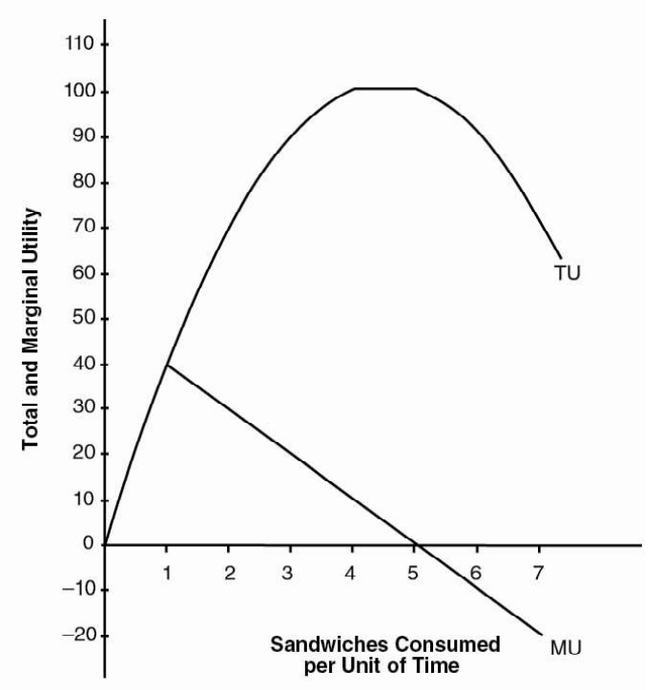

Fig. 2.5 *Total and Marginal Utility*

Assumptions

The law of diminishing *MU* holds only under certain given conditions. These conditions are often referred to as the *assumptions* of the law.

First, **the unit of the consumer goods must be standard,** e.g., a cup of tea, a bottle of cold drink, a pair of shoes or a shirt and so on. If the units are excessively small or large, the law may not apply. For example, a sip of tea or a bite of sandwich may increase your desire for more tea or sandwich. It means that *MU* increases.

Secondly, **consumer's taste and preference remains unchanged** during the period of consumption. If taste and preference change during the period of consumption, the law may not apply.

Thirdly, **there must be continuity in consumption** and where break in continuity is necessary, it must be appropriately short.

Fourthly, **the mental condition of the consumer remains normal** during the period of consumption. For, if a person is eating and also drinking alcohol the utility pattern will not be certain.

Given these conditions, the law of diminishing *MU* holds universally. In some cases, e.g., accumulation of money, collection of hobby items like stamps, old coins, rare paintings and books, and melodious songs, etc., *MU* may initially increase rather than decrease, but it does decrease eventually. That is, the law of *MU* generally operates universally.

2.5.1 Consumer's Equilibrium: Cardinal Utility Approach

NOTES

A consumer attains his equilibrium when he maximizes his *TU* given his income, consumption expenditure and prices of commodities he consumes. Analysing consumer's equilibrium requires answering the question 'how does a consumer allocate his money income to the various goods and services he consumes to arrive at his equilibrium?' In this section, we explain how a consumer attains his equilibrium by applying the cardinal utility approach, under (i) a single commodity case, and (ii) the multiple commodity case.

The cardinal utility approach or what is also called as the Marshallian approach to consumer's equilibrium is based on the following assumptions.

Assumptions

1. ***Rationality:*** It is assumed that the consumer is a rational being in the sense that he satisfies his wants in order of their merit and the necessity. It means that he buys first a commodity which yields the highest utility and he buys last a commodity which gives the least utility.
2. ***Limited Money Income:*** The consumer has a limited money income to spend on the goods and services he chooses to consume.
3. ***Maximization of Satisfaction*:** Every rational consumer intends to maximize his satisfaction from his given money income. That is, he chooses the commodities and spends his income on each of the commodity in such a way that his *TU* is maximized.
4. ***Utility is Cardinally Measurable:*** The cardinalists assume that utility is cardinally measurable, i.e., it can be measured in absolute terms and in cardinal numbers.
5. ***Diminishing MU:*** The consumption is subject to the law of diminishing marginal utility. That is, the utility derived from successive units of a commodity consumed decreases as a consumer consumes more and more units of it.
6. ***Constant Utility of Money:*** The *MU* of money remains constant whatever the level of consumer's income and each unit of money has utility equal to one.
7. ***Utility is Additive*:** Cardinalists maintain that utility derived from different goods can be added up. The additivity of the utility can be expressed through a utility function. Suppose that the basket of goods and services consumed by a consumer contains *n* items, and their quantities may be expressed as $x_1, x_2, x_3, \ldots, x_n$. The utility function of the consumer may be expressed as

$$U = f(x_1, x_2, x_3, \ldots, x_n)$$

Given the utility function, the *TU* obtained from *n* items may be expressed as

$$U_n = U_1(x_1) + U_2(x_2) + U_3(x_3) + \ldots + U_n(x_n)$$

NOTES

Single Commodity Case

Having noted the assumptions of cardinal utility approach, we turn to analyse consumer's equilibrium. As a general rule, a utility maximizing consumer consuming several commodities reaches his equilibrium when he maximizes his *TU*. However, for the sake of simplicity, we illustrate first consumer's equilibrium with a simple *one-commodity case.*

Suppose that a consumer with a given money income consumes only one commodity, *X*. Since both his money income and commodity *X* have utility for him, he can either spend his money income on commodity *X* or retain it with himself. If he has total money and no commodity *X*, the *MU* of money will be lower than that of commodity *X* because $MU_m = 1$. But *MU* of commodity is supposed to be greater than 1. So long as *MU* of commodity *X* (i.e., MU_x) is greater than *MU* of money income (MU_m), *TU* can be increased by exchanging money for the commodity. Therefore, a utility maximizing consumer exchanges his money income for the commodity as long as $MU_x > MU_m$. As assumed earlier, *MU* of commodity of *X* is subject to the law of diminishing returns (assumption 5), whereas *MU* of money income (MU_m) remains constant (assumption 6). Therefore, a utility maximizing consumer will exchange his money income for commodity *X* as long as $MU_x > MU_m$. The consumer reaches his equilibrium at the level of consumption at which $MU_x = MU_m$.

In reality, however, the price of most goods is more than Re 1. In that case, the consumer's equilibrium can be expressed as

$$MU_x = P_x (MU_m) : (\text{where } MU_m = 1) \tag{2.8}$$

It implies that the consumer reaches equilibrium where,

$$\frac{MU_x}{P_x(MU_m)} = 1 \tag{2.9}$$

Consumer's equilibrium in a *single commodity case* is graphically illustrated in Figure 2.6. The horizontal line $P_x(MU_m)$ shows the constant utility of money weighted by P_x (the price of commodity *X*) and MU_m curve represents the diminishing *MU* of commodity *X*. The $P_x(MU_m)$ line and MU_x curve intersect at point *E*, where $MU_x = P_x(MU_x)$. Therefore, consumer is in equilibrium at point *E*. At any point above *E*, $MU_x > P_x(MU_m)$. Therefore, if a consumer exchanges his money income for commodity *X*, he increases his satisfaction per unity of commodity. At any point below *E*, $MU_x < P_x(MU_m)$, the consumer can therefore increase his satisfaction by reducing his consumption of commodity *X*. That is, at any point other than *E*, the consumer gets satisfaction less than maximum. Therefore, point *E* is the point of consumer's equilibrium.

NOTES

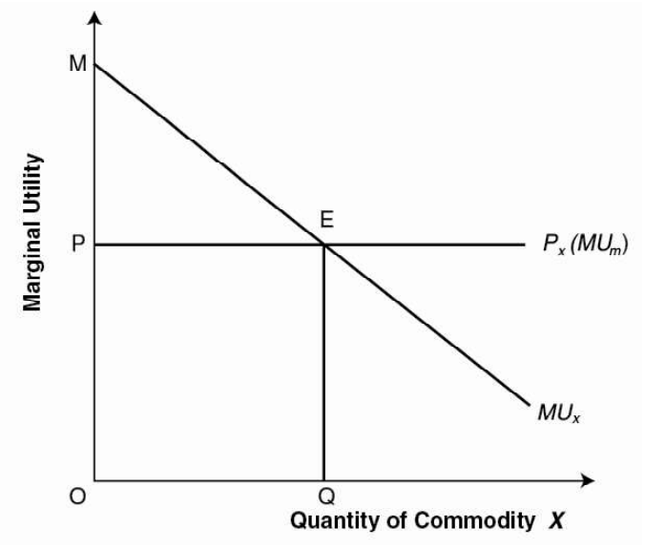

Fig. 2.6 *Consumer's Equilibrium: One Commodity Case*

The theoretical fact that the consumer is in equilibrium at point *E* can be proved by the data shown in Figure 2.6. As the figure reveals, the *TU* that the consumer derives by consuming *OQ* units of *X* equals the area *OMEQ*. The total money that consumer pays for *OQ* units equals $OP \times OQ = OPEQ$. This is the total utility of money paid for consuming *OQ* units. When the total utility paid (*OPEQ*) is subtracted from the total utility gained (*OMEQ*), it gives the net utility gained. That is, *OMEQ* " *OPEQ* = *MPE* = net utility gain. The net utility gained (*MPE*) is maximum. It can be checked that any consumption less than or more than *OQ* units will reduce the area *MPE*. So the consumer maximizes his utility at point *E* where $MU_x = MU_m$.

The Multiple Commodity Case

We have explained above the determination of consumer's equilibrium in a single commodity case. In reality, however, a consumer consumes a large number of goods. Let us now see how a consumer consuming a large number of goods and services attains his equilibrium.

We know that the *MU* schedules of various commodities may not be the same. Some commodities yield higher utility and some lower. The *MU* of some goods decreases at a higher rate and of some at lower rate. A rational and utility maximizing consumer consumes commodities in the order of their utilities. He picks up the commodity which yields the highest utility and next he picks up the commodity which yields the second highest utility and so on. The consumer switches his expenditure from one commodity to another in accordance with their *MU*. He continues to switch his expenditure from one commodity to the other until he reaches a stage where *MU* of each commodity *per unit of money expenditure* is the same. This is called the **law of equi-marginal utility**.

The Law of Equi-Marginal Utility: Let us now present the law of equi-*MU* in a simple two-commodity case. Let us suppose that a consumer consumes

NOTES

only two commodities X and Y, and their prices are given as P_x and P_y, respectively. Following the equilibrium rule of single commodity case, the consumer distributes his expenditure between commodities X and Y in such a way that

$$MUx_x = P_x(MU_m)$$

and

$$MUx_y = P_y(MU_m)$$

or alternatively, consumer is in equilibrium where

$$\frac{MU_x}{P_x(MU_m)} = 1 \tag{2.10}$$

and

$$\frac{MU_y}{P_y(MU_m)} = 1 \tag{2.11}$$

Equations (2.10) and (2.11) may be combined to express consumer's equilibrium condition under two-commodity case as follows.

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

or $$\frac{MU_x}{P_x(MU_m)} = 1 = \frac{MU_y}{P_y(MU_m)} \tag{2.12}$$

Since, by assumption 5, MU of each unit of money remains constant, Eq. (2.12) may be rewritten as

$$\frac{MU_x}{MU_y} = \frac{P_x}{P_y} \tag{2.13}$$

or $$\frac{MU_x}{P_x} = \frac{MU_y}{P_y} \tag{2.14}$$

Equation (2.14) gives the utility maximization rule that the consumer reaches his equilibrium when the MU derived from each unit of money spent on the two commodities X and Y is the same.

The two-commodity case provides the basis for generalizing the consumer's equilibrium by the cardinal utility approach in a multi-commodity case. In fact, a consumer consumes a large number of goods and services with his given income and at different prices. Supposing a consumer consumes A to Z goods and services, his equilibrium condition may be expressed as follows:

$$\frac{MU_A}{P_A} = \frac{MU_B}{P_B} = \frac{MU_C}{P_C} = \ldots = \frac{MU_Z}{P_Z} \tag{2.15}$$

Thus, according to the law of equi-marginal utility, a utility maximizing consumer consuming several goods and services intends to equalize the MU of each unit of his money spent on various goods and services.

2.6 CONSUMER SURPLUS

NOTES

The consumers' willing to pay for a commodity depends on the utility they expect to derive from the commodity. The price which a consumer is willing to pay may not match with the market price of the commodity. It may be greater or less than the market price. If market price is less than what the consumer is willing to pay, then he saves some money. In economics terminology, this saving is called as **consumer surplus**.

The concept of *consumer's surplus* is believed to have been originated by a French engineer, Arsene Julis Dupuit in 1844, in his effort to measure social benefit of such collective goods as roads, canals and bridges. In his opinion, the value of the benefit of such collective goods was greater than the price actually charged because most people would be willing to pay a higher price than they actually paid. The concept was later refined by Marshall who also provided a measure of consumer's surplus. His premise of measuring consumer's surplus was, however, rejected by the *ordinalists*, especially J.R. Hicks, who attempted to provide a different method of measuring consumer's surplus through their indifference curve technique.

There are various methods of measuring consumer's surplus and their merits and demerits. In this section, you will learn about the Marshallian concept and measure of consumer surplus and its drawbacks.

2.6.1 Marshallian Concept of Consumer Surplus

Although the concept of consumer surplus was originated by Dupuit as early as 1844, it remained an immeasurable concept until Marshall suggested, as late as 1920, a method of measuring consumer's surplus in money terms. Marshall defined consumer's surplus as "the excess of the price which [a consumer] would be willing to pay rather than go without the thing, over that which he actually does pay." According to the Marshallian theory of demand, what a consumer is willing to pay for one unit of a commodity measures the money value of his expected utility and what he actually pays gives the measure of the monetary cost of the expected utility. According to Marshall, the difference between the two values is the *'***consumer surplus***'*. For example, if you are prepared to pay ₹500 for a ticket to watch a cricket match and you pay only ₹200, the actual price of the ticket, you have a consumer surplus of ₹300.

The concept of consumer's surplus can be expressed also in terms of utility (or satisfaction). Recall that Marshall assumed marginal utility (*MU*) of money to remain constant. Under this condition, what a consumer is willing to pay for a commodity indicates the utility that he expects to derive from the commodity and what he actually pays gives the measure of the loss of utility (of money). The difference between the utility gained and the utility lost in acquiring the commodity is the consumer's 'surplus satisfaction' which Marshall called 'consumer's surplus'.

Measurement of Consumer Surplus

Having defined the concept of consumer surplus, Marshall provided a systematic method of measuring the consumer surplus, on the basis of certain assumptions.

NOTES

Assumptions The Marshall's method of measuring consumer's surplus is based on the following assumptions.

- First, it is assumed that the market price is given so that neither the sellers nor the buyers can affect the price. The consumer's surplus will not exist if there is a monopolist and he adopts first degree price discrimination in his pricing policy.
- Secondly, the utility is cardinally measurable and *MU* of consumer's money income remains constant throughout.
- Thirdly, the utility of each commodity is absolute and is independent of other goods and services consumed by the consumer.
- Fourthly, there is no close substitute for the commodity in question. For, if close substitutes are available, there may not be any difference between 'what the consumer would be willing to pay' and 'what he actually pays' for the commodity in question.

The Marshallian concept of consumer's surplus and its measurement are graphically illustrated in Figure 2.7. Suppose the consumer's demand curve for a commodity *X* is given by the demand curve *MN*. The curve *MN* also indicates the utility derived from each successive unit of a commodity and the price that the consumer is willing to pay at different levels of his purchases. Suppose that the market price, i.e., the price which a consumer actually pays, is given by *OP*. At price *OP*, the consumer buys *OQ* units. The total utility derived by the consumer from *OQ* units is shown by the area *OMBQ*, for which the consumer pays $OPBQ = OQ \times OP$. Thus, in the Marshallian sense, total consumer surplus equals $= OMBQ \times OPBQ = MPB$. That is, the shaded area *MPB* represents the consumer's surplus in the Marshallian sense when the consumer buys *OQ* units of a commodity *X*.

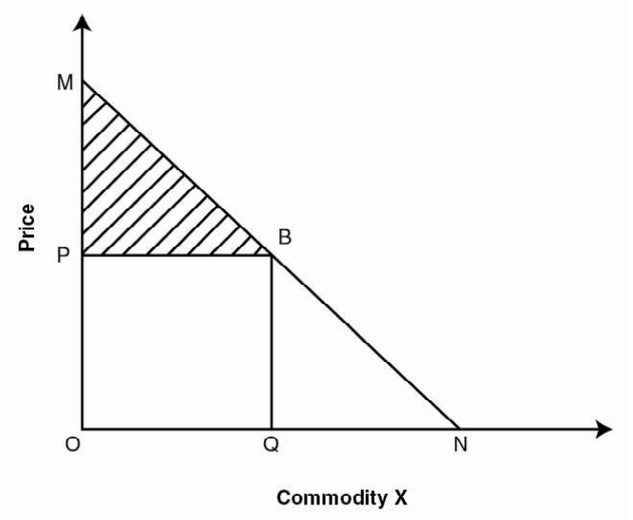

Fig. 2.7 *Consumer's Surplus*

2.6.2 Critical Appraisal

NOTES

The Marshallian concept and measurement of consumer's surplus have been criticized on many grounds, though the criticisms are equally questionable. The criticism of Marshallian concept of consumer surplus and its validity are discussed below.

- **First**, economists have pointed out difficulties in measuring the consumer's surplus as defined by Marshall and represented by 'a triangle'. A triangle cannot be formed because consumer's willingness to pay for zero unit is unknown. So demand curve cannot be extend to price axis. However, Mark Blaug rejects this criticism. In the words of Mark Blaug, 'It is sometimes objected that demand curves are usually asymptotic to the price axis. If the individual's offer for the first unit is not defined so that the demand curve does not touch the *Y*-axis, the integral under the demand curve is infinite. But this objection is easily overcome by measuring consumer's surplus from some selected value of $qx > 0$'.
- **Secondly**, a 'more fatal objection' to Marshall's method of measuring consumer's surplus as 'the triangle' under the demand curve is that real income does not remain constant along the demand curve even for 'unimportant' commodities. As the price falls along the demand curve (as shown in Figure 2.7), real income makes the estimate of consumer's surplus as ambiguous one. This criticism too does not hold because increase in demand due to decrease in price is caused also by its income effects.
- **Thirdly**, it is generally alleged that Marshallian assumptions on which the measurement of consumer's surplus is based are unrealistic. It is argued that MU of money does not remain constant; cardinal measurement of utility is not possible; utilities of various goods consumed by a consumer are not independent of each other; most goods have their substitutes—close or remote, and so on. Therefore, it is alleged that the Marshallian concept of consumer's surplus is imaginary and hypothetical. However, this criticism too does not hold in literal sense. Although utility may not be measurable cardinally or ordinally, consumers do have a mental perception of the usefulness of a commodity and, accordingly, they have a willingness to pay an amount for a commodity they need. It is not hypothetical.
- **Fourthly**, in the ultimate analysis of the consumer's purchases of various goods and services, consumer's surplus is reduced to zero. For, a consumer's willingness to pay (i.e., 'potential price') cannot exceed his income, i.e., what he actually pays out. It means that, when all purchases have been made, the consumers willingness to pay (which equals his income) equals what he actually pays (i.e., his income). This criticism is more hypothetical than the concept of consumer surplus as claimed by some economists.
- **Fifthly,** the concept of consumer's surplus cannot be convincingly applied to 'essential' and prestigious goods. For example, a hungry affluent person may be willing to pay thousands of rupees for a piece of bread whereas he may be required to pay only ten rupees. As such, his consumer's surplus

NOTES

will be equal to ₹99,990 which seems ridiculous. In case of prestigious goods, e.g., rare paintings, diamonds, jewellery, etc., what a buyer is willing to pay, generally, equals what he actually pays. It means there is no consumer's surplus. Thus, Marshallian concept of consumer surplus becomes illusory. However, these cases may be exceptions and exceptions prove the rule.

Although criticisms of Marshallian concept of consumer surplus are not strong enough to reject the concept, Samuelson considers this concept as of only 'historical and doctrinal interest' and suggests that 'the economists had best dispense with it'. Hicks has, however, tried to rehabilitate the consumer's surplus as, in his opinion, this concept is of great importance in the *economics of welfare* and also from pricing policy point of view.

Check Your Progress

7. What are the assumptions on which the Marshallian approach to consumer equilibrium is based?
8. Who designed the concept of 'consumer surplus'?

2.7 INDIFFERENCE CURVE: CHARACTERISTICS

The *indifference curve* is defined as *the locus of points each point representing a different combination of two goods yielding the same utility or level of satisfaction.* Since utility expected from the different combinations of the two goods is the same, a rational consumer is indifferent between any two combinations of goods when it comes to making a choice between them. Such a situation arises because a consumer consumes a large number of goods and services and often finds that one commodity can be used as substitute for another. This gives the consumers an opportunity to substitute one commodity for another. In that case, they are able to form various combinations of two substitute goods that give them the same level of satisfaction. When a consumer is faced with such combinations of goods, he would be indifferent between the combinations. When such combinations are plotted graphically, it appears in the form of a curve. This curve is known as the **indifference curve**. Indifference curves are also called *iso-utility* or *equal utility curves*.

For example, let us suppose that a consumer forms five combinations *a, b, c, d* and *e* of two commodities, *X* and *Y*, as presented in Table 2.6. All these combinations yield the same level of satisfaction *(U)*. The consumer is, therefore, indifferent to the choice between them. The five combinations of the two commodities *X* and *Y* may be called as an *indifference schedule*.

Table 2.6 shows five combinations of two goods, *X* and *Y*, which give the same utility. The last column of the table shows an unquantified utility *(U)* derived from each combination of *X* and *Y*. Utility *(U)* is unquantified because, under the ordinal utility approach, utility is not measurable quantitatively.

Table 2.6 *Indifference Schedule of Commodities X and Y*

Combination		Commodity *X*	+	Commodity *Y*		Utility
a	=	25	+	5	=	*U*
b	=	15	+	7	=	*U*
c	=	10	+	12	=	*U*
d	=	6	+	20	=	*U*
e	=	4	+	30	=	*U*

When the combinations *a, b, c, d* and *e* given in Table 2.6 are plotted and joined by a smooth curve (as shown in Figure 2.8), the resulting curve *IC* is known as the **indifference curve**. On this curve, one can locate many other points showing many other combinations of *X* and *Y*, which yield the same level of satisfaction. Therefore, the consumer is indifferent to make choice between the points on the indifference curve. Therefore, the curve is called the 'indifference curve'.

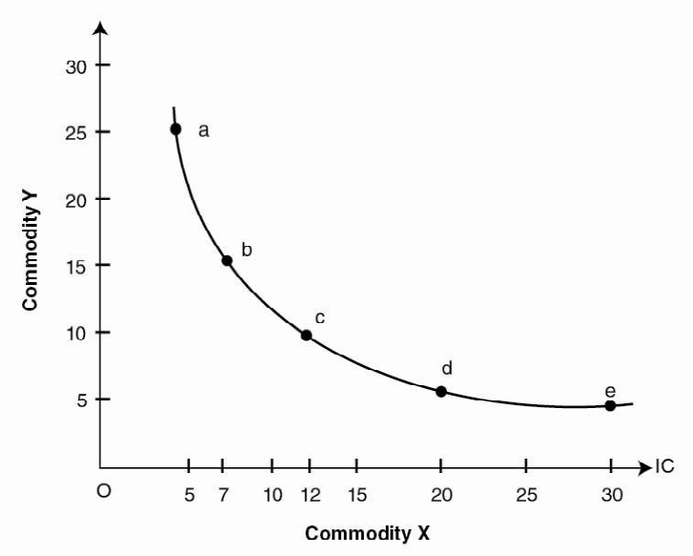

Fig. 2.8 *Indifference Curve*

Indifference Map

Figure 2.8 presents a single indifference curve *IC* drawn on the basis of the indifference schedule given in Table 2.6. The consumer can similarly frame many other combinations of *X* and *Y* with less amounts of both the goods such that each combination yields the same level of satisfaction but less than the level of satisfaction indicated by the indifference curve *IC* in Figure 2.8. Similarly, a consumer can concoct many other combinations with more of both the goods—each combination yielding the same satisfaction, but yielding a greater level of satisfaction than the smaller combination. Thus, another indifference curve can be drawn above the *IC* curve. This exercise may be repeated as many times as one wants, each time generating a new indifference curve. A set of indifference curves constitute the **indifference map**, as shown in Figure 2.9.

NOTES

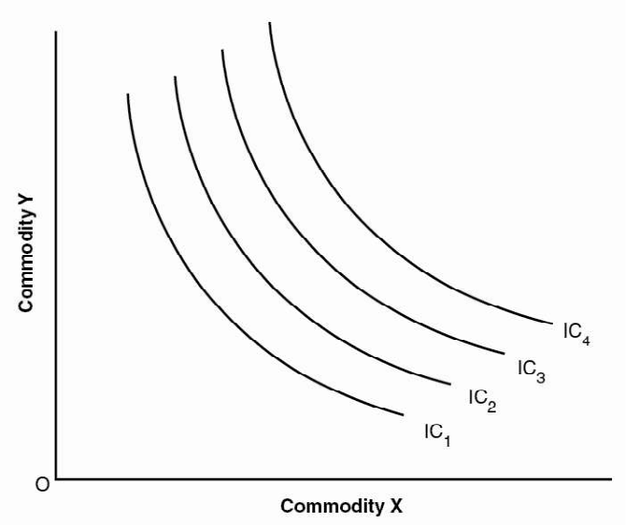

Fig. 2.9 *The Indifference Map*

In fact, the area between the *X* and the *Y* axes is known as the *indifference plane* or the *commodity space*. This plane contains finite points and each point on the plane indicates a different combination of the goods *X* and *Y*. Intuitively, it is always possible to locate two or more points indicating different combinations of the goods *X* and *Y* yielding the same level of satisfaction. It is thus possible to draw a number of indifference curves that neither intersect nor are tangent to one another, as shown in Figure 2.9. The set of indifference curves, IC_1, IC_2, IC_3 and IC_4, drawn in this manner constitute the **indifference map**. In fact, an indifference map may contain any number of indifference curves ranked in the order of consumer's preferences.

Characteristics of Indifference Curves

The indifference curve is a tool of analysis. As a tool of analysis, it has the following four basic properties.

1. Indifference curves slope downward to the right.
2. Indifference curves combining imperfect substitutes are convex to the origin.
3. Indifference curves do not intersect nor are they tangent.
4. An upper indifference curve implies a higher level of satisfaction than the lower ones.

Let's discuss each of these points here briefly.

1. **Indifference curves slope downward to the right:** In the words of Hicks, 'So long as each commodity has a positive marginal utility, the indifference curve must slope downwards to the right'. The downward slope of an indifference curve implies that in a basket of two substitute goods, if the quantity of one commodity decreases, the quantity of the other commodity must increase if the consumer has to maintain the same level of satisfaction. If the quantity of the other commodity does not increase simultaneously, the basket of commodities decreases with the decrease in

the quantity of one commodity. In that case, a smaller bundle of goods is bound to yield a lower level of satisfaction, which defies the logic of indifference curve.

2. **Indifference curves are convex to the origin:** Indifference curves for normal goods have not only a negative slope, but are also convex to the origin. The convexity of the indifference curves is caused by the following factors.
   - The two goods are *imperfect substitutes* for one another, and
   - The diminishing *MRS* incase of imperfect substitutes.
3. **Indifference curves neither intersect nor are tangent to one another:** If two indifference curves intersect or are tangential to each other, it would imply two types of inconsistencies in indifference curve logistics: (1) upper and lower indifference curves indicate the same level of satisfaction; and (2) the bigger and smaller combinations of two goods yield the same level of satisfaction. Such conditions are improbable if a consumer's subjective valuation of utility of a commodity is greater than zero. Obviously, if two indifference curves intersect, it would mean a violation of the consistency or transitivity assumption for consumers' preferences.

   Let us now prove the point graphically. Suppose two indifference curves, IC_1 and IC_2, intersect at point A, as shown in Figure 2.10. Consider two other points—point B on the indifference curve IC_1 and point C on the indifference curve IC_2, both falling on a vertical line.

   Points *A, B* and *C* represent three different combinations of commodities *X* and *Y*. Let us call these combinations, respectively, as combination *A, B* and C. Note that combination *A* is common to both the indifference curves. Since points *A* and *B* fall on the same *IC* curve IC_1, it means that in terms of utility,

   $$A = B$$

   Similarly, since points *A* and *C* fall on the same indifference curve, IC_2, it means that in terms of utility:

   $$A = C$$

   Since $A = B$ and $A = C$, it means that:

   $$B = C$$

   However, if combinations of goods at points *B* and *C* yield the some utility, it would mean that, in terms of utility,

   $$ON \text{ of } X + BN \text{ of } Y = ON \text{ of } X + CN \text{ of } Y$$

   Since *ON* of *X* is common to both the terms, it means that utility of *BN* of *Y* is equal to utility of *CN* of *Y*. However, as Figure 2.10 shows, $BN > CN$. Therefore, combinations at *B* and *C* cannot be equal in terms of utility in the subjective introspection of the consumer. The intersection of indifference curves, therefore, violates the transitivity rule, which is a logical necessity in indifference curve analysis.

NOTES

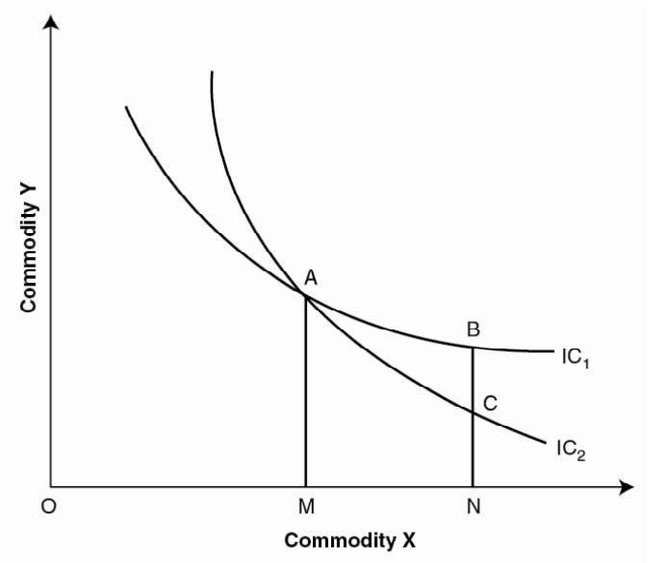

Fig. 2.10 *Intersecting Indifference Curves*

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

For example, consider the indifference curves IC_1 and IC_2 in Figure 2.11. The vertical movement from point *a* on the lower indifference curve, IC_1, to point *b* on the upper indifference curve, IC_2, means an increase in the quantity of *Y* by *ab*, the quantity of *X* remaining the same (*OX*). Similarly, a horizontal movement from point *a* to point *d* means a greater quantity of commodity *X*, the quantity of *Y* remaining the same (*OY*). A diagonal movement from point *a* to point *c* means larger quantities of both *X* and *Y*. Unless the utility of additional quantities of *X* and *Y* are equal to zero, these additional quantities will yield additional utility. Therefore, the level of satisfaction indicated by the upper indifference curve IC_2 would always be greater than that indicated by the lower indifference curve IC_1.

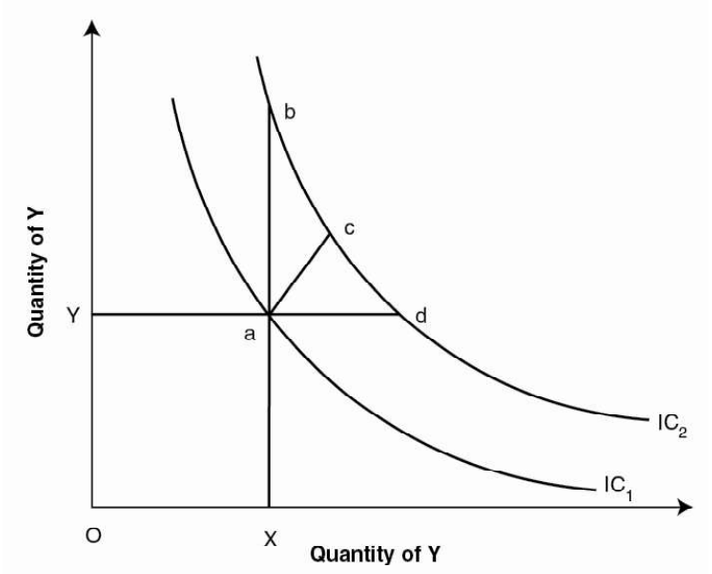

Fig. 2.11 *Comparison Between Lower and Upper Indifference Curves*

2.7.1 Consumer's Equilibrium: Ordinal Utility Approach

Before, you learn about the concept of consumer's equilibrium under the Ordinal utility approach, it is important to know a few terms.

Marginal Rate of Substitution

When a consumer makes different combination of two goods, yielding the same level of satisfaction, he substitutes one good for another. The rate at which he substitutes one good for the other is called the 'Marginal Rate of Substitution (MRS)'. One of the basic postulates of indifference curve analysis is that (MRS) diminishes. The axiomatic assumption of ordinal utility theory is analogous to the assumption of 'Diminishing Marginal Utility' in cardinal utility theory. The postulate of diminishing marginal rate of substitution states an observed behavioural rule that when a consumer substitutes one commodity (say X) for another (say Y), the 'Marginal Rate of Substitution' (MRS) decreases as the stock of X increases and that of Y decreases.

Conceptually, the *MRS* is the rate at which one commodity can be substituted for another, the level of satisfaction remaining the same. The *MRS* between two commodities, *X* and *Y*, can also be defined as the number of units of *X* which are required to replace one unit of *Y* (or number of units of *Y* that are required to replace one unit of *X*), in the combination of the two goods so that the total utility remains the same. It implies that the utility of units of *X* (or *Y*) given up is equal to the utility of additional units of *Y* (or *X*) added to the basket.

The negative slope of the indifference curve implies that two commodities are not perfect substitutes for each other. In case they are perfect substitutes, the indifference curve will be a straight line with a negative slope. Since, goods are not

perfect substitutes for each other, the *subjective value attached to the additional quantity (i.e., MU) of a commodity decreases fast in relation to the other commodity whose total quantity is decreasing.*

NOTES

Budget LineBudget Constraint and Budget Line

A utility maximising consumer would like to reach the highest possible indifference curve on his indifference map. However, the consumer is assumed to have a limited income. Limited income sets a limit to which a consumer can maximise his utility. The limitedness of income acts as a *constraint.* This is known as *budgetary constraint.* The assuming a two-commodity model, budgetary constraint may be expressed as:

$$P_x \cdot Q_x + P_y \cdot Q_y = M \tag{2.16}$$

where P_x and P_y are respective prices of X and Y, and Q_x and Q_y are their respective quantities; M is consumer's money income.

Equation (2.16) states that a consumer, given his income and prices of X and Y in the market, can buy only limited quantities of the two goods – X and Y. The maximum Q_x and Q_y can be obtained from Eq. (2.16), as follows.

$$Q_x = \frac{M}{P_x} - \frac{P_y}{P_x} Q_y \tag{2.17a}$$

and $$Q_y = \frac{M}{P_y} - \frac{P_x}{P_y} Q_x \tag{2.17b}$$

Equations (2.17a) and (2.17b) are budget equations. Given the budget equations if values of M, P_x and P_y are known, then the values of Q_y and Q_x can be easily calculated. For example, if

$Q_x = O$ then $Q_y = M/P_y$

and if $Q_y = O$, then $Q_x = M/P_x$

Similarly, Q_x and Q_y may be alternatively assigned a positive numerical value and the corresponding values of Q_x and Q_y calculated. When the values of Q_x and Q_y are plotted on X and Y axis, it gives a line with a negative slope, which is called *budget line* or *price line,* as shows in Fig. 2.12.

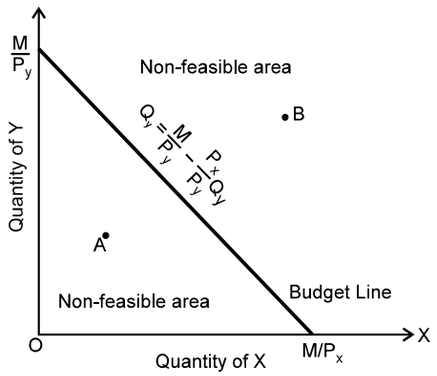

Fig. 2.12 *Budget Line and Budget Space*

An easier method of deriving the *budget line* is to be find the point M/P_y on *Y*-axis (assuming $Q_x = O$) and point M/P_x on *X*-axis (assuming $Q_y = O$). By joining these points by a line, one can obtain the *budget line* as given by the budget equation in Fig. 2.12.

The budget line divides the *commodity space* into two parts which may be termed as (*i*) feasibility area, and (*ii*) non-feasibility area. The area lying in the south-west of the budget line is *feasibility area* (Fig. 2.12). For, any combination of goods *X* and *Y* represented by a point within the area (e.g., point *A*) or on the boundary line (i.e., budget line) is a feasible combination, given *M,* P_x and P_y. The area in the north-east of the budget line is *non-feasibility area* because any point falling in this area, e.g., point *B*, is unattainable (given *M,* P_x and P_y).

Let us now look at the factors that shift budget line, and the slope of the budget line.

(a) Shifts in Budget Line

The budget line changes its position following the change in consumer's income and prices of the commodities. If consumer's income increases, prices of *X* and *Y* remaining the same, budget line shifts upwards remaining parallel to the original budget line. Likewise, income remaining the same, if prices change, the budget line changes its position.

(b) Slope of the Budget Line

The slope of the budget line is of great importance in determining consumer's equilibrium. The slope of the budget line (*AB*) in Fig. 2.13 is given by the following ratios.

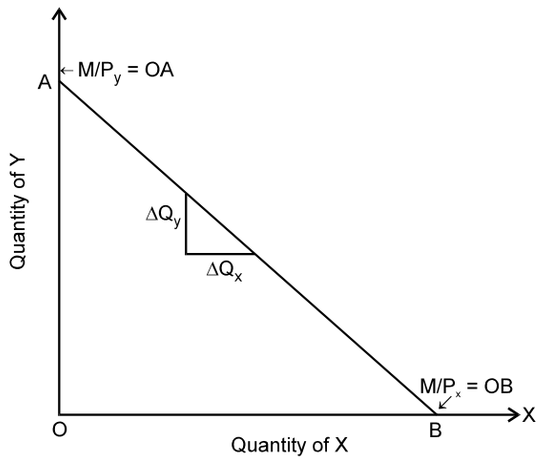

Fig. 2.13 *Slope of the Budget Line*

Since $OA = M/P_y$ and $OB = M/P_x$ (Fig. 2.13) the *slope of the budget line* may be rewritten as

$$\frac{OA}{OB} = \frac{M/P_y}{M/P_x} = \frac{P_y}{P_x} \qquad (2.18)$$

As Eq. (2.18) shows, the slope of the budget line equals the price ratio (P_y/ P_x).

Consumer's Equilibrium

NOTES

As noted above, consumer attains his equilibrium when he maximises his total utility, given his income and market prices of goods and services he consumes. Under indifference curve analysis of consumer behaviour, *necessary condition* for total utility to be maximum is that *MRS* must be equal to the ratio of commodity prices. Considering our earlier two-commodity models, the *necessary* (or the *first order*) condition may be expressed as

$$MRS_{x,y} = \frac{MU_x}{MU_y} = \frac{P_x}{P_y}$$

This is a necessary but not *sufficient condition* of consumer's equilibrium. Another condition, a *second order* or *supplementary condition* is that the necessary condition must be fulfilled at the highest possible indifference curve.

Consumer's equilibrium is illustrated in Fig. 2.14. A hypothetical indifference map of the consumer is shown by indifference curves IC_1, IC_2 and IC_3. The line *AB* is the hypothetical budget line. Both necessary and supplementary conditions of consumer's equilibrium are fulfilled at point *E*, where indifference curve IC_2 is tangent to the budget line, *AB.* Since both, the curve IC_2 and the budget line, *AB,* pass through point *E*, therefore, at this point, the slopes of the indifference curve IC_2 and the budget line (*AB*) are equal. The consumer is therefore in equilibrium at point *E*.

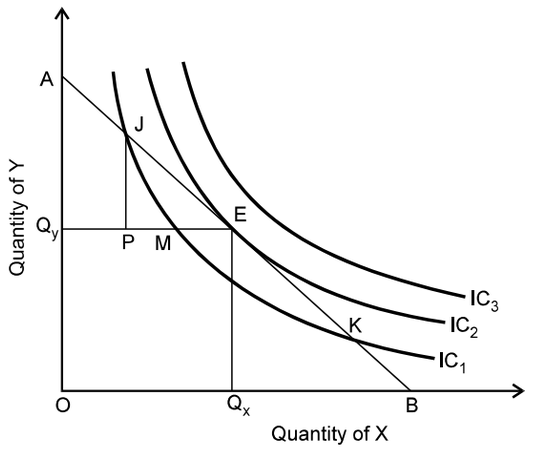

Fig. 2.14 *Equilibrium of the Consumer*

That the consumer is in equilibrium at point *E* can also be proved algebraically. We know that the slope of an indifference curve is given by

$$-\frac{\Delta Y}{\Delta X} = \frac{MU_x}{MU_y} = MRS_{y,x}$$

NOTES

We know also that the slope of the budget line is given by Eq. (2.17) as

$$\frac{OA}{OM} = \frac{P_y}{P_x}$$

At point E, $MRS_{y,x} = P_y/P_x$. Therefore, the consumer is in equilibrium at point E. The tangency of IC_2 with the budget line indicates that IC_2 is the highest possible indifference curve which the consumer can reach, given his budgetary constraint and the prices. At equilibrium point E, the consumer consumes OQ_x of X and OQ_y of Y, which yield him maximum satisfaction.

Although, the necessary condition is satisfied also on two other points, J and K, these points do not satisfy the supplementary or the second order condition of consumer's equilibrium. Indifference curve IC_1 is not the highest possible curve on which the necessary condition is fulfilled. Since, indifference curve IC_1 lies below the curve IC_2, at any point on IC_1, the level of satisfaction is lower than the level of satisfaction indicated by IC_2. So long as the utility maximising consumer has the opportunity to reach the curve IC_2, he would not like to settle on a lower curve.

From the information contained in Fig. 2.14, 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 he moves to point M, he will be equally well-off because points J and M are on the same indifference curve. If he moves from point J to M, he will have to sacrifice JP of Y and take PM of X. But in the market, he can exchange JP of Y for PE of X. That is, he gets extra ME (= $PE - PM$) of X. Since ME gives him extra utility, point E yields a utility higher than the point M. Therefore, point E is preferable to point M. The consumer will therefore have a tendency to move to point E from any point at the curve IC_1, in order to reach the highest possible indifference curve, all other things (taste, preference, and prices of goods) remaining the same.

Another fact which is obvious from Fig. 2.14 is that, due to budget constraint, the consumer cannot move to an indifference curve placed above and to the right of IC_2. For example, his income would be insufficient to buy any combination of two goods at the curve IC_3. Note that IC_3 falls beyond the budget line.

To conclude, a utility maximising consumer, given his income, taste and preferences and prices of goods, will attain his equilibrium when MRS = price ratio at the highest possible indifference curve.

Check Your Progress

9. Define an indifference curve.
10. Mention two properties of an indifference curve.

NOTES

2.8 ANSWERS TO 'CHECK YOUR PROGRESS'

1. A demand schedule is a tabular presentation of quantity demanded of a commodity at different prices per unit of time.
2. A demand curve is a graphical presentation of the demand schedule.
3. The formula for calculating the price elasticity of demand is given as follows:

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

4. When price elasticity of demand is measured between any two finite points on a demand curve, it is called arc elasticity.
5. Market supply is the sum of supply of a commodity made by all individual firms or suppliers.
6. A supply curve is the graphical presentation of the supply schedule.
7. The Marshallian approach to consumer equilibrium is based upon the following assumptions:
   - Rationality
   - Limited money income
   - Maximization of satisfaction
   - Utility is cardinally measurable
   - Diminishing MU
   - Constant utility of money
   - Utility is additive
8. The concept of consumer's surplus is believed to have been originated by a French engineer, Arsene Julis Dupuit in 1844, in his effort to measure social benefit of such collective goods as roads, canals and bridges. The concept was later refined by Marshall who also provided a measure of consumer's surplus.
9. The indifference curve is defined as the locus of points each point representing a different combination of two goods yielding the same utility or level of satisfaction.
10. Two properties of an indifference curve are the following:
    (i) Indifference curves slope downward to the right.
    (ii) Indifference curves do not intersect nor are they constant.

2.9 SUMMARY

NOTES

- For the purpose of demand analysis, a distinction is often made between the individual demand and the market demand—individual demand for analysing consumer behaviour and market demand for analysing market behaviour.
- The law of demand states the relationship between the quantity demanded and the price of a commodity.
- A demand schedule is a tabular presentation of quantity demanded of a commodity at different prices per unit of time.
- According to the law of demand, when a price of a product increases, its demand decreases and vice versa, all other demand determinants remaining constant.
- 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.
- The price elasticity of demand is defined as the degree of responsiveness 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.
- When price elasticity of demand is measured between any two finite points on a demand curve, it is called arc elasticity and elasticity measured at a point on the demand curve is called point elasticity.
- The price elasticity of demand varies from commodity to commodity depending on the nature of the commodity. While the demand for some commodities is highly elastic, for some it is highly inelastic.
- Going by the dynamic demand function, there are some other important demand determinants, viz., (i) price of the related good—substitutes and complements, (ii) income of the consumers, (iii) advertisement of the product, and (iv) future price expectation.
- Cross-Elasticity is the measure of responsiveness of demand for a commodity to the changes in the price of its substitutes and complementary goods. For instance, in case of substitutes, the cross-elasticity of demand for tea (T) is the percentage change in quantity demanded of tea due to a change in the price of its substitute, coffee (C).
- Supply means the quantity of a commodity which its producers or sellers offer for sell at a given price, per unit of time. Market supply, like market demand, is the sum of supply of a commodity made by all individual firms or suppliers.

NOTES

- The law of supply is expressed generally in terms of price–quantity relationship. The law of supply can be stated as follows: The supply of a product increases with the increase in its price and decreases with decrease in its price, other things remaining constant.
- A supply function is a mathematical statement which states the relationship between the quantity supplied of a commodity and its determinants. The short-run market supply function is based on the law of supply.
- The law of diminishing *MU* is the fundamental law on which is based the cardinal utility analysis of the consumer behaviour. This law states that as the quantity consumed of a commodity increases per unit of time, the utility derived by the consumer from the successive units goes on decreasing, provided the consumption of all other goods remains constant.
- The law of diminishing *MU* holds only under certain given conditions. These conditions are often referred to as the assumptions of the law.
- The consumers' willing to pay for a commodity depends on the utility they expect to derive from the commodity. The price which a consumer is willing to pay may not match with the market price of the commodity. It may be greater or less than the market price.
- The Marshallian concept and measurement of consumer's surplus have been criticized on many grounds, though the criticisms are equally questionable.
- The indifference curve is defined as the locus of points each point representing a different combination of two goods yielding the same utility or level of satisfaction.
- Consumer attains his equilibrium when he maximises his total utility, given his income and market prices of goods and services he consumes. Under indifference curve analysis of consumer behaviour, *necessary condition* for total utility to be maximum is that *MRS* must be equal to the ratio of commodity prices.

2.10 KEY TERMS

- **Individual demand:** It refers to the quantity of a commodity that a person is willing to buy at a given price over a specified period of time.
- **Market demand:** It 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.
- **Price elasticity of demand:** It is defined as the degree of responsiveness of demand for a commodity to the change in its price.

- **Cross-elasticity:** It is the measure of responsiveness of demand for a commodity to the changes in the price of its substitutes and complementary goods.
- **Supply:** It implies the quantity of a commodity which its producers or sellers offer for sell at a given price, per unit of time.

2.11 SELF-ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. Briefly mention the factors governing the law of demand.
2. Name the kinds of demand elasticities.
3. Define income elasticity of demand.
4. Write a short note on the application of demand elasticity.
5. Mention the significant characteristics of the indifferent curve.

Long-Answer Questions

1. Discuss the exceptions to the law of demand.
2. Explain the method of measuring arc elasticity of demand.
3. Elaborate the determinants of price elasticity of demand.
4. Explain the supply schedule and supply curve with the help of diagrams.
5. Graphically represent the concept of marginal utility and equi-marginal utility.
6. Critically analyse Marshallian concept and measurement of consumer surplus.
7. Explain consumer's equilibrium under ordinal utility approach.

2.12 FURTHER READING

Dwivedi, D. N. 2016. *Microecomics: Theory and Applications*, 3rd Edition. New Delhi: Vikas Publishing House.

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

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

NOTES

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

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

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

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

UNIT 3 PRODUCTION, COST AND REVENUE

NOTES

Structure

3.0 INTRODUCTION

In economic sense, the term 'production' means an activity by which resources (men, material, time, etc.) are transformed into a more useful commodity or value-added product. Specifically, *production means transforming inputs (labour, machines, raw materials, time and so on) into an output with value added*. This concept of production is, however, limited to only 'manufacturing'.

In this unit, you will study about the law of production, law of variable proportions, isoquant curve, law of returns to scale, concept of total, average and marginal costs and total, marginal and average revenue.

3.1 OBJECTIVES

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

- Explain the law of production
- Discuss the law of variable proportions

NOTES

- List the properties of an isoquant curve
- Elaborate the law of returns to scale
- Define total, average and marginal costs
- Elaborate producer's equilibrium

3.2 PRODUCTION

In economic sense, the *production process* may take a variety of forms other than manufacturing. Transporting a commodity in its original form from one place to another where it can be consumed or used in the process of production is *production*. For example, a sand dealer collects and transfers sand from the river bank to the construction site, and a coal company digs out and transports coal from coal mines to the market place. Similarly, a fisherman catches and transports fish from sea, lake and river to the fish market. These activities too are 'production'. Transporting men and materials from one place to another is a productive activity. For example, roadways, railways and airways produce transport *service*. Storing a commodity for future sale or consumption is also 'production'. Wholesaling, retailing, packaging, assembling are all productive activities. These activities are just as good examples of production as manufacturing. Cultivation is the earliest form of productive activity.

Besides, production process does not necessarily involve physical conversion of inputs into tangible goods. Some kinds of production involve an intangible input to produce an intangible output. For example, in the production of legal, medical, social and consultancy services are both input and output intangible; lawyers, doctors, social workers, consultants, hairdressers, musicians, orchestra players are all engaged in producing intangible goods.

Input and Output

An *input* is any thing—a good or a service—that is used in the process of production. In the words of Baumol, 'An input is simply anything which the firm buys for use in its production or other processes.' An **output** is any good or service that comes out of production process. An output may be **tangible** or **intangible**. For example, bread and butter, clothes, cars and computers are the tangible products and services produced by doctors, lawyers, consultants, teachers and social workers are intangible.

Production process requires a wide variety of inputs, called also as **factors of production**, depending on the nature of product. But economists have classified inputs under following categories.

1. *Land* including area, underground and overground resources;
2. *Labour* including physical and mental effort and skill;
3. *Capital*, machinery, equipments, tools used in production and also factory and office buildings;
4. Raw materials used for producing another good or material;

5. Entrepreneurship including management skill and risk-bearing intention and ability;
6. Technology—technique of production using different combination of labour and capital; and
7. Time—all kind of goods and services require some time for their production.

These inputs are classified as 'fixed' and 'variable' inputs in the analysis of input-output relationship.

Fixed and Variables Inputs

Inputs are classified as (i) *fixed* inputs or fixed *factors,* and (ii) *variable inputs* or *variable factors*. Fixed and variable inputs are defined in economic sense and in technical sense. In *economic sense*, given the production technology, a *fixed input* is one whose supply is limited in the short run and is used in a fixed quantity. In *technical sense*, a fixed factor or input is one that remains fixed (or constant) for a certain level of output, given the technology.

In economic sense, a *variable input* is defined as one which can be used in variable quantity in production of a commodity, e.g., labour and raw material and so on. All the users of such factors can employ a larger quantity in the short run. Technically, a variable input is one that changes with the change in output, given the technology.

It is important to note here that in the long run, all inputs are treated as variable inputs because, in the long run, supply of all the inputs keeps increasing and more of all the inputs can be used to produce a larger output.

Short Run and Long Run

The reference to *time* period involved in production process is another important concept used in production analysis. The two reference periods are *short run* and *long run.* **Short run** refers to a period of time during which the supply and the use of certain inputs (e.g., plant, building, machinery and so on) is fixed. In the short run, therefore, production of a commodity can be increased to a limited quantity by increasing the use of only variable inputs (labour).

It is important to note here that 'short run' and 'long run' are economists' jargon. They do not refer to any fixed time period. While in some industries short run may be a matter of few weeks or few months, in some others (e.g., electricity and power industry, automobiles and so on), it may mean three or more years.

The **long run** refers to a period of time during which the supply of all the inputs keeps increasing, but not long enough to permit a change in technology. That is, in the long run, all the inputs are variable. Therefore, in the long run, a firm can employ more of both the inputs—labour and capital to increase its production.

Economists use another term, i.e., (*very long run*) which refers to a period during which the technology of production is also supposed to change. In the very long-run period, the production function also changes. The technological advances result in a larger output from a given quantity of inputs.

NOTES

3.2.1 Production Function and Introduction to Laws of Production

We know that the quantity produced of a commodity depends on the quantity of inputs used to produce the commodity. It means that there is a relationship between input and output. The input–output relationship can be presented in the form of a schedule, a graph or an algebraic equation. When input–output relationship is expressed in the form of an equation, it is called **production function**. By definition, *production function is a mathematical statement which describes the technological relationship between inputs and output in physical terms*. In fact, a production function presents the quantitative relationships between inputs and output. A production function represents the technology of a firm, of an industry or of the economy as a whole.

Formation of Production Function

In this section, let us see how a complex production function is simplified and the number of inputs in the production function (used as independent variables) is reduced to a manageable number, especially in theoretical analysis or models.

An empirical production function is generally very complex. It includes a wide range of inputs, known also as factors of production, viz., (i) land, (ii) labour, (iii) capital, (iv) raw material, (v) time, and (vi) technology. All these variables enter the actual production function. The long-run production function is generally expressed as

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

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

Economists have, however, reduced the number of variables used in a production function to only two variables, viz., *capital (K)* and *labour (L)*, for the sake of analytical convenience and simplicity in the analysis of input–output relation. Production function is generally expressed as

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

The reasons for ignoring other inputs are following. *Land* and *building* (L_B), as inputs, are constant for the economy as a whole, and hence it does not enter into the aggregate production function. In the case of individual firms, land and building are lumped with 'capital'. In case of 'raw materials', it has been observed that 'this input bears a constant relation to output at all levels of production'. For example, cloth bears a constant relation to the number of ready-made garments. Similarly, for a given size of a house, the quantity of bricks, cement, steel and so on remains constant, irrespective of number of houses constructed. In car manufacturing of a particular brand or size, the quantity of steel, number of the engine, and number of tyres and tubes are fixed per car. This constancy of input–output relations leaves the methods of production unaffected. So is the case, generally, with time. That is why in most production functions, only two inputs—labour and capital—are included.

NOTES

We will illustrate tabular and graphic forms of production function when we move on to explain the laws of production. Here, we illustrate the algebraic form of production function assuming that technology of production remains unchanged over a period of time. The mathematical form of production function is the most commonly used function in production analysis.

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

$$Q = f(K, L) \tag{3.1}$$

where Q = the quantity produced per time unit, K = capital and L = labour.

The production function (Eq. 3.1) implies that quantity produced depends on the quantity of capital, K, and labour, L, employed to produce the commodity. Increasing production will require increasing K and L. Whether the firm can increase both K and L or only L depends on the time period it into account takes for increasing production, i.e., whether production decision is related to a *short run* or a *long run*.

Short-run and Long-run Production Function

By definition, short run refers to the period during which the supply of labour is variable an the supply of capital is *constant or fixed* and long run is the period during which supply of both labour and capital is *variable*. In the short run, therefore, the firm can increase production by increasing labour only since the supply of capital in the short run is fixed. In the long run, however, the firm can employ more of both capital and labour because supply of capital also becomes variable over time. Accordingly, there are two kinds of production functions.

1. Short-run production function, and
2. Long-run production function.

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

$$Q = f(\bar{K}, L) \tag{3.2}$$

where $\underline{K}$ denotes constant K.

In the **long-run production function** both K and L are treated as variable factors and the function takes the following form.

$$Q = f(K, L) \tag{3.3}$$

In Eq. (3.3), both capital (K) and labour (L) are treated as variable factors.

Assumptions A production function is formed on the basis of the following assumptions.

1. Both inputs and output are perfectly divisible;
2. There are only two factors of production—labour (L) and capital (K);
3. There is limited substitution of one factor for the other, because labour and capital are imperfect substitutes;

NOTES

4. Technology of production is given, and
5. Supply of fixed factors is constant or fixed in the short run.

These are the general assumptions on the basis of which a production function is constructed. However, if there is a change in these assumptions, the production function will have to be modified accordingly.

Introduction to Laws of Production

Having introduced the concept of production function, we now proceed to discuss the theory of production by using the production function. The traditional theory of production has been formulated under two conditions: (i) short-run conditions, and (ii) long-run conditions. Under short-run conditions, only labour is assumed to be a variable factor, all other factors assumed to remain constant. Under long-run conditions, both labour and capital are treated as the variable factors. Accordingly, there are two kinds of laws of production.

1. Short-run laws of production, and
2. Long-run laws of production.

The laws of production under short-run conditions are called '*the laws of variable proportions*', the '*laws of returns to a variable input*' and the '*law of diminishing marginal returns*'. The long-run input–output relations are studied under the '*laws of returns to scale*'.

Check Your Progress

1. Define production function.
2. State the formula which denotes the single-variable production function.

3.3 SHORT-RUN LAWS OF PRODUCTION: LAW OF VARIABLE PROPORTIONS

The laws of returns to variable proportions or variable input can be stated as when more and more units of a variable input (labour) are used, given the quantity of fixed inputs, the total output increases initially at an increasing rate and then at a constant rate but it will eventually increase at diminishing rates. The ultimate law is that the marginal increase in total output eventually decreases when additional units of a variable factor are applied to a given quantity of fixed factors. In the words of Hirshleifer, "If one factor (or group of factors) is increased while another factor (or group of factors) is held fixed, output or total product q will first tend to rise. But, eventually at least, a point will be reached where the rate of increase, the Marginal Product [$MP_L = Q/L$] associated with increments of the variable factor, begins to fall; this is the point of diminishing marginal returns." Baumol states the law of diminishing returns in similar terms: "As more and more of some input, i, is employed, all

NOTES

other input quantities being held constant, eventually a point will be reached where additional quantities of input *i* will yield diminishing marginal contributions to total output.' As such, there are three laws of returns to variable inputs (i) the law of increasing returns, (ii) the law of constant returns, and (iii) the law of diminishing returns. Before we discuss these laws, let us take note of assumptions under which these laws are formulated.

Assumptions: The laws of returns to variable input are based on the following assumptions.

1. The state of technology is given;
2. Labour is homogenous; and
3. Capital remains constant.

To illustrate the law of diminishing returns with the help of an example, let us assume that (i) a firm has a set of machinery as its capital (K), fixed in the short run, and (ii) it employs more of workers to increase its production. Thus, the short-run production function for the firm will take the following form.

$$Q = f(L, \overline{K})$$

where Q_c = quantity produced, L = labour; and $\overline{K}$ = capital (held constant).

Let us suppose that the labour–output relationship when estimated empirically reveals a **cubic production function** of the following form.

$$Q = -L^3 + \beta L^2 + \alpha L$$

When estimated with actual data, it takes the following form.

$$Q = -L^3 + 15L^2 + 10L \qquad (3.4)$$

Given the production function (3.4), the quantity (Q) that can be produced with different number of workers can be easily worked out by assigning a numerical value to the variable factor (L). For example, if $L = 5$, Q can be worked out as follows:

$$Q = -53 + 15 \times 52 + 10 \times 5$$
$$= -125 + 375 + 50$$
$$= 300$$

Similarly, by assigning different numbers to L, a series of labour and output can be generated. A tabular array of output levels associated with different number of workers from 1 to 12, in our hypothetical example is given in Table 3.1 (Columns 1 and 2).

To understand the laws of returns to the variable input labour (L), what we need now is to work out *marginal productivity of labour* (MP_L) to find the trend in the contribution of the marginal labour and *average productivity of labour* (AP_L) to find the average contribution of labour. The MP_L and AP_L can be worked out from the data given in Table 3.1. The process of working out MP_L and AP_L.

NOTES

Table 3.1 *Total, Marginal and Average Products*

Number of Workers (*L*)	Total Product (TP_L) (tonnes)	Marginal Product* (MP_L)	Average Product** (AP_L)	Stage of Returns
(1)	(2)	(3)	(4)	(5)
1	24	24	24	Stage I increasing returns
2	72	48	36	
3	138	66	46	
4	216	78	54	
5	300	84	60	
6	384	84	64	Stage II diminishing returns
7	462	78	66	
8	528	66	66	
9	576	48	64	
10	600	24	60	
11	594	- 6	54	Stage III negative returns
12	552	–42	46	

*$MP_L = TP_n - TP_{n-1}$. *Note that* MP_L *obtained by differential method will be different.*
**$AP_L = TP_L IQ$

3.3.1 Marginal Productivity of Labour

Marginal product of a factor is defined as the increase in output due to marginal increase in the factor. *Marginal productivity of labour* (MP_L) can be obtained by differentiating the production function (3.4) with respect to labour (L) as follows:

$$MP_L = \frac{\delta Q}{\delta L} = -3L^2 + 30L + 10 \tag{3.5}$$

By substituting numerical value for labour (L) in Eq. (3.5), MP_L can be obtained at different levels of labour employment. However, this method can be used only where labour is perfectly divisible and $\partial L \to 0$. Since, in our example, each unit of $L = 1$, calculus method cannot be used. Alternatively, where labour can be increased at least by one unit, MP_L can be obtained as

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

The MP_L worked out by this method is presented in column 3 of Table 3.1.

Average Productivity of Labour

Average productivity of labour (AP_L) can be obtained by dividing production function by L.

$$AP_L = \frac{Q}{L} = \frac{-L^3 + 15L^2 + 10L}{L} = -L^2 + 15L + 10 \tag{3.6}$$

Now AP_L can be obtained by substituting numerical values for L in Eq. (3.6). AP_L obtained by this method is given in column 4 of Table 3.1.

3.3.2 The Three Stages in the Law of Diminishing Returns

Table 3.1 and Figure 3.1 present the three general stages in the application of the law of diminishing returns. In *Stage I*, TP_L increases at increasing rate. This is indicated by

NOTES

the rising MP_L until the employment of the 5th worker. Given the production function (3.4), the 5th and 6th workers represent an intermediate stage of constant returns to the variable factor, labour. In Stage II, TP_L continues to increase but at diminishing rates, i.e., MP_L begins to decline. This stage in production shows the law of diminishing returns to the variable factor. Total output reaches its maximum level at the employment of the 10th worker. Beyond this level of labour employment, TP_L begins to decline. This marks the beginning of *Stage III* in production.

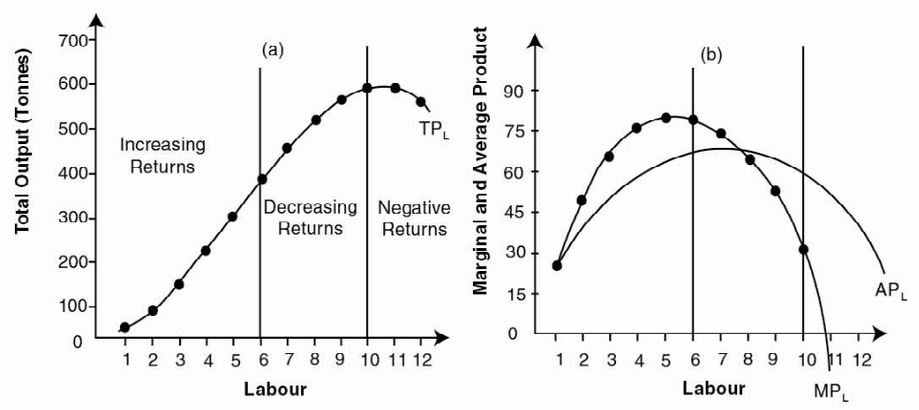

Fig. 3.1 *Total, Average and Marginal Products*

The laws of variable proportions can be illustrated graphically also. The labour-output data given in Table 3.1 is presented graphically in panels (a) and (b) of Figure 3.1. Panel (a) of Figure 3.1 presents the total product curve (TP_L) and panel (b) presents the marginal product (MP_L) and average product (AP_L) curves. The TP_L schedule demonstrates the law of diminishing returns. As the curve TP_L shows, the total output increases at an increasing rate until the employment of the 6th worker, as indicated by the increasing slope of the TP_L curve. Beyond the 6th worker, TP_L increases (until the 10th worker) but the rate of increase in TP_L begins to fall and turns negative from 11th worker onwards. The TP_L curve shows the operation of the law of diminishing returns.

The MP_L and AP_L indicated by the trend by the TP_L curve are presented in panel (b) of Figure 3.1. The MP_L tends to increase in the stage of 'increasing returns' and it tends to decline in the stage of decreasing returns to variable input–labour.

To conclude, the law of diminishing returns can be stated as follows. Given the fixed factor (capital), when more and more labour is employed, the return from the additional labour, i.e., MP_L, may initially increase but does eventually decrease.

Factors Behind the Laws of Returns to Variable Input

The questions that we seek to answer here are: (i) what factors cause increasing returns to labour? and (ii) what factors cause decreasing returns to labour? The factors behind the increasing returns to the variable input—labour—can be described as follows.

NOTES

Factors Behind Increasing Returns: *One* of the important factors causing increasing returns to a variable factor is the **indivisibility of fixed factor** (capital). Each capital unit requires an optimum number of labour. If labour is less than the optimum number, it results in underutilization of capital and therefore lower productivity of labour. Let us suppose that optimum capital–labour combination is 1:6. If capital is indivisible and less than six workers are employed, then capital remains underutilized. When more and more workers are added, utilization of capital increases and also the productivity of additional worker. *Another* factor for the increase in labour productivity is that employment of additional workers gives the advantages of **division of labour**, until optimum capital–labour combination is reached.

Factors Behind Decreasing Returns: Once the optimum capital–labour ratio is reached, employment of additional workers amounts to *substitution of capital with labour*. But, technically, one factor can substitute another only to a limited extent. In other words, there is a limit to which one input can be substituted for another for the same marginal output. Hence, to replace the same amount of capital, more and more workers will have to be employed because per worker marginal productivity decreases. Besides, increase in labour beyond optimum combination of inputs, results in underutilization of labour. Consequently, MP_L of labour decreases.

Applicability of the Law of Diminishing Returns

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

3.4 ISO PRODUCT OR ISOQUANT CURVE

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

An isoquant 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 *isoquant*

NOTES

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 isoquant represents a measurable quantity of output of a product.

The probability of a given quantity of a commodity to 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 isoquant curve.

3.4.1 Derivation of Iso Product Curve

Assumptions: Isoquant 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) are imperfect substitute and can be substituted for one another though at a diminishing rate;
3. Labour and capital can be substituted for one another only upto a certain limit;
4. Production function is continuous, implying that labour and capital are perfectly divisible and can be substituted in any small quantity.

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 isoquant curve ($Q_x = 100$) in Figure 3.2. The curve Q_1 all along its length represents a fixed quantity – 100 units of product X. This quantity of output can be produced with a number of labour–capital combinations. For example, points A, B, C and D on the isoquant Q_x represent four different combinations of inputs, K and L, as given in Table 3.2, all yielding the same output—100 units. Note that movement from A to D indicates decreasing capital (K) and increasing labour (L). This implies substitution of labour for capital such that all the input combinations yield the same quantity of commodity X, i.e., $Q_x = 100$, whatever the combination of labour and capital. Similarly, if labour decreases and capital increases, output remains the same.

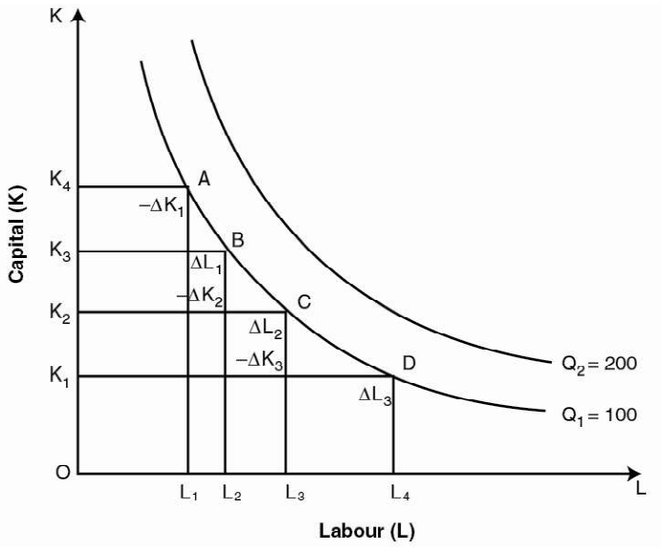

Fig. 3.2 *Isoquant Curves*

Table 3.2 *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

3.4.2 Characteristics of Iso Product Curve

Isoquants 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. **Isoquants have a negative slope:** An isoquant has a negative slope in the *economic region*. In other words, isoquants slope downward from left to right. The economic region is the region on the isoquant plane in which substitution between inputs is technically efficient. It is also known as the *product maximizing region* with different combinations of labour and capital. The negative slope of the isoquant 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 so increased that the total output remains remains the same. For example, movement from *A* to *B* on Q_x (Figure 3.2) means that if K_4K_3 units of capital are removed from the production process, L_1L_2 units of labour have to be added to maintain the same level of output. The substitution of one input for another gives isoquant a negative slope.

2. **Isoquants are convex to the origin:** Convexity of isoquant means that it has a bend towards the point of origin. Isoquants are convex to origin because the rate at which one input can be 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** (*MRTS*). The *MRTS* is defined as:

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

In plain words, *MRTS* is the rate at which labour can substitute capital at margin, or other way round, without affecting the total output. This rate is indicated by the slope of the isoquant. The *MRTS* decreases for two reasons.

(i) No factor is a perfect substitute for another, and

(ii) Productivity of inputs is subject to the law of 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 is the reason for the diminishing marginal rate of substitution. That *MRTS* goes on diminishing along the isoquant can be proved by deriving the *MRTS* from $Q_1 = 100$ in Figure 3.2. Suppose that in Figure 3.2, $K_4K_3 = K_3K_2 = K_2K_1$, it means that $K_1 = K_2 = K_3$. But in case of labour, as the figure shows, $L_1L_2 < L_2L_3 < L_3L_4$, i.e., $\Delta L_1 < \Delta L_2 < \Delta L_3$.

Given these conditions, let us workout the $MRTS = K/L$ at different levels of input combination.

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

This shows that *MRTS* goes on decreasing.

3. **Isoquants do not intersect nor are tangent to each other:** The intersection of or tangency between any two isoquants implies two inconsistent production possibilities.

(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.

These conditions contradict the laws of production unless marginal productivity of inputs is zero or less than zero. In Figure 3.3, two isoquants 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 isoquants ($Q_1 = 100$ and $Q_2 = 200$), it means that the same combination of inputs can produce 100 units and also 200 units of the commodity. This is practically impossible.

NOTES

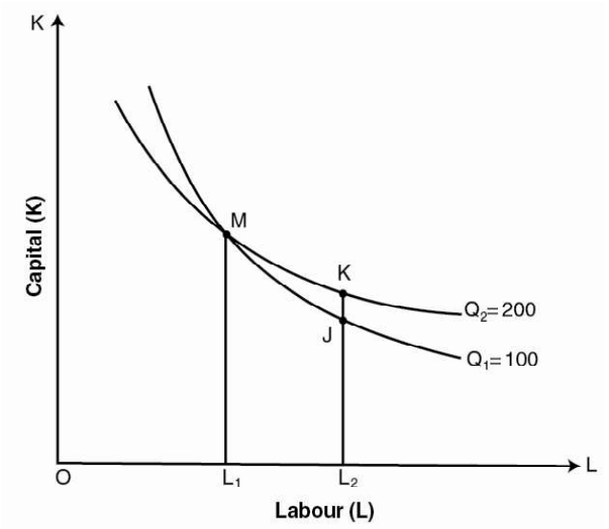

Fig. 3.3 *Intersecting Isoquants*

To prove inconsistency, consider two other points—point *J* on isoquant Q_1 = 100 and point *K* on isoquant 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 isoquant Q_1 = 100, factor combinations at point *M* and *J* are equal in terms of their output. On the isoquant Q_2 = 200, factor combinations at *M* and *K* are equal in terms of their output. Since point *M* is common to both the isoquants, it follows that input combinations at *J* and *K* are equal in terms of output. This implies that, in terms of output, $OL_2 + JL_2 = OL_2 + KL_2$. 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 is not the fact because, as can be seen in Figure 3.3, $JL_2 < KL_2$.

However, the intersection of the isoquants 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 isoquant will not intersect nor be tangent to one another. If they do, it violates the law of production.

4. **Upper isoquants represent a higher level of output:** Between any two isoquants, the upper one represents a higher level of output than the lower one. The reason is that any point on an upper isoquant implies a larger input combination, which, in general, produces a larger output. Therefore, upper isoquants indicate a higher level of output.

For instance, isoquant Q_2 in Figure 3.4 will always represent a higher level of output than isoquant Q_1. For, any point at isoquant Q_2 consists of more of either capital or labour or both. For example, consider point *a* on isoquant Q_1 and compare it with any point at isoquant Q_2. Point *b* on isoquant Q_2 indicates more of capital (*ab*) given the labour at *OL*, and point *d* shows more of labour (*ad*) given the capital *OK*, and point *c* shows more of both labour and capital. Therefore, isoquant Q_2 represents a higher level of output (200 units).

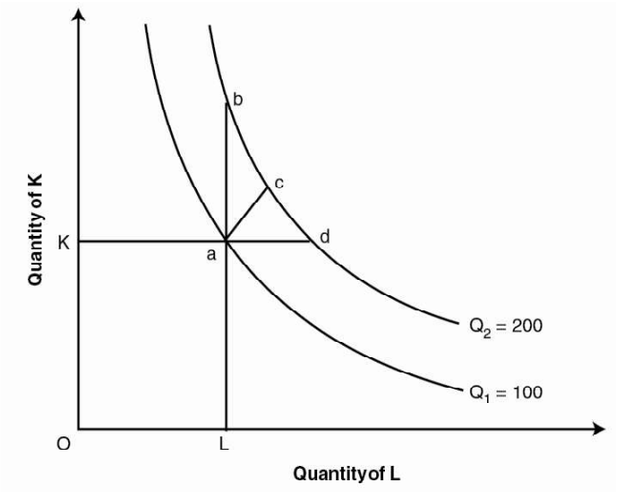

Fig. 3.4 *Comparison of Out at Two Isoquants*

Check Your Progress

3. Mention the assumptions which govern the laws of return to variable input.
4. What is an isoquant?

3.5 LONG-RUN LAWS OF PRODUCTION: THE LAWS OF RETURNS TO SCALE

In this unit so far, we have discussed the derivation and the properties of isoquants—the tool of analyzing the input–output relationship under increasing scale of production. In this section, we discuss the laws of returns to scale, i.e., input–output relationships under the condition that both the inputs (labour and capital) are variable and their quantity increases *proportionately and simultaneously*. When both the inputs (labour and capital) are increased proportionately, the scale of production, i.e., the size of the firm, increases. The laws of production, i.e. the input–output relationships under the condition of increasing scale of production, are called the *laws of returns to scale*. As mentioned earlier, *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 use more of both labour and capital to increase their production. The question that we will answer here is: how does total output change when both the inputs are increased proportionately and simultaneously? The answer to this question lies in what law of returns 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 inputs,
2. Output may increase proportionately to increase in inputs, and
3. Output may increase less than proportionately to increase in inputs.

NOTES

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.

These three law of returns to scale are explained below first graphically by applying the isoquants and then through the *production function*.

3.5.1 The Law of Increasing Returns to Scale and Economies of Scale

When both the inputs—labour and capital—are increased proportionately and simultaneously and output increases more than proportionately increase in inputs, 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 3.5. 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 3.5 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.

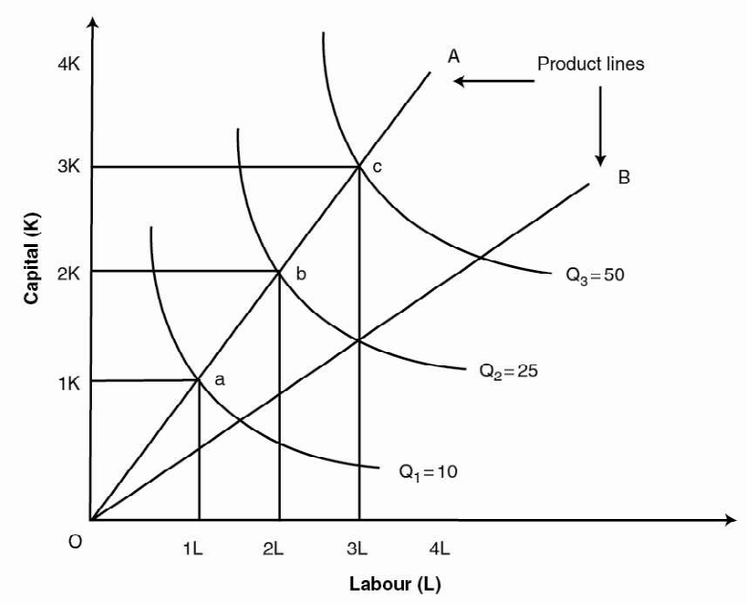

Fig. 3.5 *Increasing Returns to Scale*

NOTES

Similarly, movement from point *b* to *c* indicates 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 kind of input-ouput relationship reveals that the *law of increasing returns to scale is in operation*.

Factors Behind the 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.

1. **Technical economies:** Certain inputs, particularly machinery, used in the process of production are available in a given size. Such inputs are indivisible. That is, capital—the mechanical device—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. All these technical devices have to be acquired in full size. If planned production is low, the machinary will remain underutilized causing low return. When scale of production is enhanced, more labour will be employed ensuring full utilization of machinary. As a result, production increases at increasing rate.
2. **Managerial economies:** Managerial economies arise due to hiring managers of different specialization and efficiency. In case planned production is of small scale, only a few mangers have to be employed. Some of them may be assigned functions outside the area of their specialization. In that case, managerial productivity is found to the low. For instance, a production manager cannot manage sales very efficiently. However, when the level of production is enhanced, managerial manpower is employed according to their specialization. As a result, production increases at an increasing rate.
3. **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 managerial manpower increases productivity per unit of inputs. Their cumulative effects contribute to the increasing returns to scale. Managerial specialization contributes a great deal to increasing production.
4. ***Dimensional* economies:** Increasing returns to scale is also a matter of dimensional relations. For example, when the size of a room ($15' \times 10' = 150$ sq. ft.) is doubled to $30' \times 20'$, the area of the room is more than doubled, i.e., $30' \times 20' = 600$ 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.

NOTES

All these factors make increasing returns to scale.

3.5.2 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 3.6. The lines *OA* and *OB* represent two expansion path lines indicating two hypothetical techniques of production. The isoquants, $Q_1 = 10$, $Q_2 = 20$ 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 $1K$ to $2K$ and labour increases from $1L$ to $2L$. 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 $2K$ to $3K$ and L from $2L$ to $3L$. 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.

$$1K + 1L = Q = 10$$
$$2K + 2L = 2Q = 20$$
$$3K + 3L = 3Q = 30$$

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

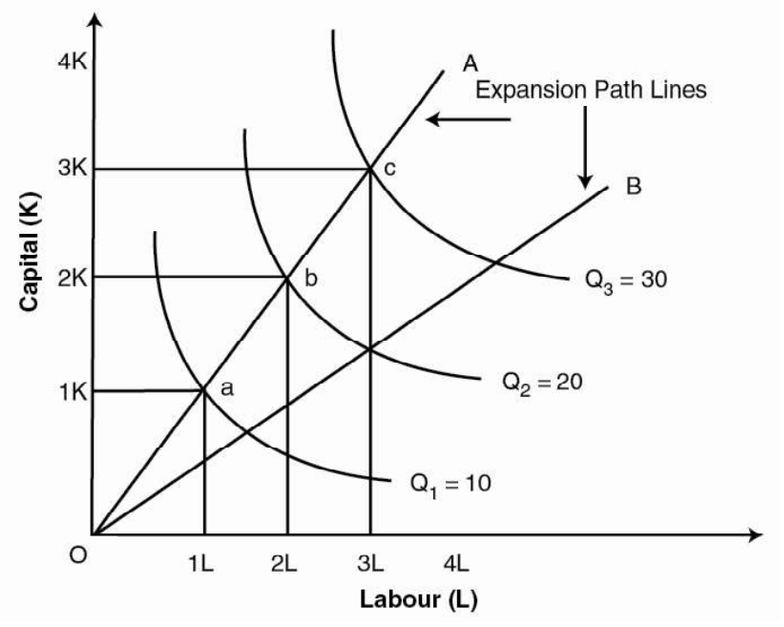

Fig. 3.6 *Constant Returns to Scale*

The Reason for Constant Returns to Scale? The constant returns to scale are attributed to the *limits of the economies of scale*. With the expansion of the scale of production, economies of scale arise from such factors as indivisibility of certain inputs, greater possibility of specialization of capital and labour, use of

NOTES

labour-saving techniques of production and so on. But, there is a limit to the economies of scale. When economies of scale reach their limit 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 one like Cobb–Douglas production function.

3.5.3 The Law of Decreasing Returns to Scale

When inputs, *K* and *L*, are increased and output increases less than proportionately and the rate of rise in output goes on decreasing it reveals the law of **decreasing returns to scale**. Decreasing returns to scale are illustrated in Figure 3.7. 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 (100 per cent) 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.

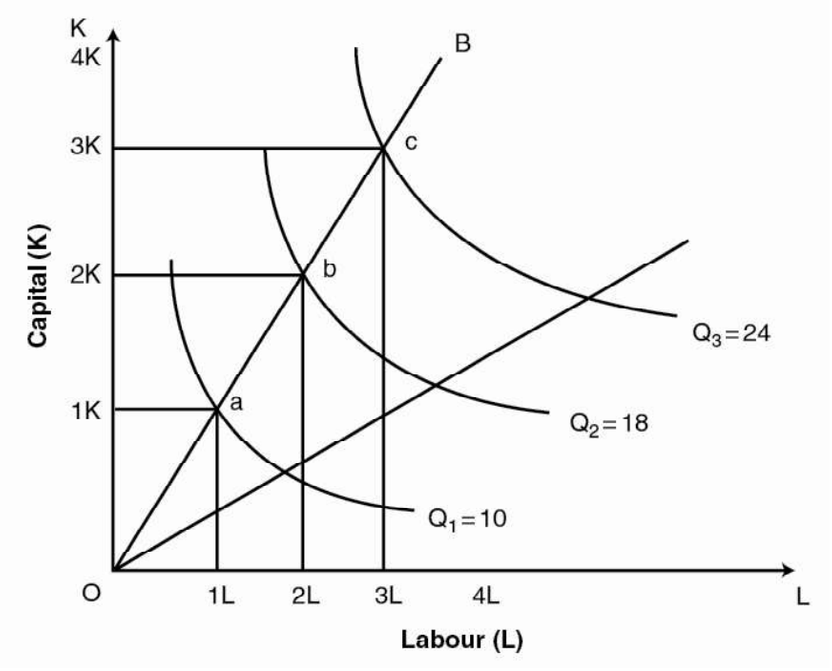

Fig. 3.7 *Decreasing Returns to Scale*

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., the **managerial diseconomies**. As the size of the firm expands, managerial efficiency decreases causing decrease in the rate of increase in output.

NOTES

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.

3.6 CONCEPT OF COST AND REVENUE: TOTAL, MARGINAL AND AVERAGE

In this section, you will learn about the total, marginal and average costs and revenues. Let's begin the discussion with concept of costs.

Concept of Costs

Total cost represents the value of the total resources used in the production of goods and services. It refers to the total outlays of money expenditure, both explicit and implicit on the resources used to produce a given output. For theoretical purpose, total cost includes payments for labour, capital, land and opportunity cost. The total cost for a given output is obtained from the cost function.

Average cost is of statistical nature. It is obtained simply by dividing the total cost (TC) by the total output (Q), i.e., TC/Q = average cost.

Marginal cost is the addition to the total cost on account of producing one additional unit of the product. Or, marginal cost is the cost of marginal unit produced. Marginal cost (MC) is also defined as $\Delta TC/\Delta Q$.

Total, average and marginal cost concepts used in the economic analysis of the firm's productive activities shall be discussed in detail in the following section.

Fixed and Variable Costs

Fixed costs are those which are fixed in volume for a certain given output. Fixed costs do not vary with the variation in the output between zero and a certain level of output. The costs that do not vary over a certain level of output are known as fixed cost. Fixed costs include cost of (i) managerial and administrative staff; (ii) depreciation of machinery, building and other fixed assets; and (iii) maintenance of land, etc. The *concept of fixed is associated with short-run*.

Variable costs are those which vary with the variation in the total output. Variable costs are the function of the output. Variable costs include cost of raw materials, running cost of fixed capital, such as fuel, ordinary repairs, routine maintenance expenditure, direct labour charges associated with the level of output, and the costs of all other inputs that vary with output.

These cost concepts are economic in nature and are used in economic analysis of costs behaviour in relation to output.

NOTES

Short-run and Long-run Costs

Two other important cost concepts which are analogous with variable and fixed costs and often figure in economic analysis are **short-run** and **long-run costs.** Short-run costs can be defined as the costs which vary with the variation of output, the size of the firm remaining the same. In other words, short-run costs are the same as variable costs. Long-run costs, on the other hand, can be defined as the costs which are incurred on the fixed assets, like plant, building machinery, land, etc. Such costs have long-run implication in the sense that these costs are not used up in the single batch of production, and are used over time in the process of production. Long-run costs are, by implication, the same as fixed costs. In the long-run, however, even the fixed costs become variable costs as the size of the firm or scale of production increases. Broadly speaking, the short-run costs are those associated with variable costs in the utilisation of fixed plant or other facilities, whereas long-run cost-behaviour encompasses changes in the size and kind of plant.

Short-Run Cost-Output Relations

In this section, we discuss the relationship between output and costs and the behaviour of cost in relation to the change in output. Cost-output relations are expressed through a *cost function.* Therefore, before we proceed, let us have a look at the cost function and 'how cost function is constructed'.

Cost functions depend on (i) production function; and (ii) market-supply function of inputs. Production function specifies the technical input-combination and its relation to the output. Production func-tion of a firm combined with the supply function of inputs or prices of inputs determines the cost function of the firm. Thus, cost function is a function derived from the production function and the market supply function.

Cost-output relations depend on the nature of cost function. Change in cost function causes a change in the cost-output relations. Since cost function is dependent of the production function, it may change due to change in the latter. Since a production function can take different forms depending on what variables are held constant, cost functions can also take different forms. Whether certain variables in the function can be held constant or not depends on whether short-run or long-run is considered for constructing the production function. Accordingly, there are two kinds of cost functions: (a) *short-run cost function*, and (b) *long-run cost-function.*

The cost-function may be symbolically written as

$$C = f(Q, T; P_f, K)$$

where C = total cost

Q = quantity produced

T = technology

P_f = factor price

K = capital, the fixed factor

NOTES

Since, in the *short-run,* all determinants of cost other than Q are constant, the *short-run cost- function* may be specified as

$$C = f(Q)$$

Let us now explain the cost-output relations in the short-run, known as the *traditional theory of cost of production.*

Relationship between Cost and Output in Short Run

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

$$TC = TFC + TVC$$

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

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

(i) $$AC = \frac{TC}{Q}$$

(ii) $$AFC = \frac{TFC}{Q}$$

(iii) $$AVC = \frac{TVC}{Q}$$

Since $$TC = TFC + TVC,$$

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

$$= AFC + AVC$$

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

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

Since $\Delta TC = \Delta TFC + \Delta TVC$ and in short run $\Delta TFC = 0$,

$\Delta TC = 0 + \Delta TVC$. Thus, in short run,

$$MC = \frac{\Delta TVC}{\Delta Q}$$

NOTES

The Relationship between *AC* and *AVC*

(a) Since $AC = AFC + AVC$, *AC* falls when *AFC* and *AVC* fall.

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

(i) if decrease in *AFC* > increase in *AVC, AC* falls;

(ii) if decrease in *AFC* = increase in *AVC, AC* remains constant; and

(iii) if decrease in *AFC* < increase in *AVC, AC* increases.

The Relationship between *MC* and *AC*

Much more important is the relationship between *AC* and *MC.* It may be described as follows:

(i) When *MC* falls, *AC* follows. But the rate of fall in *MC* is greater than that of *AC* because, in case of *MC,* the decreasing marginal cost is attributed to a single marginal unit while, in case of *AC,* the decreasing marginal cost is distributed over the whole output. So long as *MC* curve lies below the *AC* curve, *MC* pulls *AC* downwards and when *MC* is above *AC,* it pulls the latter upwards.

(ii) Similarly, when *MC* increases, *AC* also increases but at a lower rate for the reason given above. There is however an exceptional range of output where this relationship does not exist—*MC* increases while *AC* continues to fall. This range of output lies between 20 and 40 units. Over this range of output, *MC* starts increasing while *AC* continues to decrease. The reason is when *MC* falls, it falls at a rate higher than the rate of fall in the *AC.* When *MC* starts increasing, it does so at a relatively lower rate which is not sufficient to push the *AC* up. That is why *AC* continues to fall over some range of output even if *MC* falls.

(iii) *MC* intersects *AC* at its minimum. The reason is, when *MC* decreases it pulls *AC* down, and when *MC* increases, it pushes *AC* up. And when *AC* is at its minimum, it is neither being pulled down nor being pushed up, by the *MC.* It follows that *MC* = *AC* at its minimum.

Long-Run Cost-Output Relations

From cost theory point of view, long run is the period during which all the inputs become variable. In the short run, at least one input (especially capital) remains fixed. The variability of inputs is based on the assumption that, in the long-run, supply of all the inputs, including those held constant in the short-run, becomes elastic. The firms are, therefore, in a position to expand the scale of their production in the long-run by hiring or purchasing larger quantities of all the inputs. The long-run cost-output relations therefore imply the relationship between the total costs and the total output, whereas in the short-run this relationship is essentially one between the total output and the variable costs.

NOTES

To understand the long-run cost-output relations and to derive long-run cost curves it is helpful to imagine that, operationally, a long-run is composed of a series of short-run production decisions. As a corollary of this, long-run cost curves would be composed of a series of short-run cost curves.

Concept of Revenue

Revenue is the income a firm retains from selling its products once it has paid indirect tax. Revenue provides the income which a firm needs to enable it to cover its costs of production, and from which it can derive a profit. Profit can be distributed to the owners, or shareholders, or retained in the business to purchase new capital assets or upgrade the firm's technology.

Revenue is measured in three ways:

- **Total revenue:** Total revenue (TR), is the total flow of income to a firm from selling a given quantity of output at a given price, less tax going to the government. The value of TR is found by multiplying price of the product by the quantity sold.
- **Average revenue:** Average revenue (AR), is revenue *per unit*, and is found by dividing TR by the quantity sold, Q. AR is equivalent to the price of the product, where P x Q/Q = P, hence AR is also *price*.
- **Marginal revenue:** Marginal revenue (MR) is the revenue generated from selling one extra unit of a good or service. It can be found by finding the change in TR following an increase in output of one unit. MR can be both positive and negative.

Revenue Function

Revenue is equal to the number of units sold times the price per unit. To obtain the revenue function R(x), multiply the output level by the price function P(x).

$R(x) = X- P(x)$

Relationship between Total Revenue Profit and Total Costs

Total revenue is the income a business receives from the sale of all the goods produced. Total profit is determined by subtracting total costs from revenues. Total revenue is determined by multiplying the price received for each unit sold by the number of units sold. Total revenue profits are a product of subtracting total costs from total revenue.

Price elasticity measures consumer responsiveness in relationship to quantity demanded and price per unit purchased. If producers can increase total revenue by lowering price, demand is considered elastic. If producers can increase total revenue by increasing price, demand is considered inelastic. Businesses receive maximum total revenue at the point when the greatest number of units can be sold for the highest possible price. Economists plot demand and price data on a graph to determine at what point price and demand will yield the highest total revenues.

NOTES

Profit and Revenue

In classical economics, it is assumed that firms will seek to maximise their profits. This occurs when the difference between TR (total revenue) – TC (total cost) is the greatest. Profit maximisation will also occur at an output where MR (marginal revenue) = MC (marginal cost). When MR> MC, the firm is increasing its total profit. When MR< MC total profit starts to fall. Therefore, profit is maximised where MR = MC.

Check Your Progress

5. Name the kinds of laws of returns to scale.
6. What are TFC and TVC?

3.7 PRODUCER'S EQUILIBRIUM

In general sense, the term *equilibrium* means a state of conditions under which opposite forces are in balance. In case of firm's equilibrium, there are two opposite factors—cost and revenue. A firm is said to be in equilibrium at an output at which its cost and revenue are so in balance that its objective is maximized. This is known as producer's equilibrium. A profit maximizing firm attains its equilibrium at the output at which the difference between its total revenue (*TR*) and total cost *(TC)* is maximized. In traditional theory of firm, there are two approaches to analyze the equilibrium of the firm.

1. Total Revenue (*TR*) and Total Cost *(TC)* approach, and
2. Marginal Revenue *(MR)* and Marginal Cost *(MC)* approach.

The determination of equilibrium of a profit maximizing firm has been explained under the two approaches under the following assumptions.

Assumptions

The short-run equilibrium of the firm is analysed under the following conditions.

1. Capital of the firm is fixed and capital cost is also fixed;
2. Labour is the only variable factor and labour cost is the only variable cost;
3. Input prices, especially price of labour, are constant;
4. Product price is constant and *TR* of the firm increases at a constant rate; and
5. Total cost *(TC)* of the firm increasing initially at decreasing rate and at increasing rate beyond a level of output and firms face a U-shaped average cost *(AC)* and marginal cost *(MC)* curves.

Producer's equilibrium can be achieved under two different conditions: when the prices are constant (under Perfect competition) and when prices fall with rise in output (under Imperfect Competition). In this section, you will only learn about

NOTES

the condition when the price is constant, i.e., perfect competition. The conditions of imperfect competition and equilibrium will be taken up in Unit 4.

Given these assumptions, the short-run equilibrium of the firm under perfect competition has been discussed here under the two approached.

Short-Run Equilibrium of the Firm: *TR-TC* Approach

The short-run equilibrium of the firm is illustrated here graphically in Figure 3.8.

The trend in short-run total cost (*STC)* and total revenue *(STR)* of an individual firm are shown by he *STC* and *STR* curves, respectively, in Figure 3.8. As shown in Figure 3.8, with increase in output, *TC* increases initially at decreasing rate and at increasing rate beyond output OQ_1. The trend in total revenue (*TR*) of the firm is shown by straight line *STR.* It means that *STR* increases at constant rate because price is assumed to remain constant.

As Figure 3.8 show, *STC* and *STR* curves intersect at two points, *A* and *B.* These points mark the two ***break-even points***—the points at which firm's cost of production is equal to its total revenue, i.e., *STC=STR.* A perpendicular drawn from point *A* to the output axis shows output OQ_1 and a perpendicular drawn from point *B* to the output axis shows output OQ_3. As the figure shows, at output *less* than OQ_1 and at output more than OQ_3 firm's *STC >STR.* It means that the firm makes losses if its output is less than OQ_1 and more than OQ_3. It means that firm's profitable range of output lies between OQ_1 and OQ_3 as over this range of output firm's *STR >STC.* However, amount of profit varies over this range of profitable output. The profit at different levels of output can be measured by measuring the difference between *STR* and *STC* at different levels of output. This is shown by the total profit curve Π.

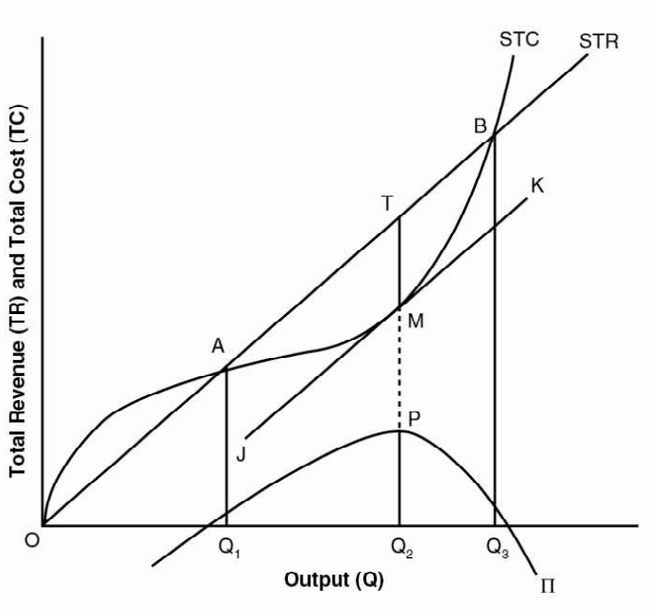

Fig. 3.8 *Short-Run Equilibrium of the Firm by TR-TC Approach*

NOTES

Our objective is now to find the profit maximizing level of output as it will determine the short-run equilibrium of the firm. The profit maximizing output can be obtained by drawing a tangent to *STC* curve and parallel to the *STR* line, as indicated by line *JK*. The line *JK* is tangent to the *STC* curve at point *M*. It gives the maximum profit at *TM* at output OQ_2. The total profit trend is indicated by the *total profit* curve Π. Thus, the profit of the firm is maximized at PQ_2 at the output level of OQ_2. This is how the short-run equilibrium of a firm is determined by *TC-TR approach.*

Short-Run Equilibrium of the Firm: *MR-MC* Approach

The short-run equilibrium of a firm under perfect competition has been discussed here under the *MR-MC* approach. Under *MR-MC* approach, a profit maximizing firm is said to be in equilibrium at the level of output at which its marginal revenue (*MR*) is equal to its marginal cost (*MC*). In other words, given the price and cost conditions, a profit maximizing firm attains equilibrium at output at which its *MR* = *MC*. The short-run equilibrium of a firm under *MR-MC* approach has been discussed in this section.

The firm's equilibrium in the short run is illustrated in Figure 3.9. The determination of market price is shown in panel Figure 3.9(a). As shown in Figure 3.9(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 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, in panel (b) of the figure.

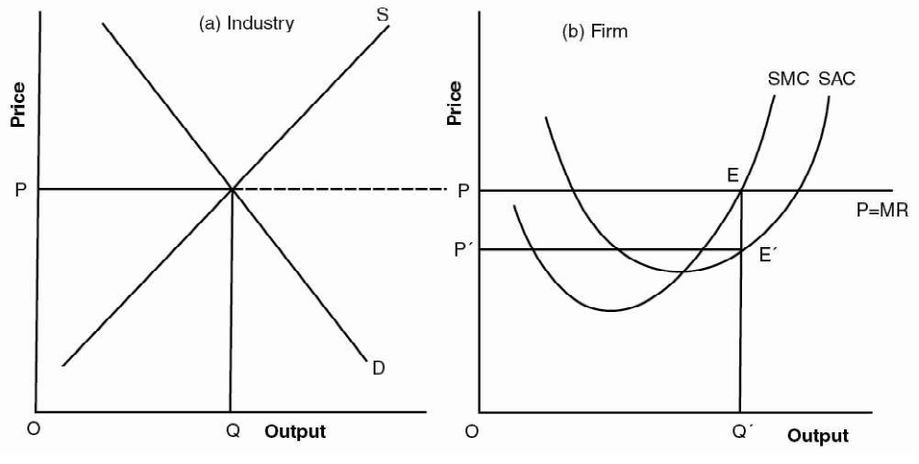

Fig. 3.9 *Short-Run Equilibrium of Firm: MR-MC Approach*

Firm's short-run equilibrium is illustrated in panel (b) of Figure 3.9(b). As can be seen in Figure 3.9(b), *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

NOTES

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 $PP' \times OQ'$ $(=PE)$ where PP' is the per unit super normal profit at output OQ'.

Do Firms Make Supernormal Profit in the Short-Run?

Figure 3.9(b) shows that the firm makes supernormal profit in the short run. A question arise here: Does a firm make necessarily a supernormal profit in the short run? In short-run, a firm may not always make profits. Given its cost and revenue conditions, the firm may make just a normal profit or even make a loss. Whether a firm makes abnormal profits, normal profits or makes losses depends on its cost and revenue conditions. If its short-run average cost (SAC) is less than the price ($P = MR$) at equilibrium, as shown in Figure 3.9(b), the firm makes abnormal profit. If its SAC is equal to $P = MR$, as shown in Figure 3.10, the firm makes only a normal profit as it covers only its SAC which includes *normal profit*. But, if price falls down from OP to OP', firm's equilibrium shifts from E to E', the firm makes losses as shown in Figure 3.10. The per unit loss = (SAC " AR) = $OT - OP'$. As Figure 3.10 shows, at equilibrium output OQ', per unit loss = $CQ' - E'Q' = CE'$. The total loss equals $CE' \times OQ'$. It is shown by the area $TCE'P' = P'E' \times CE'$.

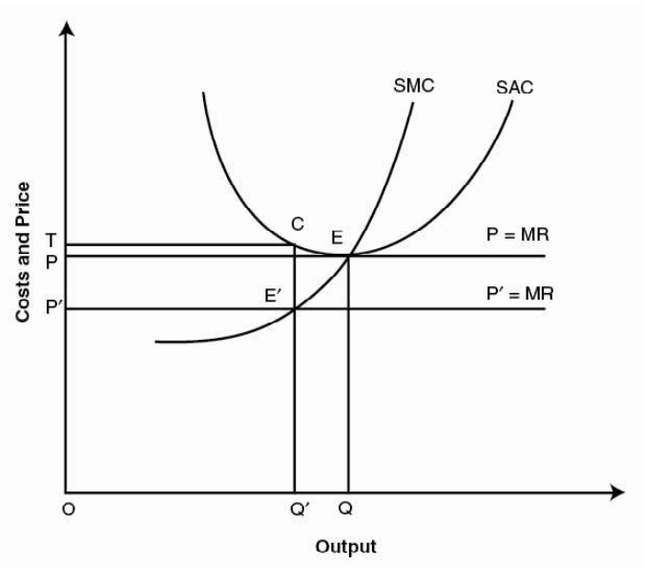

Fig. 3.10 *Short-Run Equilibrium of Firm with Normal Profit and Losses*

Shut-Down or Close-Down Point

In case a firm is making loss in the short run, it tries to minimize its losses. In order to minimize its losses, it must cover its *short-run average variable cost* ($SAVC$). The behaviour of *short-run average variable cost* is shown by the curve $SAVC$ in Figure 3.11. A firm unable to recover its minimum $SAVC$ will have to close

down. Its *SAVC* is minimum at point *E* where it equals its *MC*. Note that *SMC* intersects *SAVC* at its minimum level as shown in Figure 3.11. At point *E*, therefore, firm's loss is minimum.

Another condition that must be fulfilled for loss minimization is that *P* = *MR* = *SMC*. That is, for loss to be minimum, *P* = *MR* = *SMC* = *SAVC*. This condition is fulfilled at point *E* in Figure 3.11. Point *E* denotes the *shut-down point* or *break-down point* because at any price below *OP*, the firm has to close down as it fails to recover even its variable cost and begins to make loss.

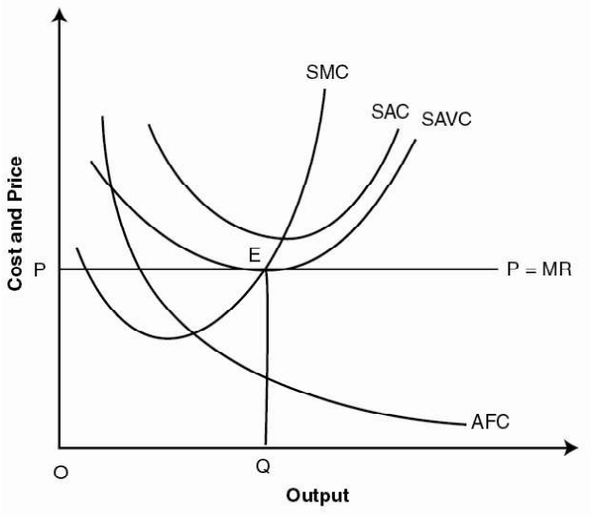

Fig. 3.11 *Shut-Down Point*

Check Your Progress

7. Define marginal revenue.
8. When is a firm said to be in a state of equilibrium?

3.8 ANSWERS TO 'CHECK YOUR PROGRESS'

1. Production function is a mathematical statement which describes the technological relationship between inputs and output in physical terms.
2. The formula which denotes the single-variable production function is given as follows:

   $Q = f(\overline{K}, L)$

3. The laws of return to variable input is governed by the following assumptions:
   - The state of technology is given
   - Labour is homogenous
   - Capital remains constant

NOTES

4. An isoquant is locus of points representing different combinations of two inputs (labour and capital) yielding the same output.
5. The three kinds of laws of returns to scale are the following:
   (i) The law of increasing returns to scale,
   (ii) The law of constant returns to scale, and
   (iii) The law of decreasing returns to scale.
6. *TFC* (i.e., the cost of plant, building, equipment, etc.) remains fixed in the short-run for certain level of output, whereas *TVC* varies with the variation in the output.
7. Marginal revenue is the addition to the total revenue (TR) as a result of sale of one additional unit.
8. A firm is said to be in equilibrium at an output at which its cost and revenue are so in balance that its objective is maximized.

3.9 SUMMARY

- In economic sense, the term 'production' means an activity by which resources (men, material, time, etc.) are transformed into a more useful commodity or value added product.
- In economic sense, the production process may take a variety of forms other than manufacturing. Transporting a commodity in its original form from one place to another where it can be consumed or used in the process of production is production.
- Production process requires a wide variety of inputs, called also as factors of production, depending on the nature of product.
- Short run refers to a period of time during which the supply and the use of certain inputs (e.g., plant, building, machinery and so on) is fixed. In the short run, therefore, production of a commodity can be increased to a limited quantity by increasing the use of only variable inputs (labour).
- In fact, a production function presents the quantitative relationships between inputs and output. A production function represents the technology of a firm, of an industry or of the economy as a whole.
- The laws of production under short-run conditions are called 'the laws of variable proportions', the 'laws of returns to a variable input' and the 'law of diminishing marginal returns'. The long-run input–output relations are studied under the 'laws of returns to scale'.
- The law of diminishing returns is an empirical law, often observed in various production activities. This law, however, may not apply universally to all kinds of productive activities since the law is not as true as the law of gravitation.

NOTES

- An isoquant is locus of points representing different combinations of two inputs (labour and capital) yielding the same output.
- When both the inputs—labour and capital—are increased proportionately and simultaneously and output increases more than proportionately increase in inputs, it gives the law of increasing returns to 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.
- Marginal cost (MC) is the addition to the total cost on account of producing one additional unit of product.
- 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 the long run.
- Marginal revenue is the addition to the total revenue (TR) as a result of sale of one additional unit.
- A profit maximizing firm attains its equilibrium at the output at which the difference between its total revenue (TR) and total cost (TC) is maximized.

3.10 KEY TERMS

- **Normal profit:** It is a necessary minimum earning, in addition to alternative cost, which a firm must earn to remain in its present occupation.
- **Explicit costs:** These are those which are actually incurred by the business firms and are entered in the books of accounts.
- **Production:** It means transforming inputs (labour, machines, raw materials, time and so on) into an output with value added.

3.11 SELF-ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. Mention the various categories of inputs used in the production process.
2. Write a short note on the formation of production function.
3. What is the difference between an isoquant curve and a consumer indifference curve?
4. Write short notes on the following: (a) cost concepts and (b) revenue concepts.

NOTES

Long-Answer Questions

1. Discuss the kinds of production functions.
2. Explain the law of variable proportions with the help of diagrams.
3. Elaborate the properties of an isoquant curve.
4. Critically analyse the three kinds of law of returns to scale.
5. What is producer's equilibrium? Discuss.

3.12 FURTHER READING

Dwivedi, D. N. 2016. *Microeconomics: Theory and Application*, 3rd Edition. New Delhi: Vikas Publishing House.

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

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

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

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

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

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

UNIT 4 MARKET AND PRICE DETERMINATION

NOTES

Structure

4.0 INTRODUCTION

The 'Theory of Firm', i.e., firm's decision-making behaviour discusses how a firm decides 'how much to produce' and 'what price to charge'. This behaviour of the firm is discussed also under the 'Theory of Price'. The firm's decision regarding price and output determination depends, at least theoretically, on the following two factors:

(i) Objective of the firm, and

(ii) The kind of market structure in which they work.

In this unit, you will learn about the market structure factor. The modern theory of price came into existence during the 1930s with Joan Robinson's *The Economics of Imperfect Competition* and Edwin H. Chamberlin's *The Theory of Monopolistic Competition,* both written independently in 1933. Earlier, the theory of price determination was in the form of the 'Theory of Value' attributed to Alfred Marshall and his *Principles of Economics.* The theory of value propounded by Marshall, on the assumptions of perfect competition and a static equilibrium system, was regarded to provide answer to all questions regarding price and output

NOTES

determination. The existence of perfect competition was however challenged by Piero Sraffa. He showed that perfect competition was not logically consistent with partial equilibrium analysis. This led to the abandonment of the assumption of the perfect competition. Joan Robinson and Chamberlin developed independently the theory of imperfect competition and theory of monopolistic competition, respectively. Joan Robinson and Chamberlin demonstrated that price and output are determined by individual decisions under the condition of imperfect competition. They had, however, retained the earlier assumption of profit maximization. This assumption was later challenged, and many new theories of firms were suggested. None of the theories has however received a universal acceptance. This unit discusses the concepts of market structure including perfect competition, monopoly, monopolistic competition and oligopoly.

4.1 OBJECTIVES

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

- Examine the meaning and types of market
- Discuss price and output determination for firm and industry under perfect competition
- Explain price and output determination for firm and industry under monopoly
- Elaborate price and output determination for firm and industry under monopolistic competition

4.2 MEANING AND TYPES OF MARKET

A market is one of many varieties of systems, institutions, procedures, social relations and infrastructures whereby parties engage in exchange. It refers to any structure that allows buyers and sellers to exchange any type of goods, services and information. Most markets rely on sellers offering their goods or services (including labour) in exchange for money from buyers. It can be said that a market is the process by which the prices of goods and services are established.

Markets vary in form, scale (volume and geographic reach), location, and types of participants, as well as the types of goods and services traded. Examples include:

(i) Physical retail markets, such as local farmers' markets (which are usually held in town squares or parking lots on an ongoing or occasional basis),shopping centers and shopping malls

(ii) Non-physical markets, such as Internet markets

(iii) Markets for intermediate goods (i.e., goods used in production of other goods and services)

(iv) International currency and commodity markets

(v) Labour markets

(vi) Stock markets, which deal with the exchange of shares in corporations

(vii) Illegal markets (markets for illicit drugs, arms or pirated products, etc.)

NOTES

There are two roles in markets, buyers and sellers. The market facilitates trade and enables the distribution and allocation of resources in a society. Markets allow any tradable item to be evaluated and priced. A market emerges more or less spontaneously or may be constructed deliberately by human interaction in order to enable the exchange of rights of services and goods.

Apart from the firm's objective, another factor that plays an important role in firm's choice of price and output is the **market structure**. The term market structure refers to the organizational features of an industry that influence the firm's behaviour in its choice of price and output. Market structure is an economically significant feature of the market. It affects the behaviour of firms in respect of their production and pricing behaviour. Market structure is classified on the basis of organizational features of the industry, more specifically, on the basis of degree of competition among the firms. In general, the organizational features include the number of firms, distinctiveness of their products, elasticity of demand and the degree of control over the price of the product. In this section, we present a brief description of the market structure, the playing field of the firms.

4.2.1 Types of Market Based on Competition: Perfect, Imperfect and Monopoly Markets

The market structure is generally classified on the basis of the degree of competition as follows:

(i) Perfect Competition

(ii) Imperfect Competition

  (a) Monopolistic Competition

  (b) Oligopoly with and without product differentiation

  (c) Duopoly

(iii) Monopoly

The basic features of these kinds of market are summarized below. However, a brief description of each kind of market is given below:

Table 4.1 *Kinds of Market Structure*

Type of Market	*No. of Firms*	*Nature of Product Over Price*	*Firm's Control*
(i) **Perfect Competition**	Very large	Homogeneous (wheat sugar, vegetables....)	None
(ii) **Imperfect Competition**			
(a) **Monopolistic Competition**	Many (most retail trade)	Real or perceived difference in product	Some
(b) **Oligopoly**	Few	(i) Product without differentiation, e.g. aluminium, steel, and chemicals, etc. (ii) Differentiated products (tea, TV Refrigerator, toothpastes, soaps, detergents, automobiles)	Some
Monopoly	Single	Products without close substitutes, like gas, electricity and water supply	Full but usually regulated

NOTES

(i) Perfect Competition

Perfect competition is a market situation in which a large number of producers offer a homogeneous product to a very large number of buyers of the product. The number of sellers is so large that each seller offers a very small fraction of the total supply, and therefore, has no control over the market price. Likewise, the number of buyers is so large that each buyer buys an insignificant part of the total supply and has no control over the market price. Both buyers and sellers are 'price takers', not 'price makers'. The price of a commodity is determined in this kind of markets by the market demand and market supply. Each seller faces a horizontal demand curve (with $e = ¥$), which implies that a seller can sell any quantity at the market determined price. **Each firm is in competition with so large a number of firms that there is virtually no competition.** This kind of market is however more of a hypothetical nature rather than being a common or realistic one. Some examples of a perfectly competitive market include stock markets, vegetable markets, wheat and rice mandis where goods are sold by auction.

(ii) Imperfect Competition

Perfect competition, in strict sense of the term, is a rare phenomenon. In reality, markets for most goods and services have imperfect competition. Imperfect competition is said to exist when a number of firms sell identical or differentiated products with some control over the price of their product. Barring a few goods like shares and vegetable markets, you name any commodity, its market is imperfect. In spite of a large number of dealers *(arhatias)* in the wheat market, the Food Corporation of India is the biggest buyer and seller of wheat in India, with a great degree of control over wheat prices.

Imperfect competition creates two different forms of markets with different number of producers and with different degrees of competition, classified as (a) monopolistic competition, (b) oligopoly (c) monopoly.

(a) Monopolistic Competitions

Monopolistic competition is a kind of market in which a large number of firms supply differentiated products. The number of sellers is so large that each firm can act independently of others, without its activities being watched and countervailed by others. Besides, it is not only extremely difficult to keep track of competitors' strategy, but also it is not of any avail. In this respect, it is similar to perfect competition. It differs from perfect competition in that the products under monopolistic competition are somewhat differentiated whereas they are identical under perfect competition. There is free entry and free exit.

(b) Oligopoly

Oligopoly is an organizational structure of an industry in which a small number of firms supply the entire market, each seller having a considerable market share and control over the price. Most industries in our country are oligopolistic. A small number of companies supply the entire tea, medicines, cosmetics, refrigerators, TV and VCRs, cars, trucks, jeeps, and so on. The producers of all these goods have some control over the price of their products. Their products are somewhat

differentiated, at least made to look different in the consumers' perception. Therefore, demand curve for their product has high elasticity, but less than infinity, unlike under perfect competition.

(c) Monopoly

Monopoly is the market of a single seller with control over his price and output. Monopoly is antithesis of perfect competition. Absolute monopolies are rare these days. They are found mostly in the form of government monopolies in public utility goods, e.g. electricity, radio broadcasting, water, rail and postal services.

Why Markets Are Imperfect?

Imperfect competition arises mainly from the barriers to entry. Barriers to entry are created by several factors.

One, the large size firms which enjoy economies of scale can cut down their prices to the extent that can eliminate new firms or prevent their entry to the industry, if they so decide.

Two, in some countries, like India, licencing policy of the government creates barrier for the new firms to enter an industry.

Three, patenting of rights to produce a well-established product or a new brand of a commodity prevents new firms from producing that commodity.

Four, sometimes entry of new firms to an industry is prevented by a law with a view to enabling the existing ones to have economies of scale so that prices are low.

4.2.2 Supply and Demand

We have already discussed the notions of supply and demand in the previous units. To briefly recapitulate, the law of supply and demand defines what effect the relationship between the availability of a particular product and the desire (or demand) for that product has on its price. When the demand for the product is high, and the supply is low, prices will increase. On the other hand, when demand for the product is low and the supply is high, prices will decrease.

4.2.3 Equilibrium

The equilibrium price of a commodity in a free market is determined by the market forces of demand and supply. In order to analyse how equilibrium price is determined, we need to integrate the demand and supply curves. For this purpose, let us use the example of shirts. Let us suppose that the market demand and supply schedules for shirts are given as shown in Table 4.2.

Table 4.2 *Monthly Demand and Supply Schedules for Shirts*

Price per Shirt (₹)	Demand ('000 Shirts)	Supply ('000 Shirts)	Market Position	Effect on Price
100	80	10	Demand exceeds supply	Rise
200	55	28	Demand exceeds supply	Rise
300	40	40	Equilibrium	Stable
400	28	50	Supply exceeds demand	Fall
500	20	55	Supply exceeds demand	Fall
600	15	60	Supply exceeds demand	Fall

NOTES

NOTES

As the table shows, there is only one price of shirts (₹300) at which the shirt market is in equilibrium, i.e., the quantity demanded equals the quantity supplied at 40,000 shirts per month. At all other prices, the shirt market is in *disequilibrium*. That is, there is imbalance between supply and demand. When market is in the state of disequilibrium, either demand exceeds supply or supply exceeds demand. As the table shows, at all prices below ₹300 per shirt, demand exceeds supply which causes rise in shirt price. Likewise, at all prices above ₹300, supply exceeds demand showing *excess supply* causing fall in shirt price.

In a free market, disequilibrium itself creates the condition for equilibrium When demand and supply are not in balance, the market is technically in disequilibrium. It means either supply exceeds demand and or demand exceeds supply. Disequilibrium itself creates conditions for equilibrium. Suppose there is excess supply. It means there is unsold stock. The unsold stock causes a loss to the firms. This forces firms to cut down their supply and price. Thus, excess supply itself forces downward adjustments in the price and the quantity supplied. The process of downward adjustments continues till supply equals demand. Similarly, when there is excess demand, it forces upward adjustments in the price and quantity demanded. When there is excess demand, firms take the advantage of the market situation and increase supply. When they increase production, cost of production goes up. But consumers, given their demand curve, are willing to pay a higher price. This process continues until demand equals supply. In our example, the process of downward and upward adjustments in price and quantity continues till the price reaches ₹300 and quantities supplied and demanded per month balance at 40,000 shirts. This process is automatic. Let us now look into the process of price and quantity adjustments called *market mechanism*.

Market Mechanism and Market Equilibrium

Market mechanism is a process by which market forces of demand and supply interact to determine equilibrium price. To understand how it works, let us explain the process through our own example. Suppose price of the shirts be initially set at ₹100. As shown in Table 4.2, at this price, the quantity demanded of shirts is 80,000 and the quantity supplied is 10,000 shirts. Thus, at price ₹100 per shirt, demand exceeds the quantity supplied by 70,000 shirts. This gives sellers an opportunity to raise the price as increase in price enhances the profit margin. This induces firms to produce and sell more in order to maximize their profits. On the demand side, consumers accept a higher price to meet their need for shirts. This trend continues till price rises to ₹300. As Table 4.2 shows, at price ₹300, the buyers are willing to buy 40,000 shirts. This is exactly the number of shirts that sellers would like to sell at this price. At this price, there is neither shortage nor excess supply of shirts in the market. In fact, demand for shirts is equal to supply of shirts. Therefore, ₹300 is the equilibrium price. The market is, therefore, in equilibrium.

Similarly, at all prices above ₹300, supply exceeds demand showing excess supply of shirts in the market. The excess supply forces the competing sellers to cut down the price. Some firms find low price unprofitable and go out of market and some cut down their production. Therefore, supply of shirts goes down. On

the other hand, fall in price invites more customers. This process continues until price of shirts falls to ₹300. At this price, demand and supply are in balance and market is in equilibrium. Therefore, the price at ₹300 per shirt is equilibrium price.

NOTES

Graphical Illustration of Price Determination

The determination of equilibrium price is illustrated graphically in Figure 4.1. The demand curve *DD*' and the supply curve *SS*' have been drawn by plotting the demand and supply schedules, respectively, (given in Table 4.2) on the price and quantity axes.

As Figure 4.1 shows, demand and supply curves intersect at point *E* determining the *equilibrium price* at ₹300. At this price, the quantity demanded (40,000 shirts) per month equals the quantity supplied. Thus, the equilibrium price is ₹300 and the equilibrium quantity is 40,000 shirts. The equilibrium condition is not fulfilled at any other point on the demand and supply curves. Therefore, if price is set at any price other than ₹300, there would be either excess supply or excess demand for shirts in the market.

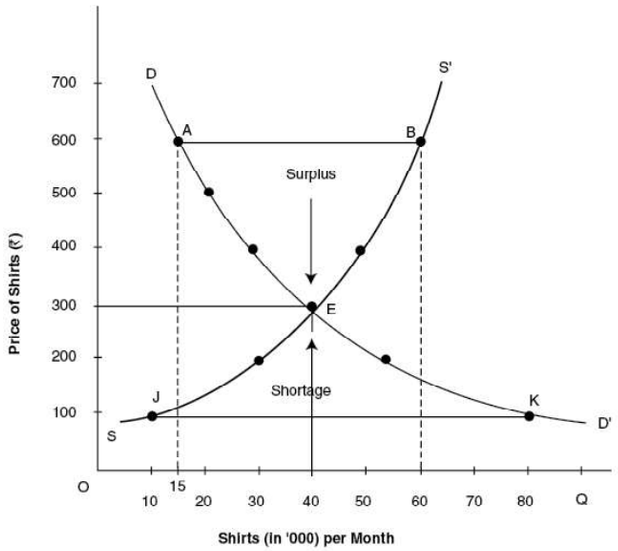

Fig. 4.1 *Graphical Illustration of Price Determination*

Let us now see how market works to bring about balance in demand for and supply of shirts. Let the price be initially set at ₹600. At this price, suppliers bring in a supply of 60,000 shirts, whereas buyers are willing to buy only 15,000 shirts. The supply, obviously, far exceeds the demand. As Figure 4.1 shows, the excess supply equals, AB = 60,000 – 15,000 = 45,000 shirts. The suppliers would, therefore, lower down the price gradually in order to get rid of the unsold stock and cut down the supply simultaneously. Besides, when price falls, demand for shirts increases too. In this process, the supply–demand gap is reduced. This process continues until price reaches ₹300 at point *E*, the point of equilibrium where demand and supply equal at 40,000 shirts. At this price, the market is in equilibrium and there is no inherent force at work which can disturb the market equilibrium.

NOTES

Likewise, if price is initially set at ₹100, the buyers would be willing to buy 80,000 shirts, whereas suppliers would be willing to supply only 10,000 shirts. Thus, there would be a shortage of 70,000 shirts as shown by the distance *JK* in Figure 4.1. The shortage will force the buyers to bid a higher price. This will lead to increase in price which will encourage the suppliers to increase their supply. This process of adjustment will continue as long as demand exceeds supply. When price rises to ₹300, the market reaches its equilibrium.

Mathematical Analysis of Determination of Equilibrium Price

In the preceding section, we have illustrated graphically how market equilibrium is determined at the point of intersection of the demand and supply curves. If demand and supply functions are known, the equilibrium quantity and equilibrium price can also be determined by using the demand and supply functions.

Let *demand function* for a commodity *X* be given as

$$D_x = 150 - 5P_x$$

and *supply function* as

$$S_x = 10P_x$$

We know that at market equilibrium, the quantity supplied equals the quantity demanded, i.e., $D_x = S_x$. Thus, the equilibrium price can be determined as follows.

$$S_x = D_x$$
$$10P_x = 150 - 5P_x$$
$$P_x = 10$$

Given the supply and demand functions, equilibrium price is obtained at $P_x = 10$ and the quantity supplied and demanded are also in equilibrium.

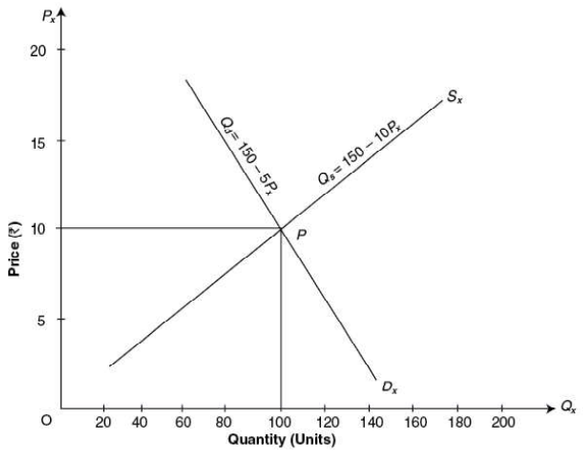

Fig. 4.2 *Algebraic Determination of Equilibrium Price and Quantity*

The algebraic determination of equilibrium price and quantity is illustrated graphically in Figure 4.2. The demand curve D_x' has been drawn by using the demand function $D_x = 150 - 5P_x$ and the supply curve S_x' by using the supply function $S_x = 10S_x$. As the figure shows, demand and supply curves intersect at point *P*. A perpendicular drawn from point *P* to the quantity axis determines the

equilibrium quantity at 100 units and a line drawn from point *P* to the price axis determines the equilibrium price at ₹10. At this price, the quantity demanded equals the quantity supplied and hence the product market is in equilibrium.

Check Your Progress

1. What is a market structure?
2. Name the kinds of market structure.
3. Which form of market structure is the antithesis of perfect competition?

4.3 PRICE AND OUTPUT DETERMINATION FOR FIRM AND INDUSTRY UNDER PERFECT COMPETITION

A perfectly competitive market is characterized by complete absence of rivalry among the individual firms. In fact, under perfect competition as conceived by the economists, competition among the individual firms is so widely dispersed that it amounts to no competition. Perfect competition is characterized by the following assumptions.

1. **Large Number of Buyers and Sellers:** Under perfect competition, the number of sellers is assumed to be so large that the share of each seller in the total supply of a product is so small that no single firm can influence the market price by changing its supply. 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 so small 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 approximately identical. Homogeneity of the product implies that buyers do not distinguish between products supplied by the various firms of an industry. Product of each firm is regarded as a perfect substitute for the products of other firms. Hence, no firm can gain any competitive advantage over the other firms. This assumption limits the power of any firm to charge a price which is even slightly higher than the market price.
3. **Perfect Mobility of Factors of Production:** Another important characteristic of perfect competition is that the factors of production (especially, labour and capital) are freely mobile between the firms. Labour can freely change the firms as there is no barrier on labour mobility—legal, language, climate, skill, distance or otherwise. There is no trade union. Capital can also move freely from one firm to another. No firm has any kind of monopoly over any industrial input. This assumption guarantees that factors of production—labour, capital, and entrepreneurship—can enter or quit a firm or the industry whenever it is found desirable.

NOTES

NOTES

4. **Free Entry and Free Exit:** There is no legal or market barrier on entry of new firms to the industry. Nor is there any restriction on exit of the firms from the industry. That is, a firm may enter the industry and quit it at its will. Thus, when normal profit of the industry increases, new firms enter the industry and if profits decrease and better opportunities are available, firms leave the industry.
5. **Perfect Knowledge about the Market Conditions:** There is perfect knowledge about the market conditions. All the buyers and sellers have full information regarding the prevailing and future prices and availability of the commodity. As Marshall puts it, '... though everyone acts for himself, his knowledge of what others are doing is supposed to be generally sufficient to prevent him from taking a lower or paying a higher price than others are doing.' Information regarding market conditions is available free of cost. There is no uncertainty.
6. **No Government Interference:** Government does not interfere in any way with the functioning of the market. There are no taxes or subsidies; no licencing system, no allocation of inputs by the government, or any kind of other direct control. That is, the government follows the *free enterprise* policy. Where there is intervention by the government, it is intended to correct the market imperfections.
7. **Absence of Collusion and Independent Decision-Making:** Perfect competition assumes that there is no collusion between the firms, i.e., they are not in league with one another in the form of *guild* or *cartel.* Nor are the buyers in collusion between themselves. There are no consumers' associations, etc. This condition implies that buyers and sellers take their decisions independently and they act independently.

Perfect vs Pure Competition

Sometimes a distinction is made between **'perfect competition'** and **'Pure Competition**'. The difference between the two is a matter of degree. While 'perfect competition' has all the features mentioned above, 'pure competition' does not assume *perfect mobility* of factors and *perfect knowledge*. That is, *perfect competition* less *perfect mobility and knowledge* is pure competition. 'Pure competition' is 'pure' in the sense that it has absolutely no element of monopoly.

The perfect competition, as characterized above, is considered as a rare phenomenon in the real business world. However, the actual markets that approximate the conditions of perfectly competitive market include the security markets for stocks and bonds, and agricultural markets like local vegetable markets. Despite its limited scope, perfect competition model has been the most popular model used in economic theories due to its analytical value.

NOTES

Equilibrium of the Firm

According to the traditional theory of firm, a firm is in equilibrium when its profit is maximum. Maximization of profits depends on the revenue and cost conditions. Revenue and cost conditions vary according to whether the period under reference is short or long. The equilibrium of the firm under *short-run* and has already been discussed in Unit 3. Let's briefly recapitulate the main points here.

Equilibrium of the Firm in the Short-Run

Here, short-run refers to a period of time during which (i) price of the product is given in the market and the firm can sell any quantity at the prevailing price; (ii) plant-size of the firm is given; and (iii) the firm is faced with given short-run cost curves.

- The prices in Perfect competition is fixed by the market, the demand curve therefore is a horizontal line where P=MR.
- The equilibrium in short run will be at the point where SMC curve intersects the MR line. This point indicates maximum profit.

In the short-run equilibrium, a firm may not always make profits. In the short-run, it may earn just a normal profit or even make losses. Whether a firm makes abnormal profits, normal profits, or losses depends on its cost and revenue conditions.

In case a firm is making loss in the short-run, it must minimize its losses. In order to minimize its losses, it must cover its short-run average variable cost (*SAVC*).

Another condition which must be fulfilled is $P = MR = SMC$. That is, for loss to be minimum, $P = MR = SMC = SAVC$. Point *E* denotes the 'shut-down point' or 'breakdown point' because at any price below *OP*, it pays firms to close down as it minimizes its losses.

Derivation of Supply Curve of the Firm

The supply curve of an individual firm is derived on the basis of its equilibrium output at different levels of the price. The equilibrium output is determined by the intersection of *MR* and *MC* curves. The derivation of supply curve of a firm is shown in Fig. 4.3(a) and 4.3(b). The equilibrium level of output in the short-run is given at point *M*—the point of intersection between *MC* and *SAVC*. This is 'breakdown point' which gives the minimum supply of the firm in the short-run. The equilibrium level of output at this point is OQ_1. Let us suppose that price increases to OP_2. As a result, the equilibrium point shifts to *R* and output increases to OQ_2. Let the price further increase to OP_3 so that equilibrium output increases to OQ_3. When price rises to OP_4, the equilibrium output rises to OQ_4. It may thus be concluded that as price increases, firm's supply goes on increasing. That is, there is positive relation between price and supply. The price and output information contained in Fig. 4.3(a) is presented in the form of a supply curve, *MS*, in Fig. 4.3(b).

NOTES

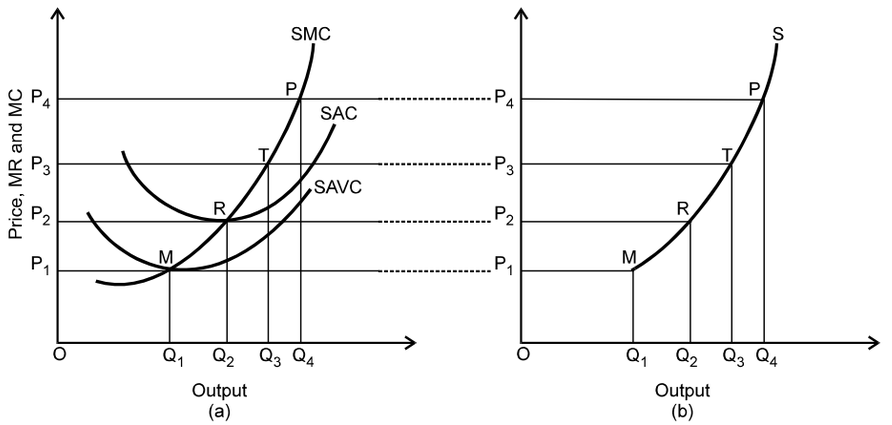

Fig. 4.3 *Derivation of Firm's Supply Curve*

Derivation of Supply Curve of the Industry

Just as the market demand curve is a horizontal summation of individual demand curves, the industry supply curve or market supply curve is the horizontal summation of the supply curves of the individual firms. If cost curves of the individual firms of an industry have identical shape, their individual supply curves would also be identical. In that case, industry supply curve can be obtained by multiplying the individual supply at various prices by the number of firms. In the short-run, however, the individual supply curves may not be identical. If so, the market supply curve can be obtained by summing horizontally the individual supply curves. Let us suppose that there are only two firms having their individual supply curves and S_1 and S_2 as shown in Fig. 4.4(a). At price OP_1, the industry supply equals $P_1A + P_1B$ which equals P_1M in Fig. 4.4(b).

Similarly, at price OP_2, the industry supply equals $P_2C + P_2C$ or $2(P_2C)$ which equals P_2N in Fig. 4.4(b). In the same way, point T is located. By joining the points M, N and T, we get the market or industry supply curve, SS'. The market supply curve so derived is used to show the determination of market price.

Now that we have derived the market supply curve, we may explain the price determination in a perfectly competitive market.

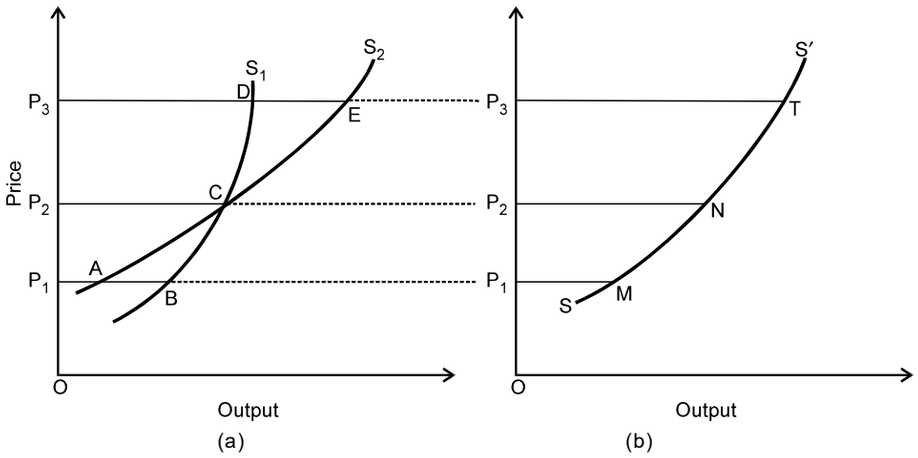

Fig. 4.4 *Derivation of Industry Supply Curve*

NOTES

Price and Output Determination

Under perfect competition, market price in a perfectly competitive market is determined by the market forces, viz., demand and supply. Here, market demand refers to the demand for the industry as a whole. It is equal to the sum of the quantity demanded by the individuals at different prices. Similarly, market supply is the sum of quantity supplied by the individual firms in the industry at a given price. The market price is therefore determined for the industry as a whole and is given for each individual firm and for each buyer. Thus, every seller in a perfectly competitive market is a 'price-taker', not a 'price-maker'.

In a perfectly competitive market, therefore, the main problem of a firm is not to determine the price of its product but to find its output at the given price so that profit is maximized.

The role of market forces and the mode of price determination depends on the time taken by supply position to adjust itself to the changing demand conditions. Price determination is analysed under three different time periods: (i) Market period or very short-run; (ii) short-run; and (iii) long-run. We will discuss below the price determination in the three periods.

Price Determination in Very Short-Run

The market period or very short-run refers to a time period in which quantity supplied is absolutely fixed or, in other words, supply response to change in price is nil. In the market period, therefore, the total output of the product is fixed. Each firm has a given quantity of commodity to sell. The aggregate supply of all the firms makes the market supply. The supply curve is *perfectly inelastic*, as shown by line *SQ* in Fig. 4.5. In this situation, price is determined entirely by the demand conditions. For instance, suppose that the number of marriage-houses (or tents) available per month in a city is given at *OQ* (Fig. 4.5), so that the supply curve takes the shape of a vertical straight line *SQ*. Let us also suppose that the monthly demand curve for marriage-houses is given by the demand curve, D_1. Demand and supply curves intersect each other at point *M*, determining the rental at *MQ*. Let us now suppose that during a particular month demand for marriage-houses suddenly increases because a relatively large number of parents decide to celebrate the marriage of their daughters and sons due to, say, non-availability of auspicious dates for some time to come. Consequently, the demand curve shifts upward to D_2. The demand curve D_2 intersects the supply curve at point *P*. The equilibrium rate of rental is thus determined at *PQ*. This becomes parametric price for all the buyers. Note that the rise in the rental from *MQ* to *PQ* is caused by the upward shift in the demand curve and that market supply curve remains perfectly inelastic in the market period. The other example of very short-run markets may be of perishable commodities like fish, milk, vegetable, etc. and of non-perishable commodities like shares and bonds.

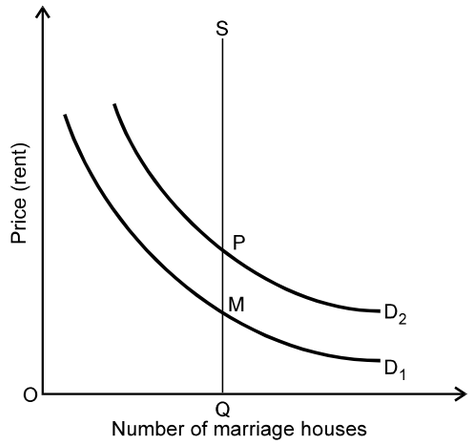

Fig. 4.5 *Determination of Market Price*

Pricing in the Short-Run

While in market period (or very short-run), supply is absolutely fixed, in the short-run it is possible to increase (or decrease) the supply by increasing (or decreasing) the variable inputs. In the short-run, therefore, supply curve is elastic, unlike a straight vertical line in the market period. Supply curve in the short-run approximates the *SMC* curve.

Under competitive conditions the process of price determination and output adjustment in the short-run is given in Fig. 4.6(a) and 4.6(b). Figure 4.6(a) shows demand curve *DD* and supply curve *SS* intersect at point *P* determining the price at OP_1. This price is fixed for all the firms in the industry.

Given the price PQ ($= OP_1$), in Fig. 4.6(a), an individual firm can produce and sell any quantity at this price. But any quantity will however not yield maximum profit. The firms will have to adjust their output to the price OP_1. The process of output determination is presented through Fig. 4.6(b).

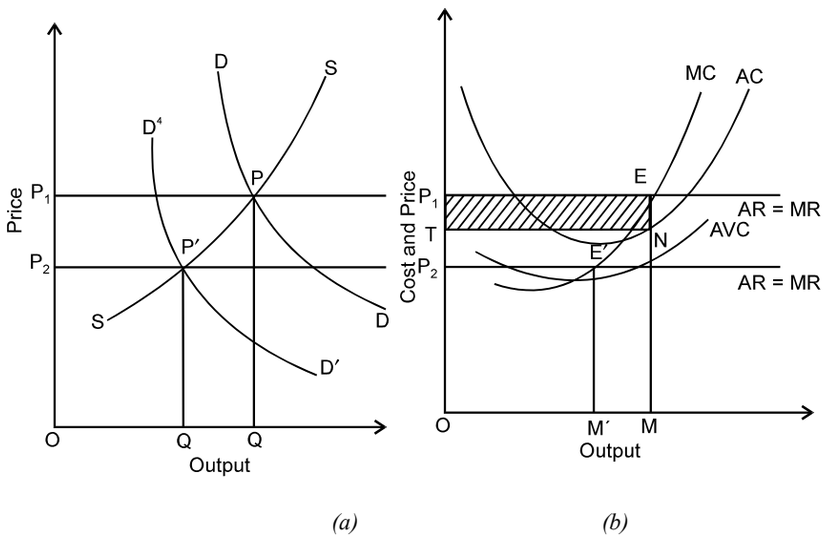

Fig. 4.6 *Pricing under Perfect Competition: Short-run*

Since a firm can sell any quantity at price OP_1, the demand for the firm's product is given by a horizontal straight line, $AR = MR$. Price being constant, its average revenue (AR) and marginal revenue (MR) are equal. Firm's upward sloping MC curve beyond its AVC curve represents its supply curve. Firm's MR and MC curves intersect each other at point E. This is the firm's equilibrium point. The perpendicular EM determines the profit-maximizing output at OM. At this output, firm's $MR = MC$, which satisfies both the first order and the second order conditions of maximum profit. The total maximum profit is shown by the area $P_1 TNE$. The total profit (η) may be calculated as

$$\eta = (AR - AC)Q$$

In Fig. 4.6(b),

$AR = EM;$

$AC = NM;$

and $Q = OM.$

By substituting the values from Fig. 4.7(b), we get

$$\eta = (EM - NM)\ OM$$

Since $EM - NM = EN$,

$$\eta = EN \times OM$$

This is the maximum profit that a firm can make, given the cost and revenue conditions as presented in Fig. 4.6(b).

Now, if price falls to OP_2 due to downward shift in the demand curve to $D'D'$, the firm will be in equilibrium at point E'. Here again the firm's $AR' = MR' = MC$. But its $AR < AC$. Therefore, the firm incurs loss. But, in the short-run, it may not be desirable to close down so long as it covers its MC.

Short-Run Equilibrium of the Industry

We have discussed above the equilibrium of the firm in the short-run. To complete the discussion on short-run price and output determination, we discuss now the short-run equilibrium of the industry.

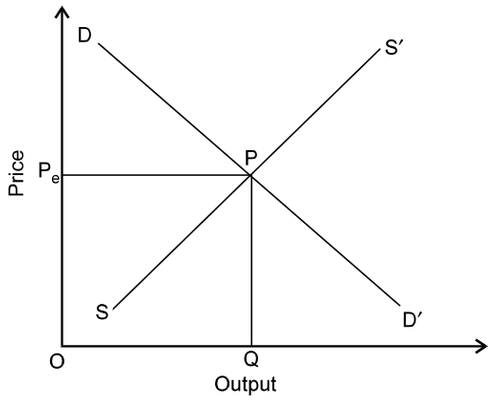

***Fig. 4.7** Equilibrium of the Industry*

NOTES

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 market is cleared is equilibrium price. The industry being in equilibrium, there is no tendency to expand or to contract the output. The equilibrium of industry is shown at point *P* in Fig. 4.7. The industry demand and supply curves intersect at point *P*, determining equilibrium price OP_e. The industry is supplying as much as consumers demand. In the short-run equilibrium of the industry, individual firms may make pure profits, normal profits or losses, depending on their cost conditions.

Price and Output determination in the Long-run

Unlike in the short-run, the supply curve in the long-run is supposed to be more elastic. Long-run brings in two additional factors in operation which make the supply curve more elastic. **First,** in the long-run, it becomes possible for the existing firms to increase their output by increasing the size of their plant. **Second,** and what is more important, new firms may enter and some existing ones may leave the industry. Entry and exit of firms bring about the long-run variation in the output. If cost and revenue conditions in the long-run are such that some firms are making losses and are not able to adjust their plant-size and cost structure to the market price, such firms leave the industry. This makes the market supply curve shift leftward causing a rise in the price. The increase in market price increases the excess profit of the profit-making firms. Under the conditions of the perfect competition (i.e., free entry and exit), the pure profit would invite many new firms to the industry. This will make supply curve shift rightward, causing a decrease in the price, which will eventually take away the excess or pure profits. All firms earn only *normal profit*. Let us now explain the price and output determination in the long-run and also the equilibrium of the firm and of the industry.

Price Determination in the Long-Run

As in the short-run, market price is determined in the long-run by the market forces of demand and supply. Let us suppose that the market demand curve is given by *DD′* which is relevant for both short-run and long-run, and short-run supply curve is given by SS_1 in Fig. 4.8(a). The market demand curve *DD′* and market supply curve SS_1 intersect each other at point P_1 and the short-run market price is determined at OP_0. At this price, the firms find their short-run equilibrium at point E_1 and each of them produces output OQ_1. The total market supply equals $OQ_1 \times$ No. of firms = ON_1 [in panel (a) of Fig. 4.8 and the industry is in short-run equilibrium.

Given the cost and revenue conditions in Fig. 4.8(b), the firms are making super normal profit of E_1M per unit. The existence of super normal profit in the short-run leads to increase in the market supply on two accounts: *one,* new firms will enter the industry attracted by the super normal profits, and *two*, the existing firms would expand their plant-size because returns to scale would increase as shown by the *LAC*. As a result, the market supply would increase so that supply curve shifts rightward to SS_2 [Fig. 4.8(a)]. The shift in supply curve brings down

the market price to *OP′* which is the long-run equilibrium price. Thus, equilibrium price is once again determined in the market.

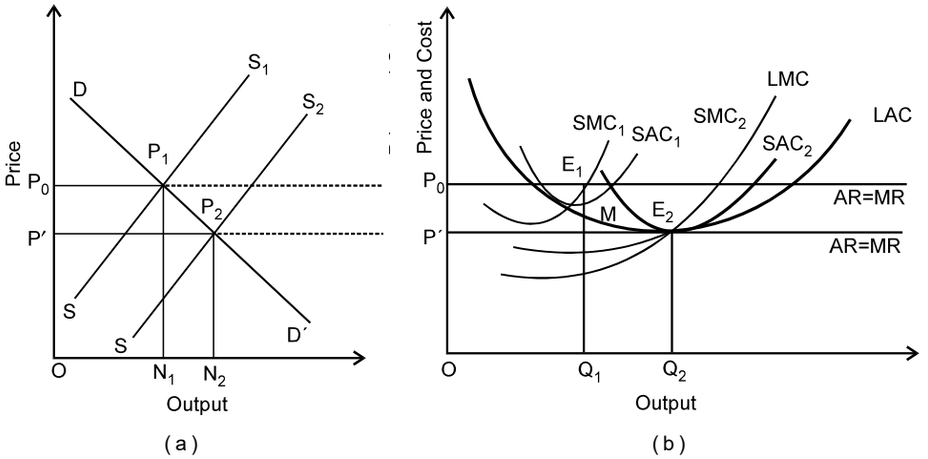

Fig. 4.8 *Long-run Equilibrium of the Firm*

Equilibrium of the Firm in the Long-Run

The firms are in equilibrium in the long-run when their

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

That is, the firms of an industry reach their equilibrium position in the long-run where both a short-run and long-run equilibrium conditions coincide. In a perfectly competitive market, the cost and revenue conditions are given for the firms. What the firms can do, therefore, is to adjust their output to the given revenue and cost conditions in order to maximize their profit. Let us now illustrate the process of adjustment of output so as to reach the equilibrium in the long-run.

Suppose that the firms are in equilibrium at point E_1 in Fig. 4.8(a) where they make excess profits $AR - SAC_1 = EM$ per unit. This gives incentives to the firms to expand their scale of production, i.e., they add more plants to the existing ones. As a result, market supply increases. Besides, supply increases also because new firms enter the industry. Therefore, the market supply curve SS_1 tends to shift rightward causing a fall in price to *OP′*. On the other hand, due to increase in demand for inputs, cost tends to rise. But so long as economies of scale are greater than the diseconomies of scale, the *LAC* tends to decrease and it pays firms to expand their plant-size. When a stage is arrives where $P < LAC$, firms incur losses. The firms which are not able to make adjustment in the plant-size or scale of production leave the industry. This works in two directions. On the one hand, supply decreases and price increases, and on the other, demand for inputs decreases which causes a decrease in the input prices. This process of adjustment continues until *LAC* is tangent to $P = AR = MR$ for each firm in the industry. This position is shown at point E_2 in Fig. 4.8(b). Eventually, at point E_2, i.e., at the point of equilibrium,

$$P = MR = LMC = LAC = SMC = SAC$$

NOTES

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 make them return to this point only.

NOTES

Equilibrium of the Industry

An industry is in equilibrium when its market demand equals its market supply. When an industry is in equilibrium, all its firms are supposed to be in equilibrium [as shown in Fig. 4.8(b)]. When an industry is in equilibrium, all its firms earn only normal profits, 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' in Fig. 4.8(a) and (b).

$$LMC = LMR = SMC = SAC = P = LAC$$

Since $P = LAC$, all the firms are earning only normal profits. At industry's equilibrium output ON_2, market demand equals market supply [Fig. 4.8(a)]. At price OP', therefore, market is cleared. The output ON_2 has a fair chance to remain stable in the long-run. For, there is no incentive for new firms to enter the industry and for existing ones to leave the industry. The industry is therefore in equilibrium.

Check Your Progress

4. How is market price determined in a perfect competitive market?
5. Name the market model which is largely used in economic theories due to its analytic value.

4.4 PRICE AND OUTPUT DETERMINATION FOR FIRM AND INDUSTRY UNDER MONOPOLY

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

Moreover, the precise definition of monopoly has been a matter of opinion and purpose. For instance, in the opinion of Joel Dean, a monopoly market is one in which 'a product of lasting distinctiveness is sold'. The monopolised product has distinct physical properties recognized by its buyers and the distinctiveness lasts over many years. Such a definition is of practical importance if one recognizes the fact that most of the commodities have their substitutes varying in degree and it is entirely for the consumers or users to distinguish between them and to accept or reject a commodity as the substitute. Another concept of pure monopoly has been advanced by D.H. Chamberlin who envisages the control of all goods and

services by the monopolist. But such a monopoly has hardly ever existed, hence his definition is unrealistic. In the opinion of some others, any firm facing a sloping demand curve is a monopolist. This definition however includes all kinds of firms except those under perfect competition. We will, however, adopt for our purpose here a general definition of a pure monopoly: a pure monopoly means an absolute power to produce and sell a commodity which has no close substitute.

4.4.1 Features of Monopoly

Some important features of monopoly are the following:

1. There is a single seller of a product which has no close substitute.
2. A monopoly firm is a price maker, not a price taker.
3. Under monopoly, there is absence of supply curve.
4. A monopoly makes a single-firm industry.

Sources and kinds of monopolies

The emergence and survival of a monopoly is attributed to the factors which prevent the entry of other firm into the industry. The barriers to entry are therefore the sources of monopoly power. The major sources of barriers to entry to a monopolised market are described here briefly.

(i) **Legal Restrictions:** Some monopolies are created by the law in the public interest. Most of the state monopolies in the public utility sector, including postal, telegraph and telephone services, radio, generation and distribution of electricity, railways, airlines and state roadways, etc. are public monopolies that are created by the public law. The state may create monopolies in the private sector also by restricting entry of other firms by law or by granting patent rights. Such monopolies are intended to reduce cost of production to the minimum by enlarging the size and investing in technological innovations. Such monopolies are known as *franchise monopolies*.

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

(iii) **Efficiency:** A primary and technical reason for growth of monopolies is the economies of scale. In some industries, long-run minimum cost of production or the most efficient scale of production almost coincides with the size of the market. Under this condition, the large-size firm finds it profitable in the long-run to eliminate the competition by cutting down its price for a short period. Once monopoly is established, it becomes almost impossible for the new firms to enter the industry and survive. Monopolies existing on

NOTES

account of this factor are known as *natural monopolies*. A natural monopoly emerges either due to technical efficiency or is created by the law on efficiency grounds.

(iv) **Patent Rights:** Another source of monopoly is the patent right 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 gives a firm exclusive rights to produce the specified commodity or to use the specified technique of production. Such monopolies are called *patent monopolies*.

4.4.2 Demand and Revenue Curves under Monopoly

Before we discuss price and output determination and the firm's equilibrium under monopoly, let us first have a look at the nature of revenue and cost curves faced by a monopoly firm. In this section, we discuss the nature of revenue curves (*AR* and *MR*). A brief account of cost and supply will follow in the next section.

Demand Curve under Monopoly

The nature of revenue curves under monopoly depends on the nature of demand curve a monopoly firm faces. We have noted earlier that in a perfectly competitive market, firms face a horizontal, straight-line demand curve. It signifies that an individual firm of an industry can sell any quantity at the prevailing price. Under monopoly, however, there is no distinction between the firm and the industry. The monopoly industry is a single-firm industry. The monopoly firm is, therefore, capable of influencing the industry price by changing the level of its production which is eventually the industry output. Besides, a monopoly firm is free to choose between price-quantity combination. It can fix higher price and sell a lower quantity and *vice versa*. For these reasons, *a monopoly firm faces a demand curve with a negative slope*. We have earlier discussed the feature of a negatively sloping demand curve, marginal revenue, and elasticity. What is important in the context of monopoly pricing is the relation between firm's average revenue (*AR*) curve and its marginal revenue (*MR*) curve. The analysis is therefore repeated here for ready reference.

Relation between *AR* and *MR*

The relationship between *AR* and *MR* plays an important role in price and output determination under monopoly. Therefore, before we explain price and output determination, let us look at technical relationship between *AR* and *MR*. The relationship between *AR* (= *P*) and *MR* can be specified in the following way.

Recall that total revenue, *TR*, equals *P* times *Q*, i.e.,

$$TR = P \cdot Q$$

and marginal revenue, (*MR*) is obtained by differentiating $TR = P \cdot Q$ with respect to *P*. Thus,

$$MR = \frac{\partial TR}{\partial P} = P + Q\frac{\partial P}{\partial Q}$$

$$= P\left(1 + \frac{Q}{P}\frac{\partial P}{\partial Q}\right) \quad \text{... (4.1)}$$

NOTES

Note that $\frac{Q}{P}\frac{\partial P}{\partial Q}$ is the reciprocal of the elasticity.

Thus, $\frac{Q}{P}\frac{\partial P}{\partial Q} = -\frac{1}{e}$.

By substituting $-\frac{1}{e}$ for $\frac{Q}{P} \cdot \frac{\partial P}{\partial Q}$ in Eq. 4.1, we get

$$MR = P\left(1-\frac{1}{e}\right) \quad \text{... (4.2)}$$

or $$MR = P - \frac{P}{e} \quad \text{... (4.3)}$$

Since $$P = AR$$

$$MR = AR - \frac{AR}{e} \quad \text{... (4.4)}$$

This relationship between *MR* and *AR* can be derived geometrically. Consider the *AR* and *MR* curves in Fig. 4.9.

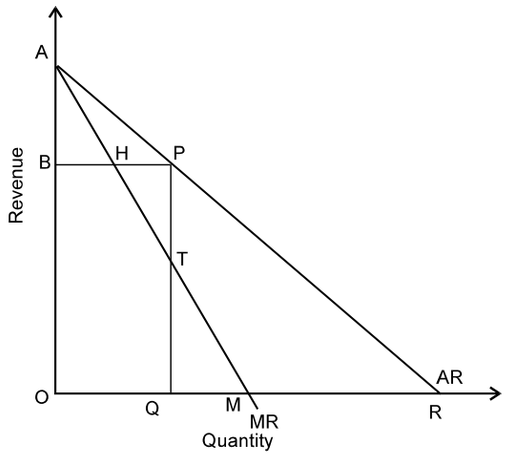

Fig. 4.9 *Relationship between AR and MR*

Let us suppose that price is given at *PQ* (=*BO*). The elasticity at point *P* on the *AR* curve can be expressed as

$$e = \frac{QR}{OQ} = \frac{PR}{AP} = \frac{OB}{AB}$$

where *e* = elasticity of demand curve.

Since $$OB = PQ,$$

$$\therefore \quad e = \frac{PQ}{AB} \quad \text{... (4.5)}$$

NOTES

It can be proved that $AB = PT$. By substituting PT for AB in Eq. (4.5), we get

$$e = \frac{PQ}{PT} \quad \text{... (4.6)}$$

Since $PT = PQ - TQ$, Eq. (4.6) may be written as

$$e = \frac{PQ}{PQ - TQ} \quad \text{... (4.7)}$$

It can be seen from (Fig. 4.10) that at price OB, $PQ = AR$ and $TQ = MR$. Therefore, Eq. (4.7) can be expressed as

$$e = \frac{AR}{AR - MR}$$

and $$MR = AR - \frac{AR}{e} \quad \text{... (4.8)}$$

Note that Eq. (4.4) is the same as Eq. (4.8).

Given the Eq. 4.7, AR can be easily obtained.

Since $$MR = AR - \frac{AR}{e}$$

or $$MR = AR\left(1 - \frac{1}{e}\right) \quad \text{... (4.9)}$$

$$AR = \frac{MR}{1 - \frac{1}{e}}$$

or $$AR = MR\left(\frac{1}{e-1}\right) \quad \text{... (4.10)}$$

The general relationships between AR and MR are given by Eq. (4.9) and Eq. (4.10). A general pattern of relations between AR and MR can be easily obtained from Eq. (4.9) as follows. Given the negative slope of the demand curve,

when

$e = 1$, $MR = 0$,	$AR > 0$	$\therefore \quad AR > MR$
$e < 1 > 0$ $MR < 0$,	$AR > 0$	$\therefore \quad AR > MR$
$e > 1 < \infty$ $MR > 0$,	$AR > 0$	but $AR > MR$
$e = 0$, $MR < 0$,	$AR = 0$	$\therefore \quad AR > MR$
$e = \infty$, $MR > 0$,	$AR > 0$	and $AR = MR$

Before we close our discussion on the relationship between AR and MR, an important aspect of relation between AR and MR curves that needs to be noted is that the slope of the MR curve is twice that of the AR curve.

Cost and Supply Curves Under Monopoly

In the short-run, cost conditions faced by a monopoly firm are, for all practical purposes, identical to those faced by a firm under perfect competitions, particularly when a monopoly firm is a competitive buyer in the input market. But in case a

monopoly firm uses specified inputs for which there is no general market and holds the position of a monopolist in the input-market, then the price of the inputs depends on the monopolist's demand for it, given the supply condition. The monopoly firm may then face a positively sloping supply curve in the input market, and its cost curves would be different from those of firms under perfect competition. In fact, the monopoly firm would face a rising supply price and its cost curves would rise rapidly. In general, however, most monopoly firms use unspecified inputs, and they are one among many buyers of the inputs. In the short-run, therefore, a monopoly firm is faced with usual U-shaped *AC* and *MC* curves.

We have noted that under perfect competition, the *MC* curve forms the basis of firm's supply curve. It is important to note here that the ***MC* curve is not the monopolist's supply curve**. In fact, under monopoly, there is no unique relation between market price and quantity supplied. Therefore, there is no supply curve under monopoly. We shall show later in this unit the absence of a precise supply curve in a monopolised industry.

Profit Maximization under Monopoly

The objective of a monopoly firm, like all other firms, is assumed to be profit maximization. Profit maximization is however not necessarily the sole objective of the firm. The monopoly firm may seek to maximize its utility function, particularly where management of the firm is divorced from its ownership. But, as mentioned earlier, most common objectives of business firm assumed in traditional theory of pricing is profit maximization. We will therefore explain the equilibrium of monopoly firm in short-run and long-run under profit maximization hypothesis.

Monopoly Equilibrium in the Short-Run

Like any other firm, a monopoly firm reaches its equilibrium where it maximizes its total profits. As noted earlier, profits are maximum where the two following conditions are fulfilled: (i) that *MC* = *MR*—the necessary condition, and (ii) that the *MC* curve must intersect the *MR* curve from below under increasing cost condition—the supplementary condition. The monopoly firm fixes its price and output in accordance with the these conditions.

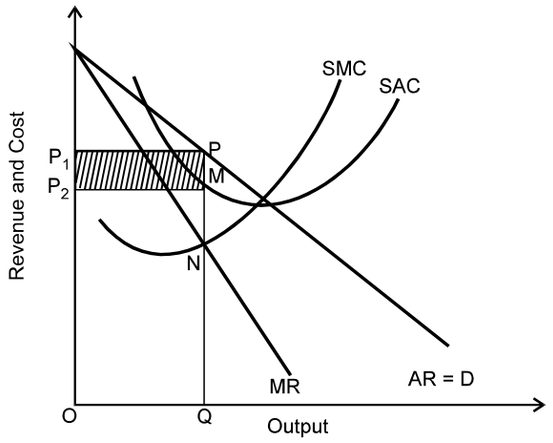

Fig. 4.10 *Price Determination under Monopoly: Short-run*

NOTES

The price and output determination under monopoly, and also the firm's equilibrium, are demonstrated in Fig. 4.10. The $AR = D$ and MR curves show the revenue conditions, while SMC and SAC present the short-run cost conditions faced by the monopoly firm. Given the revenue and cost curves, the decision rule for selecting profit maximizing output and price is the same as for a firm in the competitive industry, i.e., firm's $MR = MC$ and slope of $MC >$ the slope of MR. Therefore, the monopoly firm chooses a price-output combination for which $MR = SMC$. The MR and SMC curves intersect each other at point N. Thus, the profit maximizing output for the firm is OQ, since at this output firm's $MR = SMC$. Given the demand curve $AR = D$, the output OQ can be sold per time unit at only one price, i.e., $PQ (= OP_1)$. Thus, the determination of equilibrium output simultaneously determines the price for the monopoly firm. Once price and output are determined, the total profits are also simultaneously determined.

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

Two Common Misconceptions

There are two common misconceptions about monopoly firm which must be cleared before we proceed.

One of the misconceptions is that a monopoly firm necessarily makes super normal profits. There is however no guarantee that monopoly firm will always make profits in the short-run. In fact, whether a monopoly makes profits or losses in the short-run depends on its revenue and cost conditions. It is quite likely that its SAC lies above its AR as shown in Fig. 4.11. The monopoly firm then makes losses to the extent of $PM \times OQ = P_2MPP_1$. The firm may yet continue to produce and sell in the hope of making profits in the long-run. The monopoly firm, like a competitive firm, will however stick to the maximization rules (i.e., $MR = MC$) in order to minimize its losses.

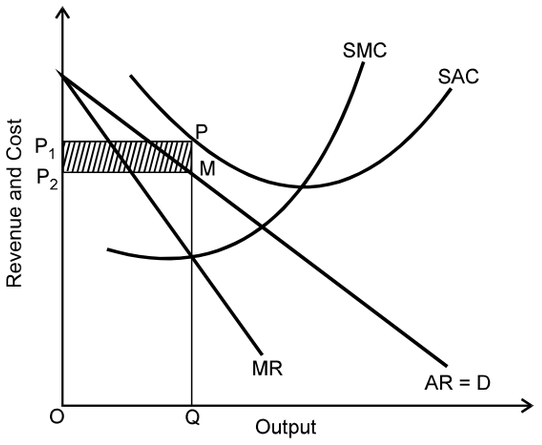

Fig. 4.11 *Monopoly Equilibrium in the Short-run: Losses*

Another common misconception about monopoly is that the demand curve faced by a monopoly firm is perfectly inelastic so that it can charge any price it likes. In fact, the demand curve faced by a monopolist is both firm's and industry's demand curve. And, most market demand curves are negatively sloped being highly elastic towards their upper end and highly inelastic towards their lower end. The equilibrium output of the monopolist that maximizes his profits will always be within the elastic region of the demand curve, if his $MC \neq 0$.

Monopoly Equilibrium in the Long-Run

The long-run conditions faced by a monopolist are different from those faced by competitive firms in an important respect, i.e., the entry of new firms into the industry. While in a competitive industry, there is free entry of new firms to the industry, a monopoly firm is protected from competition by the *barriers to entry*.

Protected by barriers to entry, a monopoly firm 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 market; (b) expected economic profits; and (c) risk of inviting legal restrictions. Assuming none of these conditions limits the expansion of monopoly firm, **the general case** of monopoly equilibrium in the long-run is illustrated in Fig. 4.12. The *AR* and *MR* curves show the market demand and marginal revenue conditions faced by the monopoly firm. The *LAC* and *LMC* curves show the long-run cost conditions. The profit maximizing monopoly firm equalises its *LMC* and *MR* at output OQ_2. The price at which the total output OQ_2 can be sold is P_2Q_2. Thus, in the long-run equilibrium, price is P_2Q_2 and equilibrium output is OQ_2. This output-price combination maximizes the monopolist's long-run profits. The total monopoly profit is shown by the area LP_2SM.

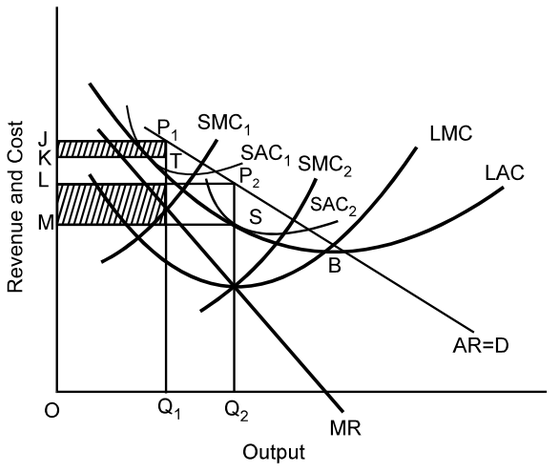

Fig. 4.12 *Monopoly Equilibrium in the Long-run*

It may be noted at the end that if there are barriers to entry, the monopoly firm would not reach the optimal scale of production in the long-run, nor will make full use of its existing capacity. This case can be verified from Fig. 4.12. The optimum size of the plant is given by point *B*, i.e., at the minimum *LAC*. But the

NOTES

monopoly firm settles at less than optimal output because optimum size of the plant will not yield the maximum profit.

Also, if the size of the market and the cost conditions permit, a profit maximizing monopoly firm may even exceed the optimum size of the plant and overutilize its long-run capacity. Figure 4.13 depicts the more-than-optimal size of the plant and its overutilization. The optimum size of the plant is given at point *B*, the point of intersection between *LAC* and *LMC*, whereas the monopoly firm chooses output at *M* where his profit is maximum. Alternatively, the monopoly firm may find its equilibrium just at the optimum size of the plant. This is possible only when the market-size is just large enough to permit optimization and full utilization of the plant size. This possibility has been illustrated in Fig. 4.14.

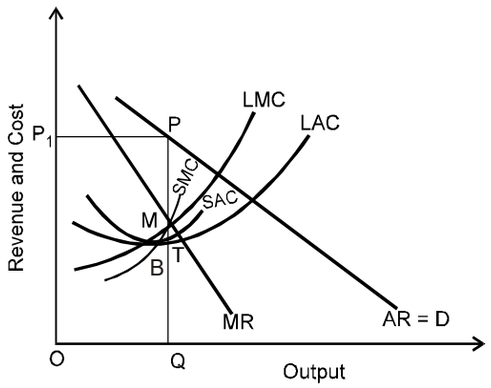

Fig. 4.13 *Monopoly Equilibrium: Overutilization of Point Size*

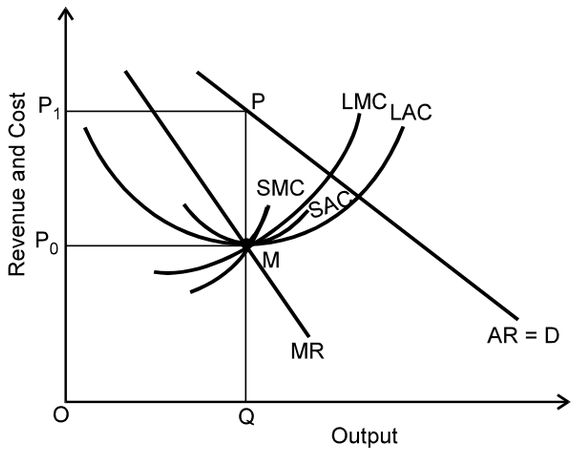

Fig. 4.14 *Monopoly Equilibrium at Optimal Size of the Plant*

Why the Absence of Supply Curve under a Monopoly?

As already mentioned, there is no unique or precise supply curve under monopoly. Let us now examine this fact by using the concept of equilibrium output. We know that supply curve presents a unique relationship between price and quantity demanded. This unique relationship between market price and quantity supplied

NOTES

does not exist under monopoly. The reason is, a profit-maximizing monopoly firm does not determine its output where $P = MC$ or where $AR = MC$. Rather, it determines its equilibrium output where $MR = MC$. Therefore, a unique relationship between price ($AR = P$) and quantity supplied cannot be traced. It is therefore quite possible to trace (i) that given the MC, the same output is supplied at different prices, and (ii) that at a given price, different quantities are supplied if the two downward sloping demand curves have different elasticities. The two cases are illustrated in Figs. 4.15 and 4.16, respectively.

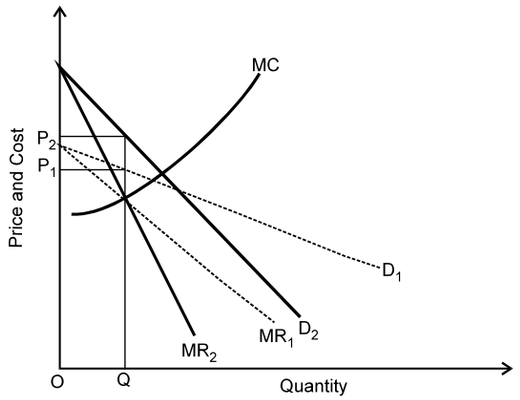

Fig. 4.15 *The Same Quantity Supplied at Two Different Prices*

As Fig. 4.15 shows, given the MC, the same quantity OQ can be supplied at two different prices—OP_1 when demand curve is D_1 and OP_2 when demand curve is D_2. Obviously, there is no unique relationship between price and quantity supplied.

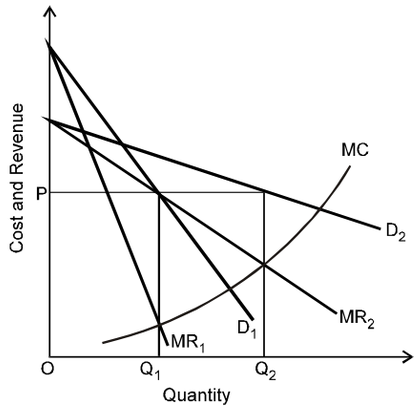

Fig. 4.16 *Different Quantities Supplied at the Same Price*

Figure 4.16 presents the case of two different quantities supplied at the same price, OP. Given the MC, quantity OQ_1 is supplied when demand curve is D_1 and quantity OQ_2 is supplied when demand curve is D_2 at the same price OP. In this case too, there is no unique relationship between price and quantity supplied. It is thus clear that there is no unique supply curve under monopoly.

NOTES

Monopoly vs Perfect Competition: Comparison of Long-run Price and Output

We have already explained the price and output determination under perfect competition and under monopoly (in this unit). We will now put two analyses together and compare the price and output in the two market structures. Comparison of long-run price and output is easier than that of short-run. We will therefore confine ourselves to only long-run price and output under monopoly and perfect competition.

Figure 4.17 presents a comparative analysis of equilibrium price and output under perfect competition and monopoly in the long-run. Let us assume that *LMC* and *LAC* are identical for both a competitive industry and a monopoly.

The equilibrium condition for a competitive industry in the long-run requires that all its firms are in equilibrium. That is, all the firms have their $AR = MR = LAC = LMC$. This condition is satisfied at point P' in Fig. 4.17. Thus, in a competitive industry, equilibrium price will be OP_1 and equilibrium output will be OQ_2. Now, if this industry were to be monopolised, the revenue conditions (*AR* and *MR*) and profit maximization rule will be different. The monopoly firm will maximize its profits at the level of output where $MR = MC$. The equilibrium condition for the monopoly firm in fulfilled at point *B*. Therefore, the equilibrium output under monopoly will be OQ_1 and the equilibrium price will be OP_2.

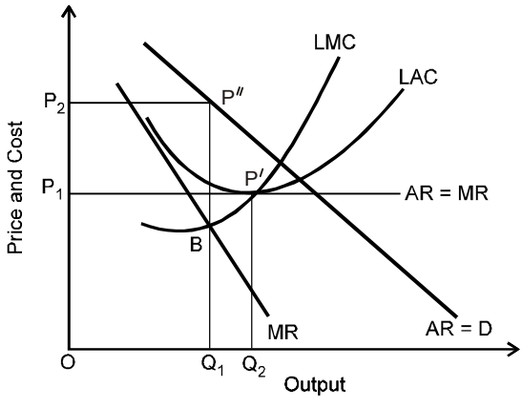

Fig. 4.17 *Comparison of Price and Output: Monopoly vs Competitive Industry*

Two Major Conclusions

Two important conclusions can be drawn from the comparison of equilibrium price and output of monopoly and competitive industry—(i) monopoly results in an inoptimal output, and (ii) monopoly causes loss of social welfare. These conclusions are illustrated in Fig 4.18, assuming a constant cost industry.

Inoptimal of Output. As Fig. 4.18 shows, if both monopoly and competitive industries are faced with identical cost conditions, the equilibrium output under competitive conditions will be higher than under monopoly and price in the

competitive industry will be lower than in monopoly. In other words, output under monopoly is lower and price higher compared to competitive industry.

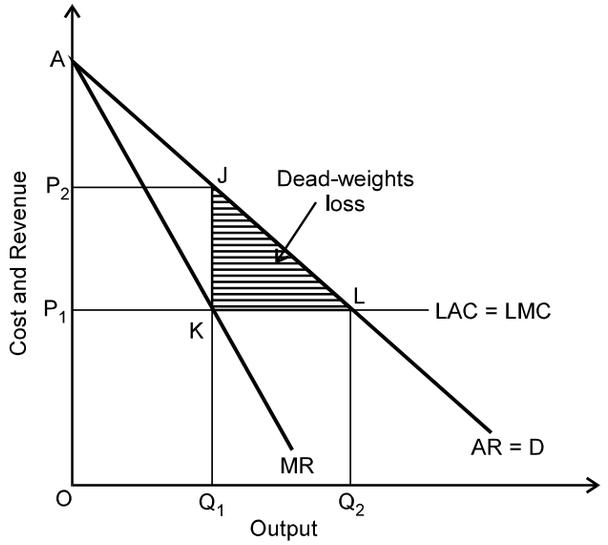

Fig. 4.18 *Price and Output under Monopoly and Perfect Competition*

For the purpose of comparison, let us suppose that both monopoly and competitive firms are faced with identical cost and revenue conditions. Given the cost and revenue conditions, the perfectly competitive industry will produce OQ_2 at which its $LAC = LMC = AR$. Its price will be OP_1. On the other hand, the monopoly firm produces and output that equalises its *LMC* and *MR*. Thus, monopoly firm produces OQ_1 and charges prices OP_2. The comparison of prices and outputs under monopoly and perfect competition stands as follows:

Variable	***Monopoly***	***Competitive***	***Comparison***
Output	OQ_1	OQ_2	$OQ_1 < OQ_2$
Price	OP_2	OP_1	$OP_2 > OQ_1$

Loss of Social Welfare

On the basis of the above conclusion, it is alleged that monopoly firms are less efficient than competitive firms. Monopoly causes loss of social welfare and distortions in resource allocation. The loss of social welfare is measured in terms of loss of consumer's surplus. The total consumer's surplus equals the difference between the total utility which society gains and the total price which it pays for a given quantity of goods. As shown in Fig. 4.18, if industry is perfectly competitive, the total output available to the society will be OQ_2 at price OP_1. The total price which society pays for OQ_2 is given by the area $OP_1LO_2 = OP_1 \times OQ_2$. The total utility which it gains from the output OQ_2 is given by the area $OALQ_2$ which, in Marshalliam terminology, is the value which society would be willing to pay for output OQ_2. Thus, consumer's surplus = area $OALQ_2$ – area OP_1LQ_2 = area AP_1L.

NOTES

NOTES

If the industry is monopolised, the consumer's surplus is reduced to AP_2J. Thus, the total loss of consumer's surplus under monopoly is

$$AP_1L - AP_2L = P_2JLP_1$$

Of this total loss of consumer's surplus, P_2JKP_1 goes to the monopolist as monopoly or pure profit. The remainder *JKL* goes to none, and therefore, it is termed as **dead-weight loss** to the society caused by monopoly.

4.4.3 Price Discrimination by Monopoly

The theory of pricing under monopoly, as discussed above, gives the impression that once a monopoly firm fixes up the price of its product, the same price is charged from all the consumers. This however may not be the case. 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 slightly differentiated) product is sold at different prices to different consumers, it is called **price discrimination**. When a monopolist sells the same product at different prices to different buyers, the monopoly is called a **discriminatory monopoly**.

Consumers are discriminated in respect of prices on the basis of their incomes or purchasing powers, geographical location, age, sex, quantity they purchase, their association with the sellers, frequency of visits to the shop, the purpose of the use of the commodity or service, and on other grounds which the seller may find suitable.

A common example of consumers being discriminated on the basis of their incomes is found in medical and legal professions. Consulting physicians and lawyers (having excess capacity) charge different fees from different clients on the basis of their paying capacity. The government charges different rates of tariffs for different grades and purpose of units of electricity consumed. Price discrimination on the basis of age is found in railways, roadways and airways: children below 15 years are charged only-half the adult-rates. 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. In case of public utility services, however, lower rates are charged when commodity or service is consumed in smaller quantity, for example, lower rates on the first few calls by the telephone owners, and no surcharge on electricity upto certain level of consumption. The most common practice of price discrimination is found in cinema shows, musical concerts, game-shows, etc.

For the purpose of price discrimination, 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 for the purpose of price discrimination. The modification in service may involve some additional cost. But price differentials are much more than is justified by cost differentials.

Although price-discrimination is the most 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 where market imperfection exists. Most business firms discriminate between their customers on

the basis of personal relationship, quantity purchased, duration of their association with the firm as buyers, and so on.

Necessary Conditions for Price Discrimination

NOTES

First, the market for different class of consumers must be separable so that buyers of low-price market are not in a position to resell the commodity in the high-price market for such reason as (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, and (iii) lack of distribution channels, e.g. transfer of electricity and gas.

Second, if market is divided into submarkets, the elasticity of demand must be different in each submarket. The purpose of price-discrimination is to maximize the profit by exploiting the markets with different price elasticities. 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 serve the objective of profit maximization.

Third, the seller must possess some monopoly 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 it in extracting the consumer's surplus. According to Pigou,there are three degrees of price-discrimination practised by the monopolists: (i) *first degree* price discrimination; (ii) *second degree* price discrimination; and (iii) *third degree* price discrimination.

(a) **First Degree Price Discrimination:** The discriminatory pricing that attempts to take away the entire consumers' surplus is called *first degree 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, he knows his buyer's demand curve for his product. Under perfect price discrimination, the seller sets the price at the highest possible level at which all those who are willing to buy the product at that price buy at least one unit each. When the consumer's surplus of this section of consumers is exhausted, he gradually lowers down the prices so that the consumer's surplus of the users of the subsequent units can be extracted. This method of pricing is continued until the whole consumer's surplus available at the price where $MR = MC$ is extracted. Also consider 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 presumably the richest patient and the lowest fee from the poorest one. The first degree of price discrimination is the limit of discriminatory pricing.

(b) **Second Degree Price Discrimination.** Under the second degree of discriminatory pricing, the firm charges different prices from different class

NOTES

of consumers—high, middle and low income consumers. The monopolist adopting the *second degree* price discrimination intends to siphon off only the major part of the consumer's surplus, rather than the entire of it. The second degree price discrimination is feasible where (i) the number of consumers is large and price rationing can be effective, as in case of utility services like telephones and natural gas; (ii) demand curves of all the consumers are identical; and (iii) a single rate is applicable for a large number of buyers. As shown in Fig. 4.19, a monopolist using a second degree price discrimination charges price OP_1 for the first few units, OQ_1 and price OP_2 for the next O_1Q_2, units, and price OP_3 for the next additional purchase of Q_2Q_3 units. Thus, by adopting a block-pricing system, the monopolist maximizes his total revenue (*TR*) as

$$TR = (OQ_1 \cdot AQ_1) + (Q_1Q_2 \cdot BQ_2) + (Q_2Q_3 \cdot CQ_3)$$

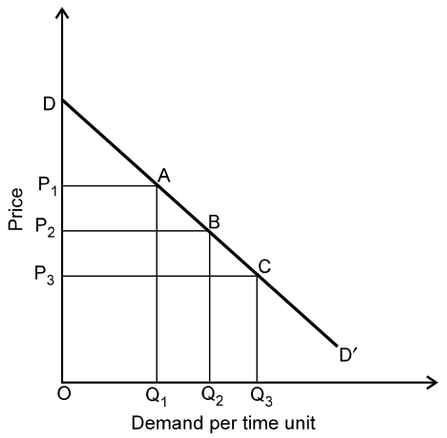

Fig. 4.19 *Second Degree Price Discrimination*

If a 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.

(c) **Third Degree Price Discrimination.** When a profit maximizing monopoly sets different prices in different markets having demand curves with different elasticities, it is using third degree price discrimination. When a monopolist is faced with two or more markets, completely separated from each other—each having a demand curve with different elasticity—a uniform price cannot be set for all the markets without loosing profits. The monopolist is therefore required to allocate total output between the different markets so that profit can be maximized in all the markets. The profit in each market would be the maximum only when the *MR* = *MC* in each market. The monopolist therefore divides total output between the markets so that in all the markets *MR* = *MC*. The process of allocation of output and determination of price for different markets is illustrated in Fig. 4.20. Suppose that a monopolist has to sell goods in only two markets, *A* and *B*. The two markets are so separated that resale of commodity is not possible. The demand curve (D_a) and marginal revenue curve (MR_a) given in Fig. 4.20(a) represent the *AR* and

NOTES

MR curves in market *A* and curves D_a and MR_b, in Fig. 4.20(b) represent *AR* and *MR* curves, respectively, in market *B*. The horizontal summation of demand curves D_a and D_b gives the total demand curve for the two markets, as shown by the curve $AR = D$, and horizontal summation of MR_a and MR_b is given by the curve *MR* (Fig. 4.20). The firm's marginal cost is shown by the curve *MC* which intersects *MR* at point *E*. Thus, optimum level of output for the firm is determined at *OQ*. At this level of output, $MR = MC$. Since the whole of *OQ* cannot be profitably sold in any one market because of their limited size, the firm has to allocate the output between the two markets.

The monopolist allocates output *OQ* between the two markets in such proportions that the necessary condition of profit maximization is satisfied in both the markets. That is, in both the markets $MC = MR$. The profit maximizing output for each market can be obtained by drawing a line from point *E* and parallel to *X*-axis, through MR_b and MR_a. The points of intersection on curves MR_a and MR_b at points *a* and *b*, respectively, determine the optimum share for each market. As shown in Fig. 4.20, the monopoly firm maximizes its revenue in market *A* by selling OQ_a units at price AQ_a, and by selling OQ_b units in market *B* at price BQ_b.

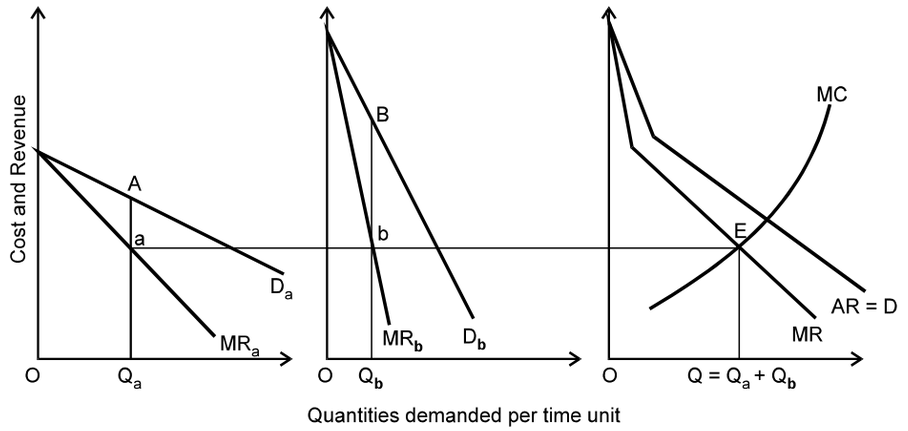

Fig. 4.20 *Third Degree Price Discrimination*

The firm's total equilibrium output $OQ = OQ_a + OQ_b$. Since at OQ_a, $MR_b = MC$ in market *A*, and at OQ_b, $MR_b = MC$ in market *B*,

$$MC = EQ = MR_a = MR_b$$

Thus, the equilibrium condition is satisfied in both market segments, 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 foreign markets. However, it need not be limited only to domestic and foreign markets. It may be suitably practised between any two or more markets separated from each other by any or more of such factors as geographical distance, transport barriers or cost of transportation, legal restrictions on the inter-regional or interstate transportation of commodities by individuals, etc.

NOTES

Whether Price Discrimination is Justified

Price discrimination has been condemned as illegal and immoral. The objection is: why charge higher price from some and lower price from others while there is no extra advantage to those who pay higher price or why benefit some at the cost of some others? In the United Kingdom and the United States, railways were prohibited from charging discriminatory rates. Discriminatory pricing has also been criticised as a destructive tool in the hands of a monopoly. For, in the past, large corporations had sought to use price discrimination to prevent the growth of competition. Besides, price discrimination may cause malallocation of resources and, hence, may be deterrent to social welfare. This is however not the case always. In some cases price discriminations is socially advantageous. In fact as Lipsey has observed, 'whether an individual judges price discrimination to be good or bad is likely to depend upon the details of the case as well as upon his own personal value judgements.' He adds, 'Certainly there is nothing in economic theory to suggest that price discrimination is always in some sense worse than non-discrimination under conditions of monopoly or oligopoly.'

Price discrimination is however considered to be desirable in certain specific cases on the following grounds:

First is the case of goods and services which are essential for the society as a whole but their production is uneconomic in the sense that long-run average cost curve (*LAC*) lies much above the aggregated market demand curve as shown is Fig. 4.21. Such goods and services cannot be produced. But, production of such goods and services can be possible if price discrimination is permitted. Price discrimination thus becomes essential for the survival of the industry.

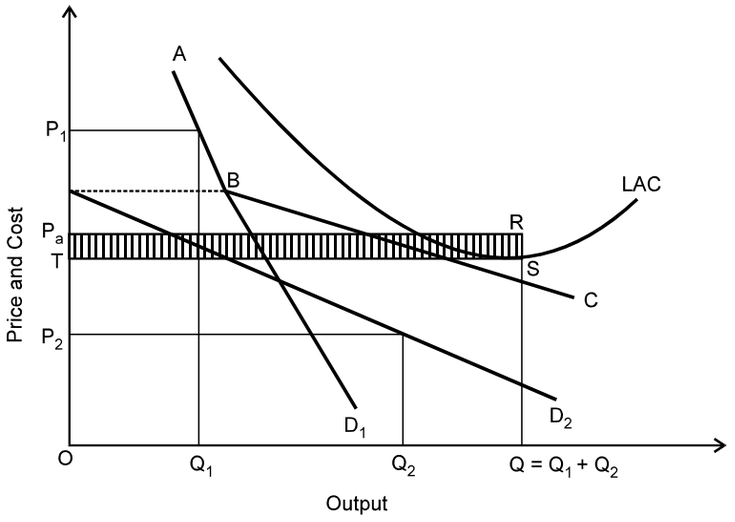

Fig. 4.21 *Price Discrimination for Industry's Survival*

Suppose, for example, (i) that there are two markets I and II, (ii) that individual demand curves for the two markets, I and II, are given as D_1 and D_2, (iii) market demand curve is given by *ABC*, and (iv) the long-run average cost curve is given by *LAC* (Fig. 4.21). Note that *LAC* lies throughout above the total demand curve *ABC*. Therefore, production is not possible if one price is to be

NOTES

charged. But, if price discrimination is adopted and prices are so charged in the two markets that the total revenue exceeds *LAC* at some level of output, then monopoly may profitably survive to the advantage of the society. Let us suppose that the monopolist sets price OP_1 in the market I in which demand is less elastic and OP_2 in market II in which demand is highly elastic. He would sell OQ_1 units at price OP_1 in market I and OQ_2 at price OP_2 in market II. His total output would then be at $OQ = OQ_1 + OQ_2$. His total revenue (*TR*) would be

$$TR = (OP_1 \times OQ_1) + (OP_2 \times OQ_2)$$

and suppose

$$AR = (OP_1 \times OQ_1 + OP_2 \times OQ_2)/OQ = OP_a$$

At output *OQ*, the *LAC* is *OT*. Thus his total cost,

$$TC = OQ \times OT = OQST$$

and his total revenue,

$$TR = OQ \times OP_a = OQRP_a$$

Since $OQRP_a > OQST$, the monopoly firms not only covers its cost but also makes excess profit. Its total profit,

$$\pi = OQRP_a - OQST = P_aRST$$

This kind of situation arises mostly in public utility services, like railways roadways, post and telegraph services, etc., in which high paying sector of the market subsidises the low paying sector. But, if low-paying sector is not subsidised, no production would be possible.

Second, discriminatory pricing can be adopted with justification where a uniform, single profitable price is likely to restrict the output and deprive many (particularly the people of lower income groups) of the essential goods or service. For example, if doctors in private practice, who often change discriminatory price for their services, are asked to charge a uniform fee from all the patients, they would charge a fee high enough to maintain the level of their income. The high fee may deprive the poor of the doctor's service and may force them to opt for inferior or inadequate treatment. The result of the uniform high fee will be that the rich patients who can pay a still higher fee gain as they pay a price lower than what they could afford, and on the other hand, poor patients are deprived of proper medical service.

Third, there may be cases where a section of consumers gains more than the people of other sections from the use of the same product. For example, from the use of electricity factory-owners gain more than the households. In such cases, uniform price would be unjustified from a normative point of view, provided the objective is not to restrain the domestic consumption of electricity and spare it for productive purposes. There is, on the other hand, full justification for discriminatory pricing of electricity.

Government Regulation of Monopoly Prices

The existence of monopolies in a market economy is criticised on the grounds that they restrict production and consumption, widen income and wealth disparities,

NOTES

exploit consumers and employees, cause distortions in allocation of resources, reduce the prospect of employment, and cause loss of social welfare. In most countries, therefore, there is general apathy towards the monopolies. Consequently, governments in the market economies attempt to control and regulate monopolies to the advantage of the society. There are various measures—direct, indirect, legal and otherwise—to control and regulate the monopolies. However, we discuss below only the price regulation of natural monopolies.

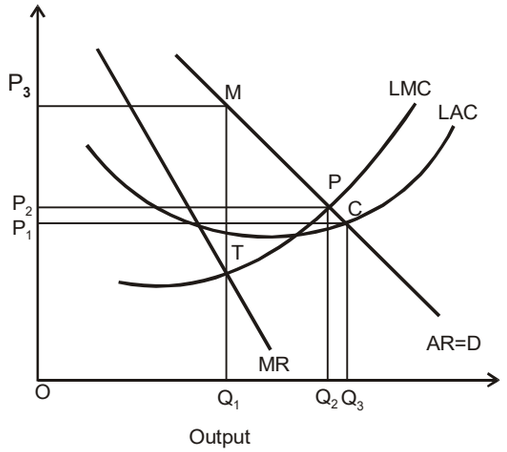

Fig. 4.22 *Government Regulated Monopoly*

Price regulation is a common feature in case of natural monopolies. When the size of the market is small relative to the optimum size of the firm, market size cannot support more than one firm of optimal size. The monopoly in such a market is a natural monopoly. The natural monopoly is thus protected by market size itself. The government may either nationalise such monopolies or regulate their prices so as to eliminate the excess profits. If the government intends to regulate the monopoly price, the question arises: what price should be fixed for the monopolist to charge? The two alternative prices that have been suggested are: **one** that allows some excess profit to the monopolist, and the **second** that allows only normal profit to the monopolist. Both the alternative prices, along with their repercussion on output, are illustrated in Fig. 4.22. An unregulated monopoly would produce OQ_1 units, charge price OP_3, and make excess profit of $MT = MQ_1 - TQ_1$ per unit. If monopoly price is regulated, one possible price is given at point P where $LMC = AR$, the price being OP_2 ($= PQ_2$). Alternatively, price may be fixed at point C where $AR = LAC$ and price = OP_1 ($= CQ_3$). When OP_1 is the price set for the monopolist, only a normal profit is allowed to the firm, but output is maximum possible under the given cost and revenue conditions. If price is fixed at OP_2, the monopolist gets some excess profit, but the output is less than that at price OP_1. In both the cases, however, the total output under regulated monopoly is much higher than that under unregulated monopoly. Which of the two alternative prices (OP_1 and OP_2) is more appropriate is a matter of debate.

NOTES

Measures of Monopoly Power

It is only in rare cases that monopolies have absolute power. Monopoly power varies from industry to industry. The degree of monopoly power matters a great deal in pricing and output decisions of a monopolist. Besides, measuring monopoly power is required also in connection with control and regulation of monopolies. We discuss here the various measures of monopoly power.

Measuring monopoly power has been a difficult proposition. The efforts to devise a measure of monopoly power have not yielded any universal or non-controversial measure. As Hunter has observed, 'The idea of devising a measure of monopoly power, with reference both to its general incidence and to particular situation has been and probably always will remain an attractive prospect for economists who wish to probe in this field.' If not for any other reason, then for 'sheer intellectual curiosity', economy theorists feel compelled to work on this problem, as they 'could not with good conscience go on talking about 'great' or 'little' monopoly power or about various degrees of monopoly without trying to ascertain the meaning of these words.'

Therefore, to devise at least a 'conceivable' measure of monopoly, even if 'practical' measurement is impossible, continues to interest the economists, for at least two reasons. **First**, apart from intellectual curiosity people would like to know about the economy in which they live, its industrial structure, and the industries from which they get their supplies. **Second**, growth of monopolies have forced governments of many countries to formulate policies and devise legislative measures to control and regulate monopolies. If the government is to succeed in its policy of restraining monopoly, it must have at least some practicable measure of monopoly and monopolistic trade practices.

Although economists have devised a number of devices to measure the degree of monopoly power, none of the measures is free from flaws. Yet, the various measures do provide an insight into the monopoly power and its impact on the market structure. Besides, they also help in formulating an appropriate publicpolicy to control and regulate the existing monopolies. We discuss here briefly the various measures of monopoly power suggested by the economists.

1. **Number-of-Firms Criterion:** One of the simplest measures of degree of monopoly power is to count the number of firms in an industry. The smaller the number of firms, the greater the degree of monopoly power of each firm in the industry, and conversely, the larger the number of firms, greater the possibility of absence of monopoly power. As a corollary of this, if there is a single firm in an industry, the firm has an absolute monopoly power. This criterion seems to have been derived from the characteristics of the perfect competition in which the number of firms is so large that each firm supplies only an insignificant proportion of the market and no firm has any control on the price.

   This criterion has however a serious **drawback**. The number of firms alone does not reveal much about the relative position of the firms within the

NOTES

industry because (i) 'firms are not of equal size,' and (ii) their number does not indicate the degree of control each firm exercises in the industry. Therefore, the numerical criterion of measuring monopoly power is of little practical use.

2. **Concentration Ratio:** The *concentration ratio* is one of the widely used criteria used for measuring monopoly power. The concentration ratio is obtained by calculating the percentage share of the largest group of the firms in the total output of the industry. "The number of firms chosen for calculating the ratio usually depends on some fortuitous element—normally the census of production arrangements of the country concerned." In Britain the share of the largest three firms of a census industry, and in the USA, the share of the largest four firms is the basis of calculating concentration ratio. Apart from the share of the largest firms in the industry-output, "[the] size of the firm and the concentration of control in the industry may be measured...in terms of production capacity, value of assets, number of employees or some other characteristics."

   These measures too are however not free from **drawbacks** as they involve statistical and conceptual problems. For example, production capacity may not be straightaway used as it may include 'unused, obsolete or excess capacity'; the value of assets involves valuation problem as accounting method of valuation and market valuation of assets may differ. Employment figures may not be relevant in case of capital intensive industries. The use of such figures may be misleading. The two other convenient measures are 'gross output value' or 'net output' (value added). But the former involves the risk of double counting and the latter, the omission of inter-establishment transfers.

   Another important objection to these measures of degree of monopoly power is that they do not take into account the size of the market. Size of the market may be national or local. A large number of firms supplying the national market may be much less competitive than the small number of firms supplying the local market. For, it is quite likely that the national market is divided among the thousand sellers so that each seller has status of a monopolist in his own area.

   The most serious defect of concentration ratio as an index of monopoly power is that it does not reflect the competition from other industries. The degree of competition is measured by the elasticity of substitution between the products of different industries. The elasticity of substitution may be different under different classification of industries. Therefore, an industry with concentration ratio under one classification of industries may have a very low elasticity of substitution and hence a high degree of monopoly. But, if classification of industries is altered, the same industry with a high concentration ratio may have a very low elasticity of substitution, and hence, may show a low degree of monopoly.

3. **Excess Profitability Criteria:** J.S. Bain and, following him, many other economists have used *excess profit* as a measure of monopoly power. If

NOTES

profit rate of a firm continues to remain sufficiently higher than all opportunity costs required to remain in the industry, it implies that neither competition among sellers nor entry of new firms prevents the firm from making a pure or monopoly profit. While calculating the excess profit, the opportunity cost of owner's capital and margin for the risk must be deducted from the actual profit made by the firm. Assuming no risk, the degree of monopoly may be obtained by calculating the divergence between the opportunity costs (O) and the actual profit, (P), as $(P - O)/P$. If $[(P - O)/P] = O$, there exists no monopoly, and it $[(P - O)/P] > O$, there is monopoly. The higher the value of $(P - O)/P$, the greater the degree of monopoly.

Another measure of degree of monopoly based on excess profitability has been provided by A.P. Lerner. According to him, the degree of monopoly power (MP) may be measured as

$$MP = \frac{P - MC}{P}$$

where P = price, MC = marginal cost. Since for a profit maximizing firm, $MR = MC$, Lerner's measure of monopoly power MP may also be expressed as

$$MP = \frac{P - MR}{P}$$

Since $P/(P - MR) = e$, $(P - MR)/P = 1/e$, i.e., MP equals to the reciprocal of elasticity. Thus, Lerner's measure of monopoly power may be expressed also as $MP = 1/e$. It may thus be inferred that lower the elasticity, the greater the degree of monopoly, and vice versa. According to Lerner's formula, monopoly power may exist even if firm's $AR = AC$ and it earns only normal profit.

Lerner's formula of measuring the degree of monopoly power is considered to be theoretically most sound. Nevertheless, it has been criticised on the following grounds.

First, it is suggested that any formula devised to measure degree of monopoly power should bring out the difference between the monopoly output and competitive output or the 'ideal' output under the optimum allocation of resources. The divergence between P and MC used in Lerner's formula does not indicate the divergence between the monopoly and the 'ideal' output. Lerner has possibly used the divergence between P and MC as the substitute for the divergence between monopoly and 'ideal' output. "This substitution of a price-cost discrepancy for a difference between actual and 'ideal' output is probably the greatest weakness of formula which is supposed to measure deviation from the optimum allocation of resources."

Second, price-cost discrepancy may arise for reasons other than monopoly, and price and cost may be equal or close to each other in spite of monopoly power.

Third, since data on MC are hardly available, this formula is of little practical use for policy purposes.

NOTES

4. **Triffin's Cross-Elasticity Criterion:** Triffin's criterion seems to have been derived from the definition of monopoly itself. According to this criterion, cross-elasticity is taken as the measure of degree of monopoly—the lower the cross-elasticity of the product of a firm, the greater the degree of its monopoly power. But, this criterion is based on the inter-relationships between the individual firms and indicates only the relative power of each firm. It does not furnish a single index of monopoly power.

Check Your Progress

6. List the important features of a monopoly.
7. Mention two important misconceptions about monopoly firm.

4.5 PRICE AND OUTPUT DETERMINATION FOR FIRM AND INDUSTRY UNDER MONOPOLISTIC COMPETITION

The model of monopolistic competition developed by Edward H. Chamberlin presents a more realistic picture of the actual market structure and the nature of competition. Let us first discuss briefly the nature of the market structure and monopolistic competition among the firms.

Monopolistic Competition: *Monopolistic Competition* is a market structure in which a *large number* of sellers sell *differentiated* products which are close, but not perfect, substitutes for one another. Monopolistic competition combines the characteristics of perfect competition and monopoly.

The assumptions of the monopolistic competition are the same as those of pure competition, with an exception of homogeneity of products. While pure competition model assumes that products are *homogeneous* in every possible dimension, monopolistic competition model assumes that products are *differentiated*. The product of each firm is so differentiated from those of other firms that consumers are able to distinguish the product of a firm from those of others. For example, consumers know for sure the difference between different brands of mobile phones, e.g., Nokia, Sony, Samsung, Reliance. Since each firm produces a product distinguishable from that of other firms, each firm holds a monopoly power over its own products.

Although products are differentiated, they remain a close substitute for one another. Product differentiation, along with other features, creates condition for competition among the firms which are monopolists in case of their own products. This kind of competition is the genesis of monopolistic competition.

4.5.1 Features of Monopolistic Competition

The main features of monopolistic competition are summarized below:

1. Under monopolistic competition, the number of firms is fairly large.

2. Firms produce and sell differentiated products-each product being a close substitute for another.
3. There is free entry and free exit of firms to and from the industry.
4. Factors of production enjoy free mobility, at least theoretically.
5. Both buyers and sellers have full knowledge about the market.
6. There is no collision between monopolistic firms.

Foundations of the Monopolistic Competition model

(i) Assumptions

Chamberlin's model of monopolistic competition is based on the following assumptions:

1. There is a large number of buyers and sellers in the market.
2. Each seller sells a product differentiated from that of others.
3. The differentiated products are close, not perfect, substitute for one another.
4. There is free entry and free exit of firms.
5. The firms seek to maximize their profits in both short and long-runs.
6. Technology and factor prices are given and the firms are aware of revenue and cost curves.

(ii) Product Differentiation and the Demand Curve

Chamberlin has defined *product differentiation* in the following words: "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, so long as it is of any importance whatever to buyers, and leads to a preference for one variety of the product over another.... Differentiation may be based upon certain characteristics of the product itself, such as exclusive patented features, trademarks, trade names, peculiarities of the package or container, if any, or singularity in quality, design, colour or style. It may also exist with respect to the conditions surrounding its sales. In retail trade, these conditions include such factors as the convenience of the seller's location, the general tone or character of his establishment, his way of doing business, his reputation for fair dealing, courtesy, efficiency, and all the personal links which attach his customers either to himself or to those employed by him."So far as these and other tangible and intangible factors create consumers' preference for one product over the others, the products are virtually differentiated.

Thus, product differentiation is primarily intended to make consumers distinguish the product of one producer from that of the other producers in the industry. When the consumers are able to distinguish one product from the others, they may develop a preference or brand loyalty for one product over the others. Once preference for a product is created, it alters the course of demand curve for the product. In the ultimate analysis, product differentiation leads to a change in demand curve for the product from a horizontal demand line (as under pure competition) to a downward sloping demand curve. The downward sloping

demand curve enables the sellers to exercise some discretion in determining the price of his product.

(iii) Cost Curves and Selling Cost

NOTES

In his model of monopolistic competition, Chamberlin has assumed the traditional U-shaped cost curves, viz., *AC, AVC* and *MC*. In addition, he has introduced a new cost, i.e., *selling cost*. "Selling costs are defined as costs incurred in order to alter the position or the shape of the demand curve for a product." Selling costs include all the expenses that are intended to promote the sales, including cost of advertisement, salesmen's salaries, expenses of sales department, margins granted to dealers—wholesalers and retailers—and on window displays and demonstration of new goods. Selling costs affect demand curve in two ways.

First, selling costs make the demand curve for the product shift upward by informing consumers about the availability of the product and by increasing consumer's preference for the product.

Second, selling cost makes the demand curve less elastic by strengthening the consumers' preference for the product.

Chamberlin assumes average selling cost to be U-shaped, that is, selling cost per unit of sales initially decreases but eventually increases. Thus, the average selling-cost curve has a shape similar to the *AC* curve.

(iv) The Concept of Industry and Product Groups

Under monopolistic competition, products are so differentiated that each product is distinguishable from others, and each firm is, in a sense, an industry in itself, exactly as a monopoly firm is an industry in itself. The heterogeneity of the products, therefore, causes a problem in analytical treatment of the industry. It may be recalled that, in case of homogeneous products, demand curve for an industry can be obtained by adding individual demand curve. But in case of heterogeneous products, the demand for individual products cannot be added to obtain market demand and supply curves.

For this reason, Chamberlin attempted to redefine the industry for his analytical purpose. He defined the monopolistically competitive industry as a 'group' of firms producing a 'closely related' commodity, referred to as *product group*. The product 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, e.g. personal computers, soaps, toothpastes, automobiles, TV sets. The two products are considered as *economic substitutes* for each other if they satisfy the same want and have *more or less* the same price. For example, all brands of TV sets are economic substitutes for one another. But flat TV sets are not economic substitutes for ordinary ones since their prices are widely different, though they remain technological substitutes. Operationally, the product group may be defined as the group of firms whose products have between themselves high *price* and *cross elasticities*. This definition, although theoretical plausible, involves the problems of measuring cross-elasticities and of determining

the degree of cross-elasticities that can make a commodity admissible to the group. Determining the product group would therefore involve subjective judgement.

Price and Output Determination under Monopolistic Competition

NOTES

Chamberlin's theory of price and output determination under monopolistic competition is basically the same as that under monopoly with a difference, of course. While under monopoly demand and cost curves are both assumed to be given, under monopolistic competition, firms are assumed to indulge in competition to change the slope of the demand curve or to make it shift rightward, given the cost curves.

They seek to make these changes in the demand curve by any or all of the following measures:

(i) change in the price of the product;

(ii) change in the nature of the product; and

(iii) change in the advertisement outlays.

As to price change, since a monopolistically competitive firm faces an elastic demand curve similar to one faced by a monopoly firm, it has the option to raise the price and sell less or to lower the price and sell more. But, it fixes a price that maximizes its profits. As to change in product, the sales can be promoted by change in the quality of product through technical changes, introduction of a new design, use of better material, use of new package or containers, prompt and courteous services, credit facilities, etc. Also, the firm may influence its volume of sales by increasing advertisement expenditure so that more consumers are attracted to the product. Increase in advertisement expenditure also increases the selling price. The firm is therefore required to so adjust its price and output that its profits is maximum.

While adjustment between price and output for profit maximization is a short-run phenomenon, changes in the quality of the product and advertisement expenses are long-run phenomena. We will therefore explain Chamberlin's theory of price and output determination under monopolistic competition under both short-run and long-run conditions.

Firm's Short-run Equilibrium

While monopolistic competition is characteristically closer to perfect competition, it is closer to monopoly in regard to pricing and output determination. Like a monopolist, a monopolistic competitor faces a downward sloping demand curve having a smaller slope. This demand curve is the product of (i) strong preference of a section of consumers for a particular product; and (ii) the quasi-monopoly of the seller over the supply. The strong preference or loyalty of the consumers gives the seller an opportunity to raise the price and yet retain some customers. And, since each product is a close substitute for another, they attract the consumers of other products by lowering down their prices.

As mentioned above, short-term analysis of pricing and output determination under monopolistic competition is similar to price and output determination under monopoly. The short-term equilibrium analysis is primarily the adjustment of price

NOTES

and output to the given cost and revenue conditions. The short-run price and output adjustment is illustrated in Fig. 4.23. The $AR = D$ and MR curves show the revenue conditions and SAC and SMC curves show the cost conditions faced by the firm in the short-run.

As shown in Fig. 4.23, the necessary condition of profit maximization, i.e., MR must be equal to MC, is fulfilled at output OQ. This output can be sold at price PQ, so the price is also determined. At this output and price, the firm earns a maximum economic profit, shown by the rectangle P_1PMP_2.

The economic profit per unit (PM) exists in the short-run because new firms cannot enter the industry. But the rate of profit would not be the same for all the firms under monopolistic competition because of difference in the elasticity of demand. For the same reason, product price will be different for the different products, though price differentials will be only marginal.

In the short run, the firms may attempt to maximize their profits by changing the nature of the product and by increasing advertisement expenditure. But, since there are many close substitutes, neither of the strategies would be of much avail in the short run. If the firms do adopt these strategies, they would do so only to maximize their profits.

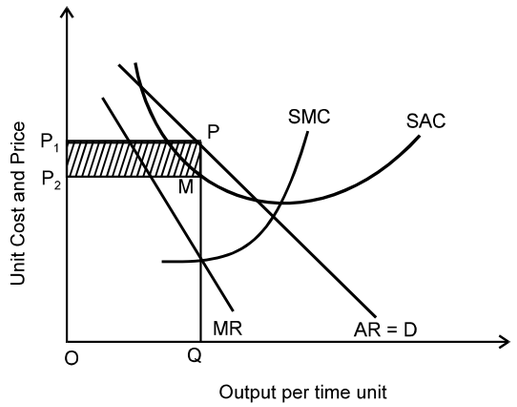

Fig. 4.23 *Short-run Equilibrium Under Monopolistic Competition*

Long-run Equilibrium under Monopolistic Competition

The conditions faced by a firm of the 'product-group' under monopolistic competitions are different from those in the short run. In the long run, a firm can expand its plant-size and if there are no barriers to entry, new firms will enter the product group, especially when existing firms are making pure profits. Besides, the firms get the opportunity to change the nature and position of the demand curve for their product by (i) changing the quality of the product; and (ii) incurring a large amount of advertisement expenditure. Since conditions change in the long run, the firms can manoeuvre their price and output in order to maximize their profits in the long run too.

Basic Assumptions

Chamberlin has made the following explicit and implicit assumptions in order to develop his theory of monopolistic competition under long-run conditions.

NOTES

First, the basic assumption is that there is a large number of firms producing differentiated products which are close substitutes for one another.

Second, the number of firms in the product group is so large that each firm expects its manoeuvering of prices and output to go unnoticed by the rival firms.

Third, one of the heroic assumptions of Chamberlin is that both demand and cost curves for all the products are uniform throughout the group.That is, all firms have identical revenue and cost curves.

Fourth, his second heroic assumption is that consumer's preferences are evenly distributed among the different varieties, and that differences between products are not such as to give rise to differences in cost.

Under these assumptions, Chamberlin develops **three models** of equilibrium:

(i) to analyse equilibrium with free entry of new firms to the industry with non-price competition

(ii) to analyse equilibrium under price competition, assuming no entry or exit

(iii) to present a combined analysis of the first and the second models.

We will explain briefly the three models in the subsequent subsections.

(i) Long-run Equilibrium with Free Entry of New Firms

The long-run equilibrium of firms under the assumption of free entry of new firms is illustrated in Fig. 4.24. The *LAC* and *LMC* are the cost curves faced by the firms. The initial *AR* and *MR* curves (i.e., prior to the entry of new firms) are given by $AR = D_2$ and MR_2. Given the cost and revenue curves, the firms will be in short-run equilibrium at point *E* where price is OP_2 and output OQ_2. The firms are making super-normal profits to the extent of *EB* per unit of output.

The existence of super-normal profits attracts new firms to the product group. With the entry of new firms, the sale of each firm in the group decreases. Consequently, the demand curve (or *AR* curve) for the firm shifts leftward and so does the *MR* curve. This forces the firms to adjust their price and output to the new *AR* and *MR* conditions, cost conditions remaining the same. If still there exists super-normal profit more new firms join the product group, and existing firms will be forced to readjust their price and output to another equilibrium position. This process will continue until the demand curve becomes tangent to the *LAC*, and all the firms earn only normal profits. The final equilibrium position of the firms in the long-run under monopolistic competition is shown at point *A*, where price is OP_1 and output OQ_1. At this price and output, all firms make only normal profit. Therefore, there is no incentive for the new firms to enter the industry. The equilibrium will therefore be stable at point *A*.

NOTES

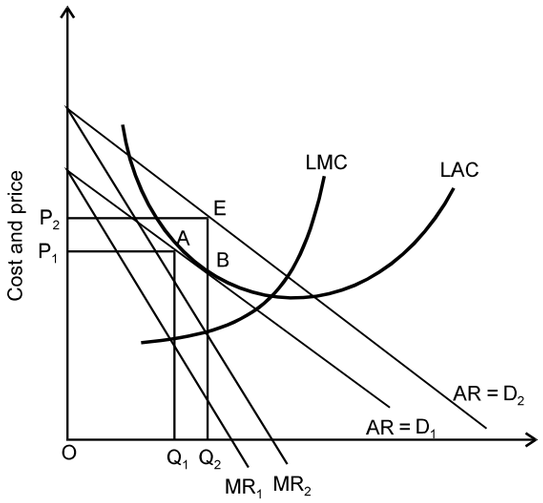

Fig. 4.24 *Introduction of Second Demand Curve*

(ii) Long-run Equilibrium with Price Competition

In his model of long-run equilibrium with price competition Chamberlin assumes that the number of existing firms in the product group is optimal. That is, the number of firms is compatible with long-run equilibrium of the industry. There is no entry or exit of the firms. In this case, the equilibrium analysis has been accomplished in two stages. In the first stage, a second demand curve is introduced to the model to incorporate the effects of competitive change in prices. In the second stage, the long-run equilibrium under the condition of optimal number of firms has been analysed.

To commence the analysis of stage one, let us assume that demand curve for the product of a firm is given by DD_1, as shown in Fig. 4.25, and the firm is in equilibrium at point *E*, with price *OP* and output *OQ*. Now suppose, that the firm contemplates a price reduction and assumes that the rival firms will not react to its price cut. By reducing its price, the firm expects to expand its sales on account of two factors: *first*, the demand for its product is elastic, and *second*, the consumers of other products will switch over to the product of this firm if other firms do not simultaneously reduce their prices. Thus, the firm can expect a substantial increase in its total sales.

For example, if the firm *reduces* its price by *EM*, the demand for its product increases by *MB*. Of this, *MN* is due to the elasticity of its original demand DD_1 and *NB* is due to the fact that some customers of other firms switch over to the product of this firm. On the other hand, if the firm *raises* its price, and the rival firms do not raise their prices it looses its market partly because its demand curve is elastic and partly because its customers switch over to other products which become automatically cheaper. Thus, another demand curve DD_2 emerges. This is called as the *second demand curve*. But, if all the firms change their prices simultaneously this advantage to an individual firm is lost and all the firms return to the original demand curve DD_1.

NOTES

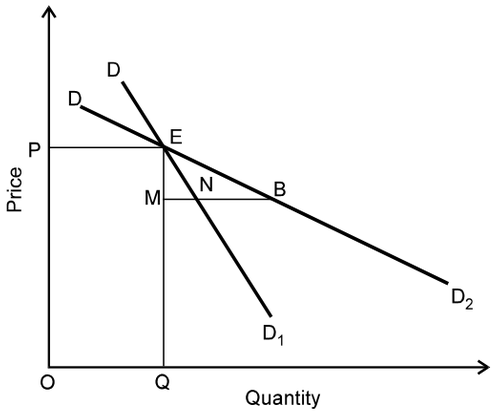

Fig. 4.25 *Introduction of Second Demand Curve*

Having introduced the second demand curve, we can proceed to present Chamberlin's long-run equilibrium with price competition, i.e., *second stage* of his analysis. The long-run equilibrium with price competition is presented in Fig. 4.26. The curves DD' and dd_1 are the two demand curves and LAC is the long-run average cost curve of a 'typical' firm of the group. Let the initial short-run equilibrium of the firms of the group be at point P with price OP_2 and output OQ_1. At this price and output, the firm makes the total abnormal profit represented by the area P_2PMC.

Although in equilibrium, each firm regards dd_1 to be its demand curve and believes that it can increase its profits by reducing the price and expanding the output. The basis of this belief is the elasticity of their assumed demand curve dd_1. In an attempt to increase profits, each firm reduces its price expecting to move along the demand curve dd_1. However, instead of moving along dd_1, the firms move along the market demand curve DD', because all of them reduce their prices simultaneously. However, according to Chamberlin, the firms do not learn from their past experience and each firm sticks to its own belief that the demand curve (dd_1) for its product is more elastic than the market demand curve (DD'). Therefore, the firms go on reducing their prices independently and their assumed demand curve (dd_1) shifts downward. This process continues until the downward shift in dd_1 makes it tangent to the LAC curve, as shown by dd_2. A further reduction in price will make firms incur loss. Therefore, reduction in price below OP_1 is not desirable. Thus, the long-run equilibrium of firms takes place at E, where each firm produces OQ_2 and fixes its price at OP_1.

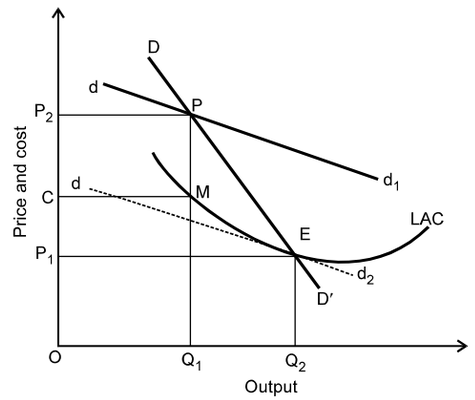

Fig. 4.26 *Long-run Equilibrium Price Competition*

NOTES

(iii) Long-run Equilibrium with Free Entry and Price Competition

We have explained above the equilibrium of the firm with free entry and with price competition separately. In this section, we bring together the two equilibrium analyses, and explain Chamberlin's third and the final model of firms' equilibrium under monopolistic competition. According to Chamberlin, the ultimate equilibrium under monopolistic competition can be achieved through an integrated analysis of effects of free entry and price adjustments. The integrated analysis of equilibrium is presented in Fig. 4.27.

In Fig. 4.27, DD_1 is assumed to represent the initial demand curve and LAC to represent the long-run average cost curve. Let us suppose that the firms are initially in equilibrium at point B, and they make abnormal profits to the extent of vertical distance between DD_1 and the LAC. Since entry to the 'product group' is free, new firms are attracted by the industry. When new firms with slightly differentiated products enter the 'product group', the market share of each existing firm is reduced. Hence, their demand curve DD_1, shifts leftward. Given the LAC, the leftward shift in the demand curve will continue until it becomes tangent to LAC, as shown by DD_3 in Fig. 4.27, because till this point of tangency is reached, firms make abnormal profits and new firms continue to enter the 'product group'.

Thus, it might seem that the long-run equilibrium is attained at point A with output QQ_1 and price OP_3. This is however not the case. This is only half of the story, i.e., the influence of free entry. Let us now consider the competitive manoeuvering of price and its role in determining to the long-run equilibrium.

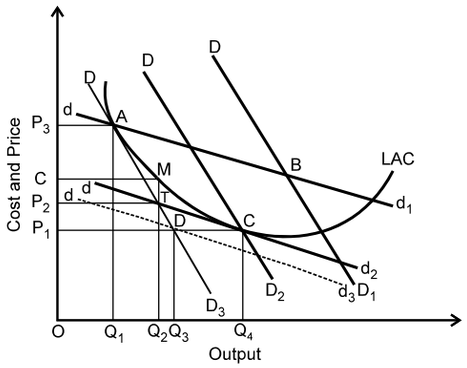

Fig. 4.27 *Long-run Equilibrium with Free Entry and Price Competition*

Once the firms reach point A, each firm thinks that its demand curve is dd_1, not DD_3. Each firm believes that it can increase its profit by reducing the price and thereby increasing the sales. Therefore, in their attempt to increase their profits, they reduce prices of their product simultaneously because each firm has the same incentive to do so. As a result, their subjective demand curve (dd_1) slides downward to dd_2 and they incur losses. For example, if price is reduced to OP_2, the total loss equals the rectangle $CMTP_2$. It might seem that the firms could eliminate their loss by reducing the price to OP_1. But when all the firms reduce their price to OP_1—

and they will do so under the assumption—their subjective demand curve dd_2 slides further down to dd_3, the dotted line that lies below the *LAC*. As a result, the firms make increasing losses. A temporary equilibrium will be attained at point *D* with output OQ_3, where all firms incur heavy losses. Consequently, the firms which are unable to sustain losses will eventually leave the industry. The remaining firms find their share in the market increasing. Therefore, DD_3 and dd_3 move to the right until DD_3 shifts to DD_2 and dd_3 shifts upward to the position of dd_2. Note that dd_2 intersects DD_2 at point *C* the point where dd_2 is tangent to *LAC*. Thus, the long-run equilibrium is attained at point *C* where it is stable because all firms earn only normal profits and, therefore, there is no entry or exit of the firms.

Analysis of Selling Cost and Firm's Equilibrium

As noted above, under monopolistic competition, products are differentiated; market is imperfect; consumers are not fully aware of existence of a particular variety of the product. Therefore, producers always have an opportunity to advertise their product, attract more customers to their product, and create brand loyalties in the minds of consumers and thereby increase their share in the market. In fact, the basic objective of advertising is to change the position and shape of the demand curve for the product of the advertising firm. Whether a firm succeeds in achieving these objectives depends also on the effectiveness of competitive advertising by the rival firms. But, one thing is obvious: advertising involves additional costs which pushes the *AC* curve upward.

Apart from advertisement expenses, monopolistically competitive firms, incur other costs on competitive promotion of their sales, e.g., salary to sales personnel, allowance to dealers, discounts to customers, expenses on displays, gifts to customers and free samples, additional costs on attractive packaging of goods. All such expenses plus advertisement expenditure make a firm's *selling cost*.

Incurring selling cost increases sales, but at a varying degree. In the initial stage, the increase in sales may be greater than the proportionate increase in the selling cost. But eventually, it decreases. Consequently, the unit selling cost or the average selling cost (*ASC*) initially decreases (because of economies of scale) but ultimately increases. The *ASC* curve is therefore U-shaped, similar to the conventional short-run *AC* curve. It implies that total sales are subject to diminishing returns to increasing selling costs.

Selling Costs and Group Equilibrium

To analyse group equilibrium of firms with selling cost, let us recall that the main objective of the firm is to maximize its total profits. When they incur selling costs, they do so with the same objective in mind. All earlier assumptions regarding cost and revenue curves remain the same. The analysis of group equilibrium is presented in Fig. 4.28. Suppose *APC* represents the average production cost and price is given at OP_3. None of the firms incurs any selling cost.

NOTES

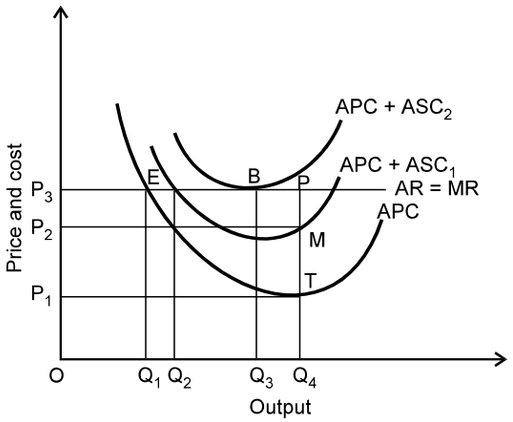

Fig. 4.28 *Selling Costs and Group Equilibrium*

Also, let all the firms be in equilibrium at point *E* where they make only normal profits. Now suppose that one of the firm's incurs selling cost so that its *APC* added with average selling costs (*ASC*) rises to the position shown by the curve $APC + ASC_1$ and its total sales increases to OQ_4. At output OQ_4 the firm makes supernormal profits of P_3PMP_2. This profit is however possible only so long as other firms do not advertise their own products. If other firms do advertise their products and incur the same amount of selling cost, the initial advantage to firm advertising first will disappear and its output will reduce to OQ_2. In fact, all the firms produce only OQ_2 units. But their short-sightedness impels them to increase their selling cost because they expect to reduce their *APC* by expanding their output. With increased selling cost, their $APC + ASC$ curve shifts further upward. This process continues until $APC + ASC_2$ becomes tangent to the $AR = MR$ line at point *B*. Beyond point *B*, advertising is of no avail to any firm, even if other firms do not advertise. The equilibrium will be stable at point *B* where each firm produces OQ_3. Note that the equilibrium output OQ_3 is greater than the initial output of OQ_1. In equilibrium, however, firms make only normal profits.

A Critical Appraisal of Chamberlin's Theory of Monopolistic Competition

Chamberlin's theory of monopolistic competition has been criticised on both theoretical and empirical grounds. Let us first look into its theoretical or methodological weaknesses.

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

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

NOTES

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

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

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

Empirical validity: So far as empirical validity of Chamberlin's concept of monopolistic competition is concerned, it is claimed that it is difficult to find any example in the real world to which his model of monopolistic competition is relevant. Most markets frequently available in the real world may be classified under prefect competition, oligopoly or monopoly. It is therefore alleged that Chamberlin's model of monopolistic competition analyses an unrealistic market. Some economists, e.g. Cohen and Cyert, hold the position that the model of monopolistic competition is not a useful addition to economic theory because it does not describe any market in the real world. Despite above criticism, Chamberlin's contribution to the theory of price cannot be denied. Chamberlin is first to introduce the concept of *differentiated product* and *selling costs* as a decision variable and to offer a systematic analysis of these factors. Another important contribution of Chamberlin is the introduction of the concept of demand curve based on market share as tool of analysing behaviour of firms, which later became the basis of the *kinked-demand* curve analysis.

Check Your Progress

8. List the main features of monopolistic competition.
9. Mention the basic assumptions designed by Chamberlin for developing his theory of monopolistic competition under long-run conditions.

4.6 PRICE AND OUTPUT DETERMINATION FOR FIRM AND INDUSTRY UNDER OLIGOPOLY

Oligopoly is a form of market structure in which a *few* sellers sell differentiated or homogeneous products. 'How *few* are the sellers' is not easy to define numerically in the oligopolistic market structure. The economists are not specified about a definite number of sellers for the market to be oligopolistic in its form. It may be two, three, four, five or more. In fact, the number of sellers depends on the size of the market. Given the size of the market, if number of sellers is such that each seller has command over a sizeable proportion of the total market supply then there exists oligopoly in the market.

NOTES

The products traded by the oligopolists may be differentiated or homogeneous. Accordingly, the market may be characterised by ***heterogeneous oligopoly*** or ***homogeneous (or pure) oligopoly***. Cooking gas of Indane and of Burshane are the examples of homogeneous oligopoly. Differentiated oligopolies include automobiles, cigarettes, refrigerators, TV industries. Pure oligopoly includes such industries as cooking gas, cement, baby food, cable wires, dry batteries, etc. Other examples of oligopolistic industries are aluminium, paints, tractors, steel, tyres and tubes.

4.6.1 Characteristics of Oligopoly

The basic characteristics of oligopolistic market structure are the following:

1. **Intensive Competition:** The characteristic fewness of their number brings oligopolist in intensive competition with one another. Let us compare oligopoly with other market structures. Under perfect competition, competition is non-existent because the number of sellers is so large that no seller is strong enough to make any impact on market conditions. Under monopoly, there is a single seller and, therefore there is absolutely no competition. Under monopolistic competition, number of sellers is so large that degree of competition is considerably reduced. But, under oligopoly, the number of sellers is so small that any move by one seller immediately affects the rival sellers. As a result, each firm keeps a close watch on the activities of the rival firms and prepares itself with a number of aggressive and defensive marketing strategies. To an oligopolist, business is a 'life' of constant struggle as market conditions necessitate making moves and counter-moves. This kind of competition is not found in other kinds of market. ***Oligopoly is the highest form of competition.***
2. **Interdependence of Business Decisions:** The nature and degree of competition among the oligopolists makes them interdependent in respect of decision-making. The reason for inter-dependence between the oligopolists is that a major policy change made by one of the firms affects the rival firms seriously and immediately, and forces them to make counter-moves to protect their interest. Therefore, each oligopolist, while making a change in his price, advertisement, product characteristics, etc. takes it for granted that his actions will cause reaction by the rival firms. Thus, interdependence is the source of action and reaction, moves and counter-moves by the competing firms. An illuminating example of strategic manoeuvering by the oligopoly firm has been given by Robert A. Meyer. To quote the example, one of the US automobile companies announces in September an increase of $180 in the list price of its new car model. Following it, a few days later, a second company announces an increase of only $80 and a third announces increase of $91. The first company makes a counter-move: it suddenly reduces the increase in list price to $71 from $180 announced earlier. One can now expect that other firms will follow the first in price-cutting. Obviously, there is a good deal of uncertainty in the behaviour of firms.

3. **Barrier to Entry:** An oligopolistic market structure is also characterized, in the long run, by strong barriers to entry of new firms to the industry. If entry is free, new firms attracted by the super-normal profits, if it exists, enter the industry and the market eventually becomes competitive. Usually barriers to entry do exist in an oligopolistic market. Some common barriers to entry are economies of scale, absolute cost advantage to old firms, price-cutting, control over important inputs, patent rights and licencing, preventive price and existence of excess capacity. Such factors prevent the entry of new firms and preserve the oligopoly.

4.7 ANSWERS TO 'CHECK YOUR PROGRESS'

1. The term market structure refers to the organizational features of an industry that influence the firm's behaviour in its choice of price and output. Market structure is an economically significant feature of the market. It affects the behaviour of firms in respect of their production and pricing behaviour.
2. The kinds of market structure are the following:
   (i) Perfect competition
   (ii) Imperfect competition
   (iii) Monopolistic competition
   (iv) Oligopoly
   (v) Monopoly
3. Monopoly is antithesis of perfect competition.
4. Market price in a perfectly competitive market is determined by the market forces, viz., demand and supply. Here, market demand refers to the demand for the industry as a whole. It is equal to the sum of the quantity demanded by the individuals at different prices.
5. Perfect competition model has been the most popular model used in economic theories due to its analytical value.
6. The important features of a monopoly are the following:
   (i) There is a single seller of a product which has no close substitute.
   (ii) A monopoly firm is a price maker, not a price taker.
   (iii) Under monopoly, there is absence of supply curve.
   (iv) A monopoly makes a single-firm industry.
7. Two important misconceptions about monopoly firm are as follows:

   One of the misconceptions is that a monopoly firm necessarily makes super normal profits. There is however no guarantee that monopoly firm will always make profits in the short-run. In fact, whether a monopoly makes profits or losses in the short-run depends on its revenue and cost conditions.

NOTES

Another common misconception about monopoly is that the demand curve faced by a monopoly firm is perfectly inelastic so that it can charge any price it likes. In fact, the demand curve faced by a monopolist is both firm's and industry's demand curve.

8. The main features of monopolistic competition are the following:
   - Under monopolistic competition, the number of firms is fairly large.
   - Firms produce and sell differentiated products-each product being a close substitute for another.
   - There is free entry and free exit of firms to and from the industry.
   - Factors of production enjoy free mobility, at least theoretically.
9. Chamberlin has made the following explicit and implicit assumptions in order to develop his theory of monopolistic competition under long-run conditions.

   First, the basic assumption is that there is a large number of firms producing differentiated products which are close substitutes for one another.

   Second, the number of firms in the product group is so large that each firm expects its manoeuvering of prices and output to go unnoticed by the rival firms.

   Third, one of the heroic assumptions of Chamberlin is that both demand and cost curves for all the products are uniform throughout the group. That is, all firms have identical revenue and cost curves.

   Fourth, his second heroic assumption is that consumer's preferences are evenly distributed among the different varieties, and that differences between products are not such as to give rise to differences in cost.

4.8 SUMMARY

- The modern theory of price came into existence during the 1930s with Joan Robinson's The Economics of Imperfect Competition and Edwin H. Chamberlin's The Theory of Monopolistic Competition, both written independently in 1933.
- A market is one of many varieties of systems, institutions, procedures, social relations and infrastructures whereby parties engage in exchange. It refers to any structure that allows buyers and sellers to exchange any type of goods, services and information.
- The term market structure refers to the organizational features of an industry that influence the firm's behaviour in its choice of price and output. Market structure is an economically significant feature of the market. It affects the behaviour of firms in respect of their production and pricing behaviour.

NOTES

- Perfect competition is a market situation in which a large number of producers offer a homogeneous product to a very large number of buyers of the product.
- Imperfect competition creates two different forms of markets with different number of producers and with different degrees of competition, classified as (a) monopolistic competition, (b) oligopoly (c) monopoly.
- The equilibrium price of a commodity in a free market is determined by the market forces of demand and supply.
- Market mechanism is a process by which market forces of demand and supply interact to determine equilibrium price.
- A perfectly competitive market is characterized by complete absence of rivalry among the individual firms. In fact, under perfect competition as conceived by the economists, competition among the individual firms is so widely dispersed that it amounts to no competition.
- According to the traditional theory of firm, a firm is in equilibrium when its profit is maximum. Maximization of profits depends on the revenue and cost conditions.
- Just as the market demand curve is a horizontal summation of individual demand curves, the industry supply curve or market supply curve is the horizontal summation of the supply curves of the individual firms.
- The term pure monopoly signifies an absolute power to produce and sell a product which has no close substitute. In other words, a monopoly market is one in which there is only one seller of a product having no close substitute.
- In the short-run, cost conditions faced by a monopoly firm are, for all practical purposes, identical to those faced by a firm under perfect competitions, particularly when a monopoly firm is a competitive buyer in the input market.
- Monopoly causes loss of social welfare and distortions in resource allocation. The loss of social welfare is measured in terms of loss of consumer's surplus.
- The degree of price discrimination refers to the extent to which a seller can divide the market and can take advantage of it in extracting the consumer's surplus.
- The existence of monopolies in a market economy is criticised on the grounds that they restrict production and consumption, widen income and wealth disparities, exploit consumers and employees, cause distortions in allocation of resources, reduce the prospect of employment, and cause loss of social welfare.
- The model of monopolistic competition developed by Edward H. Chamberlin presents a more realistic picture of the actual market structure and the nature of competition.

NOTES

- *Monopolistic Competition* is a market structure in which a *large number* of sellers sell *differentiated* products which are close, but not perfect, substitutes for one another. Monopolistic competition combines the characteristics of perfect competition and monopoly.

4.9 KEY TERMS

- **Market:** It refers to any structure that allows buyers and sellers to exchange any type of goods, services and information.
- **Perfect competition:** It is a market situation in which a large number of producers offer a homogeneous product to a very large number of buyers of the product.
- **Monopolistic competition:** It is a kind of market in which a large number of firms supply differentiated products.
- **Market mechanism:** It is a process by which market forces of demand and supply interact to determine equilibrium price.
- **Discriminatory monopoly:** When a monopolist sells the same product at different prices to different buyers, the monopoly is called a discriminatory monopoly.

4.10 SELF-ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. Prepare an overview of the market structure.
2. What are the barriers to entry in markets?
3. What is the difference between perfect and pure competition?
4. Why are monopoly firms considered less efficient than competitive firms? Give reasons for your answer.
5. Mention the necessary conditions for price discrimination.

Long-Answer Questions

1. Briefly mention the mathematical analysis of determination of equilibrium price.
2. Elaborate the assumptions of perfect competition.
3. Discuss the major sources of barriers to entry to a monopolised market.
4. Explain the nature of revenue curves under monopoly.

5. Discuss the various measures of monopoly power.
6. Examine price and output determination under monopolistic competition.
7. Critically analyse Chamberlin's theory of monopolistic competition.

NOTES

4.11 FURTHER READING

Dwivedi, D. N. 2016. *Microeconomics: Theory and Application*, 3rd Edition. New Delhi: Vikas Publishing House.

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

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

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

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

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

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

UNIT 5 FACTOR PRICING AND FACTOR MARKET

NOTES

Structure

5.0 INTRODUCTION

In this unit, we will begin our discussion with theory of factor pricing. Theories of rent, interest and profit have been discussed in this unit in detail. Since factor demand is an important determinant of factor price, let us look at the factors that determine the demand for a factor. In our analysis of demand for factors, we will consider demand for labour. However, the analysis can be applied to other factors also.

NOTES

5.1 OBJECTIVES

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

- Explain the theory of factor pricing
- Discuss the marginal productivity theory
- Define the adding-up theorem
- Elaborate on the modern theories of rent, wages, interest and profit with examples

5.2 FACTOR PRICING

The theory of factor pricing has its origin in the *classical theory of distribution*. The theory of distribution explains how national income (output) is distributed among the various factors of production. Distribution of national income between the factors of production is called *functional distribution* of national income, i.e., distribution of national income on the basis of economic functions of the people. Classical economists, Adam Smith, Ricardo and Marx were concerned with the functional distribution of national income. The basic classical proposition in this regard is that the share of a factor of production in the national income equals the *average price* of a factor multiplied by its *total employment*. For example, assuming there are only two factors, labour (L) and capital (K) with a given wage rate (w) and interest rate or capital rental (r), the share of labour and capital in the national output (Q) can be determined as follows.

$$\text{Share of labour} = \frac{w \cdot L}{Q}$$

$$\text{Share of capital} = \frac{r \cdot K}{Q}$$

and

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

In determining the functional distribution of national income, classical economists had to find how *factor price* and *employment* are determined. This led to the emergence of classical theories of *factor pricing* and *employment*. In this part of the book, we will discuss the theories of factor pricing. Before proceeding further, let us understand the concept of factor demand and factor supply—the basis of the theory of factor pricing.

Factor Demand and Supply

As noted above, the theory of factor pricing is fundamentally the same as the theory of product pricing. Both product and factor prices are essentially determined by the interaction of market forces—demand and supply forces. In case of factor pricing, demand and supply forces refer to demand for and supply of factors of production. As in case of product pricing, the basic function of the theory of factor pricing is to explain the determinants of demand for and supply of a factor and to show how demand and supply interact to determine the price of a factor. However,

NOTES

both demand and supply of factors have *certain specific nature* as compared to demand for and supply of goods. The fact is that the peculiarities of factors of production necessitate formulating separate theories of factor pricing. Let us look at some specific nature of demand for and supply of factors of production.

On the demand side, there is a difference between the demand for a product and the demand for a factor. While demand for consumer goods is, in general, *autonomous demand* based on the utility of the goods, demand for factor services is a *derived demand*. This aspect is discussed below in detail.

On the supply side, factors of production—labour, land and capital—differ in their character in respect of their supply, especially land and labour. Land is a gift of nature and labour is partly a natural factor and partly a man-made factor. Capital is a man-made factor. For this reason, perhaps, early economists developed theories of factor pricing with respect to each factor separately, known as:

(i) theory of wages,
(ii) theory of rent,
(iii) theory of interest, and
(iv) theory of profit.

Let us now take a view of the development of factor pricing theory.

Development of Theories of Factor Pricing

The theories of factor pricing were developed by different classical economists at different points of time, and are, therefore, different in approach to factor price determination. *Theory of wages*, often called as *marginal productivity theory of wages*, is attributed to John Bates Clark, an American economist, popularly known as J.B. Clark. J. B. Clark had developed the marginal productivity theory in his book, *The Distribution of Wealth* (1899, p. VIII). However, several other economists had earlier hinted at the concept of marginal productivity, including Von Thunen in his *Der Isolierte Staat* (1826), Longfield in his *Lectures on Political Economy* (1834), and Henry George in his *Progress and Poverty* (1879). Besides, in 1880s and 1990s, many other economists had made important contributions to the concept of *marginal productivity.* Important contribution were made by W. Stanley Jevons in his '*The Theory of Political Economy*'; Philip Wicksteed in his '*An Essay on the Coordination of the Theory of Distribution*'; Alfred Marshall in his '*Principles of Economics*'; Stuart Wood in his '*The Theory of Wages*'. (*AEA*, No 4, 1889); and Leon Walras in his '*Elements d'economie politique pure*' and also by some others. 'The classical theory of rent' was propounded by David Ricardo. Several classical economists have contributed to 'The Theory of Interest'. Similarly, several classical economists, including J.B. Clark, have contributed to 'The Theory of Profit'.

Factor Demand: Why a Factor is Demanded

As noted earlier, the price of factor is determined by the market forces—demand and supply. Before we discuss the theory of factor price determination, let us know: (i) why a factor of production is demanded, and (ii) what determines the demand for a factor. These aspects are discussed here briefly.

Factor Demand is a Derived Demand

A factor of production is demanded because it is productive and what it produces is demanded by the people, the consumers and producers. In other words, demand for a factor depends on its productivity and the demand for its product. A factor which is not productive—be it land, labour or capital—is not demanded for its own sake. For example, infertile land is not demanded for cultivation; an uneducated labour is not demanded for teaching; and an outdated machinery is not demanded because it cannot produce goods and services in demand.

However, productivity of a factor alone is not sufficient for its demand. For example, a labour may be very skilled and efficient in making hand fans but its service is not demanded in cities because hand fans are not demanded there. Similarly, services of a computer software expert in MS DOS are no more in demand because this software has been replaced by Windows. It implies that *what a factor can produce must be in demand*. It means that demand for a factor depends on the demand for its product. That is why **demand for a factor a derived demand***:* factor demand arises because there is demand for goods and services it can produce. For example, farm labour is demanded because people demand wheat, rice, sugar and other farm products; construction labour is demanded because people demand house and so on.

5.2.1 Marginal Revenue Productivity: The Determinant of Factor Demand

The productivity of a factor and demand for its product create the demand for the factor. But these factors do not determine the *quantity demanded* of a factor. Theoretically, the quantity demanded of a factor by a profit-maximizing firm is determined by the revenue it yields to the firm. From factor demand point of view, what matters is its *marginal revenue productivity (MRP)* and *the price of the factor*. The *MRP* of a factor can be defined as *the revenue that a firm can earn by employing one additional unit of a factor*. The concept of *MRP* can be clarified further with reference to a specific factor, e.g., labour. When a firm employs an additional unit of labour, it produces a certain quantity of a good. The quantity produced by the marginal labour is called *marginal physical productivity of labour* (MPP_L). Now, the *MRP* of labour (MRP_L) can be defined as follows.

$$MRP_L = MPP \cdot \bar{P}$$

where MPP_L = *marginal physical productivity of labour*, and $\bar{p}$ = constant price of the product.

Here, we present a simplified version of the MPP_L curve and show the derivation of MRP_L, under the following assumptions.

1. Labour is the only variable factor of production, and
2. MPP_L is subject to the law of diminishing return.

Table 5.1 shows the computation of and relation between MPP_L and MRP_L, assuming diminishing returns to a variable factor (labour). As the table shows, as more and more units of labour are employed, total production (*TP*) increases but at diminishing rates. The increase in *TP* at decreasing rate implies that the marginal physical productivity (MPP_L) of labour decreases with increase in employment of

labour. Column 3 shows the diminishing MPP_L. Column 4 shows the constant price and column 5 shows the marginal revenue productivity (MRP_L).

The MPP_L and MRP_L curves given in panels (a) and (b) Figure 5.1 have been derived by using the MPP_L and MRP_L data in Table 5.1. As the figures show, on the assumption of diminishing returns, both MPP_L and MRP_L curves have a downward slope. The MRP_L curve provides the basis of derivation of the demand curve of labour. That is why the *marginal productivity theory* is called the theory of factor demand. We now move on to discuss the derivation of demand curve for a factor of production (labour) through marginal productivity theory.

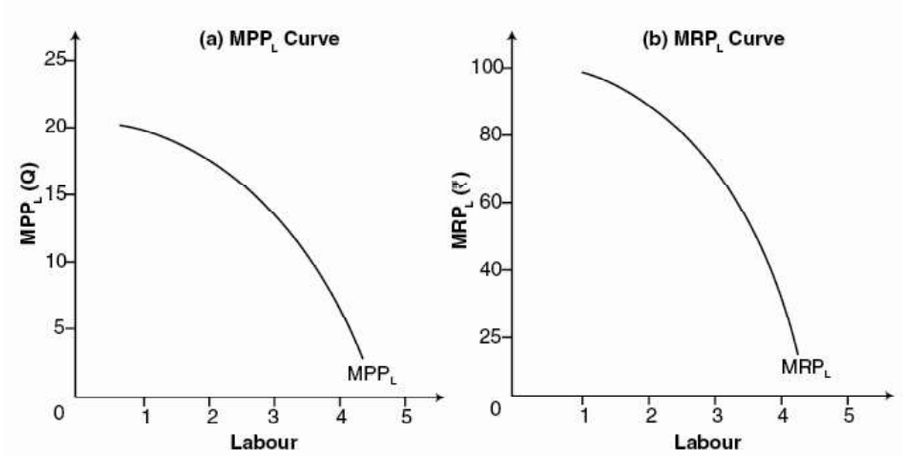

Fig. 5.1 *Derivation of MPP_L and MRP_L Curves*

Table 5.1 *Marginal Physical and Revenue Productivity of Labour*

Variable Factor (*L*)	Total Production (*TP*) (Units)	MPP$_L$ (Units)	*P* (Price) (Constant)	$MRP_L = MPP_L \times P$ (₹)
(1)	(2)	(3)	(4)	(5)
1	20	20	5	100
2	38	18	5	90
3	52	14	5	70
4	58	6	5	30
5	58	0	5	0

5.3 MARGINAL PRODUCTIVITY THEORY

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

NOTES

is widely associated with his name. Clark was however not the lone contributor to the development of this theory. In the last quarter of the 19th century, many other economists, particularly, Jevons, Wicksteed, Marshall, Stuart, Walras and Barone made significant contributions in developing the marginal productivity concept.

According to Clark, the marginal productivity principle is a complete theory of wages, which could be well applied to other factors of production also. However, many theorists including Marshall and Hicks have objected to the marginal productivity theory being regarded as theory of wages or as theory of distribution. Marshall argued that, although the marginal productivity theory 'has sometimes been put forward as a theory of wages… [there] is no valid ground for any such pretention.' For, according to him, 'demand and supply exert equally important influence on wages; neither has a claim to predominance any more than has either blade of a scissors…'

Marginal productivity has nothing to say about the supply side of the market. Hicks holds a similar view. One might argue that, in the short-run, supply of labour is fixed and hence wages are governed only by the demand for labour. But if labour is defined not as an individual worker but as a work-effort per unit of time, then the labour supply can by no means be regarded as perfectly inelastic. As such, the marginal productivity theory alone is unable to specify the hourly wage rate in the market. Thus, economists have objections to the validity of marginal productivity theory even in the short-run. It is suggested that supply of factor (or inputs) must enter the theory to develop a complete theory of distribution. Marshall-Hicks approach suggests that both demand for and supply of a factor are equally important in the theory of factor price determination.

Strictly speaking, marginal productivity theory offers only a theory of demand for a factor of production. The marginal productivity theory provides an analytical framework for deriving the demand for a factor which is widely used in modern economic analysis. Let us now explain the derivation of demand curve for a factor (say labour) on the basis of marginal productivity theory. Though we consider only labour for deriving the demand for a factor, the analysis is relevant to all other factor services.

Marginal Productivity and Factor Demand

Demand for a factor is a derived demand: It is derived on the basis of the marginal productivity of a factor. Firms demand factors of production—land, labour, capital—because they are productive. Factors are demanded not merely because they are productive but also because the resulting product has a market value. Thus, demand for a factor of production depends on the existence of demand for the goods and services that a factor of production can create. For example, land is demanded because there exists demand for what land can produce—foodgrains, vegetables, fruits, etc. Doctors, engineers, managers are in demand because there is a demand for their services. Services of coalminers are demanded because there exists demand for coal. It is in this sense that the demand for a factor of production is a derived demand.

NOTES

The demand curve for factors is derived under two different conditions:

(i) when a single variable factor is used in the process of production; and

(ii) when more than one variable factor is used in the production process.

Let us first derive the demand curve for a variable factor assuming that only one factor is involved in production process. We will next derive the demand curve for a variable factor when several factors are used in the production process.

For deriving the demand curve for a factor, we assume that *labour* is the only variable factor. One may however assume any other factor. The process of deriving the demand curve for labour, presented in this section, can be applied to any other factor also. Furthermore, we outline, in this section, the theory of wage determination under perfectly competitive conditions. The same theory can be applied to explain the determination of price of all other factors.

Demand for a Single Variable Factor: Labour

The demand for a variable factor depends on the value of its marginal productivity. Therefore, we shall first derive the *value of marginal productivity* (*VMP*) curve of labour. The curve MP_L shows diminishing marginal returns to the variable factor—labour. The MP_L at each level of labour employment multiplied by a *constant* price P_x, gives the *value of marginal physical product curve*, as shown by the curve $VMP_L = MP_L . P_x$. It is this curve which is the basis of demand curve for labour. Now the question is: how does the VMP_L curve help in deriving the demand curve for labour? Derivation of labour demand curve is shown in Fig. 5.2.

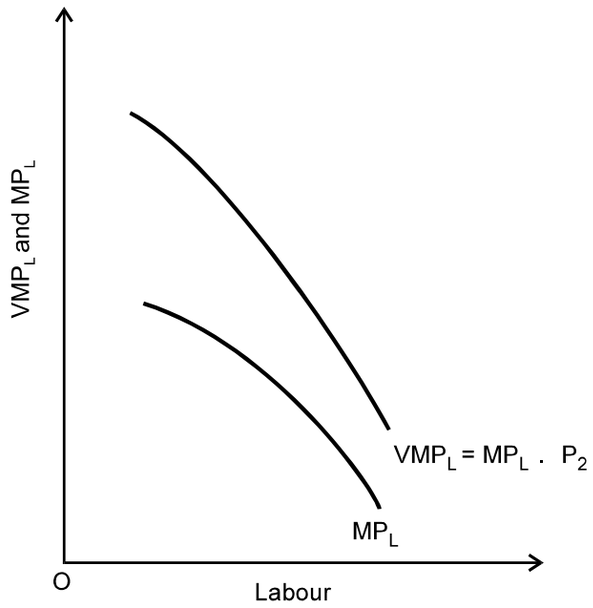

Fig. 5.2 *MP_L and VMP_L Curves*

Derivation of a Firm's Labour Demand Curve

Before we proceed to derive a firm's demand curve for labour, let us make the following assumptions for the sake of simplicity in the analysis.

NOTES

(a) Firm's objective is to maximise profit.

(b) The firm uses a single variable factor, labour, whose market is perfectly competitive and hence the price of labour, wages (w), is constant and given for all the firms. This implies (i) that supply of labour for an individual firm is perfectly elastic; and (ii) that firm's $MC = w$ which is constant.

(c) The firm produces a single commodity, say X, whose price is constant at P_x.

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

The short-run equilibrium of the profit-maximising firm is illustrated in Fig. 5.3. The VMP_L curve shows the value of marginal product of labour, the only variable factor. The S_L lines present the labour supply curves for an individual firm [assumption (b)], at the constant wage rates. The VMP_L curve and SL_3 line intersect each other at point E_3, where $VMP_L = W_3$, that is, where the value of marginal product equals marginal cost of labour. The profit-maximising firm will therefore employ only OL_1 units of labour. By employing OL_1 units of labour, the firm maximises its profit. Given these conditions, any additional employment of labour will make $W_3 > VMP_L$. Hence, the total profit will decrease by $W_3 - VMP_L$. Similarly, if one unit less of labour is employed, VMP_L will be greater than W_3 and the total profit is reduced by $VMP_L - W_3$. Thus, given the VMP_L and SL_3, the profit maximising firm will demand only OL_1 units of labour.

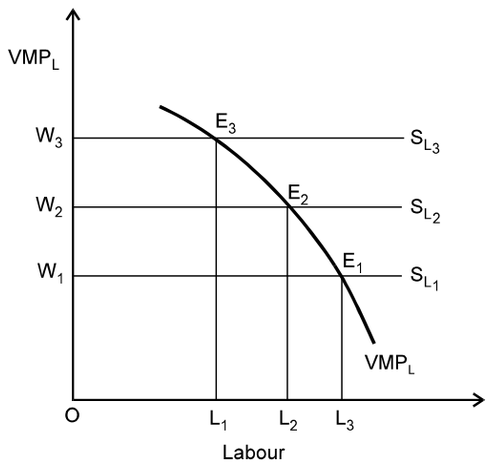

Fig. 5.3 *MP_L and VMP_L Curves*

The aforementioned analysis can be extended to derive the firm's demand curve for labour. If wage rate falls to OW_2 firm's equilibrium point shifts from point E_3 to E_2 increasing the demand for labour from OL_1 to OL_2. Similarly, when wage rate falls to OW_1, firm's equilibrium shifts downward to E_1 causing an increase in the demand for labour to OL_3. To summarise, when wage rate is OW_3, demand

for labour OL_1; when wage rate falls to OW_2, demand for labour increases to OL_2; and when wage rate falls further to OW_1, labour demand increases to OL_3. Obviously, as wage rate falls, demand for labour increases. This relationship between the wage rate and labour demand gives a usual downward sloping demand curve for labour, which is, by definition, the same as VMP_L curve. It may now be concluded that *individual demand* curve for a single variable factor (e.g., labour) is given by its value of marginal product curve (VMP_L) or its marginal revenue product curve (MRP_L).

Demand for Labour with More than One Variable Factor: Two Factor Case

When more than one variable factor is used by a firm in the process of production, the *VMP* curve of a variable factor is not the firm's demand curve for the factor. The reason is that *VMP* curve is drawn on the assumption that there is only one factor, other factors held constant. When more than one factor is variable, a change in the price of one variable factor leads to a change in the demand for other factors through its three kinds of different effects, viz., **substitution effect, output effect** and **profit effect.**

Measuring Substitution and Output Effects

Substitution effect arises due to change in relative factor prices. If price a one factor changes, relative prices of factors change: while one actor becomes relatively cheaper, others become relatively costlier. The profit maximising firms then substitute cheaper factor for the costlier ones. Therefore, the firm's demand curve for a factor will be more elastic than the *VMP* curve. How greater is the elasticity of substitution depends on the ease with which one factor can be substituted for another. The ease with which one factor can be substituted for another depends on the technical condition of production.

The influence of *substitution effect* on the demand for a variable factor is illustrated in Fig. 5.4 by using *isoquant* and *isocost* curves. For the sake of convenience, let us assume (i) that there are only two variable factors, labour (L) and capital (K) used in the production process of a commodity; (ii) that the initial wage rate is given as w_1, and price of capital as r_1; and (iii) that initial isocost is given by AB and initial isoquant by Q_1.

Suppose that the firm is in equilibrium at point P where it uses OL_1 of labour and OK_2 of capital to produce Q_1 units of a commodity. Now let the wage rate fall to w_2 so that the new isoquant is AD. Consequently, the firm moves on to a new equilibrium point R at isoquant Q_2. The movement from P to R indicates increase in use of labour by L_1L_3. This increase in labour employment is the result of *substitution effect* and *output effect* of change in wage rate. Thus, the movement from P to R can be decomposed into *substitution* and *output* effects. The two effects can be split by drawing an imaginary isocost ($A'B'$) parallel to the isocost AD and tangent to the original isoquant Q_1. The significance of the isocost $A'B'$ is that it removes the **output effect** of the fall in wage rate and specifies the *substitution effect.* It tells how much of labour is substituted for how much of capital. As can be seen from Fig. 5.4, the movement from P to Q on the isoquant

Q_1 shows substitution of L_1L_2 units of labour for K_1K_2 units of capital or MQ units of labour for PM units of capital. It is clear that L_1L_2 is the substitution effect of decrease in price of labour.

NOTES

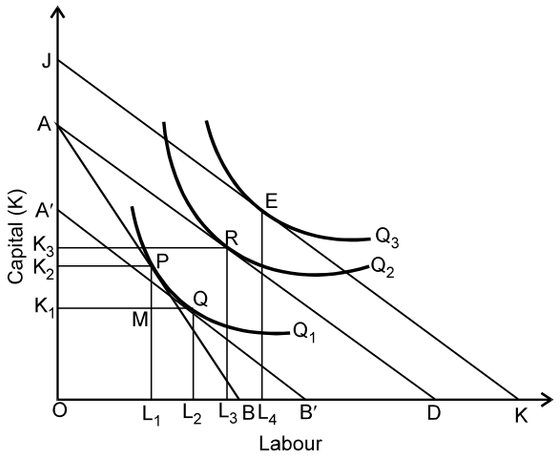

Fig. 5.4 *Effects of Change in Wage Rate*

The **output effect** can now be obtained as $L_1L_3 - L_1L_2 = L_2L_3$. Thus, the movement from point Q to R is the *output effect.* Since the firm will ultimately settle at equilibrium point R, it will use more of both labour and capital. The increase in capital employment with the increase in labour increases marginal product of labour. Consequently, the VMP_L shifts to the right due to output effect.

Measuring Profit Effect

Let us now return to the **profit effect.** The movement from point P to R (Fig. 5.6) accounts for only *substitution* and *output effects.* These effects do not account for the *profit effect* of fall in wage rate. The price effect of a fall in wage rate is illustrated in Fig. 5.5. The profit effect arises from the downward shift in the MC curve due to fall in wage rate as shown in Fig. 5.5. For example, given the commodity price at OP and the marginal cost curve MC intersecting at point E determine the profit maximising output at OQ. Let us suppose that output Q is the same output level indicated by Q_2 (Fig. 5.4). When wage rate falls, the MC curve shifts downward to the position of MC'. In a perfectly competitive market, the equilibrium shifts from E to E' and the profit maximising output increases from OQ to OQ'. The profit maximising firm will therefore have to increase its output by QQ'. The expansion of output requires additional expenditures on labour and capital. The increase in expenditure will make the isoquant AD shift upward to JK (Fig. 5.4), and the firm will be finally in equilibrium at point E at isoquant Q_3. At equilibrium E, the total demand for labour is OL_4 of which L_2L_3 is the additional demand for labour caused by the output effect, and L_3L_4 is the **profit maximisation effect.** The output and profit maximisation effects, both being positive, lead to additional employment of capital. Thus, the employment of both labour and capital increases simultaneously as a result of output and profit maximisation effects. Increase in capital together with labour leads to increase in the marginal product of labour (MP_L). Consequently, in the

final analysis, the decrease in wage rate causes a rightward shift in the VMP_L curve.

NOTES

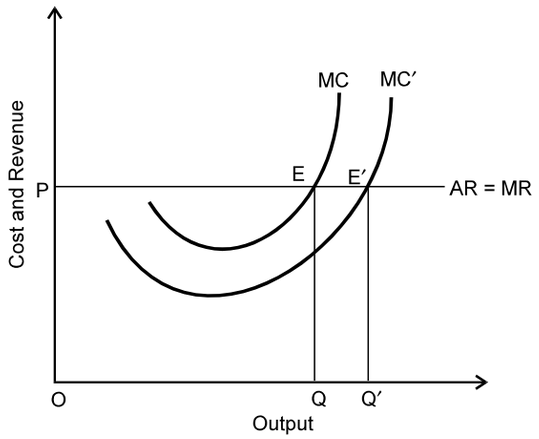

Fig. 5.5 *Profit Maximisation Effect*

Derivation of Labour Demand Curve

The derivation of demand curve for a variable factor (labour) when other factors are also variable is illustrated in Fig. 5.6. Let us suppose that the profit maximising firm is in equilibrium initially at point E_1 where VMP_{L_1} is intersected by the SL_2 line. Given the wage rate OW_2, the firm will employ OL_1 units of labour. Let the wage rate now fall to OW_1 so that the new labour-supply line for the individual firm is SL_1. Had labour been the only variable factor, the firm would have employed OL_2 units of labour. But, under the condition that both the factors, labour and capital, are variable, the fall in wage rate will make the VMP_{L_1} curve shift rightward to VMP_{L_2} as a result of its output and profit maximisation effects. The VMP_{L2} intersects the new labour to supply curve SL_1 at point E_2. The point E_2 is therefore be the equilibrium point after the fall in wage rate. A similar analysis may be repeated for further fall in the wage rate, generating new corresponding equilibrium points. By joining the equilibrium points E_1 and E_2, we get the demand curve DD_L for the variable factor (labour).

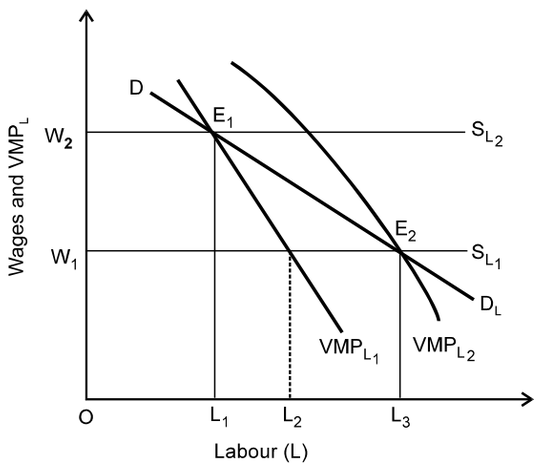

Fig. 5.6 *Labour Demand Curve when both L and K are Variable*

NOTES

Market Demand Curve for a Variable Factor: Labour

As first approximation, the market demand for a variable factor is the horizontal summation of the individual demand curves for the factor in question. However, a straightforward horizontal summation of individual demand curves for a factor does not yield its market demand curve because this process leaves out the *market* or *external* effects of change in the factor prices. The market or external effects are the effects of change in the price of a variable factor on the price of the commodity which it produces and its repercussions on the demand for the factor. In a purely competitive market, when the price of a variable factor (say, labour) decreases, all the firms employ more of labour and hence the supply of the commodity they produce (say X) increases. As a result, the supply curve of the commodity X, shifts rightward which leads to a fall in the price of the commodity, P_x. The fall in P_x causes a decrease in the VMP_L. For example, if P_{x1} and P_{x2} are the two original and new prices, respectively, $VMP_{L_1} = P_{x1}.MPP$, and $VMP_{L_2} = P_{L_2}.MPP$, then $VMP_{L_1} > VMP_{L_2}$ since $P_{x1} > P_{x2}$. Thus, the fall in P_x causes VMP_L to shift downward. Therefore, market demand curve for a variable factor cannot be obtained directly by summing the individual demand curves for it.

The derivation of market demand curve for a variable factor is illustrated in Fig. 5.7. Suppose that curve dd_1 in Fig. 5.7(a) is a typical individual demand curve for labour, and the initial wage rate is given at OW_2. The profit maximising firm is in equilibrium at point P and employs Ol_1 units of labour. Supposing all firms employ the same (Ol_1) units of labour, the market demand for labour can be obtained by multiplying Ol_1 with the number of firms which, let us suppose, equals OL_1 in Fig. 5.7(b). Thus, we get point J as one point of market demand curve for the labour. Now let the wage rate fall to OW_1. Other things remaining the same, the firms would move down to point T on demand curve dd_1 and their employment of labour increases to OL_3. But other things will not remain the same. The increase in labour employment by all firms—each using l_1l_3 additional units of labour—the total supply of the commodity increases. Consequently, commodity supply curve would shift rightward causing a fall in commodity price, P_x. Following the fall in P_x, VMP_L shifts leftward to dd_2. So the new demand curve is dd_2 and the new equilibrium point is M where demand for labour is Ol_2 [Fig. 5.7(a)]. Thus, market demand for labour equals Ol_2 multiplied by the number of firms. Suppose the total market demand for labour thus obtained equals OL_2 [Fig. 5.7(b)] we get a new point, K, which represents the market demand for labour at wage rate OW_1. By joining points J and K, we obtain the market demand curve for labour.

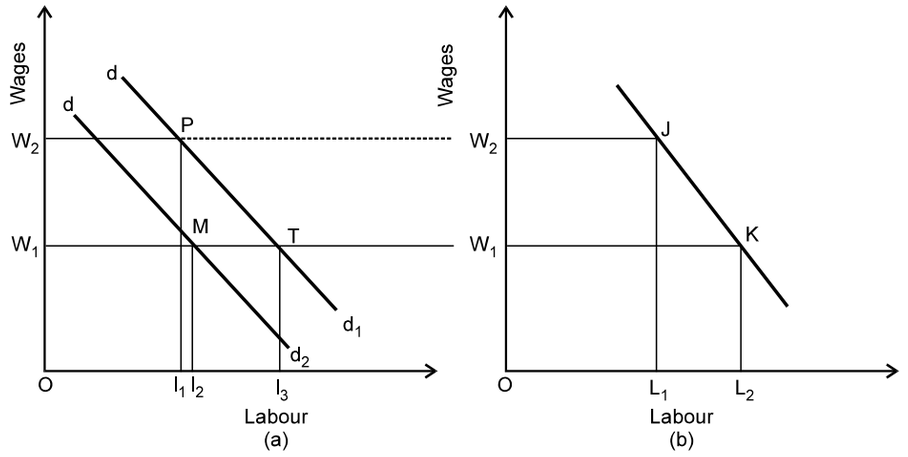

Fig. 5.7 *Derivation of Market Demand Curve for a Variable Factor: Labour*

The Factor Supply: An Overview

We have explained in the previous section how demand curve for a factor is derived. In this section, we explain the supply aspects of factor inputs.

The factors of production are traditionally classified as (i) land (including all natural resources); (ii) labour including all human (productive) efforts; and (iii) capital (including all man-made intermediate goods). Let us look into the meaning and supply aspects of each category of these factors of production.

(a) *Land*

The classical economists "treated land as a 'free gift of nature', a special factor of production distinct from man-made means of production and reproducible human labour.' Land as a means of production has been used in three different senses: (a) the dry area of land within the territories of a country; (b) the area of fertile land available for cultivation; and (c) natural resources found under and above the ground.

When the term 'land' is used in sense (a), the supply of land is traditionally assumed to be absolutely *inelastic.* Ricardo considered land as non-reproducible means of productions but this statement applies only in a physical sense and not in economic sense. In the economic sense of the term, 'land' is definded as 'cultivable fertile area' and as a 'basket of natural resources'. Land is thus not much different from man-made resources. The reason being that, if land is definded as fertile cultivable area, a considerable amount of effort and cost has to go into the improvement of land, maintenance of the fertility of soil (e.g., in manuring and soil conservation) and creation of irrigation facilities, reclamation of land. The supply of cultivable land is therefore subject to fluctuations. Its supply can be increased by means of land reclamation, soil conservation, fertilisation and expansion of irrigation facilities. On the other hand, supply of cultivable land can be contracted by reducing the cost on soil conservation, etc. Similarly, if land is defined to include all the natural resources, the exploitation of resources and converting them into usable intermediate goods, requires a heavy cost. Since the magnitude of natural resources, available in the territory of a country is limited, their supply can be

increased with increasing cost. For example, water resources, forest resources, minerals, petroleum and natural gas, etc., are available in a limited supply, scattered over the whole territory of the country. It is known by experience that exploitation of natural resources is subject to diminishing return. That is, their availability can be increased but at an increasing marginal cost. Thus, **the supply of land**, in the economic sense, **is not perfectly inelastic**. Land supply is however less elastic than the supply of other man-made intermediate goods.

(b) *Capital*

Capital is a man-made factor of production. It is reproducible. The stock of capital in a country consists of plant, machinery, building, etc. The stock of capital diminishes at the rate of its wear and tear in the process of production. Maintaining capital stock intact requires replenishment of the capital stock. The stock of capital can be increased by increasing investment in capital production. The production of capital goods, like that of consumer goods is subject to diminishing returns. That is, more of capital goods can be produced only at an increasing marginal cost. Here it suffices to note that capital supply curve is positively sloped like commodity supply curves.

(c) *Labour*

Labour may be defined as a physical and mental effort to produce goods, services, ideas and techniques. The *labour force* of a country consists of the number of its people willing to work. The total number of hours they are willing to work, given the reward per time unit, is called the *supply of labour.* The total supply of labour depends on (a) the size of population; (b) the proportion of population willing to work; and (c) number of hours which they are willing to work at a given wage rate. The supply of labour has some peculiarities in contrast to supply of other factors. It, therefore, requires a detailed treatment, which is accomplished in the following section.

Derivation of Individual Labour Supply Curve

We will use the indifference curve technique to draw the *labour supply curve*. Let us assume that all labour is homogeneous; labour units are identical; and, unlike other factors, labour finds a trade-off between hours of work or leisure and income.

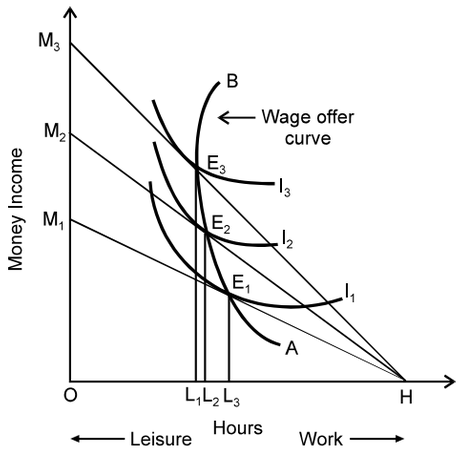

Fig. 5.8 *Labour Supply Curve*

Figure 5.8 illustrates the derivation of labour supply curve. Money income is measured on the vertical axis and the hours available to a worker per time-unit (per day or per week) are measured on the horizontal axis. One can also read the hours of work on the horizontal axis. The hours of work equals the total hours minus the hours of leisure. Let us now suppose that the total number of hours available to an individual is *OH* which he can use either for leisure or for work. If he works for *OH* hours and enjoys no leisure, then he makes OM_1 income. The wage rate (w_1) may then be obtained as

$$w_1 = \frac{OM_1}{OH}$$

$$= \text{Slope of } M_1H \text{ line}$$

The indifference curve I_1 represents the leisure-income preference function of the individual. Let the individual be in equilibrium at point E_1 by working for HL_3 hours and enjoying leisure for OL_3 hours. Note that $HL_3 + OL_3 = OH$, the hours available to the individual. When the wage rate rises to $w_2 = OM_2/OH$, the individual moves to a new equilibrium point E_2. At this equilibrium point, he works for HL_2 hours and has leisure of OL_2 hours and earns income E_2L_2. Note that with increase in wage rate the supply of labour by the individual increases. Similarly, when wage rate increases further to $w_3 = OM_3/OH$, the individual moves to equilibrium point E_3 where he works for HL_1 hours, reduces his leisure from OL_2 to OL_1. By joining the equilibrium points E_1, E_2 and E_3, we get the **wage-offer curve**. The curve *AB* is essentially the labour supply curve, though its shape is apparently unusual. The information contained in curve *AB* may be regraphed as in Fig. 5.9(a), to obtain the normal labour supply curve. The curve *SS*′ is the normal positively sloped supply curve of labour. If wage-rate continues to increase above w_2, the labour supply curve *SS*′ will be a backward bending one, as shown by curve *SL* in Fig. 5.9(b). This is so because higher wages create disincentive for longer hours of work and incentive for increasing hours of leisure.

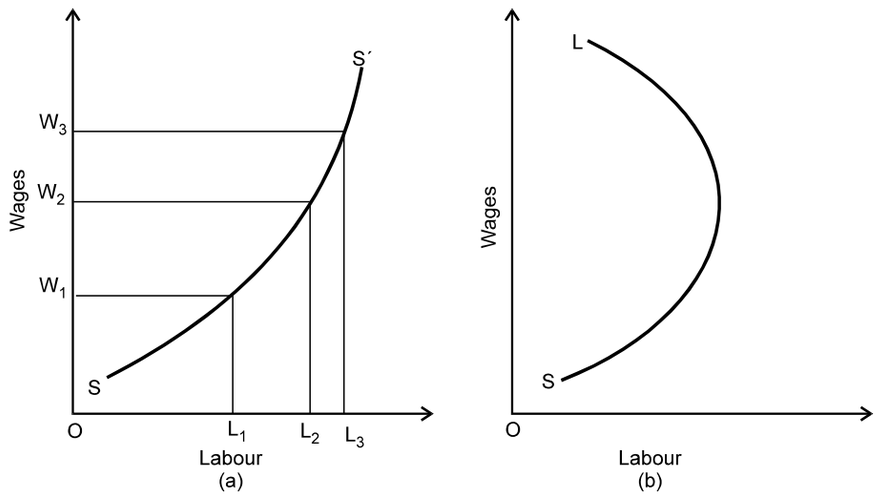

Fig. 5.9 *Labour Supply Curve*

NOTES

Derivation of Market Labour Supply Curve

The market labour supply curve is the sum of individual labour supply curves. Economists however disagree on the shape of the market supply curve for labour. Different shapes have been suggested for short- and long-run labour supply curves, depending on the occupational mobility, the type of labour used and level of economic growth. In the short-run, if a firms uses specific type of labour, "nothing can be said about the slope or shape of the labour supply curve. It may be negative, or it may have segments of positive and negative slope.' In the long-run, however, the market supply of specialised labour is likely to have a positive slope because (a) occupational mobility of labour increases; (b) with the increase in population and expansion of education and training facility, supply of all types of labour *increases*; and (c) with the increase in information about job market, young people are able to plan their education and career in accordance with job market conditions.

In case of non-specialised labour, the shape of labour supply curve is more clear. It is generally positively sloping for two reasons. *First*, industries which plan to employ more labour, can draw it from other industries by offering higher wages. Hence the labour supply to such industries increases only at higher wage rates. *Second*, the output decreases in the industries which lose their labour. The fall in the output of such industries leads to a rise in the price of their products causing an upward pressure on the demand-price of labour. Thus, such industries too obtain larger supply of labour at increasing wage rates. For these reasons, the industries attempting to employ larger units of labour must face a labour supply curve with a positive slope. There is a likelihood that the market labour supply curve in an affluent country is a backward-bending one. For, when individuals become richer and richer, they begin to prefer smaller number of working hours, longer holidays, and fewer hours of work per day, and so on, as they need longer time to enjoy their high earnings. This likelihood would however be limited to the rich nations in which the scope for further expansion is limited. It is therefore generally believed that labour supply curve in the long-run must have a positive slope.

Check Your Progress

1. Define functional distribution of national income.
2. Mention the different conditions under which the demand curve for factors is derived.

5.4 ADDING UP THEOREM

The ultimate aim of the distribution theory is to explain how the share of factors of production in total output is determined.

According to the marginal productivity theory, the share of each factor in national income is determined by the marginal productivity of a factor and the number of units of the factor employed, i.e., national income = $VMP_L \cdot L + VMP_L \cdot K$. However,

there has been a controversy on as to how is the share of each factor in the national income determined. The controversy is known as 'adding-up controversy'. This controversy is discussed first to focus on the nature of the problem. This is followed by Euler's theorem of distribution and then 'relative factor share'.

Adding-up Controversy and Solution

When the marginal productivity theory first gained acceptance by the end of the 19th century, a controversy arose whether distribution of national income among the various factors of production according to their marginal productivity was morally justifiable. In the course of the debate, another question came up, i.e., whether the sum of total labour income and of capital income equals the total product.

In other words, the controversy was, if each factor is paid the value of its marginal product (*VMP*), does this mean that the entire output is exhausted and nothing is left that falls into the hands of exploiting capitalists? Some economists attempted to show that, if each factor is paid its *VMP*, the distribution of income under free enterprise or capitalist system must be equitable. Precisely, they attempted to demonstrate that

$$Q = (MP_l)L + (MP_k)K \qquad ...(5.1)$$

In terms of value, national income is equal to $P \cdot Q$, and:

$$P \cdot Q = (MP_l \cdot P)\ L + (MP_k \cdot P)\ K \qquad ...(5.2)$$

where P is the average price of the products.

Since, $MP_l \cdot P = VMP_l$ and $MP_k \cdot K = VMP_k$

$$P \cdot Q = VMP_l + VMP_k \qquad ...(5.3)$$

It is, thus, proved that national income in distributed between the factors of production according to their marginal productivity.

5.4.1 Euler's Product Exhaustion Theorem

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

$$Q = \frac{\partial Q}{\partial L} \cdot L + \frac{\partial Q}{\partial K} \cdot K \qquad ...(5.4)$$

Since $\partial Q/\partial L = MP_l$ and $\partial Q/\partial K = MP_k$, Eq. (5.4) takes the form of Eq. (5.1), i.e.,

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

This may be proved as follows.

A production function, $Q = f(L, K)$, is homo-geneous of degree v if

$$f(\lambda L, \lambda K) = \lambda^v \cdot f(L, K) \qquad ...(5.5)$$

NOTES

By differentiating Eq. (5.5) with respect to λ, we get

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

When return to scale is constant, ν = 1, and then Eq. (5.5) may be written as

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

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

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

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

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

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

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

Clark-Wicksteed-Walras Product Exhaustion Theorem

Euler's product exhaustion theorem assumes a homogeneous production function, i.e., constant returns to scale. Clark, Wicksteed and Walras have, however, shown that the assumption of homogeneous production function is not necessary for the product exhaustion theorem. It holds for all types of production functions. That is, according to Clark-Wicksteed-Walras theorem, if each factor is paid its *VMP*, then the total factor payments will exhaust the value of total output. A graphical proof of Clark-Wicksteed Walras theorem of product exhaustion is given in Figure 5.10.

Let us assume (i) an economy consists of *n* identical firms, (ii) each firm employs the same number of homogeneous labour, (iii) the marginal physical product of labour is given by the curve *MPL* in Figure 5.10, and (iv) each firm employs *OL* number of workers. The total output of each firm will then be represented by the area *OMBL*. Suppose also that each labour is paid a real wage of *OQ* = *BL* and that the total wages equal the area *OQBL*. That is, the share of labour in total output *OMBL* is *OQBL*. The residual (*OMBL* – *OQBL* = *QMB*) goes to land as *rent*. The rent so computed is merely a residual. But, Clark and others proved that *QMB* is not merely a residual: it is also the marginal physical product of land. By proving this, they had established the product exhaustion theorem. Note that, given *n* firms, the total output of the industry is *n* × *OMBL*. Now suppose that the number of firms increases to *n* + 1, the number of workers remaining the same, i.e., *n* × *OL*. The new firm gets its labour supply from the old firms. Suppose that *n* × *OL* workers are so distributed between *n* + 1 farms that each farm again has the same number of workers, say *OL′*. Note that number of workers employed by each firm decreases from *OL* to *OL′* so that

$$(n + 1)\,QL' = n \cdot OL$$

NOTES

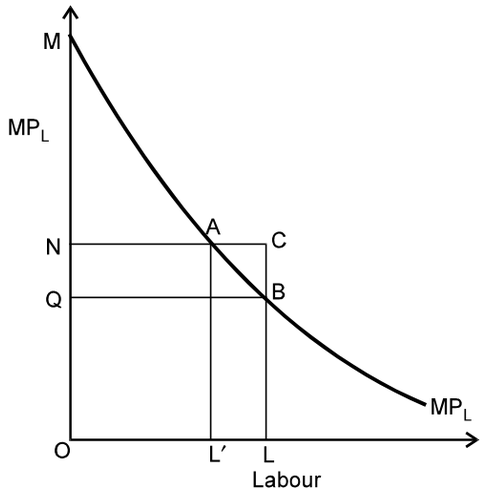

Fig. 5.10 *Product Exhaustion Theorem*

When each firm employs OL' workers, output per firm is $OMAL'$, and the total output of the industry is

$$(n + 1) \cdot OMAL' = n \cdot OMAL' + OMAL' \quad ...(5.6)$$

The initial output of the industry with n firms can be written as

$$n \cdot OMBL = n \cdot OMAL' + n \cdot L'ABL \quad ...(5.7)$$

The difference between Eqs. (5.6) and (5.7) is the marginal product of land (MP_{LD}) That is,

$$MP_{LD} = (n \cdot OMAL' + OMAL') - (n \cdot OMAL' + n \cdot L'ABL)$$
$$= OMAL' - n \cdot L'ABL \quad ...(5.8)$$

As can be seen from Figure 5.10,

$$OMAL' = NMA + ONAL' \quad ...(5.9)$$

and $$L'ABL = L'ACL - ABC \quad ...(5.10)$$

By substituting Eqs. (5.9) and (5.10) in Eq. (5.8), we get

$$MP_{LD} = NMA + ONAL' - n(L'ACL - ABC)$$
$$= NMA + ONAL' - n \cdot L'ACL + n \cdot ABC \quad ...(5.11)$$

Since $n \cdot L'ACL = ONAL'$, by substitution, we can write Eq. (5.11) as

$$MP_{LD} = NMA + ONAL' - ONAL' + n \cdot ABC \quad ...(5.12)$$
$$= NMA + n \cdot ABC$$

Consider the last term, $n \cdot ABC$. As $n \to \infty$, the share of each firm in the given supply of labour tends to be zero. Therefore the last term $n \cdot ABC \to 0$. Consequently,

$$MP_{LD} = NMA = \text{rent of land}$$

It is the same residual, for all firms with OL' number of workers, calculated earlier as rent. Thus, Clark-Wicksteed-Walras theorem is proved.

Relative Factor Shares and Income Distribution

Now we will discuss how a *change* in relative factor prices affect the relative factor shares and income distribution. When relative factor prices change, one factor becomes relatively cheaper and the other becomes relatively costlier. This impels the profit maximizing firms to substitute the cheaper factor for the costlier one. As a result, factor ratio changes. For example, suppose there are only two variable factors, labour (L) and capital (K) and factor ratio is given as K/L. This factor ratio changes, at margin when one factor is substituted for another. When factor ratio changes, relative factor-share changes.

5.4.2 Elasticity of Factor Substitution and Relative Factor Shares

The extent to which relative **factor shares** in income are affected by the change in relative factor prices depends on the *elasticity of factor substitution*. The concept of the elasticity of factor substitution was developed by J.R. Hicks. It is regarded as the foundation of the modern neo-classical theory of distribution and relative factor shares. Ferguson remarks that the concept of *elasticity of substitution* lies at the heart of the neoclassical theory of distribution. The *elasticity of substitution* (σ) is defined as

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

Recall that, in a perfectly competitive input market, a firm is in equilibrium when it chooses a labour-capital combination at which *MRTS* is equal to the ratio of factor prices (w/r). That is, under perfectly competitive conditions, a firm is in equilibrium when

$$MRTS_{l,k} = \frac{w}{r} \qquad ...(5.13)$$

where w = price of labour (wage rate) and r = price of capital (interest). Thus, in a prefectly competitive factor market, the firm's equilibrium condition given in Eq. (5.13), may also be written as

$$\sigma = \frac{\partial(K/L)/(K/L)}{\partial(w/r)/(w/r)} \qquad ...(5.14)$$

The elasticity of substitution (i.e., the value of σ) is always positive, though in some cases, $\sigma = 0$. Thus, *the value of* σ *ranges from zero to infinity.* The value of σ yields useful information regarding the degree of substitutability between the factors. If $\sigma = 0$, it means that substitution between factors, say labour and capital, is impossible; the two factors can be used only in a fixed proportion; and that isoquant is L-shaped.

The positive range of s may be classified and interpreted as follows:

$\sigma < 1$: **Inelastic substitutability:** The degree of substitutability between the two factors is very low.

$\sigma = 1$: **Unit elasticity of substitution:** The two factors can be proportionately substituted for one another (see properties of Cobb-Douglas production function).

$\sigma > 1$: **Highly elastic substitutability:** One factor can substitute another to a large extent.

$\sigma = \infty$: **Perfect substitutability:** One factor can substitute another to any extent.

Let us now examine the relationship between the value of σ and the relative shares of factors in the total output. Consider a two-factor model in which the total income $P \cdot Q$ is the sum of labour-share ($w \cdot L$) and capital-share ($r \cdot K$). That is,

$$P \cdot Q = \quad w \cdot L + r \cdot K \qquad ...(5.15)$$

The relative share of *labour* in the total value of output is then given by

$$\frac{w \cdot L}{P \cdot Q}$$

Similarly, the relative share of *capital* in the total value of the output is given by

$$\frac{r \cdot K}{P \cdot Q}$$

Thus, the *ratio* of relative share of L and K

$$= \frac{wL}{PQ} + \frac{rK}{PQ}$$

$$= \frac{wL}{rK} \qquad ...(5.16)$$

$$= \frac{w/r}{K/L} \qquad ...(5.17)$$

Eq. (5.17) can be used to show the effect of change in relative factor price (w/r) on the relative shares of L and K in the value of total output.

Suppose w/r increases by 10 per cent, i.e., labour becomes costlier by 10 per cent. This will lead to a substitution of capital (the relatively cheaper factor) for the labour (the relatively costlier factor). The extent of substitution depends on the value of σ (i.e., the elasticity of substitution). Suppose $\sigma = 0.5$, i.e., the elasticity of substitution is less than one. Then a 10 per cent increase in w/r will result in a 5 per cent increase in the capital-labour ratio (K/L). This will alter the relative shares of K and L.

The new relative shares can be obtained as

$$\left(\frac{wL}{rK}\right)^{*} = \frac{(w/r)(1+0.10)}{(K/L)(1+0.05)}$$

$$= \frac{1.10}{1.05} \cdot \frac{(w/r)}{(K/L)}$$

where * denotes the new relative factor share.

NOTES

Obviously,

$$\left(\frac{wL}{rK}\right)^{*} > \left(\frac{wL}{rK}\right)$$

That is, new relative factor-share ratio is greater than the initial ratio. One may show by the same reasoning that if $\sigma > 1$, and other things remain the same, then

$$\left(\frac{wL}{rK}\right)^{*} < \left(\frac{wL}{rK}\right)$$

There is a two-way causation in the change in relative factor shares caused by a change in relative factor prices. Changes in the relative supply position of factors and changes the relative factor prices which, in turn, changes the factor-ratios depending on the value of σ. This leads to a change in the relative shares of the factors in the total output.

It is clear that the concept of elasticity of substitution plays a very important role in neo-classical theory of income distribution.

5.4.3 Technical Progress and Factor Shares

We have so far analysed factor price determination and relative factor shares on the basis of a tacit assumption that production function is given and technology of production remains unchanged over the reference period. In the real world, however, technological progress does take place. Technological progress means a given quantity of output can be produced with less quantity of inputs or a given quantity of inputs can produce a greater quantity of output. This means a downward shift in the production function (the isoquant) towards the point of origin (*O*).

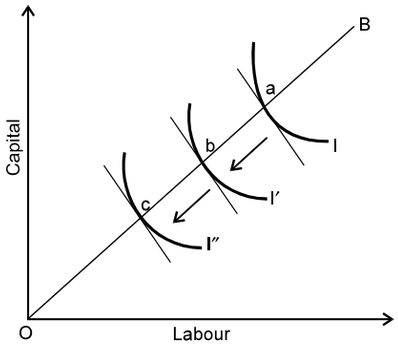

Fig. 5.11 *Technological Progress-Neutral*

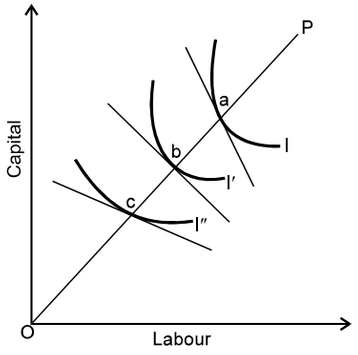

Fig. 5.12 *Capital Deepening Technological Progress*

Technological progress is graphically shown in Figure 5.11. A given level of output is shown by isoquants *I*, *I′* and *I″*. That is, all three isoquants, *I*, *I′*, *I″* represent the same level of output.

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

According to J. R. Hicks, technological progress may be classified as neutral, capital-deepening and labour-deepening. Technological progress is *neutral* if, at constant *K/L*, the marginal rate of technical substitution of capital for labour i.e., $MRTS_{l,k}$) remains constant. The neutral technological progress is illustrated in Figure 5.11. At each equilibrium point, $MRTS_{l,k} = w/r$. When technological progress is neutral, both *K/L* and *w/r* remain unchaged. It follows that relative factor share remains unchanged when technological progress is neutral.

Capital-deepening technological progress is illustrated in Figure 5.12. Technological progress is capital-deepening when, at a constant capital/labour ratio (*K/L*), $MRTS_{l,k}$ declines. It implies that, at constant *K/L*, MP_k increases relative to MP_l. Therefore, at equilibrium *w/r* declines, as *r* increases relative to *w*, because $w = VMP_l$. Consequently, the relative factor share changes in favour of *K*. That is, share of capital in the total output increases while that of labour decreases.

Technological progress is labour-deepening when, at a given *K/L*, the $MRTS_{l,k}$ increases. Labour-deepening technological progress is illustrated in Figure 5.13. It can be shown, following the above reasoning, that under labour-deepening technological progress, the share of labour in the total output increases while that of capital increases.

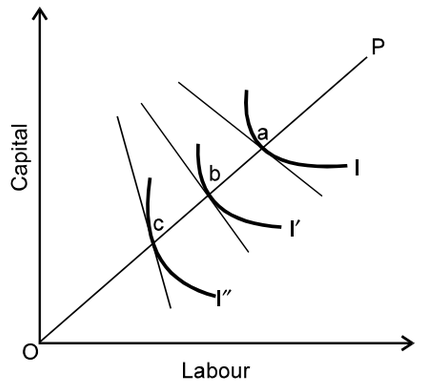

Fig. 5.13 *Labour Deepening Technological Progress*

5.5 MODERN THEORY OF RENT

The marginal productivity theory applies to all factor prices, including land rent. In this saection, however, we will describe the Ricardian theory of **rent** and the Marshallian concept of **quasi-rent**. The Ricardian theory of rent is the earliest known rent theory and is generally known as the classical theory of rent. The point of distinction between Ricardian and modern theories of rent is that while Ricardo considered rent as 'surplus produce' attributable solely to land as a factor of production, modern economists consider rent as 'economic surplus' which accrues as well to all other factors in fixed supply in the short-run.

Antecedents of Rent Theory

Ricardian theory of rent has an interesting antecedent. In the early 19th century, food prices in Britain had considerably increased partly due to Napoleonic War

and partly due to increase in population and the consequent increase in demand for food. This caused a great deal of anxiety to the British Government. So both House of Lords and the House of Commons appointed a Committees to find the cause of rise in food price. The Committees reported that 'food prices were high because rents were high'. The contemporary economists, namely, West, Torrens, Malthus and Ricardo reacted to this suggestion and offered, separately, an alternative explanation to the problem. In their opinion, food prices were high not because rents were high, rather, rents were high because food prices were high. According to them, food prices had gone up due to Napoleonic 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 land, which caused rise in rents. Ricardo, who was said to be a new bourgeoisie, 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 holding this view, Ricardo was criticised as being anti-landed aristocracy. However, Ricardo's theory of rent emerged out of his effort to establish his argument.

Ricardian Theory of Rent

Ricardo defined rent as 'that portion of the produce of earth which is paid to the landlord for the use of the *original* and indestructible powers of soil'. 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 and its limited productivity. Land as a factor of production proves scarce with the growth of population. Growth of population forces extension of cultivation to inferior lands. According to Ricardo, rent arises due to differential in surplus accruing to the cultivators resulting from the differences in fertility of soil of different grades of land. In simple words, rent arise because of difference in surplus produce of land of different productivity.

Ricardian theory of rent is based on the principle of demand and supply. If, in a country, the fixed supply of land 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 shown that *rent arises in both extensive and intensive cultivation of land.* Let us first explain the rent on **extensive cultivation.** When land is cultivated extensively, rent on superior land equals the excess of its produce over that of the inferior land. Suppose there are three grades of land—*A*, *B* and *C* and if an equal amount of capital and labour is used to cultivate the same area of each grade of land, and 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 existing population, no rent is payable till the demand for land exceeds the supply of *A* grade land. When population

increases, demand for land increases, beyond grade *A* land, the land of grade *B* will be brought under cultivation. But, compared to the yields from land *A,* (i.e., 100 quintals), land *B* yields only 80 quintals of wheat, even if the same quantities of capital and labour are used. This difference in the yields from lands of grade *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, which yields only 70 quintals of wheat. This gives rise to rent on land *B* and raises rent on land *A.* According to Ricardian theory, rent on land of different grade is worked out by the following formula.

Rent = yield from a land *less* yield from the lowest grade of land.

For example, the rent on land of grade *A* and *B* can be worked out as follows.

Rent on land *A* = 100 – 70 = 30 quintals of wheat

Rent on land *B* = 90 – 70 = 20 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'.

In case of **intensive cultivation,** Ricardo observes that it often happens that before land *B* is brought under cultivation, more of capital can be employed to increase productivity of land *A.* But, it is quite likely that doubling the capital on land *A*, the produce is not doubled. It may yield only 95 quintals instead of 100 quintals, which is greater than the produce of land *B.* The cultivators would therefore intensify cultivation of land *A,* instead of employing their capital on land *B* or on any inferior land. In this case, the rent on land *A* would be 5 quintals = 100 – 95 quintals. Thus, in case of intensive cultivation, capital and labour will not be employed on land *B* till the yields from subsequent units of factors used on land *A* are greater than that of land *B.* 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. This has two repercussions: *one,* rent on land *A* increases and *two,* the inferior land, i.e., land *B,* is brought under cultivation. It shows that the Ricardian concept of rent is based on the *law of diminishing return.*

Critical Evaluation

Ricardian theory has been criticised on the following grounds.

First, Ricardo's concept of rent is based on the assumption that powers of soil are 'original and indestructible', which can hardly be accepted. Fertility can be created through techniques of soil conservation and land reclamation and can be destroyed through the continuous use of the soil. Destruction of 'power of soil' has become particularly easy due to growth of atomic energy.

Second, 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 acrues also to other factors—labour, capital and entrepreneurship—as well as to land.

Third, 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

NOTES

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 envisaged by Ricardo.

5.5.1 Quasi-Rent: The Short-Term Rent on Fixed Factors

The *quasi-rent,* a concept used by Marshall, refers to the short-term earnings of factors which are in fixed supply in the short run. To explain the concept of quasi-rent, let us make a distinction between the short run and the long run. In the long run, all inputs are variable in large quantities as their supply is elastic. In the short run, however, the supply of certain inputs is fixed. For example, the supply of plant and machinery in the short run is inelastic.

In the short run, variable factors can be transferred to their alternative uses if they are paid an amount less than their transfer earning (or opportunity cost). Therefore, if variable factors are to be retained in their current use in the short run, they must be paid equal to their transfer earning. Otherwise, variable factors shall be transferred to their alternative uses. On the contrary, the fixed factors cannot be transferred to their alternative uses in the short run. Therefore, in the short run, fixed factors are paid what is left after the variable factors are paid their opportunity cost. That is, fixed factors are paid, in the short run, the *residual* of the total revenue. This residual payment to a factor fixed in the short run is called quasi-rent. The quasi-rent may thus be defined as *TR—TVC.*

The determination of quasi-rent is illustrated in Fig. 5.14. Suppose, given the *AVC, AC* and *MC* curves, price is *OP,* and the firm is in equilibrium at point *E.*

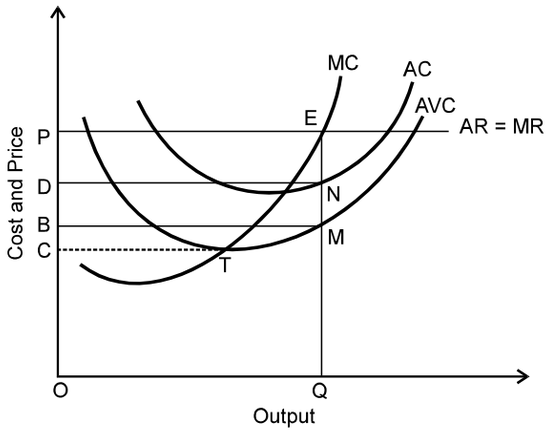

Fig. 5.14 *Determination of Quasi-Rent*

At equilibrium, 'firms' total revenue is

$$OP \times OQ = OPEQ$$

and $$TVC = OB \times OQ = OBMQ$$

NOTES

The firm must pay a total sum of *OBMQ* to retain the variable factors. Under perfectly competitive conditions, this sum equals their transfer earnings the earning that a factor expects from its second best use. The quasi-rent may be obtained as

Quasi-Rent = *OPEQ* – *OBMQ* = *PBME*

The quasi-rent will always be a non-negative quantity. For example, so long as price is greater than *OC,* the quasi-rent will be greater than zero. When price is *OC,* total revenue (*TR*) equals total variable cost (*TVC*), i.e.,

TR = *OC* × *CT* and *TVC* = *OC* × *CT*

Since *TR* – *TVC* = 0, quasi-rent = 0. When price falls below *OC,* there will be no production. There is therefore no question of quasi-rent.

The quasi-rent can be divided into two components: (i) opportunity cost; and (ii) economic profits. We have seen that when prices is *OP,* quasi-rent is represented by the area *PBME.* Of this, the area *DPEN* represents the difference between the *TR* and *TC* (= *OQ* × *OD*). Therefore, the area *DPEN* represents the total pure or economic profits. The area *BDNM* represents the *total fixed cost, TFC* = (*AC* – *AVC*) *OQ* = (*OD* – *OB*) *OQ.* The fixed factors would have earned the same amount in another firm of the same industry, under competitive conditions. Therefore, the area *BDNM* is the *opportunity cost* of fixed factors. Thus

Quasi-rent = *TFC* + Economic Profit

5.5.2 Factor Price, Transfer Earning and Economic Rent

The equilibrium price of a factor service can be divided into two components:

(i) *Transfer Earning;* and

(ii) *Economic Rent*

Transfer earning or what is also known as *opportunity cost,* may be defined as *the amount that a factor must earn to remain in its present occupation.* Or, transfer earning is the amount that a factor expects to earn if transferred to its second best use. For example, suppose a doctor earns ₹ 10,000 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 ₹ 8,000 per month. Thus, doctor's transfer earning is ₹ 8,000 per month. He must earn a minimum of ₹ 8,000 per month to remain in his private practice. So long as he earns ₹ 8,000 per month from his private practice, he has no incentive to join a hospital as an employee.

Economic rent is the excess of actual earning of a factor over its transfer earning. *Economic rent* may thus be defined as factor's actual earning minus its transfer earning. Consider the factor supply curve, S_f in Fig. 5.15, which has less positive slope. It implies that more and more units of factor shall be supplied to an industry if factor payments increase, and, conversely, less and less units will be supplied to the industry if factor payments decrease. That is, when factor payment decreases, factors are transferred to their alternative uses. For example, given the

NOTES

demand curve D_f, the market factor price is determined at OP_3, where equilibrium supply of factor is OM. Note that, given the supply curve, S_f, all but one of OM units (i.e., $OM - 1$) of factor are willing to remain in this industry at factor prices lower than OP_3. That is, minimum payments that must be made to all but the last factor unit, in order to prevent transfer of factors to alternative uses, is less than the equilibrium price OP_3.

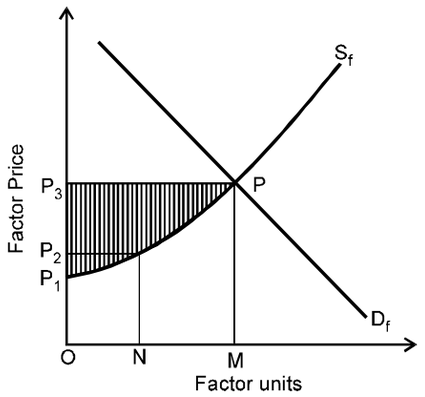

Fig. 5.15 *Economic Rent*

In other words, the transfer earning of all factor units, *excepting the last one,* is less than their actual earning. For example, the *transfer earning* of *ON*th unit of the factors is only OP_2, whereas the actual earning, i.e., the market factor price, is OP_3. Thus, economic rent earned by the *ON*th unit is $OP_3 - OP_2 = P_2O_3$. The same exercise may be performed for all the OM factor units, and economic rent computed. The shaded area, $P_1 P P_3$, represents the *total economic rent of OM* units. Note that the *OM*th unit, i.e., the last unit, of the factor does not earn economic rent because its actual earning equals its transfer earning. The total transfer earning is shown by the area below the shaded area.

Note that the terms 'economic rent' means differently from the term 'rent' in common parlance. In its common usage, the term 'rent' means the *actual* payment to the landlord, much of which is transfer earning. But, when an economist speaks of 'rent' he means 'economic rent', i.e., the excess of payment over transfer earning.

Elasticity of Factor Supply and Economic Rent

The existence of economic rent depends on the elasticity of factor supply. Economic rent may be zero or equal to transfer earning depending on whether factor supply is perfectly elastic or perfectly inelastic. These are the two limiting cases of economic rent.

When factor supply is perfectly elastic, economic rent is zero: Perfectly elastic factor supply (i.e., $e_s = ¥$) means that an individual factor-owner can supply his factor as much as he wishes, and an individual firm or industry can buy as many units of the factor as it wants to, *at a given price.* In such a case, *the whole price paid to the factor*, i.e., its actual earning, equals its transfer earning. There is no excess payment over the transfer earning. Hence economic rent is zero. This case

NOTES

is illustrated in Fig. 5.16. The market factor price is determined at *OP,* i.e., the actual earning of the factor. No factor owner can charge a price in excess of *OP.* Hence economic rent is zero.

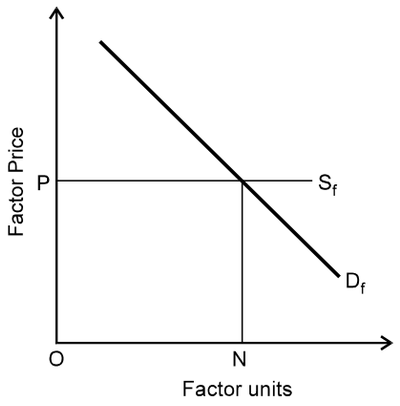

Fig. 5.16 *Zero Economic Rent*

Factor supply is perfectly inelastic, economic rent equals actual earning: If factor supply is fixed and factor has only one use, the factor owners would have to put their factors on the market for whatever they can earn. Even if factor owners are not satisfied with what the market offers, they cannot transfer their factors to other uses, since there is none. Therefore, in such cases *transfer earning* is zero. Thus, the whole factor price is *economic rent.* This case is illustrated in Fig. 5.17. The market price is fixed at *OM.* The whole of which is economic rent.

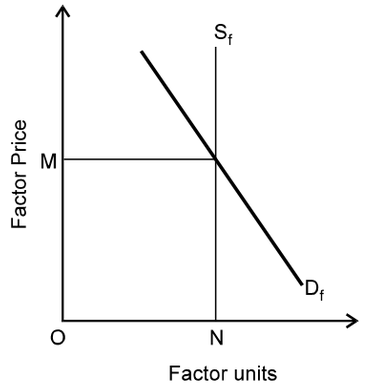

Fig. 5.17 *Zero Transfer Earning*

When factor supply curve has a positive slope, economic rent equals factor price.

Check Your Progress

3. Who developed the concept of the elasticity of factor substitution?
4. Define quasi-rent.

NOTES

5.6 MODERN THEORY OF WAGE

Let us begin by discussing money and real wages.

5.6.1 Money and Real Wages

There are two types of wages: Nominal and real wages. Nominal wages are also known as money wages. Nominal wages are the types of wages which do not take into account inflation rate, purchasing power. It is the amount that is promised to the worker when he/she is hired. Essentially, the economic conditions of the market do not affect the wage that is paid to the worker. Any increase in the amount too solely depends on the employee and not due to any market conditions.

Real wages are the wages that are paid to the worker after taking into account the inflation rate. This is why real wages can be defined as the purchasing power of the worker as it reflects the amount of goods and services that can be bought with the wages after adjusting inflation.

Real wages have an indirect impact on money wages. This is so because an increase or decrease in the purchasing power will affect the money wage that a worker demands from his employer. There are a number of factors which determine real wages. These include the inflation rate, the job security, the food and accommodation benefits, working hours, future prospects etc. Real wages are a better indicator to be used in economic analysis as it gives an accurate picture of cost of living. Nominal or money wages on the other hand are decided on the basis of an agreement, do not get changed frequently and therefore do not represent the purchasing power of the workers.

In this section, we will discuss the theory of wage determination in different kinds of labour and product markets. The labour and product markets are classified from wage determination points of view as follows.

(i) Perfect competition in both product market and labour market,

(ii) Monopoly in product market and perfect competition in labour market,

(iii) Monopoly in product market and monopsony in perfectly competitive labour market,

(iv) Perfect competition in product market and monopoly (labour union) in labour market, and

(v) Bilateral monopoly–monopoly in both product market and labour market.

5.6.2 Wage Determination under Perfect Competition in Product and Labour Markets

Under the condition of a perfectly labour market and a perfectly product market, wages are determined by the demand for and supply of labour. Wage determination in this kind of market structure is illustrated in Fig. 5.18.

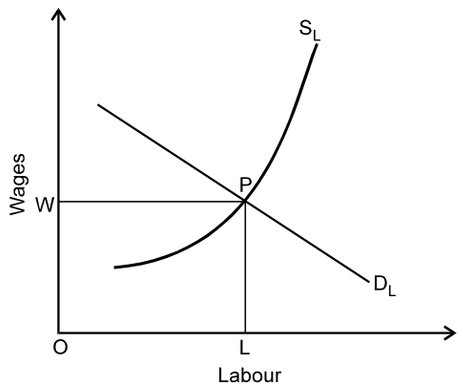

Fig. 5.18 *Determination of Wages in a Perfectly Competitive Market*

As shown in the figure, the labour demand curve D_L and labour supply curve S_L intersect at point *P.* At this point, demand for and supply of labour are equal at *OL,* and wage rate is determined at *OW.* This wage rate will remain stable in a competitive market so long as demand and supply conditions do not change. This analysis of labour price determination is similar to commodity price determination in a perfectly competitive product market. What distinguishes the analysis of factor price determination from the commodity price determination is the method of deriving demand and supply schedules for a variable factor of production.

5.7 MODERN THEORY OF INTEREST

It may be noted at the outset that the theoretical problems associated with capital and interest are much more complicated and difficult to resolve than those pertaining to other factors and factor prices. The complications arise for at least two reasons. *First,* unlike land and labour, capital is a man-made factor—a produced means of production. Its supply is, therefore, well under human control. It is produced only when it is needed. That is, production of each capital asset is a matter of economic decision. Besides, there are various types of capital assets with varying productivity which can be used to produce a commodity. Furthermore, producing capital goods require a cut in the present consumption, which brings in the question of time preference. So producing each capital asset requires a deliberate and careful decision. This complicates the matter. *Second,* production of capital is a round about process which is time consuming. For example, to produce a lathe requires a foundry and mining of iron ore. To mine iron requires mining machinery and equipment, and so on. 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 complicate the analysis. Let us begin our discussion on the theories of interest and investment decisions with brief review of development of the interest theory.

NOTES

Some early thoughts on Interest Theory

The earliest form of interest theory appeared in the writings of Locke, Petty, and Law in the 18th century. These authors considered interest as a monetary phenomenon and believed that interest varies inversely with supply of money: the larger the availability of money, the lower the interest, and *vice versa*. They argued that when money supply increases, purchasing power of money decreases, demand for money decreases and, hence, interest rate decreases. Adam Smith, following Hume, criticised this theory. He argued that interest is merely a ratio of two sums of money, which remains unaffected by the change in money supply.

Cantillon, Hume and Turgot criticised the monetary theory of Locke, *et. al.* very strongly. They argued that when money supply increases, prices rise, investors are therefore required to borrow a larger sum to finance a given project. Hence, interest rises with the rise in money supply. They also opposed the idea of interest being uniquely related to money supply. They themselves believed that interest depends on the demand for investment goods. This belief marks the beginning of real theories of interest. However, not until recently, a consistent theory of interest was formulated. The classical economists, Adam Smith and David Ricardo, could offer no real solution to the problems of interest determination. In fact, they did not provide a consistent theory of profits or pure interest at all. They treated rate of interest simply as the rate of return on capital invested.

Senior was the first to offer a broad outline of modern theory of capital and interest. According to Senior's theory, the rate of interest is determined by the forces of demand for and supply of capital goods. Capital goods are demanded because they are productive and can produce goods and services that are in demand. Thus, he suggested the idea that demand for a factor is a derived one. In his view, the supply of capital goods depends on the consumer's willingness to abstain from consumption and to save a part of resources that could be used to produce capital goods. Thus, according to Senior, creation of capital goods required abstinence from consumption and the interest was the 'reward for abstinence.' Senior's theory was later improved by Mill, which came to be known as 'Abstinence Theory of Interest'.

The 'abstinence theory of interest' was later subjected to severe criticism. It is said to be an incomplete theory of interest as it considers only supply of savings and does not explicitly relate demand for capital to its productivity.

Since it is mostly the rich who can save, Marx objected to the idea that the rich made sacrifices by abstaining from current consumption. According to him, rich are not required to forego their current consumption as they possess surplus income. Besides, many owners of productive assets might have inherited them rather than procured by reducing their consumption. Senior however argued that income from inherited property is of the nature of rent, not interest. To avoid such criticism, Marshall substituted a natural word, 'waiting', for the term 'abstinence'. The term 'waiting' implies that an investor has to wait until he receives the returns from his investment. If the present value of the total expected returns is greater than the cost of investment, it yields a surplus to the investor. This surplus is a

reward for 'waiting.' Thus, according to Marshall, interest is the reward for 'waiting'.

5.7.1 Classical Theories of Interest

NOTES

As noted here, classical views on interest vary from economist to economist. However, two most important theories of interest were propounded by two classical economists–Bohm-Bawerk and Irving Fisher. We will first describe their theories of interest and then present Keynesian formulation of classical theory of interest.

Bohm-Bawerk's Theory of Interest

Bohm-Bawerk's theory of interest is also referred to as Austrian theory of interest. Bohm-Bawerk, an Austrian economist, emphasised that land and labour are 'original' or primary factors of production, while 'capital' is an intermediate factor whose supply depends on land and labour used for its production. This distinction is fundamental to Bohm-Bawerk's theory of interest and is the source of 'origin of interest.' Bohm-Bawerk assumed that 'round about' method of production (i.e., the use of capital in the process of production) is more productive (though at diminishing rates), than the direct method of production, i.e., by the direct use of primary factors (labour and land) to produce final goods. But the 'round about' method is more time consuming than the direct method.

According to Bohm-Bawerk, interest arises out of the process of lending present income against the promise of future income. Interest arises because people prefer present goods to future ones. That is, they prefer present consumption of goods to their future consumption. They are therefore willing to pay a premium on present goods for the privilege of spending it when they want to spend. Or, conversely, they discount the future goods. The people would therefore be willing to lend a certain quantity of present goods only if they are sure of being repaid with a greater quantity of goods of the same kind and quantity. It is this preference which forms the basis of interest.

Bohm-Bawerk gives **three reasons** why people prefer the present to the future goods. ***First,*** *the circumstances of want and provision in the present and in the future are different.* If income rises continuously, it becomes subject to the law of diminishing marginal utility. If individuals act so as to maximise the sum of utilities of all future incomes, they would be willing to pay a premium on present goods (when income is rising). The reason for this is that increase in present consumption adds more to the total utility than the increase in the future consumption. That is why people prefer present to future goods. ***Second****, people underestimate future* because of (a) deficiency of imagination, i.e., myopia in respect of future; (b) limited will power; and (c) the shortness and uncertainty of life. These factors lead to discounting of future wants. ***Third****, present goods are technically superior to future ones.* This is so 'because present goods can be invested now and reinvested as they accrue tomorrow, but available tomorrow can be invested tomorrow' only. Besides, present goods invested today in round about method of production yield a larger physical output in the future than an equal quantity invested in future either in direct or in round about production.

The reason for this is that, even if other things remain the same, yields from investment today compound over a longer period by one time unit. This is a very strong reason why people prefer present goods to future ones and why they are willing to pay a premium on present goods and discount on future goods. This third reason, according to Bohm-Bawerk, constitutes an independent ground for a positive rate of interest.

Irving Fisher, however, argued that anyone reason is not sufficient to make interest rate positive. It is in fact the interaction of the three reasons which is fundamental to the interest being greater than zero.

Nonetheless, as Mark Blaug remarks, Bohm-Bawerk's three 'reasons' and their interaction provide an exhaustive explanation of the existence of interest in a stationary as well as in a dynamic economy. According to Bohm-Bawerk's theory, rate of interest is determined by forces of demand for and supply of funds. The demand for funds comes solely from the capitalists and supply comes primarily from the retained earning. The total capital consists of subsistence fund advanced to the workers, which is fixed. The marginal productivity of this capital decreases over time. The rate of interest is determined where present value of expected marginal product equals the wage rate.

Fisher's Theory of Interest: The Time Preference Theory

According to Fisher, interest arises because people prefer present to future income. Thus, Fisher's notion of interest is the same as that of Bohm-Bawerk. The rate of interest, according to Fisher's theory, equals the price that people are willing to pay for income now rather than income at some future date. This price (or the rate of interest) is determined by the interaction of 'willingness to give up present consumption in favour of a larger consumption in future, and "opportunity" to invest. The opportunity to invest is measured by the rate of return over cost.' By cost is meant the loss of withdrawing an income stream, and by return is meant the gain that results from substituting a new income stream. Fisher's 'rate of return over cost' is that critical discount rate at which two or more alternative investment opportunities have the same present net values.

Fisher's theory of interest determination is presented graphically in Fig. 5.19. Horizontal axis measures *income today* and vertical axis measures *income tomorrow.* The 45º line represents a constant flow of income, i.e., income available today is available tomorrow also. The indifference curves represent the people's willingness curve which shows their time preference with respect to income today and incomes tomorrow. The convexity of indifference curves implies the diminishing marginal utility of income. A very important aspect of the indifference curve (i.e., willingness curve) is its slope at the point of intersection by the constant income stream line: it reflects the people's time preference. If slope of the indifference curve at the point of intersection is greater than 1, it reflects a positive time preference, if slope = 1, time preference is neutral, if slope < 1, time preference is negative. For example, indifference curve I_1 shows positive time preference; and indifference curve I_2 shows negative time preference; and indifference curve I_3 shows a neutral time preference.

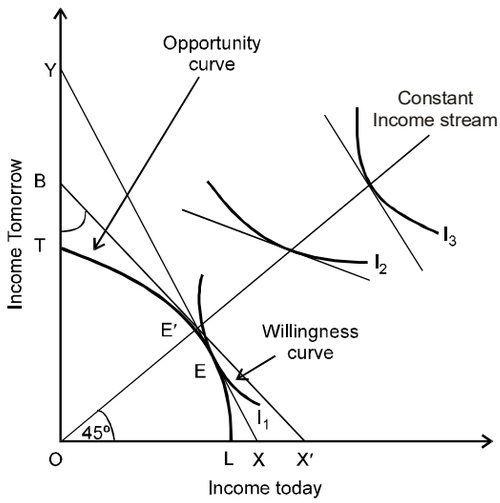

Fig. 5.19 *Determination of Interest Rate: Fisher's Theory*

The opportunity curve, or technical transformation curve, *TL,* reveals the rates at which present income can be transformed into future. Thus, it reveals the net productivity of capital. The concavity of the opportunity curve implies diminishing returns to capital, invested present income. In other words, the concavity of opportunity curve indicates diminishing marginal productivity.

The rate of interest is determined at point *E* where the willingness curve, I_1, and opportunity curve, *TL* are tangent with one another. At point *E,* their slopes are equal, the slope of willingness curve being positive. The interest rate is therefore positive and equals the slope of the line *YX.* If, at the point of tangency, slope of transformation curve (i.e., the opportunity curve) and that of the indifference curve are equal to unity, the rate of interest will be equal to zero because at point *E,* net productivity of capital is zero and time preference is neutral.

The economists have pointed out several **weaknesses** in Fisher's theory of interest. ***First,*** it places too much emphasis on consumption rather than on output. ***Second,*** this theory does not consider the influence of banking system on the rate of interest. ***Third,*** Fisher's theory ignores the role of expectations.

Schumpeter has, however, commended Fisher's theory as 'the peak achievement' of 'perfection' in interest theory within its framework.

Keynesian Formulation

The classical theory of interest refers, according to Keynes, to the theories (or views) of Marshall, Cassel, Tausig, Walras, etc. In fact, none of these authors, whom Keynes calls modern classical school, has given a precise or an explicit account of the interest theory.

Keynes provides an excellent summary of classical theory of interest. 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; savings 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 that rate of interest is equal to the amount of saving at that rate.' In simple words, rate of interest is determined where demand for investible funds equals the supply of funds, i.e., supply of savings. The classical theory of interest is presented in Fig. 5.20. The investment demand schedule, I_1 is negatively related to interest, *r,* whereas supply of savings schedule, *S,* is positively related to interest. The investment demand schedule I_1 intersects the supply of savings schedule at point *P.* The interest 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 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.

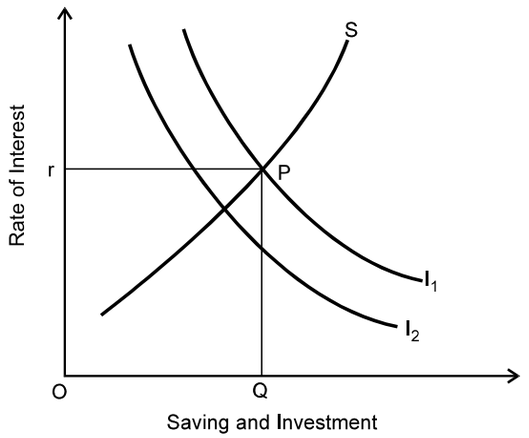

Fig. 5.20 *The Classical Theory of Rate of Interest*

Keynes's Criticism of Classical Theory

Keynes criticised classical theory of interest on the following grounds.

First, the classical theory implicitly assumes *income to be given* and saving to be a unique function of interest. Keynes argues that the classical assumption of *income to be given* itself implies that there exists an important relationship between savings and income, i.e., savings are not the function of interest alone but also of income. But the classical school neglects this important relationship between income and savings, which leads to a formal error in the analysis.

Second, the aforementioned error leads to the conclusion that, given the investment and saving schedule, interest will be determined by the shifts in investment demand schedule, as shown in Fig. 5.21. Suppose that rate of interest was initially determined at Or_1. If investment demand schedule shifts left ward to I_2, due to, say, a fall in the marginal efficiency of capital, interest rate will be determined at Or_3. Similarly, if both investment demand and savings schedules shift leftward, interest will be determined at Or_2. As Keynes puts it, 'the classical theory of the

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 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 curve.' (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, at the same time, 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 because income, *(Y)= f(I).* When income changes, savings change too because $S = f(Y)$ too. 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, **the classical theory of interest is indeterminate.**

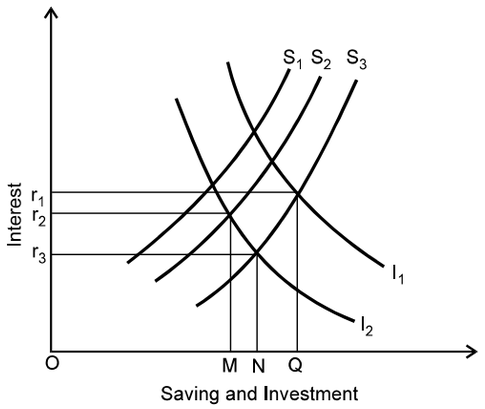

Fig. 5.21 *Change in Interest Rate and Levels of Income*

5.7.2 Neo-classical Theory of Interest

A variant of classical theory is the loanable fund theory of interest also called as neo-classical theory of interest. The neo-classical 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 schedule for loanable funds with the supply schedule of loanable funds. 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 three kinds of demand for funds are inversely related to the interest. The horizontal summation of the three kinds of demand gives the aggregate demand for loanable funds (D_L) as shown in Fig. 5.22.

NOTES

On the supply side, the supply of loanable funds, in Robertsonian sense, consists of:

(i) voluntary savings (V_s), i.e., savings out of disposable income;

(ii) bank credits (B_c), i.e., borrowing from banks; and

(iii) activated idle cash balance or dishoardings (D_h).

All the three components of the supply of loanable funds are deemed to be 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) (Fig. 5.22).

The loanable fund theory of interest is presented in Fig. 5.22. 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*.

The neo-classical theory is superior to the classical theory of interest in so far as it considers the demand for funds other than the demand for investment, and also the other sources of supply of funds than voluntary savings. Besides, the loanable fund theory, as presented in Fig. 5.22, reveals an 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 Fig. 5.22 shows, at equilibrium rate of interest, *Or,* planned savings, *ra,* are less than planned investment, *rb.*

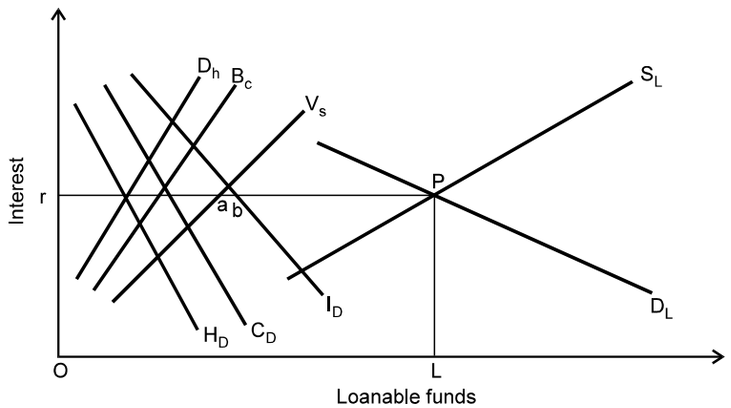

Fig. 5.22 *Loanable Fund Theory of Interest*

Criticism

Keynes's criticism of classical theory applies to the loanable-fund theory also. It means that loanable-fund 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 investible funds, 'savings' accounts for the largest proportion in the total. Saving is function of the 'disposable incomes'. It implies that the total supply of investible funds also depends on the disposable income. Therefore, interest cannot be known unless income is known. And, income cannot be known unless interest is known, since interest influences the investment. As in classical theory, the error in the neo-classical theory lies in the implicit assumption that demand for investible fund 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 criticised 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. It is therefore more appropriate to discuss Keynesian theory of interest in a macro-monetary framework.

Keynesian Theory of Interest

Keynes's theory of interest is also known as **'liquidity preference' theory of interest**. *Liquidity preference theory* states that market rate of interest is determined by aggregate demand for money (M_D) and total supply of money (M_S). That is, the equilibrium rate of interest is determined where

$$M_D = M_S$$

To understand Keynesian theory of interest, we need to know what determines the aggregate supply of money (M_S) and aggregate demand for money (M_D) and the relationship between (a) money supply and interest, and (b) money demand and interest. We will look first at Keynes's view on the demand for money and the relation between money demand and interest. This will be followed by brief discussion on the Keynesian view on the supply of money. We will finally explain liquidity preference theory of interest.

Demand for Money: The Keynesian View

In Keynesian analysis of demand for money, the terms 'demand for money', 'money holding' and 'liquidity preference' have the same meaning. In this section, we will discuss Keynesian theory of demand for money. Keynes has classified the total demand for money under the following three categories.

(i) Transaction Demand for Money;

(ii) Precautionary Demand for Money, and

(iii) Speculative Demand for Money.

Let us now look at the determinants of these kinds of demand for money and derive money demand curve.

(i) Transaction Demand for Money

The most important function of money is to acts as a medium of exchange. Money being the most common and convenient medium of exchange is demanded to carry out *planned, routine* or known transactions. Therefore, both households and business firms hold money to carry out their routine and planed transactions. The need for holding cash balance arises because there is a time gap between the receipt of income and the need for expenditure. Income is generally received periodically–weekly, monthly or annually–whereas people need money to buy

goods and services regularly as and when need arises, for example, to buy grocery, to pay electricity and telephone bills, to pay house rent, etc. Firms, on the other hand, are supposed to know their income (profits) at the end of the financial year. But they need money to pay salaries, buy material inputs, pay electricity and telephone bills, etc., several times a year. Therefore, they demand money or hold money to meet these routine payment needs. This is called *transaction demand for money.*

According to Keynes, aggregate *transaction demand for money* is function of national income, i.e.,

$$M_t = f(Y)$$

where M_t is transaction demand for money and Y is income.

In the Keynesian system, the proportion of income held for transaction purpose is *constant* in the short-run. Therefore, Keynesian transaction money demand function can be written as

$$M_t = kY$$

where k is the proportion of income demanded for transaction purpose.

The relationship between national income and transaction demand for money is illustrated in Fig. 5.23 by the line $M_t = kY$. The slope of the money demand function $M_t = kY$ is expressed as

$$\Delta Mt / \Delta Y = k \text{ and } k > 0$$

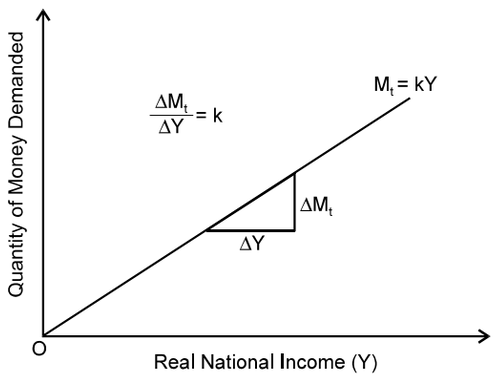

Fig. 5.23 *Transaction Demand for Money (M_t)*

In the Keynesian system, money demanded for transaction purpose is assumed to be *interest inelastic* because, whatever the rate of interest, people cannot afford shutting down their kitchen, not paying house-rent, electricity and telephone bills, school fees, and medical bills, and so on. However, some economists2 argue that when the rate of interest goes very high, even in the short-run, transaction demand for money begins to decline. The relationship between *Mt* and interest rate is illustrated in Fig. 5.24. As the figure shows, *Mt* is interest inelastic till interest rate raised to *i*3. Beyond interest rate *i*3, *Mt* becomes interest-elastic.

NOTES

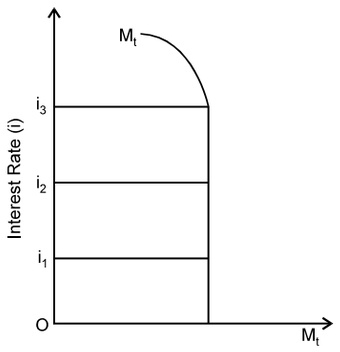

Fig. 5.24 *Interest and Transaction Demand for Money (M_t)*

(ii) Precautionary Demand for Money

Both households and business firms hold some money in excess of their transaction demand to meet their contingent expenditure and to also make provision for unforeseen opportunities. The contingent need for expenditure arises due to unpredictable events like fire, theft, sickness, loss of job, accidents, death, etc. Likewise, unpredictable opportunities arise which people take advantage of to promote their interest, e.g., a sudden temporary fall in prices of bonds and consumer durables. To protect and to promote their interest against such contingencies and unforeseen opportunities, people do hold some idle cash balance. The money held for this motive is called *precautionary demand for money.*

Precautionary demand for money (M_p) is also closely related to the level of income. We know that rich people hold a larger amount of cash in their purse as safety margin than the poor people. As a simple rule, the higher the level of income, the higher the demand for money for precautionary motive. This relationship is also expressed in functional form as $M_p = f(Y)$. Precautionary demand for money is interest-inelastic to a large extent. Beyond a certain level of interest, however, it becomes interest-elastic. In Keynesian system, however, M_p is also assumed to be interest-inelastic.

Note that both transaction and precautionary demands for money are a function of the national income. The kinds of transaction demands for money can, therefore, be lumped together and expressed as $M_t + M_P = M_T$, while M_T is interest-inelastic. When M_T is drawn against the interest rate, it takes a form of a straight vertical line.

(iii) Speculative Demand for Money

According to Keynes, people hold a part of their income in the form of idle cash balance in addition to their transaction and precautionary demand for money with a view to take advantage of changes in money market, especially changes in asset market. Buying bonds when their prices are low is advantageous. Some people hold idle cash balance to take advantage of fluctuations in the bond prices. There is, however, an element of uncertainty in regard to change in bond prices. If bond prices do not decline, it involves loss of interest on the idle cash balance. In this kind of behaviour of the people there is an element of speculation. Keynes called this kind of demand for money as *speculative demand for money.*

NOTES

Those who hold money for speculative purpose assume a risk of losing interest on the idle cash balance. The higher the interest rate, the higher the risk. Therefore, *speculative demand for money* (M_{SP}) and *interest are inversely related.* This relationship can be expressed as

$$M_{SP} = f(i),\ \Delta M_{SP} / \Delta i < 0$$

The nature of relationship between the speculative demand for money and interest rate is illustrated in Fig. 5.25.

In Fig. 5.25, M_{SP} curve shows the speculative demand for money in relation to the market rate of interest. It shows an inverse relationship between the rate of interest and the speculative demand for money: the speculative demand for money increases when interest rate decreased and *vice versa*.

It can be seen in Fig. 5.25 that a part of the Keynesian M_{SP} curve is horizontal straight line. By this, Keynes hinted at a remote possibility of a market situation in which market rate of interest falls below a 'critical' minimum level, say to $i1$ as shown in Fig. 5.25. At a certain law rate of interest, $i1$, people's desire to hold idle cash balance becomes infinite. It implies that the speculative demand for money is infinite or very large when the rate of interest goes below a certain 'critical' minimum level, i.e., a level below which banks pull down their shutter. Keynes calls this kind of a situation as *'liquidity trap'*.

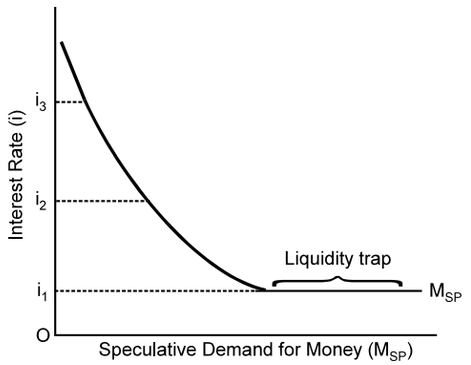

Fig. 5.25 *Speculative Demand for Money*

(iv) The Total Demand for Money

We have discussed above various kinds of demand for money. Let us now add them together and derive the demand curve for total demand for money. Total demand for money consists of two components:

(i) transaction (including precautionary) demand for money ($M_T = M_t + M_P$), and

(ii) speculative demand for money (M_{SP}).

The total demand for money (M_D) at a point in time can be expressed as

$$M_D = M_T + M_{SP}$$

As noted above, $M_T = kY$ and $M_{SP} = f(i)$, given the income and interest rate, the total demand for money can be expressed as

$$M_D = kY + f(i)$$

The relationship between the total demand for money and the interest rate is crucial to the Keynesian theory of demand for money and the theory of interest. The relationship between the two is shown by a total-money-demand curve in relation to interest rate. The derivation of the total-demand-for money curve is illustrated in Fig. 5.26. Part (a) of the figure shows the ***transaction demand for-money*** (M_T) in relation to the rate of interest. The transaction demand for money is a straight vertical line. It implies that change in the interest rate does not influence the transaction demand for money. As shown in the figure, whether the interest rate is i_1, i_2 or i_3, the transaction demand for money remain constant constant at *MT*.

Part (b) of Fig. 5.26 presents the ***speculative demand for money*** (M_{SP}) in relation to the interest rate. The speculative demand for money is inversely related to the interest rate—it increases with decrease in the interest rate and *vice versa*. Part (c) of Fig. 5.26 presents the ***total demand for money (M_D)***. The total-money-demand curve, i.e., *MD* curve, is simply the horizontal summation of the M_T and M_{SP} curves. The M_T and M_{SP} curves of parts (a) and (b) have been reproduced in part (c) shown by the dotted lines. Their horizontal summation gives the total demand for money (M_D) in relation to the interest rates. The curve M_D represents the Keynesian demand curve for money. The curve shows an inverse relationship between the total demand for money and the interest rate.

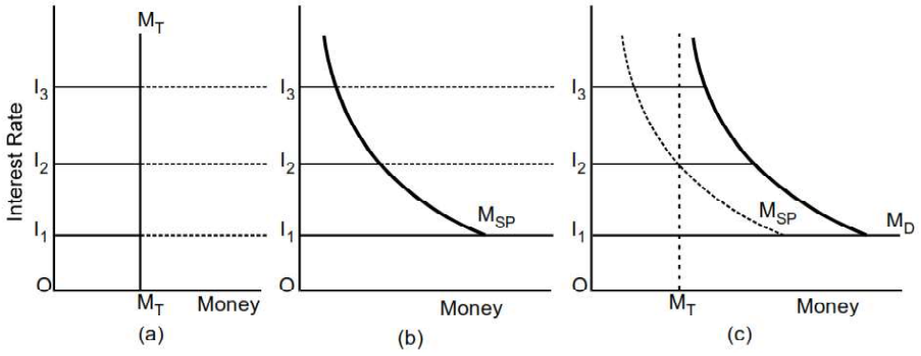

Fig. 5.26 *Total Demand for Money*

The Aggregate Supply of Money

In his theory of interest rate determination, Keynes assumed money supply to be given in the short-run. He argued that money supply in any country is determined exogenously by the monetary authority, generally the Central Bank of the country in view of the currency needs of the economy. The justification for this assumption lies in the fact that actual money supply in any country is not determined by the market forces. Money supply in India, for example, is determined by the Reserve Bank of India. Monetary authorities, in general, do not change money supply in response to change in market conditions, especially the change in the interest rate.

It must, however, be borne in mind that monetary authorities are capable and they do influence the money supply as a matter monetary policy from time to time in order to keep the demand for and supply of money in balance. But they do not increase or decrease supply of money in response to variation in market

NOTES

rate of interest. If necessary, central banks use their weapons of monetary control to change money supply in view of the monetary needs of the country in the long run.

However, in Keynesian theory of interest rate determination, supply of money is assumed to remain constant in any short period of time, as shown by the vertical line M_S in Fig. 5.27. As the figure shows, money supply is assumed to be constant and *interest-inelastic,* i.e., money supply does not vary with variation in the interest rate.

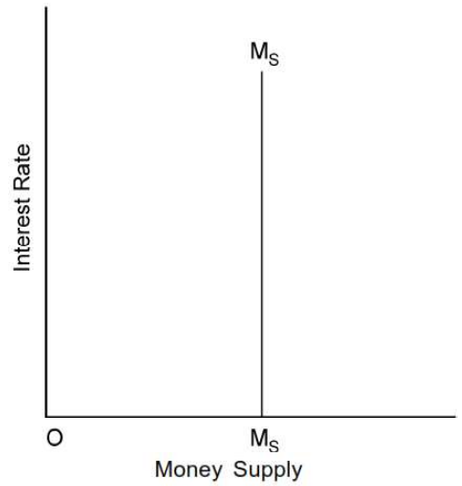

Fig. 5.27 *Constant Supply of Money*

The Keynesian Theory Of Interest

As noted in the beginning, according to Keynes, market rate of interest is determined where aggregate demand for money equals the aggregate supply of money, i.e., interest rate is determined where $M_D = M_S$. We have derived in the previous section the Keynesian money-demand curve and money supply curve. In this section, we put money demand and supply curves together and illustrate the determination of interest rate. The determination of the equilibrium rate of interest is illustrated in Fig. 5.28.

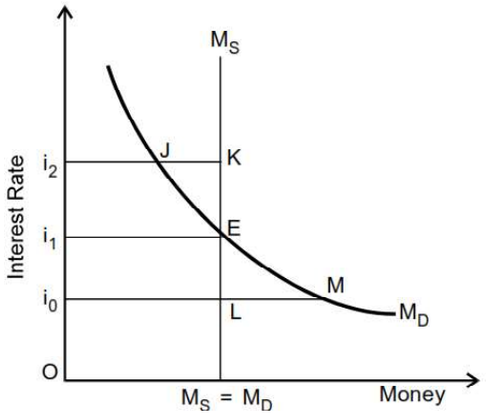

Fig. 5.28 *Determination of the Interest Rate*

NOTES

As shown in Fig. 5.28 money demand curve (M_D) and money supply curve (M_S) intersect at point E determining equilibrium rate of interest at i_1. In the framework of the Keynesian analyze, the rate of interest is supposed to be stable at i_1. If market rate of interest goes up or below the equilibrium rate of interest, this will itself create conditions for the market rate of interest to return to the equilibrium point E. For example, if interest rate rises to i_2 for some reason in any period of time, M_D will decrease by JK, money supply remaining constant. Obviously, money supply exceeds money demand. Under this condition, the interest rate will fall until it reaches the equilibrium rate i_1.

The process of restoration of equilibrium interest rate i_1 can be explained as follows. With the increase in the interest rate, bond prices go down and the speculative demand for money decreases. This creates a discrepancy of JK between M_S and M_D—supply of money exceeding demand for money by JK. This situation of disequilibrium sets the market forces in operation to restore the equilibrium. When interest rate goes up, people prefer to buy bonds because of its low price. As a result, bond prices begin to go up and interest rate goes down. The process continues until the equilibrium point E is reached.

Similarly, when the rate of interest falls, for some reason, from i_1 to i_0, the speculative demand for money increases because at a lower rate of interest the preference for cash holding increases. As a result, a discrepancy of LM is created between the demand for and supply of money. Since there is shortage of money in the money market, people begin to expect a rise in the interest rate and, therefore, the demand for money begins to decrease. The demand for money continues till the equilibrium point E is restored. Thus, in the short-run, given the money supply and the demand schedule, an equilibrium rate of interest will be determined where $M_D = M_S$. This rate will remain stable in the short-run. This is what is conveyed by the Keynesian theory of interest rate determination.

Drawbacks of the Keynesian Theory of Interest

The Keynesian theory of interest is undoubtedly superior to the classical and neoclassical theory of interest. However, Keynesian theory has its own drawbacks for which it has been criticized by economists. Ironically, Keynesian theory has been criticized on the ground Keynes criticized the classical theory.

Keynesian theory is indeterminate. Recall Keynesian argument that classical theory of interest is *indeterminate.* Keynes's argument against the classical theory of interest can be summarized as follows. Since $S = f(Y)$ saving schedule cannot be known unless income (Y) schedule is known. Since $Y = f(I)$, income schedule cannot be known unless investment function is known. Since $I = f(i)$, investment schedule cannot be known unless interest rate (i) is known. And, interest (i) cannot be known unless saving and investment schedules are known. Thus, according to Keynes, the indeterminateness of the variables make the classical theory indeterminate. According to Hansen, 'exactly the same criticism applies to Keynesian theory in its simpler form'. He reiterated, Keynes's criticism of the classical theory applies equally to his own theory." His argument may be summarized as follows. 'According to the Keynesian theory the rate of interest

is determined by the intersection of the supply schedule of money…and the demand schedule for money…'This theory 'also is indeterminate' because, even if money supply is fixed by the monetary authority, 'the liquidity preference schedule shifts up or down with changes in the income level'. In the Keynesian system, we cannot know the liquidity preference schedule unless we know the income level. Income level cannot be known unless we know the speculative demand for money or the asset holding. Thus, 'the Keynesian theory, like the classical, is indeterminate'.

To Leijonhufvud, Keynesian theory of interest is 'incredibly tortuous formulation.' According to him, the main trouble lies in his definition of 'savings' as 'non-consumption' taken from the 'pure' theories of interest. This definition might be appropriate in the pure theories of interest, but not in the Keynesian system. Keynes's 'ex-ante savings' is not clearly distinguished from the demand for money for speculative purpose and demand for non-monetary assets.

However, economists agree that Keynesian theory of interest marks a radical improvement over the classical and neo-classical theories of interest, in spite of its weaknesses and shortcomings.

Check Your Progress

5. What is the impact of real wages on money wages?
6. Mention the three reasons given by Bohm-Bawerk as to why people prefer the present to future goods?

5.8 MODERN THEORY OF PROFIT

While discussing the 'marginal productivity theory', we have noted that, in the long-run competitive equilibrium, the reward for each factor, including the reward for 'entrepreneurship' equals the value of its marginal product. It implies that, according to the marginal productivity theory, 'profit', which is the reward for 'entrepreneurship' equals the value of its marginal product. There are however a number of other profit theories developed by various economists over time. In this chapter, we will briefly review some important profit theories. We will however have first a brief look into the meaning of 'pure profit' and the question 'to whom belongs the pure profit'.

The Meaning of Pure Profit

The meaning and source of 'profit' have always been a centre of controversy. 'The word 'profit' has different meanings to businessmen, accountants, tax collectors, workers and economists...' For example, 'profit to a layman means all incomes that go to the capitalist class'. To an accountant, profit means the excess of revenue over all paid-out costs including both manufacturing and overhead expenses. For all accounting purposes, businessmen also use accountants' definition of profit. But, on the question as to whether a businessman should stay in his present business or move to another, his concept of profit

differs from the one used in accountancy. The term 'profit' in the accounting sense does not include the **opportunity cost**— the earning that a businessman foregoes to earn a given profit in his present occupation. But a businessman does consider his *opportunity cost* in his calculation of his satisfactory profit that must be large enough to cover his opportunity cost. All such costs are termed as 'opportunity costs'. Essentially, it includes all the expected incomes which he might earn from the second best alternative use of his *own resources*— labour and capital.

Concept of Pure Profit. Economists' concept of profit is of '**pure profit**'. It is also called '**economic profit**' or '**just profit**'. The word 'profit' in this unit means 'pure profit'. 'Pure profit' is a return over and above 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 than can prevent them from withdrawing their capital from its current use. The pure profit, so defined, may not be necessarily positive for a single firm in a single year; rather there may be negative profit (i.e., loss). What is important is the return over time. In the long-run, in a competitive system, however, pure profit is presumed to be equal to zero. That is, pure profit is non-existent in the long-run. 'To discover whether such profit exists, take the revenue for the firm and deduct the costs of all factors of production other than capital. Then deduct the pure return on capital and any risk premium necessary to compensate the owner of capital for the risks associated with its use in this firm and industry. Anything that remains is pure profit.'

An important question regarding 'pure profit' is 'to whom does it belong and in what form?' It is common knowledge that pure profit belongs to the entrepreneur, the owner of the firm. But the question arises: how does it accrue to the entrepreneur? For, if an entrepreneur is treated as a separate factor of production, pure profit must equal the value of its marginal product. But marginal value of product cannot be logically equated to pure profit, because as concluded above, pure profit is a 'residual'. In fact, this problem has been the source of controversy which led to various profit theories. We now turn to discuss the various theories of profit.

5.8.1 Walker's Theory of Profit: Profit as Rent of Ability

One of the most widely known theories advanced to explain the nature of profit was formulated by F.A. Walker. According to him, profit is rent of the exceptional abilities that an entrepreneur may possess over the least entrepreneur. Just as rent on land is the difference between the yields of the least fertile and super lands, pure profit is the difference between the receipts of the least efficient entrepreneur and that of those with greater efficiency or managerial ability.

NOTES

Assumptions: In formulating his profit theory, Walker visualised a state of perfect competition in which all firms (or entrepreneurs) are presumed to possess equal managerial ability or entrepreneurship. There being no barrier to prevent the entry of new firms to the industry, the number of firms would increase until the remuneration of each was just enough to keep them in the industry. Each firm would then receive only the wages of management which, in Walker's view, formed no part of (pure) profit. He regarded wages of management as ordinary wages. Thus, under perfectly competitive conditions, there would be no pure profits and all firms would be no-profit firms.

However, when one departs from the realm of perfect competition, one finds, in almost every economic activity, some firms making only a bare living while other firms in the same industry are making pure profits. Walker regarded profits of profit-making firms arising from what a more efficient firm is able to produce over and above what the least efficient firm i.e., able to produce with same amount of capital and labour. Walker attributed this surplus wholly to the greater efficiency of a firm, which distinguishes it from the least efficient ones.

Thus, to Walker, profit is reward for exceptional business ability over and above the ordinary ability required for management of the organisation which could be rewarded by a wage or salary. Just as rent is a reward for a higher productivity of land, so is the profit reward for superior managerial ability of an entrepreneur.

A natural corollary of this view is that profit did not enter the cost of production as is the case with rent. Therefore, according to Walker, profit does not enter the price determination. The logic that Walker gives for his argument runs as follows. Market price is determined by the cost of production of that portion of supply which is produced by the least efficient firms. Prices so determined make allowance for only wages of management not the surplus that accrues to the firms with greater efficiency.

Clark's Theory of Profit: Profit as Reward for Dynamic Entrepreneurship

The dynamic theory of profit is associated with the name of J.B. Clark, which he propounded in 1900. According to Clark, profits accrue in a dynamic world, not in a static world.

The Static World: As visualised by Clark, a static world is one in which there exists absolute freedom of competition; but population and capital are stationary; there are no inventions; production process does not change; and the goods continue to remain homogeneous. Besides, in a static state there is perfect mobility of factors of production but there is no motion because marginal products of labour and capital are equal in all groups of industries. Also, in a static state, there is no uncertainty and hence, no risk. Whatever risks might arise due to natural calamities are covered by insurance.

No Profit in Static Society: To show how profits were eliminated in a static state, Clark draws a distinction between the work of an entrepreneur and that

of a manager of business. He believed that the task of a manager could be described as labour which can be paid for by wage. In a static state, profit would not arise because competition would not permit any business manager to earn more than his actual wages which would be equal to marginal value his product. Therefore, there would be no surplus available which could be called as profit.

The Dynamic World: In contrast to static word, dynamic world is one in which the factors that remain constant in a static world undergo the process of change. Clark indicated certain generic changes that mark the transition of a society from a static to a dynamic state. Briefly speaking, generic changes include

(a) increase in population;

(b) increase in capital;

(c) improvement in production techniques;

(d) changes in forms of business organisation; and

(e) multiplication of consumer's wants.

Profit as Reward for Dynamic Enterprise: In Clark's view, the major functions of an entrepreneur in a dynamic society are related to these changes, i.e., to take the advantage of generic changes, promote their business, expand their sales, and reduce their cost of production. The typical changes that emerge out of this special effort of some entrepreneur are inventions and improvement in the methods of production. Such changes lead to increase in production given the costs or reduction in costs given the output, which results in emergence of profits to the initial inventors.

Profits in Dynamic World are not There for Ever: With the passage of time, profits resulting from the inventions and improvements in production methods disappear. What happens, in fact, is that competition forces other entrepreneurs to imitate or innovate the new technology. This leads to rise in demand for labour and capital. Consequently, wages and interest rise and cost of production increases. On the other hand, with larger employment of labour and capital, production increases leading to fall in product prices. The ultimate result is that profits disappear. In Clark's own words, 'profit is an elusive sum which entrepreneurs grasp but cannot hold. It slips through their figures and bestows itself on all members of the society.'

Profits Disappear to Reappears: This however should not mean that, in a dynamic society, profits arise only once and disappear never to emerge again. In fact, under dynamic conditions, the generic changes continue to take place: it is a continuous process. The process of dynamic change gives entrepreneurs opportunities time and again to adjust their business to the changing conditions, make inventions and improve production methods, with a view to make pure profit. In fact, emergence and disappearance of profits is a continuous process.

On the question of risk involved in making inventions and improving production methods, Clark was of the view that profit docs not arise due to risk.

NOTES

If risk is there, it affects capitalists because risk-income accrues to them. Profit, on the other hand, is the result of entrepreneurial functions under dynamic conditions. Therefore, profit does not result from risk-bearings.

To sum up, according to J.B. Clark, profit is a reward for coordinating managerial functions of entrepreneurs under dynamic conditions. It is a reward for dynamism. It is not a reward for risk bearing. Pure profit, according to him, is a residue that remains after interest and wages are paid. That is, the difference between the gross receipts and payments for wages and interest represents profit.

Criticism of Clark's Theory

Clark's theory, though impressive, has failed to win unqualified acceptance and has been criticised on the following grounds.

First, to some economists the division of firm's earning between the wage of management and profits is not acceptable. It has been contended, for instance, that even the routine conduct of a business calls for a prudent judgement and administrative ability just as these qualities are calls for in the exploitation of a new invention or in any other manifestation of economic change. Clark's definition was therefore a matter of phraseology and no clear line could be drawn to show the functions which give wages of management and those which were remunerated by profits.

Secondly, even if it is accepted that profits are accounted for by the coordinating functions of entrepreneur, it poses special difficulties in explaining the profits in the practical world. For instance, profits of companies are mainly paid to the shareholders. But these shareholders exercise no coordinating functions. One may say, for the sake of argument, that shareholders receive only a fair interest on their investment and that the profit is what remains after paying this 'interest'. Still, this sum after deducting the 'interest' paid to shareholders would continue to be their property, because they are the owners of retained earnings. Thus, Clark's theory fails to explain the profits in practice.

Thirdly, the basic tenet of Clark's theory is that profits result from change in business conditions and are reward for dynamism and Clark's entrepreneur is the pioneer of this change. But in practice, one finds that profit exists under different conditions. There are many profitable business concerns engaged in forms of activity in which dynamic stage is long since past and in which no change takes place. In many lines of activity business has settled down to almost routine conditions and yet profits continue to be made despite competition.

Fourthly, it has been argued by F.H. Knight that all changes would not give rise to profits. Certain changes are predictable and others are not. So far as predictable changes are concerned they pose no managerial problems or uncertainty. Therefore, such changes cannot give rise to profit. Only the unpredictable changes would require the use of managerial talent and, hence, give rise to uncertainty. Clark's theory thus misses an important element of uncertainty and risk and their relation to profit.

NOTES

5.8.2 Hawley's Risk Theory of Profit: Profit as Reward for Risk-Bearing

The risk theory of profit was propounded by F.B. Hawley in 1893. Hawley regarded risk-taking as the inevitable accompaniment of dynamic production. and those who take risk have a sound claim to a separate reward, known as profit. Thus, according to Hawley, profit is simply the price paid by society for assuming business risks. In his opinion, businessmen would not assume risk without expecting an adequate compensation in excess of acturial value. That is, the entrepreneur would always look for a return in excess of the expected losses. The reason why Hawley maintains that profit is over and above the acturial risk is that the assumption of risk is irksome; it gives rise to trouble, anxiety and disutilities of various kinds, which gives a claim to reward for all these pains in excess of acturial value of risk. Profit, according to Hawley, consists of two parts: *one*, represents compensation for acturial or average loss incidental to the various classes of risks necessarily assumed by the entrepreneur; and *second* the remaining part represents, an inducement to suffer the problems of being exposed to the risk.

Hawley recognises that the coordination which Clark spoke of was important, but he believes that profit is attendant upon profit only when coordination happens to be an incident of ownership; and that profit arises from ownership only so long as ownership involves risk. Thus, risk has to be assumed to qualify for profit. If an entrepreneur shifts his risks by insuring against them, he would cease to be an entrepreneur and would not receive any profit. It is only from the uninsured risks that profits arise, and until the uncertainty ends with the sale of entrepreneur's products, the amount of the reward cannot be determined. Profit, therefore, is a residue. Hawley's theory is also called as a **residual theory of profit.**

Hawley was conscious that his theory did not offer a full explanation of all the gains arising from business activities. In monopoly undertakings, for example, many a time profit could not be attributed to the risks which were undertaken: profits in monopoly firms arise from the very fact of not undertaking the risks. Thus, monopoly gains fall outside his theory. To meet this flaw he placed monopoly gains in a distinct, separate category of business gains which might arise to other factors also. According to his view, monopoly gains could occur also to labour, landlords, capital suppliers. But since their respective incomes—wages, rent and interest—do not arise from the operation of productive forces, these are merely economic gains.

Criticism of Hawley's Theory

Perhaps no other theory of profit has attracted so much attention and generated so much discussion as Risk Theory of Profit. It ranks today as one of the most widely accepted theories of profits. Nevertheless, Hawley's risk theory of profit has been criticised on the following grounds.

NOTES

First, in his reaction to risk theory of profit, Clark remarked that the profit visualised by Hawley was nothing but an interest on capital. Risk, in Clark's view, was risk of loss of capital. Therefore, the reward for assuming risk (of loss of capital) was interest: it is not profit.

Secondly, it has also been argued that Hawley stressed only the risk in terms of loss of capital: he did not give due consideration to the fact that risks arise also in the use of factors of production other than capital.

Thirdly, Hawley's theory of profit concentrates only on risk-bearing element, and ignores other entrepreneurial functions, viz., organisation and coordination, which also lead to emergence of profit.

Fourthly, Hawley fails to make a distinction between predictable and unpredictable risks. While predictable (or foreseeble) risks are insurable, unpredictable (or unforeseeable) risk are not. Since predictable risks can be insured, such risks do not give rise to profit because the risk is shifted on to the insurer. As Knight put it, it is in fact the uninsurable risk, which is uncertain and gives rise to profit. Thus, in his view, profit is a reward for uncertainty bearing rather than a reward for risk-bearing.

Fifthly, Carver observed that profits are reward for avoiding risk and not for bearing risk, because only those entrepreneurs who are able to avoid risk make profits.

Finally, if profits were the reward for risk bearing, then the greater the risk undertaken, the greater the profits. But, there is no empirical support to this inference which can be drawn from Hawley's theory.

Knight's Theory of Profit: Profit as a Return to Uncertainty Bearing

Frank H. Knight treated profit as a residual return to uncertainty bearing—not to risk bearing. Obviously, Knight made a distinction between **risk** and **uncertainty**. He divided risks into calculable and non-calculable risks. Calculable risks are those whose probability of occurrence can be statistically calculated on the basis of available data, e.g., risks due to fire, theft, accidents, etc. Such risks are insurable. There remains, however, an area of risks in which probability of risk occurrences cannot be calculated. For instance, there may be a certain element of cost which may not be accurately calculated; and the strategies of the competitors may not be accurately guessed. The risk element of such incalculable events are not insurable. The area of incalculable risks is thus marked by 'uncertainty'. It is in this area of uncertainty that decision becomes a peculiar responsibility of an entrepreneur. If his decisions are proved right by the subsequent events, the entrepreneur makes profit, and *vice versa*. Obviously, profit arises from the decisions taken and implemented under the conditions of uncertainty, as visualised by Knight. The profits may arise as a result of (a) decisions concerning the state of market; (b) decisions which result in increasing the degree of monopoly; (c) decisions with respect to holding stocks that give rise to windfall gains when prices increase; and (d) decisions taken to introduce new techniques or innovations that give rise to profit.

NOTES

Criticism of Knight's Theory of Profit

Several objections have been raised against Knight's theory of profit too.

First, it has been contended that Knight's uncertainty theory lacks scientific precision. Uncertainty is a difficult concept to handle. Tausig, for instance, has shown that though certain risks are in the area of uncertainly, many are not. For example, suppose that a person is betting in a horse race. If he has the knowledge of age, training, rearing, etc., of different horses and their jockeys, he would be operating in the region of *risk.* And, if he does not have the knowledge about the horses and jockeys participating in the race, he would be regarded as operating in the area of *uncertainty.* But, if he has some knowledge about the horses and/or jockeys, it will be difficult to decide whether the person is operating in the area of risk or in the area of uncertainty.

Secondly, by considering profit as a reward exclusively for uncertainty bearing, Knight has implicitly accorded it (uncertainty bearing) the status of a factor of production, whereas it is simply an element of real cost as distinguished from money cost. Therefore, uncertainty bearing cannot be accepted as a factor of production, and hence the sole cause of profit.

Thirdly, Knight's attempt to explain profits only by 'uncertainty' makes his theory unconvincing if one examines it in the light of real experience of the business world. If his theory is accepted, it would mean the greater the degree of uncertainty, the greater the profits, and *vice versa*. But there are enterprises, e.g., agriculture, which are known for their high uncertainty and low returns.

5.8.3 Schumpeter's Innovation Theory of Profit: Profit as Reward for Innovations

The Innovation Theory of Profit was developed by Joseph A. Schumpeter. Throughout his life as an economist, he was preoccupied with the study of economic evaluation and development in capitalist system. He was of the opinion that issues like interest, profit, trade cycles and many others were only incidental to a distinct process of economic development: and certain principles which could explain this process would also explain these economic variables. His theory of profit is thus embedded in his theory of economic development.

The Stationary Equilibrium: The Starting Point: To explain the phenomenon of economic development (and therefore of profit) Schumpeter starts from the state of stationary equilibrium which is characterised by full equilibrium in all the spheres. He assumes a closed, commercially organised, capitalist economy in which private property, division of labour and free competition prevail, along with constant population level. Everybody sells all his produce and insofar as be himself consumes, he is his own customer. The productive services may also be included in the same category of marketable things which are sold. But anyone who wants to purchase these goods or productive services must also have his own products or services to offer. Thus all goods and services are exchanged for one another. 'Hence it follows that somewhere in the economic system a demand is, so to say, ready awaiting

NOTES

every supply, and nowhere in the system are there commodities without complements...' It, therefore, follows that sellers of all the commodities appear as buyers to acquire the goods. This maintains their consumption and also productive capacity in the next period at the existing level. As a result, there emerges, 'an unchanging economic process which flows on at constant rates in time and merely reproduces itself.'

Profit as the Reward for Innovations: Under these conditions of stationary equilibrium, total receipts from the business are exactly equal to the total outlay: there is no profit. Profit can be made by introducing innovations in manufacturing and methods of supplying the goods. Innovations include:

(i) introduction of a new good or a new quality of good;

(ii) introduction of a new method of production;

(iii) creating or finding a new market;

(iv) finding new sources of raw material; and

(v) organising the industry in a different manner.

When an entrepreneur introduces an innovation, there will be a surplus over cost provided following conditions are fulfilled.

1. When a new supply comes forth as a result of innovation, the price of commodity should not fall to such an extent that it eliminates all the gains from the larger product.
2. The cost per unit of output with new technique should be less than that of older method.
3. The increase in demand for the productive services due to innovation should not lead to such a rise in remuneration to the productive services that it pushes per unit cost of the commodity beyond the expected revenue per unit.

If these conditions are fulfilled, the surplus realised will *ipso facto* become a net profit.

Profits Disappear Due to Imitation: The profits resulting from innovations exist only temporarily. This is so because when an entrepreneur introduces an innovation, others are likely to imitate it for its profitability. First a few and then many follow the lead, and produce the commodity in the same manner. This causes a keen competition for the productive services to be employed with the new techniques. Their supply remaining the same, their remuneration tends to increase. As a result, cost of production increases. On the other hand, with other firms adopting the innovations, supply of goods and services increases resulting in fall in their prices. Thus, on the one hand, cost per unit of output goes up and, on the other, revenue per unit decreases. Ultimately, a time comes when the difference between cost and receipt disappears. So the profit disappears. In the process, however, the economy reaches higher level of stationary equilibrium.

It is however quite likely that profits exist in spite of the process of profits being wiped out. Such profits are in the nature of *quasi-rent* arising due to some special characteristics of productive services. Furthermore, where profits arise due to factors like patents, trusts, cartels, etc., such profits would be in the nature of monopoly revenue rather than entrepreneurial profits.

It may be inferred from the aforementioned that **profit is the child as well as victim of economic development.** Economic development consists of increase in national output. When innovations occur the national output increases because the same output can be produced at lower costs, or what is the same thing, with the same amount of resources greater output can be produced. But producing at lower cost or producing more output with the same total cost results in profits. Thus, economic development gives birth to profits. But, when other producers also adopt the technique introduced by the innovator, the total national output increases, i.e., economic development catches pace. The widespread use of innovation, however, results in wiping out of profits, as was explained earlier. Hence, economic development itself is responsible for the disappearance of profits.

Criticism of Innovation Theory of Profit

The major criticism against Schumpeter's innovation theory of profit is that he ignores the risk and uncertainty, the two major sources of profit as shown in the traditional theories of profits. Although in his book *Capitalism, Socialism and Democracy,* he admits that innovations are made by the risk-taking entrepreneurs, he ignores uncertainty altogether. Besides, it has also been argued that innovation is not the only function of the entrepreneurs. As delineated in the dynamic theory of profit, entrepreneur's functions include organisational and coordinational activities also in response to the changing conditions and needs of the society.

Check Your Progress

7. Define the term 'pure profit'.
8. Who is credited for developed the dynamic theory of profit?

5.9 ANSWERS TO 'CHECK YOUR PROGRESS'

1. Distribution of national income between the factors of production is called functional distribution of national income, i.e., distribution of national income on the basis of economic functions of the people
2. The demand curve for factors is derived under two different conditions:
   (i) when a single variable factor is used in the process of production; and
   (ii) when more than one variable factor is used in the production process
3. The concept of the elasticity of factor substitution was developed by J.R. Hicks. It is regarded as the foundation of the modern neo-classical theory of distribution and relative factor shares.

NOTES

4. The quasi-rent, a concept used by Marshall, refers to the short-term earnings of factors which are in fixed supply in the short run.
5. Real wages have an indirect impact on money wages. This is so because an increase or decrease in the purchasing power will affect the money wage that a worker demands from his employer.
6. Bohm-Bawerk gives three reasons why people prefer the present to the future goods.

   First, the circumstances of want and provision in the present and in the future are different.

   Second, people underestimate future because of (a) deficiency of imagination, i.e., myopia in respect of future; (b) limited will power; and (c) the shortness and uncertainty of life. These factors lead to discounting of future wants.

   Third, present goods are technically superior to future ones.
7. 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.'
8. The dynamic theory of profit is associated with the name of J.B. Clark, which he propounded in 1900. According to Clark, profits accrue in a dynamic world, not in a static world.

5.10 SUMMARY

- The theory of factor pricing has its origin in the classical theory of distribution.
- The theory of distribution explains how national income (output) is distributed among the various factors of production.
- The theories of factor pricing were developed by different classical economists at different points of time, and are, therefore, different in approach to factor price determination.
- The productivity of a factor and demand for its product create the demand for the factor. But these factors do not determine the quantity demanded of a factor.
- Theoretically, the quantity demanded of a factor by a profit-maximizing firm is determined by the revenue it yields to the firm.
- Marginal productivity theory is regarded as the general micro-theory of factor price determination. It provides an analytical framework for the analysis of determination of factor prices.
- As first approximation, the market demand for a variable factor is the horizontal summation of the individual demand curves for the factor in question. However, a straightforward horizontal summation of individual demand curves for a

factor does not yield its market demand curve because this process leaves out the market or external effects of change in the factor prices.

- The market labour supply curve is the sum of individual labour supply curves.
- Economists however disagree on the shape of the market supply curve for labour.
- The ultimate aim of the distribution theory is to explain how the share of factors of production in total output is determined.
- Euler's product exhaustion theorem assumes a homogeneous production function, i.e., constant returns to scale. Clark, Wicksteed and Walras have, however, shown that the assumption of homogeneous production function is not necessary for the product exhaustion theorem.
- The concept of the elasticity of factor substitution was developed by J.R. Hicks. It is regarded as the foundation of the modern neo-classical theory of distribution and relative factor shares.
- The Ricardian theory of rent is the earliest known rent theory and is generally known as the classical theory of rent.
- Ricardian theory of rent has an interesting antecedent. In the early 19th century, food prices in Britain had considerably increased partly due to Napoleonic War and partly due to increase in population and the consequent increase in demand for food.
- The quasi-rent, a concept used by Marshall, refers to the short-term earnings of factors which are in fixed supply in the short run.
- Economic rent is the excess of actual earning of a factor over its transfer earning. Economic rent may thus be defined as factor's actual earning minus its transfer earning.
- The existence of economic rent depends on the elasticity of factor supply. Economic rent may be zero or equal to transfer earning depending on whether factor supply is perfectly elastic or perfectly inelastic.
- There are two types of wages: Nominal and real wages. Nominal wages are also known as money wages. Nominal wages are the types of wages which do not take into account inflation rate, purchasing power.
- It may be noted at the outset that the theoretical problems associated with capital and interest are much more complicated and difficult to resolve than those pertaining to other factors and factor prices.
- The earliest form of interest theory appeared in the writings of Locke, Petty, and Law in the 18th century. These authors considered interest as a monetary phenomenon and believed that interest varies inversely with supply of money: the larger the availability of money, the lower the interest, and vice versa.
- Senior's theory was later improved by Mill, which came to be known as 'Abstinence Theory of Interest'.

NOTES

NOTES

- Bohm-Bawerk's theory of interest is also referred to as Austrian theory of interest. Bohm-Bawerk, an Austrian economist, emphasised that land and labour are 'original' or primary factors of production, while 'capital' is an intermediate factor whose supply depends on land and labour used for its production.
- According to Fisher, interest arises because people prefer present to future income.
- Thus, Fisher's notion of interest is the same as that of Bohm-Bawerk. The rate of interest, according to Fisher's theory, equals the price that people are willing to pay for income now rather than income at some future date.
- The classical theory of interest refers, according to Keynes, to the theories (or views) of Marshall, Cassel, Tausig, Walras, etc. In fact, none of these authors, whom Keynes calls modern classical school, has given a precise or an explicit account of the interest theory.
- A variant of classical theory is the loanable fund theory of interest also called as neo-classical theory of interest. The neo-classical economists who have contributed to the growth of this theory include Wicksell, Ohlin, Robertson, Pigou and Viner.
- The Keynesian theory of interest is undoubtedly superior to the classical and neoclassical theory of interest. However, Keynesian theory has its own drawbacks for which it has been criticized by economists. Ironically, Keynesian theory has been criticized on the ground Keynes criticized the classical theory. Economists' concept of profit is of 'pure profit'.
- It is also called 'economic profit' or 'just profit'. The word 'profit' in this unit means 'pure profit'. 'Pure profit' is a return over and above opportunity cost, i.e., the payment that would be 'necessary to draw forth the factors of production from their most remunerative alternative employment.'
- One of the most widely known theories advanced to explain the nature of profit was formulated by F.A. Walker. According to him, profit is rent of the exceptional abilities that an entrepreneur may possess over the least entrepreneur.
- The dynamic theory of profit is associated with the name of J.B. Clark, which he propounded in 1900. According to Clark, profits accrue in a dynamic world, not in a static world.

5.11 KEY TERMS

- **Nominal wages:** These are the types of wages which do not take into account inflation rate, purchasing power. It is the amount that is promised to the worker when he/she is hired.

- **Calculable risks:** These are those whose probability of occurrence can be statistically calculated on the basis of available data, e.g., risks due to fire, theft, accidents, etc. Such risks are insurable.
- **Capital:** It is a man-made factor of production. It is reproducible. The stock of capital in a country consists of plant, machinery, building, etc.
- **Supply of Labour:** The total number of hours they are willing to work, given the reward per time unit, is called the supply of labour.
- **Economic rent:** It may be defined as factor's actual earning minus its transfer earning.
- **Technological progress:** It means a given quantity of output can be produced with less quantity of inputs or a given quantity of inputs can produce a greater quantity of output.
- **Pure profit:** It is a return over and above opportunity cost, i.e. the payment that would be 'necessary to draw forth the factors of production from their most remunerative alternative employment.

5.12 SELF-ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. Write a short note on the development of theories of factor pricing.
2. State the concept of marginal revenue productivity.
3. What are the differences between nominal wage and real wage?

Long-Answer Questions

1. Discuss the marginal productivity theory with examples.
2. Explain profit maximization effect with the help of a diagram.
3. Critically analyse the 'adding-up controversy'.
4. Elaborate the modern theory of rent.
5. Examine the classical theories of interest.
6. Analyse Keynes' criticism of classical theory of interest.
7. Discuss the modern theories of profit with examples.

5.13 FURTHER READING

Dwivedi, D. N. 2016. *Microeconomics: Theory and Application*, 3rd Edition. New Delhi: Vikas Publishing House.

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

NOTES

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

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

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

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

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