ALAGAPPA UNIVERSITY

ONLINE PROGRAMMES

B.B.A.

I - Semester

200313

PRINCIPLES OF ECONOMICS

Reviewer	
Dr. K. Alamelu	Professor, Department of Bank Management, Alagappa University, Karaikudi

Author

DN Dwivedi, *Professor of Economics, Maharaja Agrasen Institute of Management Studies, Delhi*

Units (1-14)

SYLLABI-BOOK MAPPING TABLE
Principles of Economics

BLOCK 4: Theory of Wages and Economic Analysis
UNIT 12: Distribution: Wages – Marginal productivity–Theory of wages-Collective bargaining – Wage differentials – Wages and productivity Wage regulation.
UNIT 13: Rent – Scarcity Vs Differential rents – Quasi rent– Rent as surplus over transfer earnings – Rent as economic surplus.
UNIT 14: Macro Economic Analysis- Theory of income and employment, Classical, Modern (Keynesian), Approach. Macro-Economic Variables, Circular flow of income, National Income Concepts, definition and its measurement.

CONTENTS

INTRODUCTION

The natural curiosity of a student who begins to study a subject or science is to know its nature and scope. Such as it is, a student of economics would like to know 'What is economics' and 'What is its subject matter'. Surprisingly, there is no precise answer to these questions. Attempts made by economists over the past 300 years to define economics have not yielded a precise and universally acceptable definition of economics. Economists right from Adam Smith—the 'father of economics'—down to modern economists have defined economics differently, depending on their own perception of the subject matter of economics of their era. Thus, economics is fundamentally the study of choice-making behaviour of the people. The choice-making behaviour of the people is studied in a systematic or scientific manner. This gives economics the status of a social science.

The basic function of business managers is to take appropriate decisions on business matters, to manage and organize resources, and to make optimum use of the available resources with the objective of achieving predetermined business goals. In today's world, business decision-making has become an extremely complex task due to the ever-growing complexity of the business world and the business environment. The dominant feature of the modern business environment is the ever-increasing inter-firm and inter-industry competition among domestic and international corporations. In a highly competitive business world, not only achieving the business goals but even the very survival and growth of the business firm depends largely on the appropriateness of business decisions and their effective implementation. Therefore, the techniques of business decision-making have changed tremendously of late.

This book, *Principles of Economics* has been divided into fourteen units. The book has been written in keeping with the self-instructional mode or the SIM format wherein each Unit begins with an Introduction to the topic, followed by an outline of the Objectives. The detailed content is then presented in a simple and organized manner, interspersed with Check Your Progress questions to test the student's understanding of the topics covered. A Summary along with a list of Key Words, set of Self-Assessment Questions and Exercises and Further Readings is provided at the end of each Unit for effective recapitulation.

BLOCK - I
BASICS OF MANAGERIAL ECONOMICS

UNIT 1   BASIC CONCEPTS OF ECONOMICS

Structure

1.0   INTRODUCTION

The basic function of business managers is to take appropriate decisions on business matters, to manage and organize resources, and to make optimum use of the available resources with the objective of achieving the predetermined business goals. In today's world, business decision-making has become an extremely complex task due to the ever-growing complexity of the business world and the business environment. The dominant feature of the modern business environment is the ever-increasing inter-firm and inter-industry competition among both domestic and international corporations. In a highly competitive business world, not only achieving the business goals, but also the survival and growth of the business firm depends largely on the appropriateness of the business decisions and their effective implementation. Therefore, the technique and process of business decision-making has of late changed tremendously. One of the prominent elements of the modern technique of business decision-making is the increasing application of economic logic, methodology, concepts, theories, and tools of economic analysis in arriving at an appropriate and feasible business decision. Therefore, a working knowledge of economic science has become essential for managers. This unit will begin with a discussion on the subject matter as well as the scope and method of economics. It will also discuss concepts such as scarcity and choice.

1.1 OBJECTIVES

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

- Define economics
- Describe the need for studying economics
- Discuss the scope and method of economics
- Explain how to read and work with graphs

1.2 SUBJECT MATTER OF ECONOMICS

A natural curiosity of a student who begins to study a subject or a science is to know the nature and scope of his subject of study. Such as it is, a student of economics would like to know 'what is economics' and 'what is its subject matter'. Surprisingly, there is no precise answer to these questions. Attempts made by economists over the past 300 years to define economics have not yielded a precise and universally acceptable definition of economics. Economists right from Adam Smith—the 'father of economics'—down to modern economists have defined economics differently depending on their own perception of the subject matter of economics of their era. For example, Adam Smith (1776) defined economics as 'an inquiry into the nature and causes of the wealth of the nations'. Nearly one-and-half century later, Alfred Marshall, an all time great economist, defined economics differently. According to Alfred Marshall (1922), "Economics is the 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". Lionel Robbins (1932) has defined it more precisely: "Economics is the science which studies human behaviour as a relationship between ends and scarce means which have alternative uses". One can find a number other definitions in economics literature. None of the definitions of economics, however, captures the entire subject matter of modern economics, though they do throw some light on what economics is about.

The study of economic science, or of any science for that matter, must commence with a working definition of it. In this regard, most modern texts follow Robbins' definition of economics, even though modern economics goes far beyond what Robbins thought to be the subject matter of economics. Let us begin with Robbins' view on subject matter of economics and then look how far it goes beyond his view.

Economics is a Social Science

Economics as a *social science* studies *economic behaviour* of the people and its consequences. What is economic behaviour? *Economic behaviour* is essentially the process of evaluating economic opportunities open to an individual or a society

and, given the resources, making 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 the government) 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 the people. The choice-making behaviour of the people is studied in a systematic or scientific manner. This 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, factor 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, the 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.*

Check Your Progress

1. How did Adam Smith define economics?
2. What does economics as a social science study?

1.3 SCOPE AND METHOD OF ECONOMICS

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-War period. *Modern economics* is how 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 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 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 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 relationship between the national aggregate variables and how aggregate variables interact with one another to determine one another. It also studies 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 the

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 growing need of 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.

2. ***Public Economics*** examines economic role of the 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, inter-national monetary and financial institutions, balance of payments, and international payment system.

5. ***Industrial Economics*** is concerned with the working, growth and structures of the industrial sector (firms and industries) of the country, management and organization of the 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 the 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 the unknown facts to the 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 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 the rural economy.

12. ***Industrial Finance*** is concerned with the development and working of 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 affects the working and growth of the industrial sector.

13. ***Environmental Economics*** examines how industrial growth affects, rather destroys, 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 business environment of the country.

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

1.3.1 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.

Economics as a Positive Science

Economics as a positive science seeks to analyze 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 response to change in economic conditions; and it brings out the relationship between the change in economic conditions and economic decision of the people. In fact, *the main function of economics is to establish cause-and-effect relationship, if there is any, between two or more economic events and to provide basis for prediction.* Emphasizing the positive character of economics, Friedman says, "Economics as a positive science is a body of tentatively accepted generalizations about economic phenomena that can be used to predict the consequences of change in circumstances." One of the main tasks of economics is 'to provide a system of generalizations' or, more precisely, economic theories, capable of being used to predict economic 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 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.

1.3.2 Methodology of Economics: What Economists Do and How

The basic function of economists is to observe and analyse economic phenomena and to formulate economic theories. *An economic theory is the statement of a general tendency*. Specifically, *an economic theory is the statement of cause-and-effect relationship between two or more observed facts of real economic life*. To formulate an economic theory, economists use a scientific method of study. Scientific method of investigation involves *observation of economic phenomena or events, collection and analysis of relevant data and prediction of economic phenomena. Predictive statements give the cause-and-effect kind of relationship between two or more economic variables*. When the relationship between the selected variables is established with a high degree of confidence, it is presented in the form of a theory or a hypothesis. This process is called *theorization* or *formulation of theory*. The process of theorization involves a scientific investigation which involves model building. Let us now look briefly at the method of model building in economics.

Model Building and Formulation of Economic Theory

An important element of scientific method of inquiry is *model building*. Conceptually, a *model* is an abstraction from reality. It represents reality in a simplified form. Practically, a model is a logically consistent analytical framework made for analyzing facts of life in an abstracted form. Economic models may take the form of a logical statement, graph or mathematical equations specifying the relationship between the economic variables. Models are used to specify the relationship between the selected variables, to deduce the consequences of the changes in the variables and to make predictions. *Economic variables* are measurable quantities, e.g., consumer goods, output, inputs, money, income, etc. The economic variables assumed to remain constant are called *parameters*. The general process of model building and theorization in economics is described below.

Model building and economic theorization consists of the following steps:

(i) Specifying the problem of study,

(ii) Formulating a testable hypothesis,

(iii) Making assumptions and making postulates,

(iv) Collection of related data and other relevant facts,

(v) Deducing the testable predictions,

(vi) Testing the validity of predictions.

Check Your Progress

3. What is microeconomics concerned with?

4. What is the objective of economics as a normative science?

1.4 THE ECONOMIC PROBLEM: SCARCITY AND CHOICE

As noted above, the economic behaviours of the people is essentialy the choice-making behaviour. 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.

Here a question arises: *Why does the problem of choice making arise?*

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

1. Human Wants, Desires and Aspirations are Unlimited. 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 mule and camel to supersonic jet planes; demand for means of communication has rises from messengers and postal services to cell phones with camera; need for computational facility from manual calculation to superfast computers; and so on. For an individual, only the end of life brings the 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*) increasing knowledge about inventions and innovations of new goods and services with greater convenience, efficiency and serviceability; (*iv*) multiplicative nature of some want (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 kind of wants. For these reasons, human wants continue to increase endlessly.

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 or important, at a point 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) and some can be postponed, e.g., purchase of a car, a computer, a good house, a tour abroad, etc. Also, while satisfying some wants gives a greater satisfaction than others. Therefore, the question arises as to 'which want to satisfy first' and

'which the last'. Besides, goods available to satisfy a want vary in their quality, efficiency and cost. For example, one can travel a distance by a bike, a car, a plane, etc. but their cost and efficiency differ. Therefore, the consumers have to make choice 'what to consume' and 'how much to consume'. Economics studies how consumers (individuals and household) make choice of their wants to be satisfied first 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 choice from 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, any thing 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, equipments, tools, technology and building, called together *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 some 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*.

3. Alternative Uses of Resources. 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 center, business office, a 'public school', a hospital or for residential purpose. But the return from the building varies from the different uses of the building. Therefore, a return maximizing owner has to make 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.

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 evaluates the costs and benefits of the alternatives while deciding on the final use of the resources. Economics studies the process of making choices between the alternative uses.

4. **People are Gain Maximizers.** Another and equally 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 maximizing behaviour of the people as a part of their rational economic behaviour. This assumption is based on observed facts. As *consumers,* people want to maximize their utility or satisfaction; as *producers,* people 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 the 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.

Economics Goes Far Beyond Choice-Making Behaviour

The foregoing description of economics may give the impression that economics ends at the study of choice-making behaviour of the people. Not quite so. Economics goes far beyond the study of choice-making behaviour. Had economics been confined to the study of choice-making behaviour of the people, many other and more important economic issues that constitute a major part of modern economic science would have not fallen within the scope of economics. Look at some of the major national and international economic issues.

- How is the level of output and employment determined in a country?

- Why are some countries very rich and some countries very poor?

- What are the factors that increased India's economic growth from about 3.5% before 1990 to 9% in 2008?

- What causes fluctuations in the national output, employment and the general price level?

- How has rupee appreciation against US dollar affected India's international trade and how have international capital flows (FDI and FII) affected the domestic economy of the country?

- What has caused inflation in India in 2008 and what are its effects on economy's growth and employment?

- Why is there large scale unemployment in India and why have the efforts to solve the problem of unemployment failed?

- Why does the government including the Finance Minister and the central bank need to intervene with the market system and adopt measures to control and regulate production and consumption, saving and investment, export and imports, wages and prices, and so on?

1.5 READING AND WORKING WITH GRAPHS

Graphical Presentation

The data we collect can often be more easily understood for interpretation if it is presented graphically or pictorially. Diagrams and graphs give visual indications of magnitudes, groupings, trends and patterns in the data. These important features are more simply presented in the form of graphs. Also, diagrams facilitate comparisons between two or more sets of data.

The diagrams should be clear and easy to read and understand. Too much information should not be shown in the same diagram otherwise it may become cumbersome and confusing. Each diagram should include a brief and self-explanatory title dealing with the subject matter. The scale of the presentation should be chosen in such a way that the resulting diagram is of appropriate size. The intervals on the vertical as well as the horizontal axis should be of equal size, otherwise distortions would occur.

Diagrams are more suitable to illustrate the data which is discrete while continuous data is better represented by graphs. The following are the diagrammatic and graphic representations that are commonly used.

(A) Diagrammatic representation

    (a) Bar diagram

    (b) Pie chart

    (c) Pictogram

(B) Graphic representation

    (a) Histogram

    (b) Frequency polygon

    (c) Cumulative frequency curve (Ogive)

Check Your Progress

5. How are human wants gradable?

6. What are the graphical representation commonly used in economics?

1.6 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. Adam Smith defined economics as an inquiry into the nature and causes of the wealth of nations.

2. Economics as a social science studies economic behavior of the people and its consequences.

3. Microeconomics is concerned with microscopic study of the various elements of the economic system and not with the system as a whole.

4. The objective of economics as a normative science is to examine 'what actually happens' from moral and ethical points of view and to judge whether 'what happens' is socially desirable.

5. Human wants are gradable since not all human wants are equally urgent and pressing, at a point time or over a period of time.

6. The graphical representation commonly used in economics are (a) Histogram; (b) Frequency polygon; (c) Cumulative frequency curve.

1.7 SUMMARY

- Economics is the science which studies human behaviour as a relationship between ends and scarce means which have alternative uses.

- Economic behaviour is essentially the process of evaluating economic opportunities open to an individual or a society and, given the resources, making choice of the best of the opportunities.

- Modern economics is now divided into two major branches: Microeconomics and Macroeconomics.

- Microeconomics is concerned with microscopic study of the various elements of the economic system and not with the system as a whole. 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 country's balance of payments.

- Economics as a positive science seeks to analyze 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.

- 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.

- 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 need for making choice from 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.

- The data we collect can often be more easily understood for interpretation if it is presented graphically or pictorially. Diagrams and graphs give visual indications of magnitudes, groupings, trends and patterns in the data.

1.8   KEY WORDS

- **Microeconomics:** It is the part of economics concerned with single factors and the effects of individual decisions.

- **Macroeconomics:** It is the branch of economics concerned with large-scale or general economic factors, such as interest rates and national productivity.

- **Labour Economics:** The branch of economics that seeks to understand the functioning and dynamics of the markets for wage labour.

- **Histogram:** It is a diagram consisting of rectangles whose area is proportional to the frequency of a variable and whose width is equal to the class interval.

1.9   SELF ASSESSMENT QUESTIONS AND EXERCISES

Short Answer Questions

1. Define economics.
2. What is the need for studying economics?
3. Write a short note on the scope of economics.
4. Mention the use of graphs in economics.

Long Answer Questions

1. The origin of economics lies in endless human wants and scarcity of resources. Elaborate.
2. Discuss the methodology of economics.

3. 'Scarcity of resources is the mother of all economic problems.' Discuss with examples.

1.10 FURTHER READINGS

Dwivedi, D. N. 2008. *Principles of Economics*, Seventh Edition. New Delhi: Vikas Publishing House.

Weil. David N. 2004. *Economic Growth*. London: Addison Wesley.

Thomas, Christopher R. and Maurice S. Charles. 2005. *Managerial Economics: Concepts and Applications*, Eighth Edition. New Delhi: Tata McGraw-Hill Publishing Company Limited.

Mankiw, Gregory N. 2002. *Principles of Economics*, Second Edition. India: Thomson Press.

UNIT 2 INTRODUCTION TO MANAGERIAL ECONOMICS

Structure

2.0 INTRODUCTION

This unit will introduce you to managerial economics. Managerial economics, as defined by Edwin Mansfield, is 'concerned with the application of economic concepts and economic analysis to the problems of formulating rational managerial decisions'. Also sometimes referred to as business economics, it is a branch of economics that applies microeconomic analysis to business decision-making. This unit will discuss the nature and scope of managerial economics, as well as the application of managerial economics to business.

2.1 OBJECTIVES

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

- Discuss the nature and scope of managerial economics
- Define opportunity cost
- Differentiate between marginalism and incrementalism
- Explain the concept of time value of money, market forces and equilibrium

2.2 NATURE, SCOPE, DEFINITION OF MANAGERIAL ECONOMICS

The subject matter of economic science consists of the logic, tools and techniques of analyzing economic phenomena as well as, *evaluating economic*

options, optimization techniques and economic theories*. The application of economic science in business decision-making is all pervasive. More specifically, economic laws and tools of economic analysis are now applied a great deal in the process of business decision-making. This has led, as mentioned earlier, to the emergence of a separate branch of study called **managerial economics.** The application of economic concepts and theories in combination with quantitative methods is illustrated in Fig. 2.1.
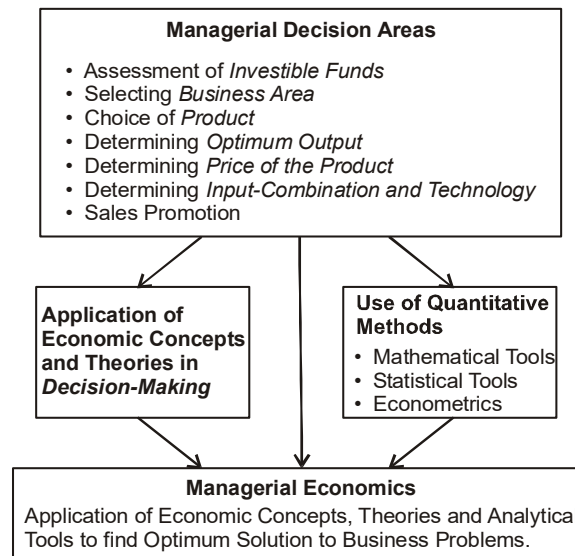
Fig. 2.1 Application of Economics to Managerial Decision-Making

Managerial economics can be broadly defined as the study of economic theories, logic and tools of economic analysis that are used in the process of business decision-making. Economic theories and techniques of economic analysis are applied to analyze business problems, evaluate business options and opportunities with a view to arriving at an appropriate business decision. Managerial economics is thus constituted of that part of economic knowledge, logic, theories and analytical tools that are used for rational business decision-making.

Some Definitions of Managerial Economics

Let us look at some definitions of managerial economics offered by a few eminent economists.

"Managerial economics is concerned with the application of economic concepts and economics to the problems of formulating rational decision making".
—Mansfield

"Managerial economics ... is the integration of economic theory with business practice for the purpose of facilitating decision making and forward planning by management".   *—Spencer and Seigelman*

"Managerial economics is concerned with the application of economic principles and methodologies to the decision-making process within the firm or organization. It seeks to establish rules and principles to facilitate the attainment of the desired economic goals of management".   —*Douglas*

"Managerial economics applies the principles and methods of economics to analyze problems faced by the management of a business, or other types of organizations and to help find solutions that advance the best interests of such organizations".   —*Davis and Chang*

These definitions of managerial economics together reveal the nature of the discipline even though they do not provide its perfect definition.

Why do Managers Need to Know Economics?

Economics contributes a great deal towards the performance of managerial duties and responsibilities. Just as biology contributes to the medical profession and physics to engineering, economics contributes to the managerial profession. All other professional qualifications being the same, managers with a working knowledge of economics can perform their functions more efficiently than those without it. The basic function of the managers of a business firm is to achieve the objective of the firm to the maximum possible extent with the limited resources placed at their disposal. The emphasis here is on the maximization of the objective and limitedness of the resources. Had the resources been unlimited, the problem of economizing on resources or resource management would have never arisen. But resources at the disposal of a firm, be it finance, men or material, are by all means limited. Therefore, the basic task of the management is to optimize their use.

How Economics Contributes to Managerial Functions

We have noted above why managers need to know economics. Let us now see how economics contributes to the managerial task of decision-making. As mentioned above, *economics,* though variously defined, *is essentially the study of logic, tools and techniques of making optimum use of the available resources to achieve the given ends.* Economics thus provides analytical tools and techniques that managers need to achieve the goals of the organization they manage. Therefore, a working knowledge of economics, not necessarily a formal degree, is essential for managers. In other words, managers are essentially practicing economists.

In performing their functions, managers have to take a number of decisions in conformity with the goals of the firm. Many business decisions are taken under conditions of **uncertainty** and **risk**. These arise mainly due to uncertain behaviour of the market forces, changing business environment, emergence of competitors with highly competitive products, government policy, international factors impacting the domestic market mainly due to increasing globalization as well as social and political changes in the country. The complexity of the modern business world

adds complexity to business decision-making. However, the degree of uncertainty and risk can be greatly reduced if market conditions are predicted with a high degree of reliability. Economics offers models, tools and techniques to predict the future course of market conditions and business prospects.

The prediction of the future course of business environment alone is not sufficient. What is equally important is to take appropriate business decisions and to formulate a business strategy in conformity with the goals of the firm. Taking a rational business decision requires a clear understanding of the technical and environmental conditions related to the business issues for which decisions are taken. Application of economic theories to explain and analyze the technical conditions and the business environment contributes a good deal to rational decision-making. Economic theories have, therefore, gained a wide range of application in the analysis of practical problems of business. With the growing complexity of business environment, the usefulness of economic theory as a tool of analysis and its contribution to the process of decision-making has been widely recognized.

Baumol has pointed out **three main contributions** of economic theory to business ecomomics.

First, 'one of the most important things which the economic [theories] can contribute to the management science' is building analytical models, which helps to recognize the structure of managerial problems, eliminate the minor details that might obstruct decision-making, and help to concentrate on the main issue.

Secondly, economic theory contributes to the business analysis 'a set of analytical methods', which may not be applied directly to specific business problems, but they do enhance the analytical capabilities of the business analyst.

Thirdly, economic theories offer clarity to the various concepts used in business analysis, which enables the managers to avoid conceptual pitfalls.

Application of Economics to Business Decisions

We have discussed above – in general terms, of course – how economics can contribute to business decision-making. In this section, we show this application in some hypothetical business issues.

Business decision-making is essentially a process of selecting the best out of alternative opportunities open to the firm. The process of decision-making comprises four main phases:

(i) determining and defining the objective to be achieved;

(ii) collections and analysis of business related data and other information regarding economic, social, political and technological environment and foreseeing the necessity and occasion for decision;

(iii) inventing, developing and analyzing possible courses of action; and

(iv) selecting a particular course of action, from the available alternatives.

This process of decision-making is, however, not as simple as it appears to be. Steps (*ii*) and (*iii*) are crucial in business decision-making. These steps put managers' analytical ability to test and determine the appropriateness and validity of decisions in the modern business world. Modern business conditions are changing so fast and becoming so competitive and complex that personal business sense, intuition and experience alone may not prove sufficient to make appropriate business decisions. Personal intelligence, experience, intuition and business acumen of the decision-makers need to be supplemented with quantitative analysis of business data on market conditions and business environment. It is in this area of decision-making that economic theories and tools of economic analysis contribute a great deal.

Economic theories state the functional relationship between two or more economic variables, under certain given conditions. Application of relevant economic theories to the problems of business facilitates decision-making in at least three ways.

First, it gives a clear understanding of various economic concepts (e.g., cost, price, demand, etc.) used in business analysis. For example, the concept of 'cost' includes 'total', 'average', 'marginal', 'fixed', 'variable', 'actual', and 'opportunity'. Economics clarifies which cost concepts are relevant and in what context.

Second, it helps in ascertaining the relevant variables and specifying the relevant data. For example, it helps in deciding what variables need to be considered in estimating the demand for two different sources of energy—petrol and electricity.

Third, economic theories state the general relationship between two or more economic variables and also events. Application of the relevant economic theory provides consistency to business analysis and helps in arriving at right conclusions. Thus, application of economic theories to the problems of business not only guides, assists and streamlines the process of decision-making but also contributes a good deal to the validity of the decisions.

The Scope of Managerial Economics

The areas of business issues to which economic theories can be directly applied may be broadly divided into two categories: (*a*) microeconomics applied to operational or internal issues, and (*b*) macroeconomics applied to environmental or external issues.

Micro-economics Applied to Operational Issues

Operational issues are of internal nature. Internal issues include all those problems which arise within the business organization and fall within the purview and control of the management. Some of the basic internal issues are: (*i*) choice of business and the nature of product, i.e., what to produce; (*ii*) choice of size of the firm, i.e., how much to produce; (*iii*) choice of technology, i.e., choosing the factor-

combination; (*iv*) choice of price, i.e., how to price the commodity; (*v*) how to promote sales; (*vi*) how to face price competition; (*vii*) how to decide on new investments; (*viii*) how to manage profit and capital; (*ix*) how to manage an inventory, i.e., stock of both finished goods and raw materials. These problems may also figure in forward planning. Microeconomics deals with such questions confronted by managers of business enterprises. The following microeconomic theories deal with most of these questions.

Theory of Demand

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.

The theory of demand has been discussed in detail in Unit 4.

Theory of Production and Production Decisions

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.

Analysis of Market-Structure and Pricing Theory

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 Analysis and Profit Management

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.

Theory of Capital and Investment Decisions

Capital like all other inputs, is a scarce and expensive factor. Capital is the foundation of business. Its efficient allocation and management is one of the most important tasks of the managers and a determinant of the success level of the firm. The major issues related to capital are (*i*) choice of investment project, (*ii*) assessing the efficiency of capital, and (*iii*) most efficient allocation of capital. Knowledge of capital theory can contribute a great deal in investment-decision making, choice of projects, maintaining the capital, capital budgeting, etc.

Macro-economics Applied to Business Environment

Environmental issues pertain to the general business environment in which a business operates. They are related to the overall economic, social and political atmosphere of the country. The factors which constitute *economic environment* of a country include the following:

  (*i*) the type of economic system in the country,

 (*ii*) general trends in national income, employment, prices, saving and investment, etc.,

(*iii*) structure of and trends in the working of financial institutions, e.g., banks, financial corporations, insurance companies, etc.,

 (*iv*) magnitude of and trends in foreign trade,

  (*v*) trend in labour supply and strength of the capital market,

 (*vi*) government's economic policies, e.g., industrial policies, monetary, fiscal, price and foreign trade,

(*vii*) social factors like value system of the society, property rights, customs and habits,

(*viii*) socio-economic organizations like trade unions, consumers' associations, consumer cooperatives and producers' unions,

 (*ix*) *political environment*, which is constituted of such factors as political system—democratic, authoritarian, socialist, or otherwise, State's attitude towards private business, size and working of the public sector and political stability, and

(*x*) the degree of globalization of the economy and the influence of MNCs on the domestic markets.

It is far beyond the powers of a single business firm, howsoever large it may be, to determine and guide the course of economic, social and political factors of the nation. However, all the firms together or giant business houses can jointly influence the economic and political environment of the country. For the business community in general, however, the economic, social and political factors are to be treated as business parameters.

The *environmental factors* have a far-reaching bearing upon the functioning and performance of the firms. Therefore, business decision-makers have to take into account the present and future economic, political and social conditions in the country and give due consideration to the environmental factors in the process of decision-making. This is essential because business decisions taken in isolation of environmental factors may not only prove infructuous, but may also lead to heavy losses as has happened in case of establishing Special Economic Zones (SEZs) in Nandigram and Tata's small make it car project in Singur district of West Bengal. Consider also, for example, the following kinds of business decisions—

- a decision to set up a new alcohol manufacturing unit or to expand the existing ones ignoring the impending prohibition—a political factor—would be suicidal for the firm;

- a decision to expand the business beyond the paid-up capital permissible under Monopolies and Restrictive Trade Practices Act (MRTP Act) amounts to inviting legal shackles;

- a decision to employ a highly sophisticated, labour-saving technology ignoring the prevalence of mass open unemployment—an economic factor—may prove to be self-defeating;

- a decision to expand the business on a large scale, in a society having a low per capita income and hence a low purchasing power stagnant over a long period may lead to wastage of resources.

The managers of business firms are, therefore, supposed to be fully aware of the economic, social and political conditions prevailing in the country while taking decisions on business issues of wider implications.

Managerial economics is, however, concerned with only the **economic environment**, and in particular with those economic factors which form the business climate. The study of political and social factors falls out of the purview of managerial economics. It should, however, be borne in mind that economic, social and political behaviour of the people are interdependent and interactive. For example, growth of monopolistic tendencies in the industrial sector of India led to the enactment of the Monopolies and Restrictive Trade Practices Act (1961), which restricts the proliferation of large business houses. Similarly, various industrial policy resolutions formulated until 1990 in the light of the socio-political ideology of the government

restricted the scope and area of private business and had restrained the expansion of private business in India. Some of the major areas in which politics influences economic affairs of the country are concentration of economic power, growth of monopoly, state of technology, existence of mass poverty and unemployment, foreign trade, taxation policy, labour relations, distribution system of essential goods, etc.

Macro-economic Issues

The major macroeconomic or environmental issues that figure in business decision-making, particularly with regard to forward planning and formulation of the future strategy, may be described under the following three categories.

1. *Issues Related to Macroeconomic Trends in the Economy.* Macro-economic trends are indicated by the trends in macro variables, e.g., the general trend in the economic activities of the country, the level of GDP, investment climate, trends in national output (measured by GNP or GDP) and employment, as well as price trends. These factors not only determine the prospects of private business, but also greatly influence the functioning of individual firms. Therefore, a firm planning to set up a new unit or expand its existing size would like to ask itself 'What is the general trend in the economy? What would be the consumption level and pattern of the society? Will it be profitable to expand the business?' Answers to these questions and the like are sought through macroeconomic studies.

2. *Issues Related to Foreign Trade.* Most countries have trade and financial relations with other countries. The sectors and firms dealing in exports and imports are affected directly and more than the rest of the economy. Fluctuations in the international market, exchange rate and inflows and outflows of capital in an open economy have a serious bearing on its economic environment and, thereby, on the functioning of its business undertakings. The managers of a firm would, therefore, be interested in knowing the trends in international trade, prices, exchange rates and prospects in the international market. Answers to such problems are obtained through the study of international trade and monetary mechanism. These aspects constitute a part of macroeconomic studies.

3. *Issues Related to Government Policies.* Government policies designed to control and regulate economic activities of the private business affect the functioning of the private business undertakings. Besides, firms' activities as producers and their attempt to maximize their private gains or profits leads to considerable social costs, in terms of environment pollution, traffic congestion in the cities, creation of slums, etc. Such social costs not only bring a firm's interests in conflict with those of the society, but also impose a social responsibility on the firms. The government's policies and its regulatory measures are designed, by and large, to minimize such social costs and conflicts. The managers should, therefore, be fully aware of the aspirations of the people and give such factors a due consideration in their decisions. The forced closure of polluting industrial units set up in the residential areas of Delhi and the consequent loss of business worth billion of rupees in 2000 is an

example of the result of ignoring the public laws and the social responsibility by the businessmen. The economic concepts and tools of analysis help in determining such costs and benefits.

Concluding Remarks

Economic theories, both micro and macro, have a wide range of applications in the process of business decision-making. Some of the major theories which are widely applied to business analysis have been mentioned above. It must, however, be borne in mind that economic theories, models and tools of analysis do not offer readymade answers to the practical problems of individual firms. They provide only the logic and methods to find answers, not the answers as such. It depends on the managers' own understanding, experience, intelligence, training and their competence to use the tools of economic analysis to find reasonably appropriate answers to the practical problems of business.

Briefly speaking, **microeconomic theories** including those of demand, production, price determination, profit and capital budgeting, and **macroeconomic theories** including of national income, those economic growth and fluctuations, international trade and monetary mechanism, and the study of state policies and their repercussions on the private business activities, by and large, constitute the scope of managerial economics. This should, however, not mean that only these economic theories form the subject-matter of managerial economics. Nor does the knowledge of these theories fulfill wholly the requirement of economic logic in decision-making. An overall study of economics and a wider understanding of economic behaviour of the society, individuals, firms and state would always be desirable and more helpful.

Check Your Progress

1. How does economics contribute to the managerial functions?

2. What are the areas of business issues to which economic theories can be directly applied?

3. What is the most common objective of all business undertakings?

2.3 APPLICATION OF MANAGERIAL ECONOMICS TO BUSINESS

Before we proceed to the central theme of business economics, it will be useful to understand some basic concepts which are often used in business analysis and business decision making. In this section, we explain certain economic concepts that are used in business decision-making . The economic concepts which are explained here include:

(*i*) Opportunity cost,

(*ii*) Time value of money,

(*iii*) Marginalism and incrementalism

(*iv*) Contribution analysis

(*v*) Optimization Rule, and

(*vi*) Risk, return and profit

A brief explanation of these economic concepts follows.

1. Opportunity Cost

The concept of opportunity cost is one of the most important concepts used in economic and business analyses. The opportunity cost concept is used in making choice from the alternative opportunities available to a person or to a business firm. We explain here a briefly the concept of opportunity cost.

The concept of opportunity cost is related to *scarcity of resources* and their alternative uses. As pointed out earlier, resources available to any person, firm or society are limited. But resources have alternative uses with different productivity, i.e., income or returns from the alternative uses of resources are different. While one kind of use yields a higher income or return, the other uses yield a lower return or income. Due to income maximizing behaviour, people (individuals, households and firms) put their scarce resources to the use that yields highest income, return or benefit. When they put their resources to the use yielding highest income, they sacrifice the income expected from the next, the second best, use of the resources. In economics terminology, this sacrifice is called, ***opportunity cost*** of earning from the best use of resources. Such as it is, *opportunity cost* may be defined as income expected from the second best use of the resources which is sacrificed for the best use of the resources. Thus, *opportunity cost is opportunity lost.* From a firm's point view, opportunity cost of using a resource is what the firm must give up to use the resource as it is used.

Opportunity cost is also called *alternative cost.* It is alternative cost because it arises due to the possibility of alternative uses of the resources. If a resource has only one use, i.e., it has no alternative use, there would not be any alternative or opportunity cost.

The concept of opportunity cost can be explained further by some examples. Suppose Abhishek has a sum of ₹ 10 million for which he has only two uses: (*i*) time deposit in the bank at 8% interest rate, and (*ii*) buy mutual funds with an annual average return at the rate of 10%. Being an income maximizer, Abhishek would invest his money in mutual funds. It means he foregoes the return expected from the time deposit, i.e., the second best alternative use of his money. The loss of bank interest is the *opportunity cost* of earning from mutual funds.

Consider another example. Suppose Sanjiv, a businessman, has an investible fund of ₹ 100 million for which it has three alternative uses listed as follows:

(*i*) investing his money in his own business yielding an annual return of ₹ 12 million,

(*ii*) investing money in stock market expecting an annual average earning of ₹ 10 million, and

(*iii*) investing in real estate with a return of ₹ 9 million.

A businessman, being a profit maximizer, will put his money in his own business, i.e., the first best option. It means that he foregoes an annual income of ₹ 10 million expected from the share business, the next best use of money. According to the concept of opportunity cost, the foregone annual return of ₹ 10 million is the *opportunity cost* of earning from the use of money in businessman's own business. It must be borne in mind that in working out opportunity cost or alternative cost, both explicit and implicit costs are taken into account.

Economic Profit. Another important concept which is associated with the concept of *opportunity cost,* is the concept of ***economic profit*** or ***economic rent.*** ***Economic profit*** is defined as the *actual earning less opportunity cost.* In our example of businessman having an investible fund of ₹ 100 million, his *economic profit* can be worked out as follows.

Economic profit = Actual earning – Opportunity cost

= ₹ 12 million – ₹ 10 million = ₹ 2 million.

2. Time Value of Money

Money has a time value. The concept of time value of money is well represented by a proverb, 'a bird in hand is preferable to two in bush'. That money has a time value can be understood by the fact that a penny received today is preferable to a penny receivable tomorrow. In general, present money is preferable to future money for the following reasons.

(*i*) It gives convenience of spending as the need arises,

(*ii*) It gives liquidity and confidence, and

(*iii*) Given time, money can earn more money: it can be invested to earn a return or interest.

In fact, the interest aspect of the time value of money is most important in economic analysis, especially in investment decision-making.

The concept of the ***time value of money*** is very often applied to investment decisions. Generally, there is time lag between investment and return. When an investment is made today, it begins to yield returns at some future date. The time gap between investment and return is called ***time lag***. During the period of time lag, the investor loses interest on the return receivable in future. It means that a rupee expected in future, say, a year hence, is worth less than a rupee today. This implies that present money has a ***future value***, if invested, and future money has a ***present value***. The concepts of *future value* and *present value* of money can

be clarified by using the method of working out the future value of present money and present value of future money.

The Future Value of Present Money. The future value of a sum of present money (money in hand) is simply the *money invested plus interest.* In simple words, the future value, called *amount, equals principal plus* interest, i.e., *amount = principal + interest.* Given this definition, the *future value* can be calculated by the method of calculating the *amount.* For example, suppose a sum of ₹ 100 is deposited with a bank for a period of one year at the interest rate of 10 per cent per annum. The formula for finding the *future value of this money,* i.e., the *amount,* is given as follows.

$$\text{Amount} = X + X(r)$$

where *X* is principal (₹ 100) and *r* is the rate of interest (10%).

By using this formula, the 'amount' (the future value of ₹ 100), can be found as follows.

$$
\begin{aligned}
\text{Future value} \quad &= 100 + 100\,(0.10) \\
&= 100 + 10 \\
&= 110
\end{aligned}
$$

Thus, given the rate of interest, the future value of a sum of money can be easily found.

The Present Value of Future Income. The present value *(PV)* of future income, an income expected at some future date, can be obtained by using the following formula.

$$PV = \frac{X}{(1+r)}$$

where *X* is future income and *r* is the market rate interest.

Given the formula, the *PV* of ₹ 110 at 10% rate of interest can be obtained as follows.

$$PV = \frac{X}{(1+r)} = \frac{110}{(1+0.10)} = \text{Rs } 100$$

Thus, the *present value* of ₹ 110 expected one year hence is ₹ 100, at interest rate of 10%. The formula for working out *PV* of income in any future year *(n)* is given below.

$$PV = \frac{X}{(1+r)^n} \quad (n \text{ being the } n\text{th year}).$$

3. Marginalism and Incrementalism

The terms Marginalism and Incrementalism refer to two principles widely used in economic analysis and business analysis. The two principles, called also as rules,

are known as *marginal principle* and *incremental principle,* respectively. Marginalism and Incrementalism rules are also used to work out the *marginal* and *incremental returns from a business activity,* respectively. Let us look at these principles and at their application in some detail.

Marginalism—The Marginal Analysis: One of the most important contributions that economics has made to business analysis and managerial decisions is the application of *marginal analysis.* The marginal analysis uses marginal change in the dependent variable resulting from a unit change in its determinant, the independent variable, given the functional relationship between the dependent and independent variable.

For example, look at the following cost function.

$$TC = f(Q)$$

Given the cost function, *TC* is the dependent variable and *Q* is the independent variable. The relevant marginality concept here is marginal cost *(MC). Marginal cost* can be defined as the change in *TC* due to change in *Q*. There are two ways of expressing the marginal cost *(MC):* (*i*) *MC* per marginal unit of output, and (*ii*) *MC* with a very small change in output.

When a series of *Q* and *TC* data is given, *MC* per unit can be calculated as

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

For example, suppose Cost-Output Table shows that *TC* of 100 units of output *(Q)* of a commodity is ₹ 2500 and *TC* of output 101 units is ₹ 2550. In that case,

$$MC = TC_{101} - TC_{101-1}$$
$$= ₹\ 2550 - ₹\ 2500$$
$$= ₹\ 50$$

Thus, marginal cost (*MC*) of a product is ₹ 50.

The concepts of marginal value and marginal analysis are used widely in analyzing many behavioural issues, e.g., marginal utility *(MU)* in consumer behaviour; marginal cost *(MC)* in cost analysis; marginal revenue (*MR*) in revenue maximization problem; and marginal productivity (*MP*) in assessing the output effect of a business decision to employ one additional unit of an input.

The application of marginal principle will be shown ahead when we deal with equilibrium of consumer and firm and price determination.

Incrementalism—The Incremental Analysis: Similar to the concept of 'marginal' value is the concept of 'incremental' value. As mentioned above, marginal principle is applied only where the independent variable, or the determinant, changes by only one unit. In general, however, most business firms produce and sell their products in bulk, not unit by unit unless, of course, it is the case of producing

such large-unit goods as airplanes, ships, buildings, turbines, etc. Where production activities are carried out on the bulk basis, and where both *fixed* and *variable* costs are subject to change, business managers use the *incremental principle* or *incremental analysis* in their business decisions.

The *incremental principle* is applied to business decisions which involve a large increase in total cost and total revenue. Such increase in total cost and total revenue is called 'incremental cost' and 'incremental revenue', respectively. Let us first explain the concept of incremental cost. **Incremental costs** can be defined as the costs that arise due to a business decision. For example, suppose a firm decides to increase production by using a larger quantity of raw materials and using its idle capacity or by adding a new plant to the existing capacity or by setting up a new production unit. This decision increases the firm's total cost of production from ₹ 100 million to ₹ 115 million. Then ₹ 115 million – ₹ 100 million = ₹ 15 million is the *incremental cost*. Thus, an increase in the total cost of production due to a business decision is *incremental cost*. Incremental cost includes both fixed and variable costs. However, it does not include the costs already incurred on the excess capacity or what is called the *sunk cost* or the *cost of unused material*.

The ***incremental revenue***, on the other hand, is the increase in revenue due to a business decision. When a business decision is successfully implemented, it does result in a significant increase in its total revenue. The increase in the total revenue resulting from a business decision is called *incremental revenue*. Suppose, for example, that after the installation of the new plant, the total production increases and the firm is able to sell the incremental product. As a result, the firm's total sales revenue increases. For example, let us suppose that total revenue increases from ₹ 130 million to ₹ 148 million. Thus, the post-decision total revenue of ₹ 148 million *less* the pre-decision total revenue of ₹ 130 million = ₹ 18 million is the *incremental revenue*.

Incremental Reasoning in Business Decision. The use of the incremental concept in business decisions is called *incremental reasoning*. The incremental reasoning is used in accepting or rejecting a business proposition or option. For instance, suppose that in our foregoing example, the firm is considering whether or not to install a new plant. As noted above, the firm estimates an incremental cost of installing a new plant at ₹ 15 million and an incremental revenue of ₹ 18 million. The incremental revenue exceeds the incremental cost by ₹ 3 million which means a 20 per cent return (gross of overheads) on the investment in the new plant. The firm will accept the proposition of installing a new plant, provided there is no better business proposition available to the firm.

It may be added at the end, by way of **comparison**, that the *marginal concept* (especially when defined and measured by calculus) is used in economic analysis where a high degree of precision is involved, whereas the incremental concept is used where a large amount of cost and revenue are involved. Besides,

incremental concept and reasoning are used in business decisions more frequently than the marginality concept. There are at least **two reasons** for this. *First,* marginality concept used in business analysis is generally associated with *one* (marginal) unit of output produced and sold whereas most business decisions involve large quantities and values. *Second,* the precise calculation of marginal change (defined in terms of the first derivative of a function) is neither practicable nor necessary in real life business considerations.

4. Contribution Analysis

Associated with the concept of incremental cost and incremental revenue is the concept of *contribution*. The *contribution* of a business decision can be defined as the difference between the *incremental revenue* and the *incremental cost* associated with that particular decision. Contribution analysis is generally applied to analyse the contribution made by a business decision to *overhead costs* and *revenue* to work out the net result of that particular business decision. It is a useful technique for taking a decision on:

 (*i*) whether or not to accept a project,

 (*ii*) whether or not to introduce a new product,

 (*iii*) whether or not to accept a fresh order,

 (*iv*) whether or not to add an additional plant,

 (*v*) whether to produce or to buy material inputs or parts, and so on.

For the use of contribution analysis for a business decision, it is important to know what is the relevant *incremental cost* and what is the *relevant incremental revenue*. That is, it is important to know what is included in and what is excluded from the *incremental cost* and *incremental revenue*.

The relevant incremental costs that are taken into account in contribution analysis include the following:

 (*i*) *Present period explicit costs:*

      (*a*) Explicit variable costs:

        Direct labour costs

        Direct material costs

        Direct variable overheads.

      (*b*) *Fixed costs:*

        New equipment

        New personnel.

 (*ii*) *Opportunity costs*: Foregone contribution expected from the second best alternative use of the firm's resources.

 (*iii*) *Future period incremental costs*: Expected present value of probable future costs arising out of a present business decision.

For the purpose of contribution analysis of a business decision, the following costs are considered to be the *irrelevant costs*.

(*i*) *Committed costs*, i.e., the costs which have already been committed by the firm and are bound to be incurred irrespective of whether or not the decision is implemented, e.g., payment of old debts, interest on old debts, committed raise in salaries and wages of managers and workers, etc.

(*ii*) *Sunk Costs*, i.e., the costs which have already been incurred on purchase of productive assets (building, plant and equipment) and non recoverable advance payments.

The relevant *incremental revenue* includes (*i*) explicit present period revenue, (*ii*) anticipated opportunity revenue, and (*iii*) anticipated future revenue. The *explicit present period revenue* is the revenue which arises directly from a decision in the present period. An *opportunity revenue* is, in fact, cost saving, i.e., the cost which can be avoided by taking a particular decision. For example, if a firm bids a low price for winning a contract, it avoids the cost of laying off its labour. A *possible future revenue* is the revenue that might arise due to the present decision. For example, if a firm wins a contract because of its low bids, it builds a reputation and is a favoured bidder for future projects, which yields revenue which would not exist otherwise.

5. Optimization Rule

Optimization literally means achieving a goal to the best possible extent. When the goal is achieved to the best possible extent, the result is called ***optimum.*** The techniques and methods used to achieve the best possible result are called ***optimization.*** Once goal is given, optimization may imply **'*maximization'*** or **'*minimization.*** The application of the concept of optimization to business decisions can be clarified with some examples from the business world. We know that all business firms have certain goals to achieve. It may be any or many of the following.

- Making a reasonable or a target profit,
- Maximization of profit,
- Maximization of revenue,
- Minimization of cost,
- Maximization of firm's wealth (the market value of its shares),
- Maximizing firm's growth,
- Maximization of market share,
- Elimination of strong or potential competitors, or
- Satisfaction of consumers, employees, shareholders, financers and authorities.

All these objectives of firms are subject to certain conditions and constraints. For our limited purpose here, following the conventional theory of firm, we assume

that *profit maximization* is the sole objective of business firms, and explain the concept of optimization.

Profit maximization as business objective is subject to certain market conditions, viz., demand at different prices, price elasticity, level and behaviour of production cost. Given these conditions, the level of maximum profit may vary from condition to condition and under different permutation and combinations of conditions. For example, given the cost curves, maximum profit will be different if demand curve is a downward sloping one or given by horizontal line. Similarly, given the revenue curves, maximum profit will be different if cost curves are U-shaped or a constant given by a horizontal line.

Optimization techniques used in economic analysis provide optimum solution to the problem. For instance, in case of profit maximization objective, given the average and marginal revenue curves and average and marginal cost curves, profit is maximized at the level of output where *MR = MC, under the condition of increasing MC.* Given the demand function and cost function, one can find mathematically the level of output and the level of price that maximize profit. The profit maximizing output and price are ***optimum output*** and ***optimum price,*** respectively. The technique of finding optimum output and price is shown below.

Suppose firm's total revenue *(TR)* and total cost *(TC)* functions are given as follows:

$$TR = 600\ Q - 3Q^2$$

$$TC = 1000 + 100Q + 2Q^2$$

Given the *TR* and *TC* functions, *MR* and *MC* functions can be derived by differentiating these functions, as shown below.

$$MR = \frac{\delta TR}{\delta Q} = 600 - 6\ Q$$

$$MC = \frac{\delta TC}{\delta Q} = 100 + 4\ Q$$

By equating *MR* and *MC* functions, we get profit maximizing output as follows:

$$MR = MC$$

$$600 - 6Q = 100 + 4Q$$

$$10Q = 500$$

$$Q = 50$$

Profit maximizing output is 50. Given the profit maximizing objective, 50 is the ***optimum output.***

Given the optimum output, ***optimum price*** can be obtained as follows. We know that

$$\text{Price } (P) = TR / Q$$

Thus,

$$P = \frac{TR}{Q} = \frac{600Q - 3Q^2}{Q} = 600 - 3Q = 600 - 3\,(50)$$

$$P = 450$$

Thus, profit maximizing price is 450. Therefore, ***optimum price*** is 450.

These calculations show the application of ***optimization technique*** in business decision making and a firm can optimize its price and output.

2.3.1 Micro Vs. Macro Economics

As we have noted, economics has two major branches: (*i*) Micro-economics, and (*ii*) Macroeconomics. Both micro and macro economics are applied to business analysis and decision-making—directly or indirectly. Business economics comprises, therefore, both micro and macroeconomic theories. The parts of micro and macroeconomics that constitute business economics depend on the purpose of analysis. In general, however, *business economics* is comprised of those topics or areas of microeconomics that can be applied to analyse internal business problems.

In general, the scope of business economics consists of all those microeconomic concepts, theories and tools of analysis which can be applied to analyse the demand prospects, production and cost conditions and pricing options with the objective of finding solutions to practical business problems. In other words, business economics is microeconomics applied to the analysis of business problems and decision-making. Broadly speaking, business economics is **applied microeconomics.**

Market forces and equilibrium

A free market is one in which market forces of demand and supply are free to take their own course and there is no outside control on price.

Concept of Market Equilibrium

In physical sense, the term equilibrium means the "state of rest". In general sense, it means balance in opposite forces. In the context of market analysis, *equilibrium refers to a state of market in which quantity demanded of a commodity equals the quantity supplied of the commodity.* The equilibrium of demand and supply produces an *equilibrium price.* The equilibrium price is the price at which quantity demanded of a commodity equals its quantity supplied. That is, at equilibrium price, demand and supply are in equilibrium. Equilibrium price is also called *market-clearing price.*

Determination of Market Equilibrium

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 mobiles. Let us suppose that the market demand and supply schedules for mobiles are given as shown in Table 2.1.

As the table shows, there is only one price of mobiles (₹ 3000) at which quantity demanded equals the quantity supplied at 40,000 mobiles. It means that the mobiles market in Delhi is in equilibrium at price ₹ 3000. At all other prices, the mobiles market is in *disequilibrium*—the state of 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 ₹ 3000, demand exceeds supply showing *shortage* of mobile supply in the market. Likewise, at all prices above ₹ 3000, supply exceeds demand showing *excess supply.*

Table 2.1 *Monthly Demand and Supply Schedules for Mobiles*

Price per Mobile phone (₹)	Demand ('000 Mobiles)	Supply ('000 Mobiles)	Market Position	Effect on Price
1000	80	10	Shortage	Rise
2000	55	28	Shortage	Rise
3000	40	40	Equilibrium	Stable
4000	28	50	Surplus	Fall
5000	20	55	Surplus	Fall
6000	15	60	Surplus	Fall

In a free market, disequilibrium itself creates the condition for equilibrium. When there is excess supply, it forces downward adjustments in the price and quantity supplied. When there is excess demand, it forces upward adjustments in the price and quantity demanded. The process of downward and upward adjustments in price and quantity continues till the price reaches ₹ 3000 and quantities supplied and demanded balance at 40,000 mobiles. This process is automatic. Let us now look into the process of price and quantity adjustments called 'market mechanism'.

How Market Attains Equilibrium—The Price Mechanism: Price mechanism is a process of interaction between the market forces of demand and supply to determine equilibrium price. To understand how it works, let the price of mobiles be initially set at ₹ 1000. At this price, the quantity demanded exceeds the quantity supplied by 70,000 mobiles. This gives sellers an opportunity to raise the price. Increase in price enhances the profit margin. This induces firms to produce and sell more in order to maximize their profits. Increasing production causes rise in cost of production. As a result, mobile price continues to rise. This trend continues till price rises to ₹ 3000. As Table 3.4 shows, at price ₹ 3000, the buyers are willing to buy 40,000 mobiles per month. This is exactly the number of mobiles that sellers would like to sell at this price. At this price, there is neither shortage nor

excess supply of mobiles in the market. Therefore, ₹ 3000 is the equilibrium price. The market is, therefore, in equilibrium.

Similarly, at all prices above ₹ 3000, supply exceeds demand showing excess supply of mobiles in the market. The excess supply forces the sellers to cut down the price. Some firms find low price unprofitable and go out of market and some cutdown their production. Therefore, supply of mobiles goes down. On the other hand, fall in price invites more customers. This process continues until price of mobiles falls to ₹ 3000. At this price, demand and supply are in balance and market is in equilibrium. Therefore, price at ₹ 3000 per mobile is the equilibrium price.

Graphical Illustration of Price Determination: The determination of equilibrium price is illustrated graphically in Fig. 2.2. The demand curve *DD′* and the supply curve *SS′* have been obtained by plotting the demand and supply schedules, respectively, (given in Table 2.1) on the price and quantity axes.

As Fig. 2.2 shows, demand and supply curves intersect at point *E* determining the equilibrium price at ₹ 3000. At this price, the quantity demanded (40,000 mobiles) equals the quantity supplied. Thus, the equilibrium price is ₹ 3000 and equilibrium quantity is 40 thousand mobiles. 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 ₹ 3000, there would be either excess supply or shortage of mobiles in the market.
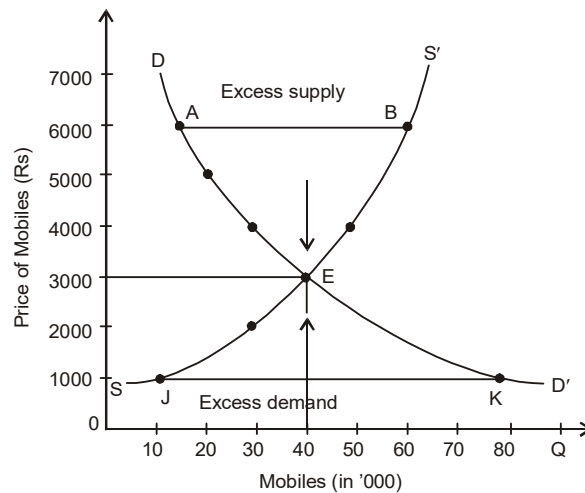
Fig. 2.2 Determination of Equilibrium Price

Let us now see how market works to bring about balance in demand for and supply of mobiles. Let the price be initially set at ₹ 6000. At this price, suppliers bring in a supply of 60 thousand mobiles whereas buyers are willing to buy only 15 thousand mobiles. The supply, obviously, far exceeds the demand. As Fig. 2.2 shows, the excess supply equals, *AB* = 60 – 15 = 45 thousand mobiles. 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 mobiles increases too. In this process, the supply-demand gap is reduced. This process continues until price reaches ₹ 3000 at point *E*, the point of equilibrium where demand and supply equal at 40 thousand mobiles. At this price, the market is in equilibrium and there is no inherent force at work which can disturb the market equilibrium.

Likewise, if price is initially set at ₹ 1000, the buyers would be willing to buy 80 thousand mobiles whereas suppliers would be willing to supply only 10 thousand mobiles. Thus, there would be a shortage of 70 thousand mobiles as shown by the distance *JK* in Fig. 2.2. The shortage will make the buyers to offer 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 ₹ 3000, the market reaches its equilibrium.

Check Your Progress

4. Why is opportunity cost also known as alternative cost?

5. Define time lag.

6. What is incremental revenue?

2.4 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. Economics provides analytical tools and techniques that managers need to achieve the goals of the organization that they manage.

2. The areas of business issues to which economic theories can be directly applied may be broadly divided into two categories: (a) microeconomics applied to operational or internal issues, and (b) macroeconomics applied to environmental or external issues.

3. Profit making is the most common objective of all business undertakings.

4. Opportunity cost is also known as alternative cost because it arises due to the possibility of alternatives uses of the resources.

5. The time gap between investment and return is known as time lag.

6. The incremental revenue is the increase in revenue due to a business decision.

2.5 SUMMARY

• The application of economic science in business decision-making is all pervasive. More specifically, economic laws and tools of economic analysis are now applied a great deal in the process of business decision-making. This has led, as mentioned earlier, to the emergence of a separate branch of study called managerial economics.

- Managerial economics applies the principles and methods of economics to analyze problems faced by management of a business, or other types of organizations and to help find solutions that advance the best interests of such organizations.

- The areas of business issues to which economic theories can be directly applied may be broadly divided into two categories: (a) microeconomics applied to operational or internal issues, and (b) macroeconomics applied to environmental or external issues.

- 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.

- Opportunity cost may be defined as income expected from the second best use of the resources which is sacrificed for the best use of the resources. Thus, opportunity cost is opportunity lost. From a firm's point view, opportunity cost of using a resource is what the firm must give up to use the resource as it is used.

- The concept of the time value of money is very often applied to investment decisions. Generally, there is time lag between investment and return. When an investment is made today, it begins to yield returns at some future date. The time gap between investment and return is called time lag.

- The terms Marginalism and Incrementalism refer to two principles widely used in economic analysis and business analysis. The two principles, called also as rules, are known as marginal principle and incremental principle, respectively. Marginalism and Incrementalism rules are also used to work out the marginal and incremental returns from a business activity, respectively.

- Contribution analysis is generally applied to analyse the contribution made by a business decision to overhead costs and revenue to work out the net result of that particular business decision.

- Optimization techniques used in economic analysis provide optimum solution to the problem.

2.6 KEY WORDS

- **Market Forces:** They are the factors that influence the price and availability of goods and services in a market economy, i.e. an economy with the minimum of government involvement.

- **Applied Economics:** It is a field that applies *economic* theories and principles to real-world situations with the desired aim of predicting potential outcomes.

- **Time Value of Money:** It is the concept that money available at the present time is worth more than the identical sum in the future due to its potential earning capacity.

- **Opportunity Costs:** It means the loss of other alternatives when one alternative is chosen.

2.7 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short Answer Questions

1. Define managerial economics.
2. Write a short note on the nature and scope of managerial economics.
3. Write a short-note on the optimization rule.

Long Answer Questions

1. Explain the concept of opportunity cost with the help of examples.
2. What is meant by time value of money? Why does the value of cash in hand have more value than the same amount of money expected to be available at some future date?
3. Distinguish between marginalism and incrementalism. Under what conditions are these concepts used in business decision-making?

2.8 FURTHER READINGS

Dwivedi, D. N. 2008. *Principles of Economics*, Seventh Edition. New Delhi: Vikas Publishing House.

Weil. David N. 2004. *Economic Growth*. London: Addison Wesley.

Thomas, Christopher R. and Maurice S. Charles. 2005. *Managerial Economics: Concepts and Applications*, Eighth Edition. New Delhi: Tata McGraw-Hill Publishing Company Limited.

Mankiw, Gregory N. 2002. *Principles of Economics*, Second Edition. India: Thomson Press.

UNIT 3   CONSUMER BEHAVIOUR

Structure

3.0   INTRODUCTION

Consumer demand is the basis of all productive activities. Just as 'necessity is the mother of invention', *demand is the mother of production*. Increasing demand for a product offers high business prospects for it in future and decreasing demand for a product diminishes its business prospect. For example, increasing demand for computers, cars, mobile phones etc. in India has enlarged the business prospect for both domestic and foreign companies selling these goods. On the other hand, declining demand for black and white TV sets and manual typewriters is forcing their companies to switch over to modern substitutes or else go out of business. It is, therefore, essential for business managers to have a clear understanding of the following aspects of demand for their products:

  (*i*) What is the basis of demand for a commodity?

  (*ii*) What are the determinants of demand?

 (*iii*) How do the buyers decide the quantity of a product to be purchased?

 (*iv*) How do the buyers respond to change in product prices, their incomes and prices of the related goods?

  (*v*) How can the total or market demand for a product be assessed and forecasted?

3.1 OBJECTIVES

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

- Explain the Cardinal Utility Approach

- Define the Diminishing Marginal Utility

- State the Law of Equi-Marginal Utility

- Describe the Ordinal Utility Approach

- List the features of Indifference Curves

3.2 CONSUMER BEHAVIOUR: AN OVERVIEW

Let us begin with the meaning of demand.

Meaning of Demand

Conceptually, the term 'demand' implies a '*desire for a commodity backed by the ability and willingness to pay for it.*' Unless a person has adequate purchasing power or resources and the willingness to spend his resources, his *desire* alone for a commodity would not be considered as his demand. For example, if a man wants to buy a car but he does not have sufficient money to pay for it, his want is not his demand for the car. And, if a rich miserly person wants to buy a car but is not willing to pay for the car, his desire too is not his demand for a car. But if a man has sufficient money and is willing to pay the price of the car, his desire to buy a car is an *effective demand*.

The desires without adequate purchasing power and willingness to pay do not affect the market, nor do they generate production activity. A want with *three attributes—desire to buy, willingness to pay and ability to pay*—becomes effective demand. Only an effective demand figures in economic analysis and business decisions.

The term 'demand' for a commodity (i.e., quantity demanded) always has a reference to 'a price', 'a period of time' and 'a place'. Any statement regarding the demand for a commodity without reference to its price, time unit and place of demand is meaningless and is of no practical use. For instance, to say 'the demand for TV sets is 50,000' carries no meaning for a business decision, nor does it have any use in any kind of economic analysis. A meaningful statement regarding the demand for a commodity should, therefore, contain the following information:

(*a*) the quantity demanded of a commodity,

(*b*) the price at which a commodity is demanded,

(*c*) the time period over which a commodity is demanded and

(*d*) the market area in which a commodity is demanded.

For example, saying, 'the annual demand for TV sets in Delhi at an average price of ` 15,000 a piece is 50,000' is a meaningful statement.

Basis of the Consumer Demand: Utility

Consumers demand a commodity because they derive or expect to derive *utility* from the consumption of that commodity. The expected utility from a commodity is the basis of demand for it. Though 'utility' is a term of common usage, it has a specific meaning and use in the analysis of consumer demand. We will, therefore, describe in this section the meaning of utility, the related concepts and the law associated with utility.

Meaning of Utility

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

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

On the other hand, from a consumer's point of view, utility is a post-consumption phenomenon as one derives satisfaction from a commodity only when one consumes or uses it. Utility in the sense of satisfaction is a 'subjective' or 'relative' concept because (*i*) a commodity need not be useful for all—cigarettes do not have any utility for non-smokers, and meat has no utility for strict vegetarians; (*ii*) utility of a commodity varies from person to person and from time to time; and (*iii*) a commodity need not have the same utility for the same consumer at different points of times, at different levels of consumption and for different moods of a consumer. In consumer analysis, only the 'subjective' concept of utility is used.

Having explained the concept of utility, we now turn to some quantitative concepts related to utility used in utility analysis, viz. *total utility* and *marginal utility*.

Total Utility

Assuming that utility is measurable and additive, total utility may be defined as the sum of the utility derived by a consumer from the various units of a good or service he consumes at a point or over a period of time. Suppose a consumer consumes four units of a commodity, *X*, at a time and derives utility from the successive units of consumption as u_1, u_2, u_3 and u_4. His total utility (U_x) from commodity *X* can be

then measured as follows.

$$U_x = u_1 + u_2 + u_3 + u_4$$

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

$$TU_n = U_x + U_y + U_z$$

Marginal Utility

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

$$MU_x = \frac{\Delta TU_x}{\Delta Q_x}$$

where TU_x = total utility, and ΔQ_x = change in quantity consumed by one unit.

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

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

3.2.1 Diminishing Marginal Utility

The law of diminishing marginal utility is one of the fundamental laws of economics. This law states that *as the quantity consumed of a commodity goes on increasing, the utility derived from each successive unit goes on diminishing, consumption of all other commodities remaining the same*. In simple words, when a person consumes more and more units of a commodity per unit of time, e.g., ice cream, keeping the consumption of all other commodities constant, the utility which he derives from each successive cup of ice cream goes on diminishing. This law applies to all kinds of consumer goods—durable and non-durable, sooner or later.

To illustrate the law of diminishing marginal utility, let us assume that a consumer consumes only one commodity *X*, and that utility is measurable in quantitative terms. Let us also suppose that total and marginal utility schedules of commondity *X* are given as in Table 3.1. The law of diminishing marginal utility is illustrated numerically in Table 3.1 and graphically in Fig. 3.1.

Table 3.1 *Total and Marginal Utility Schedules for Commodity X.*

Units of commodity X	Total utility (TU_x)	Marginal utility (MU_x)
1	30	30
2	50	20
3	60	10
4	65	5
5	60	– 5
6	45	– 15

As shown in Table 3.1, with the increase in the number of units of commodity X consumed per unit of time, TU_x increases but at a diminishing rate. The diminishing MU_x is shown in the last column of Table 3.1. Fig. 3.1 illustrates graphically the law of diminishing MU_x. The rate of increase in TU_x as a result of increase in the number of units consumed is shown by the MU_x curve in Fig. 3.1. The downward sloping MU_x curve shows that marginal utility goes on decreasing as consumption increases. At 4 units consumed, the TU_x reaches its maximum level, the point of saturation marked by point *M*. Beyond this point, MU_x becomes negative and TU_x begins to decline. The downward sloping MU_x curve illustrates the law of diminishing marginal utility.
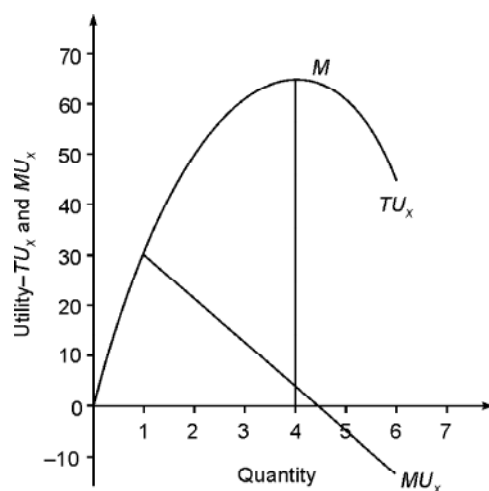
Fig. 3.1 *Total and Diminishing Marginal Utility of Commodity X*

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

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

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

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

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

Fourth, the mental condition of the consumer must remain normal during the period of consumption. A person drinking whisky may feel a greater pleasure with successive pegs because of change in his mental status due to intoxication.

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

3.2.2   Cardinal and Ordinal Concepts of Utility

Utility is a psychological phenomenon. It is a feeling of satisfaction, pleasure or happiness. Measurability of utility has, however, been a contentious issue. Classical economists, viz., Jeremy Bentham, Leon Walrus, Carl Menger, etc. and neo-classical economists, notably Alfred Marshall, believed that utility is cardinally or quantitatively measurable like height, weight, length, temperature and air pressure. This belief resulted in the *Cardinal Utility* concept. However, modern economists, most notably J.R. Hicks and R.G.D. Allen, hold the view that utility is not quantitatively measurable—it is not measurable in absolute terms. Utility can be expressed only ordinally or in terms of 'less than' or 'more than'. It is, therefore, possible to list the goods and services in order of their preferability or desirability. For example, suppose a person prefers chocolate to ice cream and ice cream to cold drink. He or she can express his/her preference as chocolate > ice cream > cold drink. But he cannot express his preference in quantitative terms. This is known as the *ordinal* concept of utility. Let us now look into the origin of the two concepts of utility and their use in the analysis of demand.

Cardinal Utility

Some early psychological experiments on an individual's responses to various stimuli led neo-classical economists to believe that utility is measurable and cardinally quantifiable. This belief gave rise to the concept of cardinal utility. It implies that utility can be assigned a cardinal number like 1, 2, 3, etc. Neo-classical economists built up the theory of consumption on the assumption that utility is cardinally measurable. They coined and used a term 'util' meaning 'units of utility'. In their measure of utility, they assumed (*i*) that one 'util' equals one unit of money, and (*ii*) that utility of money remains constant.

It has, however, been realized over time that *absolute* or cardinal measurement of utility is not possible. Difficulties in measuring utility have proved to be insurmountable. Neither economists nor scientists have succeeded in devising a technique or an instrument for measuring the feeling of satisfaction, i.e., utility. Nor could an appropriate measure of unit be devised. Numerous factors affect the state of consumer's mood, which are impossible to determine and quantify. *Utility is therefore not measurable in cardinal terms*. Yet cardinal utility concept continues to remain an essential starting point in the analysis of consumer behaviour.

Ordinal Utility

The modern economists have discarded the concept of *cardinal utility* and have instead employed the concept of *ordinal utility* for analyzing consumer behaviour. The concept of *ordinal utility* is based on the fact that it may not be possible for consumers to express the utility of a commodity in absolute or quantitative terms, but it is always possible for a consumer to tell introspectively whether a commodity is more or less or equally useful when compared to another. For example, a consumer may not be able to say that ice cream gives 5 utils and chocolate gives 10 utils. But he or she can always specify whether chocolate gives more or less utility than ice cream. This assumption forms the basis of the ordinal theory of consumer behaviour.

While neo-classical economists maintained that cardinal measurement of utility is practically possible and is meaningful in consumer analysis, modern economists maintain that utility being a psychological phenomenon is inherently immeasurable, theoretically, conceptually as well as quantitatively. They also maintain that the concept of ordinal utility is a feasible concept and it meets the conceptual requirement of analyzing consumer behaviour in the absence of any cardinal measures of utility.

Approaches to the Consumer Demand Analysis

Based on cardinal and ordinal concepts of utility, there are two approaches to the analysis of consumer behaviour.

(*i*) ***Cardinal Utility Approach***, attributed to Alfred Marshall and his followers, is also called the Neo-classical Approach.

(*ii*) ***Ordinal Utility Approach***, pioneered by J.R. Hicks, a Nobel laureate, and R.G.D. Allen, is also called the Indifference Curve Analysis.

The two approaches are not in conflict with one another. In fact, they represent two levels of sophistication in the analysis of consumer behaviour. Both the approaches are important for assessing and analyzing consumer demand for a commodity—be it for theoretical purpose or for business decision-making, depending on the level of sophistication required.

It is **important** to note in this regard that in spite of tremendous developments in consumption theory based on ordinal utility, the classical demand theory based

on cardinal utility has retained its appeal and applicability to the analysis of consumer behaviour. Besides, the study of classical demand theory serves as a foundation for understanding the advanced theories of consumer behaviour. The study of classical theory of demand is of particular importance and contributes a great deal in managerial decisions.

In the following section, we will first discuss the theory of consumer behaviour based on the cardinal utility approach.

Analysis of Consumer Behaviour: Cardinal Utility Approach

The central theme of the consumption theory – be it based on ordinal utility or cardinal utility approach – is the *utility maximizing behaviour* of the consumer. The fundamental postulate of the consumption theory is that all the consumers—individuals and households—aim at *utility maximization* and all their decisions and actions as consumers are directed towards utility maximization. The specific questions that the consumption theory seeks to answer are:

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

(*ii*) how does he or she allocate his/her disposable income between various commodities of consumption so that his/her total utility is maximized?

The theory of consumer behaviour seeks to answer these questions on the basis of the postulates that consumers seek to maximize their total utility or satisfaction.

Assumptions

The theory of consumer behaviour based on the cardinal utility approach seeks to answer the above questions on the basis of the following *assumptions*.

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

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

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

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

(*v*) **Diminishing marginal utility.** Following the law of diminishing marginal utility, it is assumed that the utility gained from the successive units of a commodity consumed decreases as a person consumes them. This is an axiom of the theory of consumer behaviour.

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

(*vii*) **Utility is additive.** Cardinalists assumed not only that utility is cardinally measurable but also that utility derived from various goods and services consumed by a consumer can be added together to obtain the total utility. Suppose a consumer consumes X_1, X_2, X_3,...X_n units of a commodity X and that she/he derives U_1, U_2, U_3 ... U_n utils respectively, from the various units of commodity X consumed. Given the assumption, the total utility that the consumer derives form n units of commodity X can be expressed as

$$U_n = U_1(X_1) + U_2(X_2) + U_3(X_3) + ... + U_n(X_n)$$

Consumer's Equilibrium: Cardinal Utility Approach

Conceptually, a consumer reaches his equilibrium position when he has maximized the level of his satisfaction, given his resources and other conditions. Technically, a utility-maximizing consumer reaches his equilibrium position when allocation of his expenditure is such that the last penny spent on each commodity yields the same utility. In order to explain how a consumer reaches his equilibrium, we begin with the simplest case of a consumer consuming only one commodity. The analysis will then be extended to explain equilibrium of a consumer consuming several goods.

(i) Consumer's Equilibrium: One-Commodity Model. 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 in the form of asset. If the marginal utility of commodity X, (MU_x) is greater than marginal utility of money (MU_m), a utility-maximizing consumer will exchange his money income for the commodity. By assumption, MU_x is subject to diminishing returns (assumption 5), whereas marginal utility of money (MU_m) remains constant (assumption 6). Therefore, the consumer will spend his money income on commodity X so long as $MU_x > P_x(MU_m)$, P_x being the price of commodity X and $MU_m = 1$ (constant). The utility-maximizing consumer reaches his equilibrium, i.e., the level of maximum satisfaction, where

$$MU_x = P_x (MU_m)$$

Alternatively, the consumer reaches equilibrium where

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

Consumer's equilibrium in a single commodity model is graphically illustrated in Fig. 3.2. The horizontal line marked $P_x(MU_m)$ shows the constant utility of money weighted by the price of commodity X (i.e., P_x) and MU_x curve represents the diminishing marginal utility of commodity X. The $P_x(MU_m)$ line and MU_x curve interest at point E. Point E indicates that at quantity OQ_x consumed, $MU_x = P_x(MU_m)$. Therefore, the consumer is in equilibrium at point E. At any point on the MU_x curve above point E, e.g., at point M, $MU_x > P_x(MU_x)$. Therefore, if the consumer exchanges his money for commodity X, he will increase his total satisfaction because his gain in terms of MU_x is greater than his cost in terms of MU_m. For example, at point M, consumer gains $MU_x = MC$ whereas the price that he pays equals NC. His marginal gain equals $MC - NC = MN$. This conditions exists till he reaches point E.
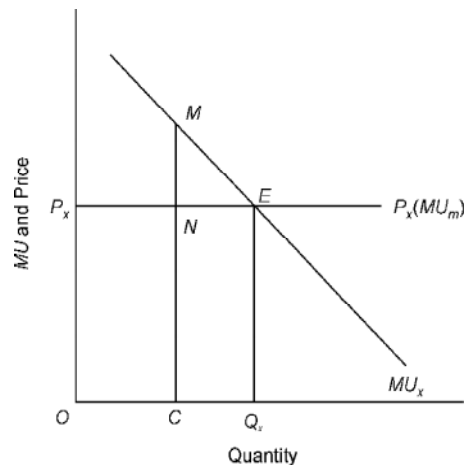
Fig. 3.2 Consumer's Equilibrium

Similarly, at any point below E, $MU_x < P_x(MU_m)$. Therefore, if he consumes more than OQ_x, he loses more utility than he gains. He is therefore a net loser. The consumer can, therefore, increase his satisfaction by reducing his consumption. This means that at any point other than E, consumer's total satisfaction is less than maximum. Therefore, point E is the point of equilibrium.

(ii) Consumer's Equilibrium: Multiple Commodity Model. In the preceding section, we have explained consumer's equilibrium making an unrealistic assumption that the consumer consumes a single commodity. In real life, a consumer consumes a large number of goods and services. So the question arises: How does a consumer consuming multiple goods reach his equilibrium? In this section, we explain consumer's equilibrium in the multi-commodity case.

We know from assumptions 2 and 5, that the consumer has limited income and that the utility which he derives from various commodities is subject to diminishing

returns. We also know that the *MU* schedules of various commodities may not be the same. Some commodities yield a higher marginal utility and some lower for the same number of units consumed. In some cases, *MU* decreases more rapidly than in case of others for the same number of units consumed. A rational and utility-maximising consumer consumes commodities in the order of their utilities. He first picks up the commodity which yields the highest utility followed by the commodity yielding the second highest utility and so on. He switches his expenditure from one commodity to the other in accordance with their marginal utilities. He continues to switch his expenditure from one commodity to another till he reaches a stage where *MU* of each commodity is the same per unit of expenditure. This is called the law of equi-marginal utility.

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

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

$$MU_x = P_x (MU_m)$$

and

$$MU_y = P_y (MU_m)$$

Given these conditions, the consumer's equilibrium can be expressed as

$$\frac{MU_x}{P_x(MU_m)} = 1 = \frac{MU_y}{P_y(MU_m)} \qquad ...(3.1)$$

Since, according to assumption (6), *MU* of each unit of money (or each rupee) is constant at 1, Eq. (3.1) can be rewritten as

$$\frac{MU_x}{P_x} = \frac{MU_y}{P_y} \qquad ...(3.2)$$

or

$$\frac{MU_x}{MU_y} = \frac{P_x}{P_y} \qquad ...(3.3)$$

Equation (3.2) leads to the conclusion that the consumer reaches his equilibrium when the marginal utility derived from each rupee spent on the two commodities *X* and *Y* is the same. Eq. (3.3) reveals that a consumer is in equilibrium when *MU* ratio of any two goods equals their price ratio.

The two-commodity case can be used to generalize the rule for consumer's equilibrium for a consumer consuming a large number of goods and services with

a given income and at different prices. Supposing, a consumer consumes A to Z goods and services, his equilibrium condition may be expressed as

$$\frac{MU_A}{P_A} = \frac{MU_B}{P_B} = \cdots = \frac{MU_Z}{P_Z} = MU_m \qquad ...(3.4)$$

Equation (3.4) gives the **Law of Equi-marginal Utility.**

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

Derivation of Individual Demand for a Commodity

The law of demand and the demand curve are based on the utility maximizing behaviour of the consumer. In this section, we explain how consumer behaviour provides the basis for the law of demand and the demand curve.

In preceding sections, we have explained the utility maximizing behaviour of the consumer and the concept of and conditions for consumer's equilibrium. Analysis of consumer's equilibrium provides a convenient basis for the derivation of the individual demand curve for a commodity. Marshall was the first economist to explicitly derive the demand curve from the consumer's utility function. As shown above, a consumer consuming only one commodity, say X, is in equilibrium where $MU_x = P_x (MU_m)$. Using this equilibrium condition, consumer's equilibrium has been illustrated in Fig. 3.2. The same logic can be used to derive consumer's demand curve for commodity X.

The derivation of individual demand for the commodity X is illustrated in Fig. 3.3 (*a*) and (*b*). Suppose that the consumer is in equilibrium at point E_1. It means that given the price of X at P_3, the equilibrium quantity is OQ_1. Now, if price of the commodity falls to P_2, the equilibrium condition will be disturbed making $MU_x > P_3$ (MU_m) at OQ_1. Since MU_m is constant, the only way to regain the equilibrium condition is to reduce MU_x, by consuming more of commodity X. Thus, by consuming Q_1Q_2 additional units of X the consumer reduces his MU_x to E_2Q_2 and reaches a new equilibrium position at point E_2 where $MU_x = P_2 (MU_m)$. Similarly, if price of X falls further, the consumer consumes more of X to maximize his satisfaction. This behaviour of the consumer can be used to derive the demand curve for commodity X.
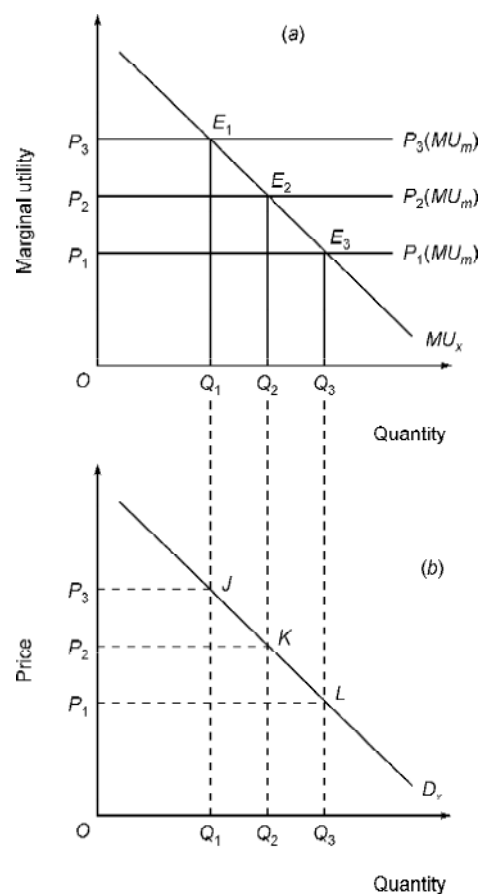
Fig. 3.3 Derivation of Demand Curve

As Fig. 3.3 (*a*) reveals, when price is P_3, equilibrium quantity is OQ_1. When price decreases to P_2, equilibrium point shifts downward to point E_2 where equilibrium quantity is OQ_2. Similarly, when price decreases to P_1 and the $P(MU_m)$ line shifts downward the equilibrium point shifts to E_3 and equilibrium quantity is OQ_3. Note that $P_3 > P_2 > P_1$ and the corresponding quantities $OQ_1 < OQ_2 < OQ_3$. This means that as price decreases, the equilibrium quantity increases. This inverse price-quantity relationship gives the law of demand and is explained below.

The inverse price-and-quantity relationship is shown in part (*b*) of Fig. 3.3 by the demand curve D_x. The demand curve D_x is drawn on the basis of information contained in panel (*a*) of Fig. 3.3. The price-quantity combination corresponding to equilibrium point E_3 is shown at point *L*. Similarly, the price-quantity combinations corresponding to equilibrium points, E_2 and E_1 are shown at points *K* and *J*, respectively. By joining points *J*, *K* and *L* we get the individual's *demand curve* for commodity *X*. The demand curve D_x is the usual downward sloping Marshallian demand curve.

Demand under Variable MU$_m$: We have explained above the consumer's equilibrium and derived his demand curve under the assumption that MU_m remains constant. This analysis holds even if MU_m is assumed to be variable. This can be explained as follows.

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

The Law of Demand: Some Additional Aspects

The two aspects that are disucssed here are: (*i*) the factors behind the operation of the law of demand, and (*ii*) exceptions to the law of demand.

(i) Factors Behind the Law of Demand. As Fig. 3.3 shows, demand curve slopes downward to the right. The downward slope of the demand curve depicts the law of demand. The factors that make the law of demand operate are the following. Incidentally, these factors answer the question 'why does demand curve slope downward to the right?'

(*a*) *Substitution Effect.* When the price of a commodity falls, prices of its substitutes remaining constant, then the substitutes become relatively costlier. Or, in other words, the commodity whose price has fallen becomes relatively cheaper. Since a utility maximizing consumer substitutes cheaper goods for costlier ones, demand for the cheaper commodity increases. The increase in demand on account of this factor is known as the *substitution effect*.

(*b*) *Income Effect.* When the price of a commodity falls, other things remaining the same, then the real income of the consumer increases in terms of that commodity. Consequently, his purchasing power increases as he is required to pay less for a given quantity. The increase in real income encourages the consumer to demand more goods and services. The increase in demand on account of an increase in real income is known as the *income effect*. Income effect is illustrated in Fig. 3.4.

Suppose demand curve for petrol is given by *DD′* in Fig. 3.4, and petrol price is given at ` 70 per litre. Given the demand curve and the petrol price, an individual consumes 5 litre of petrol per day. His daily expenditure on petrol is 70 × 5 = ` 350. Now let the petrol price go down to ` 50 per litre. This fall in petrol price increases the purchasing power of the individual in the sense that if he continues to spend ` 350 per day on petrol, he can buy 7 litre of petrol. This increase in

demand is the result of increase in purchasing power due to decrease in price. Thus, the increase in demand for petrol by 2 litre per day is the result of increase in real income (in terms of petrol). That is why it is called income effect of price change. It must be, however, borne in mind that income effect on demand for petrol may not necessarily be equal to 2 additional litre – it may be less or even more depending on the nature of the demand curve and its slope.
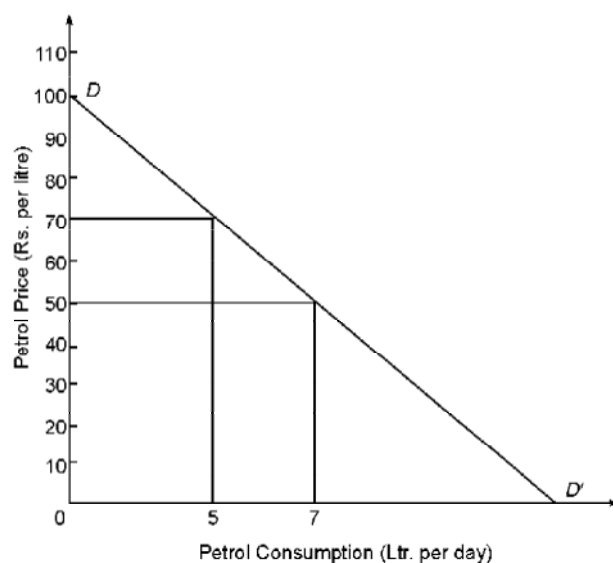
***Fig. 3.4** Income Effect of Price Change*

It should also be noted that the income effect is negative in case of ***inferior goods***. In case the price of an inferior good (accounting for a considerable proportion of the total consumption expenditure) falls substantially, consumers' real income increases and they become relatively richer. Consequently, they substitute the superior goods for the inferior ones i.e., they consume less of inferior goods and more of superior goods. As a result, consumption of the inferior goods falls. Thus, the income effect on the demand for inferior goods becomes negative—it is not negative in the case of normal goods.

(*c*) ***Utility-Maximizing Behaviour.*** The utility-maximizing behaviour of the consumer under the condition of diminishing marginal utility is also responsible for increase in demand for a normal good when its price falls. As mentioned above, when a person buys a commodity, say *X*, he exchanges his money income for the commodity in order to maximize his satisfaction. He continues to buy the commodity so long as marginal utility of his money (MU_m) is less than the marginal utility of the commodity (MU_x). Given the price of the commodity, the consumer adjusts his consumption so that

$$MU_m = P_c = MU_x$$

When price of the commodity falls, $(MU_m = P_c) < MU_x$, and consumer's equilibrium is disturbed. In order to regain his equilibrium, the consumer will have to

reduce the MU_x to the level of MU_m. This can be done only by purchasing more of the commodity. Therefore, the consumer buys the commodity till $MU_m = P_c = MU_c$. This is another reason why demand for a commodity increases when its price decreases. It is to be noted that the Law of Demand has been discussed in detail in Unit 4.

Shift in Demand Curve

When the demand curve changes its position (retaining its slope though not necessarily), the change is known as a *shift in the demand curve*. For example, suppose that the original demand curve for commodity X is given as D_1 in Fig. 3.5. As shown in the figure, at price OP_2, the consumer would buy OQ_1 units of X, other factors remaining constant. But, if any of the other factors (e.g., consumer's income or price of the substitutes) changes, it will change the consumer's ability and willingness to buy commodity X. For example, if consumer's disposable income increases due to decrease in income tax, he would be able to buy OQ_2 units of X instead of OQ_1. This is true for the whole range of prices of X; consumers would be able to buy more at all other prices. This will cause an *upward* shift in demand curve from D_1 to D_2. Similarly, decrease in disposable income of the consumer due to, say, rise in taxes may cause a *downward shift* in the demand curve from D_2 to D_1.
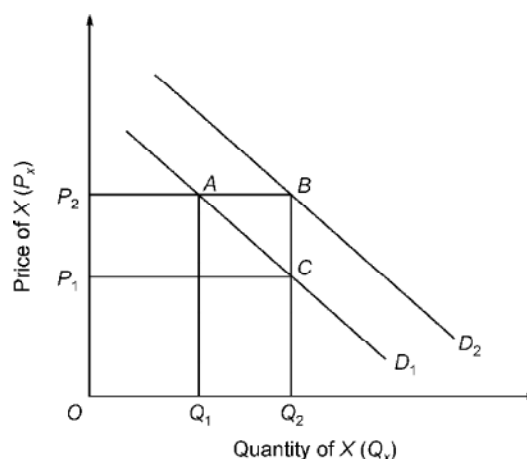
Fig. 3.5 Shift in Demand Curve

3.2.3 Factors Behind Shifts in the Demand Curve

Shifts in a price-demand curve may take place owing to the change in one or more non-price determinants of the demand for a commodity. Consider, for example, the increase in demand for commodity X by Q_1Q_2 in Fig. 3.5. Given the price OP_2, the demand for X might have increased by Q_1Q_2 for any of the following reasons.

(*i*) Increase in consumer's income so that he can buy OQ_2 of X at price OP_2: this is *income effect*;

(*ii*) Price of the substitute of *X* rises so that the consumers find it gainful to prefer Q_1Q_2 of *X* for its substitute: this is *substitution effect*;

(*iii*) Advertisement by the producer of the commodity *X* changes consumer's taste or preference in favour of commodity *X* so much that the consumer buys more of *X* or he prefers Q_1Q_2 to its substitute, again a *substitution effect*;

(*iv*) Price of a complement of *X* falls so much that the consumer can afford OQ_2 of *X*; and

(*v*) Price remaining the same, demand for *X* might increase also for such reasons as *X* gaining fashion status, improvement in its quality, change in production technology and seasonality of the product.

It is important for the business decision-makers to bear in mind the distinction between changes in demand due to (*i*) shift in price-demand curve; and (*ii*) movement along the demand curve. For instance, in Fig. 3.5, the increase in quantity demanded from OQ_1 to OQ_2 can be explained in two different ways: *one*, by moving down from point *A* to *C* along the demand curve D_1 which results from a fall in price from P_2 to P_1; and *two*, through upward shift in demand curve from D_1 to D_2. In the former case, additional demand is obtained at the cost of some revenue. In the latter case, demand increases due to a shift in the demand curve on account of some other factors, such as increase in consumer's income, increase in the price of substitutes, increase in population, etc. This kind of increase in demand results in increase in revenue. However, in case the demand curve is made to shift through advertisement or other sales promotion devices, the additional demand is not free of cost. Moreover, it is the latter kind of increase in demand which is hoped for and attempted by business firms.

Increase and Decrease vs Extension and Contraction of Demand
Economists sometimes make a distinction between (*a*) increase and decrease in demand, and (*b*) extension and contraction in demand. *Increase and decrease* in demand are associated with non-price-quantity relationships of demand whereas *extension and contraction* of demand are associated with the price-quantity relationship of demand. For example, in Fig. 3.5, movement from point *A* to *B* is an *increase in demand* and movement from *B* to *A* is a *decrease in demand*. On the other hand, movement from *A* to *C* is an *extension of demand* and movement from *C* to *A* is a *contraction of demand*. In other words, movement along the demand curve implies extension or contraction of demand.

This kind of distinction of terminology between a change in demand caused by different factors is, however, a matter of convenience. It has no theoretical basis.

Analysis of Consumer Behaviour: Ordinal Utility Approach

Unlike Marshall, the modern economists—Hicks in particular—have used the ordinal utility concept to analyze consumer's behaviour. This is called 'ordinal

utility approach'. Hicks has used a different tool of analysis called 'indifference curve' to analyze consumer behaviour. In this section, we will first explain the 'indifference curve' and then explain consumer's behaviour through the indifference curve technique. Let us first look at the assumptions of the ordinal utility approach.

1. ***Rationality.*** The consumer is assumed to be a rational being. Rationality means that a consumer aims at maximizing his total satisfaction given his income and prices of the goods and services that he consumes and his decisions are consistent with this objective.

2. ***Ordinal utility.*** Indifference curve analysis assumes that utility is only ordinally expressible. That is, the consumer is only able to express the order of his preference for different baskets of goods.

3. ***Transitivity and consistency of choice.*** Consumer's choices are assumed to be transitive. *Transitivity* of choice means that if a consumer prefers *A* to *B* and *B* to *C*, he must prefer *A* to *C*. Or, if he treats *A* = *B* and *B* = *C*, he must treat *A* = *C*. *Consistency* of choice means that if he prefers *A* to *B* in one period, he does not prefer *B* to *A* in another period or even treat them as equal.

4. ***Nonsatiety.*** It is also assumed that the consumer is never over-supplied with goods in question. That is, he has not reached the point of saturation in case of any commodity. Therefore, a consumer always prefers a larger quantity of all the goods.

5. ***Diminishing marginal rate of substitution.*** The marginal rate of substitution is the rate at which a consumer is willing to substitute one commodity (X) for another (Y) so that his total satisfaction remains the same. The marginal rate of substitution is given as $\Delta Y/\Delta X$. The ordinal utility approach assumes that $\Delta Y/\Delta X$ goes on decreasing when a consumer continues to substitute X for Y. (We will discuss marginal rate of substitution in detail in the subsequent section).

3.2.4 Meaning and Nature of Indifference Curve

An indifference curve may be defined as the *locus of points each representing a different combination of two substitute goods, which yield the same utility or level of satisfaction to the consumer. Therefore, he is indifferent between any two combinations of two goods when it comes to making a choice between them.* Such a situation arises because he consumes a large number of goods and services and often finds that one commodity can be substituted for another. The consumer can, therefore, substitute one commodity for another, and can make various combinations of two substitute goods which give him the same level of satisfaction. Since each combination yields the same level of satisfaction, he would be indifferent between the combinations when he has to make a choice. When such combinations are plotted graphically, the resulting curve is called the *indifference curve*. An indifference curve is also called *Iso-utility curve* or *Equal utility curve*.

For example, let us suppose that a consumer consumes two goods, *X* and *Y*, and he makes five combinations *a*, *b*, *c*, *d* and *e* of the two substitute commodities, *X* and *Y*, as presented in Table 3.2. All these combinations yield the same level of satisfaction.

Table 3.2 *Indifference Schedule of Commodities X and Y*

Combination		Units of Commodity Y	+	Units of Commodity X	=	Total Utility
a	=	25	+	3	=	U
b	=	15	+	5	=	U
c	=	8	+	9	=	U
d	=	4	+	17	=	U
e	=	2	+	30	=	U

Table 3.2 is an indifference schedule—a schedule of various combinations of two goods, between which a consumer is indifferent. The last column of the table shows an undefined utility (*U*) derived from each combination of *X* and *Y*. The combinations *a*, *b*, *c*, *d* and *e* given in Table 3.2 are plotted and joined by a smooth curve (as shown in Fig. 3.6). The resulting curve is known as *indifference curve*. On this curve, one can locate many other points between any two points showing different combinations of *X* and *Y* which yield the same level of satisfaction. Therefore, the consumer is indifferent between the combinations which may be located on the indifferent curve.
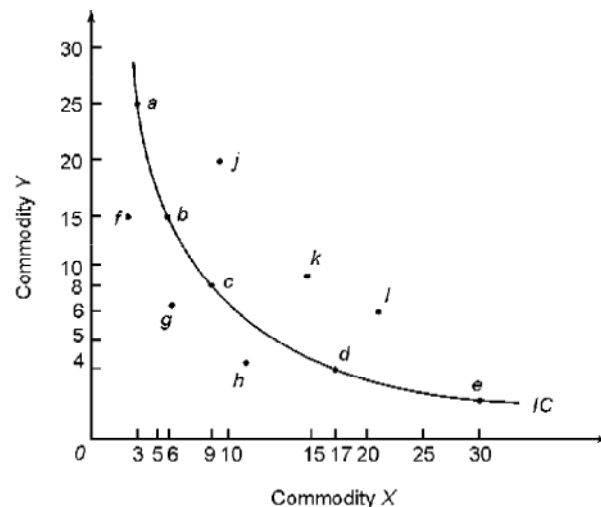
Fig. 3.6 *Indifference Curve*

Indifference Map: We have drawn a single indifference curve in Fig. 3.6 on the basis of the indifference schedule given in Table 3.2. The combinations of the two commodities, *X* and *Y*, given in the indifference schedule or those indicated by the indifference curve are by no means the only combinations of the two commodities. The consumer may make many other combinations with less of

one or both of the goods—each combination yielding the same level of satisfaction but less than the level of satisfaction indicated by the indifference curve *IC* in Fig. 3.6. As such, an indifference curve below the one given in Fig. 3.6 can be drawn, say, through points *f*, *g* and *h*. Similarly, the consumer may make many other combinations with more of one or both the goods—each combination yielding the same satisfaction but greater than the satisfaction indicated by *IC*. Thus, another indifference curve can be drawn above *IC*, say, through points *j*, *k* and *l*. This exercise may be repeated as many times as one wants, each time generating a new indifference curve.

In fact, the space between *X* and *Y* axes is known as the ***indifference plane*** or ***commodity space***. This plane is full of finite points and each point on the plane indicates a different combination of goods *X* and *Y*. Intuitively, it is always possible to locate any two or more points on the indifference plane indicating different combinations of goods *X* and *Y* yielding the same level of satisfaction. It is thus possible to draw a number of indifference curves without intersecting or being tangent to one another as shown in Fig. 3.7. The set of indifference curves *IC*₁, *IC*₂, *IC*₃ and *IC*₄ drawn in this manner make the ***indifference map***. In fact, an indifference map may contain any number of indifference curves, ranked in the order of consumer's preferences.
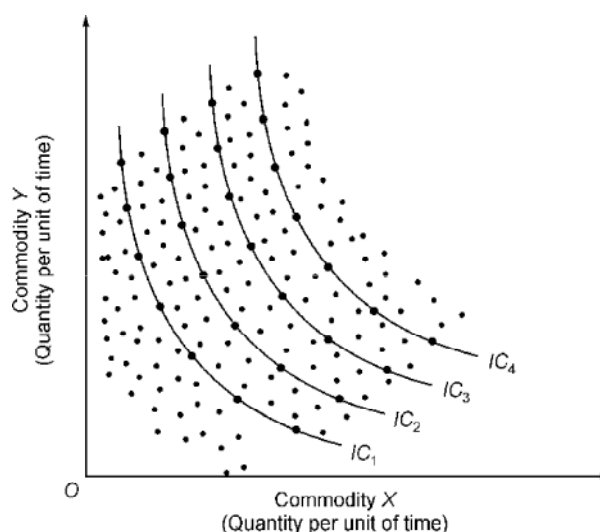
Fig. 3.7 *The Indifference Map*

3.2.5 Marginal Rate of Substitution (*MRS*)

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

Y) given up is equal to the utility of additional units of *Y* (or *X*) added to the combination. The *MRS* is expressed as $\Delta Y/\Delta X$, moving down the curve.

MRS goes on Diminishing: The basic postulate of ordinal utility theory is that $MRS_{y,x}$ (or $MRS_{x,y}$) decreases. It means that the quantity of a commodity that a consumer is willing to sacrifice for an additional unit of another goes on decreasing when he goes on substituting one commodity for another. The diminishing $MRS_{x,y}$ obtained from combinations of *X* and *Y* given in Table 3.2 is presented in Table 3.3.

Table 3.3 *The Diminishing MRS between Commodities X and Y*

Indifference Points	Combinations *Y + X*	Change in Y *(− ∆Y)*	Change in X *(∆X)*	$MRS_{y,x}$ *(∆Y/∆X)*
a	25 + 3	–	–	–
b	15 + 5	– 10	2	– 5.00
c	8 + 9	– 7	4	– 1.75
d	4 + 17	– 4	8	– 0.50
e	2 + 30	– 2	13	– 0.15

As Table 3.3 shows, when the consumer moves from point *a* to *b* on his indifference curve (Fig. 3.6) he gives up 10 units of commodity *Y* and gets only 3 units of commodity *X*, so that

$$MRS_{y,\,x} = \frac{-\Delta Y}{\Delta X} = \frac{-10}{2} = -5.00$$

As he moves down from point *b* to *c*, he loses 7 units of *Y* and gains 4 units of *X*, giving

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

Note that as the consumer moves from point *a* to *b* and from point *b* to *c*, the *MRS* decreases from –5.00 to –1.75. The $MRS_{y,x}$ goes on decreasing as the consumer moves further down along the indifference curve, from point *c* through *d* and *e*. *The diminishing marginal rate of substitution causes the indifference curves to be convex to the origin.*

Why Does MRS Diminish? The *MRS* diminishes along the *IC* curve because, in most cases, no two goods are perfect substitutes for one another. In case any two goods are perfect substitutes, the indifference curve will be a straight line having a negative slope showing constant *MRS*. Since goods are not perfect substitutes*, the subjective value attached to the additional quantity (i.e., subjective MU) of a commodity decreases fast in relation to the other commodity whose total quantity is decreasing.* Therefore, when the quantity of one commodity (*X*) increases and that of the other (*Y*) decreases, the subjective *MU* of *Y* increases and that of *X* decreases. Therefore, the consumer becomes increasingly *unwilling* to sacrifice more units of *Y* for one unit of *X*. But, if he is required to sacrifice additional units of

Y, he will demand increasing units of *X* to maintain the level of his satisfaction. As a result, the *MRS* decreases.

Furthermore, when combination of two goods at a point on the indifference curve is such that it includes a large quantity of one commodity (*Y*) and a small quantity of the other commodity (*X*), then consumer's *capacity* to sacrifice *Y* is greater than to sacrifice *X*. Therefore, he can sacrifice a larger quantity of *Y* in favour of a smaller quantity of *X*. For example, at combination *a* (see the indifference schedule, Table 3.2), the total stock of *Y* is 25 units and that of *X* is 5 units. That is why the consumer is willing to sacrifice 10 units of *Y* for 3 units of *X* (Table 3.3). This is an observed behavioural rule that the consumer's willingness and capacity to sacrifice a commodity is greater when its stock is greater and it is lower when the stock of a commodity is smaller.

These are the reasons why *MRS* between the two substitute goods decreases all along the indifference curve.

Properties of Indifference Curve

Indifference curves have the following four basic properties:

1. Indifference curves have a negative slope;
2. Indifference curves of imperfect substitutes are convex to the origin;
3. Indifference curves do not intersect nor are they tangent to one another;
4. Upper indifference curves indicate a higher level of satisfaction.

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

1. *Indifference Curves have a Negative Slope.* In the words of Hicks, "so long as each commodity has a positive marginal utility, the indifference curve must slope downward to the right", as shown in Fig. 3.6. The negative slope of an indifference curve implies (*a*) that the two commodities can be substituted for each other; and (*b*) that if the quantity of one commodity decreases, qunatity of the other commodity must increase so that the consumer stays at the same level of satisfaction. If quantity of the other commodity does not increase simultaneously, the bundle of commodities will decrease as a result of decrease in the quantity of one commodity. And, a smaller bundle of goods is bound to yield a lower level of satisfaction. The consumer's satisfaction cannot remain the same if indifference curves have a *positive slope* (i.e., $\Delta Y / \Delta X > 0$) or if slope is equal to infinity, (i.e., $\Delta Y / \Delta X = \infty$). These situations are shown in Fig. 3.8 through inconsistent indifference curves. Let us suppose that the consumer is initially at point *A* where he is deriving some utility from OX_1 of *X* and OY_1 units of *Y*.
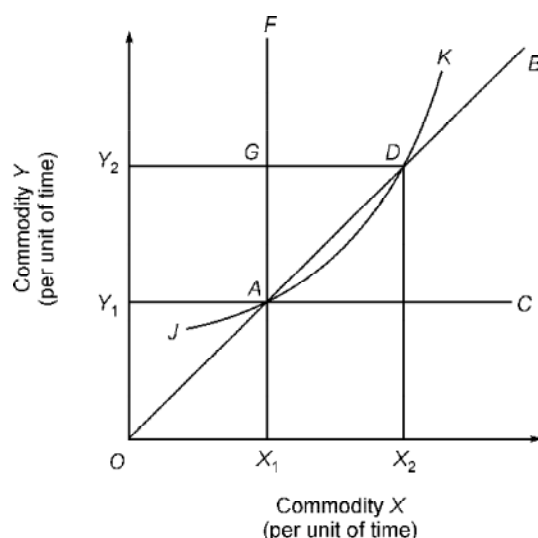
Fig. 3.8 *Inconsistent Indifference Curves*

If an indifference curve has a positive slope (i.e., $\Delta Y/\Delta X > 0$), as shown by the line *OB* and curve *JK*, it implies that upward movement along the line or the curve increases the combination of the two goods but the consumer is equally satisfied with larger and smaller baskets of *X* and *Y*. This means an irrational behaviour of the consumer. For example, if the consumer moves from point *A* to *D*, the combination of the two goods increases by $Y_1 Y_2$ of *Y* and $X_1 X_2$ of *X*. Unless *MU* of $Y_1 Y_2$ and $X_1 X_2$ are equal to zero, the level of satisfaction is bound to increase whereas on an indifference curve, the total utility is supposed to remain the same. Therefore, line *OB* and curve *JK* cannot be indifference curves.

Similarly, in the case of a vertical indifference line, FX_1, the movement from point *A* to *G* means an increase in the quantity of *Y* by $Y_1 Y_2$, while quantity of *X* remains the same, OX_1. If *MU* of $Y_1 Y_2 > 0$, the total utility will increase. So is the case if an indifference curve takes the shape of a horizontal line, like $Y_1 C$.

2. *Indifference Curves are Convex to Origin.* Indifference curves are not only negatively sloped, but are also *convex to the origin* as shown in Fig 3.6. The *convexity of the indifference curves* implies two properties:

(*i*) the two commodities are imperfect substitutes for one another, and

(*ii*) the marginal rate of substitution (*MRS*) between the two goods decreases as a consumer moves along an indifference curve. This characteristic of indifference curve is based on the postulate of diminishing marginal rate of substitution.

The postulate of diminishing *MRS*, as mentioned above, states an observed fact that if a consumer substitutes one commodity (*X*) for another (*Y*), his willingness to sacrifice more units of *Y* for one additional unit of *X* decreases, as quantity of *Y* decreases. There are two reasons for this: (*i*) no two commodities are perfect substitutes for one another, and (*ii*) *MU* of a commodity increases as its quantity

decreases and *vice versa*, and, therefore, more and more units of the other commodity are needed to keep the total utility constant.

3. *Indifference Curves can Neither Intersect Nor be Tangent with One Another.* If two indifference curves intersect or are tangent with one another, it will yield two impossible conclusions: (*i*) that two equal combinations of two goods yield two different levels of satisfaction, and (*ii*) that two different combinations— one being larger than the other–yield the same level of satisfaction. Such conditions are impossible if the consumer's subjective valuation of a commodity is greater than zero. Besides, if two indifference curves intersect, it would mean negation of *consistency* or *transitivity* assumption in consumer's preferences.
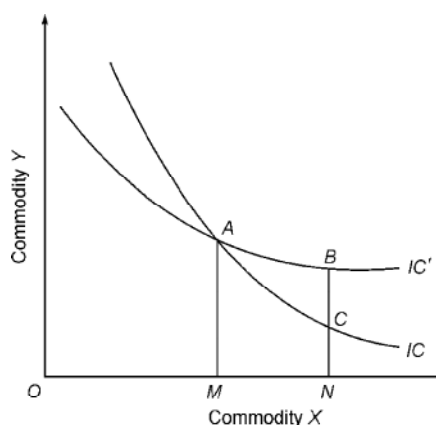
Fig. 3.9 Intersecting Indifference Curves

Let us now see what happens when two indifference curves, *IC* and *IC*′, intersect each other at point *A* as shown in Fig. 3.9. Point *A* falls on both the indifference curves, *IC* and *IC*′. It means that the same basket of goods (*OM* of *X* + *AM* of *Y*) yields different levels of utility below and above point *A* on the same indifference curve. This implies inconsistency in consumer's choice. The inconsistency that two different baskets of *X* and *Y* yield the same level of utility can be proved as follows. Consider two other points—point *B* on indifference curve *IC*′ and point *C* on indifference curve *IC* both being on a vertical line.

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

$$A = B \quad \text{(both combinations being on } IC') $$

and $\qquad\qquad\qquad\qquad\qquad A = C \quad$ (both combinations being on *IC*)

∴ $\qquad\qquad\qquad\qquad\qquad B = C \quad$ (both falling on different *IC$_s$*)

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

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

But, note that '*ON* of *X*' is common to both the sides. Therefore, it means that *BN* of *Y* = *CN* of *Y*. But as Fig. 3.9 shows, *BN* > *CN*. Therefore, combinations

B and *C* cannot be equal in terms of satisfaction. The intersection, therefore, violates the *transitivity rule* which is a logical necessity in indifference curve analysis. The same reasoning is applicable when two indifference curves are tangent with each other.

4. *Upper Indifference Curves Represent a Higher Level of Satisfaction than the Lower Ones.* An indifference curve placed above and to the right of another represents a higher level of satisfaction than the lower one. In Fig. 3.10, indifference curve IC_2 is placed above the curve IC_1. Therefore, IC_2 represents a higher level of satisfaction. The reason is that an upper indifference curve contains all along its length a larger quantity of one or both the goods than the lower indifference curve. And a larger quantity of a commodity is supposed to yield a greater satisfaction than the smaller quantity of it, provided $MU > 0$.
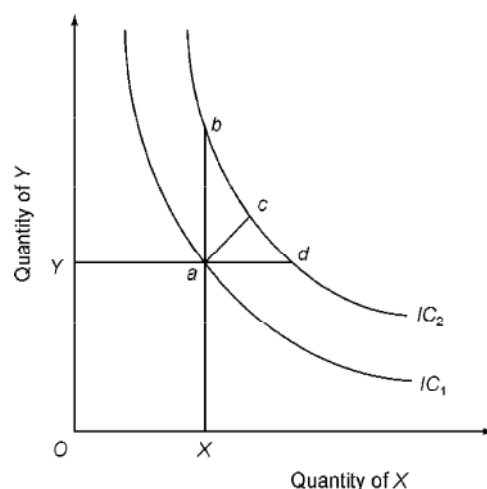
Fig. 3.10 Comparison between Lower and Upper Indifference Curves

For example, consider the indifference curves IC_1 and IC_2 in Fig. 3.10. The vertical movement from point *a* on the lower indifference curve IC_1 to point *b* on the upper indifference curve IC_2 means an increase in the quantity of *Y* by *ab*, the quantity of *X* remaining the same (*OX*). Similarly, a horizontal movement from point *a* to *d* means a greater quantity (*ad*) of commodity *X*, quantity of *Y* remaining the same (*OY*). The diagonal movement, i.e., from *a* to *c*, means a larger quantity of both *X* and *Y*. Unless the utility of additional quantities of *X* and *Y* are equal to zero, these additional quantities will yield additional utility. Therefore, the level of satisfaction indicated by the upper indifference curve (IC_2) would always be greater than that indicated by the lower indifference curve (IC_1).

3.2.6 Budgetary Constraints on Consumer's Choice: Limited Income and Prices

Given the indifference map, a utility maximizing consumer would like to reach the highest possible indifference curve on his indifference map. But the consumer has two strong constraints: (*i*) he has a limited income, and (*ii*) he has to pay a price

NOTES

for the goods. Given the prices, the limitedness of income acts as a *constraint on how high a consumer can ride on his indifference map*. This is known as *budgetary constraint*. In a two-commodity model, the *budgetary constraint* may be expressed through a ***budget equation*** as

$$P_x \cdot Q_x + P_y \cdot Q_y = M$$

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

The budget equation states that the total expenditure of the consumer on goods X and Y cannot exceed his total income, M. The quantities of X and Y that a consumer can buy, given his income (M) and prices, P_x and P_y, can be easily obtained from the budget equation, as shown below.

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

and

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

Now, Q_x or Q_y may be alternatively assigned any positive numerical value and the corresponding values of Q_y and Q_x may be obtained. When the values of Q_x and Q_y are plotted on the X and Y axes, we get a line with a negative slope, which is called the ***budget line*** or ***price line***, as shown in Fig. 3.11.
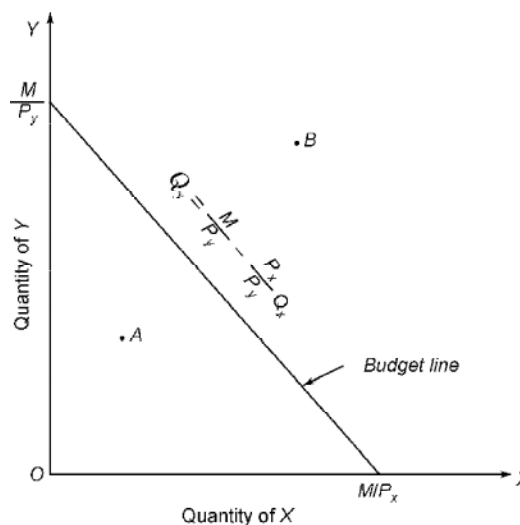
Fig. 3.11 *Budget Line and Budget Space*

An easier method of drawing the *budget line* is to mark point M/P_y on the Y axis (assuming $Q_x = 0$) and point M/P_x on X-axis (assuming $Q_y = 0$) and to join these points by a line. This gives the same budget line as given by the equation in Fig. 3.11. *The budget line shows the alternative options of commodity combinations available to the consumer given his income and the prices of X and Y.*

As can be seen in Fig. 3.11, budget line divides the *commodity space* into two parts: (*i*) feasibility area, and (*ii*) non-feasibility area. The area under the budget line (including the budget line) is *feasibility* area. Any combination of goods *X* and *Y* represented by a point within this area (e.g., point *A*) or point *P* on the boundary line (i.e., on the budget line) is a feasible combination, given *M*, P_x and P_y. The area beyond 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).

Shifts in the Budget Line

The budget line is drawn on the basis of the budget equation. Any change in the parameters of the budget equation, viz., *M*, P_x and P_y, make the budget line shift upward or downward or swivel left or right and up or down. If consumer's income (*M*) increases, prices remaining the same, the budget line shifts upwards remaining parallel to the original budget line. Suppose the original budget line is given by line *AB* in Fig. 3.12. If *M* increases (prices remaining the same), the budget line *AB* will shift to *CD*. And, if *M* decreases by the same amount, the budget line will shift backward to its original position *AB*. Income remaining the same, if prices change, the budget line changes its position and slope. For example, if *M* and P_y remain constant and P_x decreases by half then the budget line will be *AF*. Similarly, *M* and P_x remaining constant, if P_y increases, the budget line shifts to *EB*.
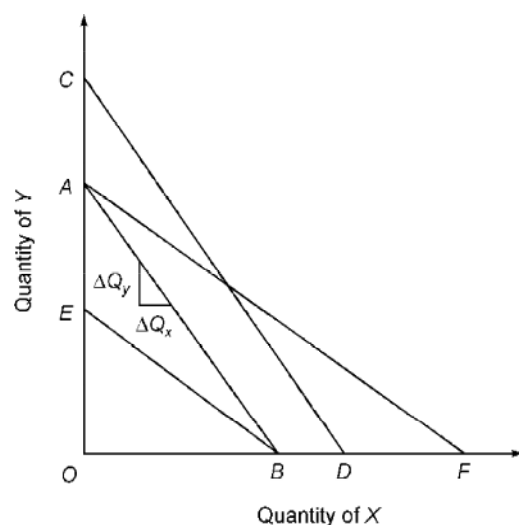
Fig. 3.12 Shift in the Budget Space

Slope of the Budget Line

Another important aspect of the budget line that matters in determining a consumer's equilibrium is its *slope*. The slope of the budget line (*AB*) in Fig. 3.12, is given as

$$\frac{\Delta Q_y}{\Delta Q_x} = \frac{OA}{OB}$$

Since $OA = M/P_y$ (when $X = 0$) and $OB = M/P_x$ (when $Y = 0$), the *slope of the budget line AB* in Fig. 3.12 may be rewritten as

$$\frac{OA}{OB} = \frac{M/P_y}{M/P_x} = \frac{P_x}{P_y}$$

Thus, the slope of the budget line is the same as the *price ratio* of the two commodities.

3.2.7 Consumer's Equilibrium: Ordinal Utility Approach

Now that we have discussed the indifference map and the budget line, we turn to analyze consumer's equilibrium. As noted earlier, a consumer attains his equilibrium when he maximizes his total utility, given his income and market prices of the goods and services that he consumes. The ordinal utility approach specifies two conditions for the consumer's equilibrium:

(*i*) necessary or the first order condition, and

(*ii*) supplementary or the second order condition.

In a two-commodity model, the necessary or the *first order condition* under ordinal utility approach is the same as equilibrium condition under cardinal utility approach. It is given as

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

Since, by implication, $MU_x/MU_y = MRS_{x, y,}$ the necessary condition of equilibrium under ordinal utility approach can be written as

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

This is a necessary but not a *sufficient condition* of consumer's equilibrium. The *second order* or *supplementary condition* requires that the necessary condition be fulfilled at the *highest possible indifference curve*.
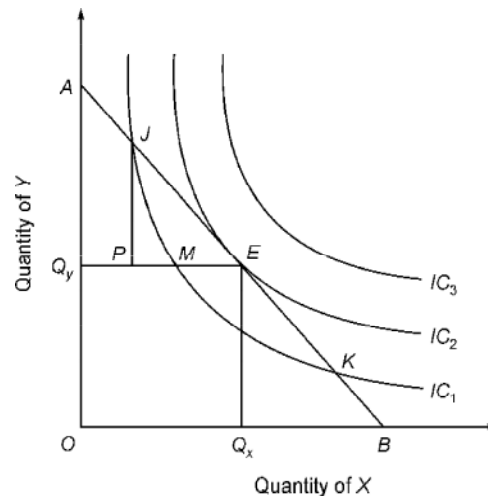
Fig. 3.13 *Equilibrium of the Consumer*

Consumer's equilibrium is illustrated in Fig. 3.13. The indifference curves IC_1, IC_2 and IC_3 present a hypothetical indifference map of the consumer. The line *AB* is the hypothetical budget line. Both the budget line *AB* and the indifference curve IC_2 pass through point *E*. Therefore, the slopes of the indifference curve IC_2 and the budget line (*AB*) are equal. Thus, both the necessary and supplementary conditions are fulfilled at point *E*. Therefore, consumer is in equilibrium at point *E*. This point can be proved as follows.

We know that between any two points on an indifferent curve,

$$\Delta Y \cdot MU_y = \Delta X \cdot MU_x$$

and, therefore, the slope of an indifference curve is given by

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

We know also that the slope of the budget line is given by

$$\frac{OA}{OB} = \frac{P_y}{P_y}$$

As shown in Fig. 3.13, at point *E*, $MRS_{y,x}$ (=) P_x/P_y. Therefore, the consumer is in equilibrium at point *E*. The tangency of IC_2 with the budget line *AB*, indicates that IC_2 is the highest possible indifference curve which the consumer can reach, given his income and the prices. At equilibrium point *E*, the consumer consumes OQ_x of *X* and OQ_y of *Y*, which yield him the maximum satisfaction.

Although the necessary condition is also satisfied on two other points, *J* and *K* (i.e., the points of intersection between the budget line *AB* and indifference curve IC_1), these points do not satisfy the *second order* condition. 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, the level of satisfaction at any point on IC_1 is lower than the level of satisfaction indicated by IC_2. So long as the utility maximizing consumer has an opportunity to reach the curve IC_2, the rationality condition demands that he would not like to settle on a lower indifference curve.

From the information contained in Fig. 3.13, it can be proved that the level of satisfaction at point *E* is greater than that on any other point on IC_1. Suppose 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, the consumer moves to point *E* which means 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 other point on 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. 3.13 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 the indifference curve IC_3 falls in the infeasibility area.

Check Your Progress

1. What is utility from the product angle?

2. What is the concept of ordinal utility based on?

3. Define indifference curve.

4. How is a budget line drawn?

3.3 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. From the product angle, utility is the want-satisfying property of a commodity.

2. The concept of ordinal utility is based on the fact that it may not be possible for consumers to express the utility of a commodity in absolute or quantitative terms, but it is always possible for a consumer to tell introspectively whether a commodity is more or less or equally useful when compared to another.

3. An indifferent curve may be defined as the locus of points each representing a different combination of two substitute goods, which yield the same utility or level of satisfaction to the consumer. Therefore, he is indifferent between any two combinations of two goods when it comes to making a choice between them.

4. The budget line is drawn on the basis of the budget equation.

3.4 SUMMARY

- The term 'demand' implies a 'desire for a commodity backed by the ability and willingness to pay for it.'

- Consumers demand a commodity because they derive or expect to derive utility from the consumption of that commodity. The expected utility from a commodity is the basis of demand for it.

- The law of diminishing marginal utility states that as the quantity consumed of a commodity goes on increasing, the utility derived from each successive unit goes on diminishing, consumption of all other commodities remaining the same.

- The concept of ordinal utility is based on the fact that it may not be possible for consumers to express the utility of a commodity in absolute or quantitative terms, but it is always possible for a consumer to tell introspectively whether a commodity is more or less or equally useful when compared to another.

- An indifference curve may be defined as the locus of points each representing a different combination of two substitute goods, which yield the same utility or level of satisfaction to the consumer. Therefore, he is indifferent between any two combinations of two goods when it comes to making a choice between them.

- The MRS is the rate at which one commodity can be substituted for another, the level of satisfaction remaining the same.

- Given the indifference map, a utility maximizing consumer would like to reach the highest possible indifference curve on his indifference map. But the consumer has two strong constraints: (i) he has a limited income, and (ii) he has to pay a price for the goods. Given the prices, the limitedness of income acts as a constraint on how high a consumer can ride on his indifference map. This is known as budgetary constraint.

3.5 KEY WORDS

- **Marginal Rate of Substitution:** It is the rate at which a consumer is ready to exchange a number of units good X for one more of good Y at the same level of utility.

- **Indifference Curve:** It is a curve on a graph (the axes of which represent quantities of two commodities) linking those combinations of quantities which the consumer regards as of equal value.

- **Demand Curve:** It is a graph showing how the demand for a commodity or service varies with changes in its price.

3.6 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short Answer Questions

1. What is the meaning of demand?
2. What are the factors behind the law of demand?
3. Graphically represent the shift in the budget line.

Long Answer Questions

1. Describe the law of diminishing marginal utility.

2. Explain the two approaches to consumer demand analysis.

3. What is indifference curve? Discuss its properties.

3.7 FURTHER READINGS

Dwivedi, D. N. 2008. *Principles of Economics*, Seventh Edition. New Delhi: Vikas Publishing House.

Weil. David N. 2004. *Economic Growth*. London: Addison Wesley.

Thomas, Christopher R. and Maurice S. Charles. 2005. *Managerial Economics: Concepts and Applications*, Eighth Edition. New Delhi: Tata McGraw-Hill Publishing Company Limited.

Mankiw, Gregory N. 2002. *Principles of Economics*, Second Edition. India: Thomson Press.

UNIT 4   DEMAND ANALYSIS

Structure

4.0   INTRODUCTION

The market for a product works on certain *market principles,* i.e., the *laws that govern the working of the market system.* The working of the market system is governed by certain *fundamental laws of market* called the *laws of demand and supply.* The *laws of demand and supply* play a crucial role in determining the market price of a commodity and the size of the market. A clear understanding of how markets work is essential in business decision-making and in chalking out an appropriate marketing strategy.

The market system works by two kinds of market forces—*demand* and *supply.* The demand and supply forces represent two sides of the market, viz., (*i*) Demand side, and (*ii*) Supply side.

In this unit, you will study about the theory of demand and concept of measurement of elasticity of demand.

4.1   OBJECTIVES

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

- Discuss the theory of demand

- State the law of demand

- Explain the concept of measurement of elasticity of demand

- List the factors affecting elasticity of demand
- Define income elasticity of demand and cross elasticity of demand

4.2 THEORY OF DEMAND

The analysis of total demand for a firm's product plays a crucial role in business decision-making. The market demand or the size of the market at a point in time at different prices gives the overall scope of business; it gives prospects for expanding business; and it plays a crucial role in planning for future production, inventories of raw materials, advertisement, and setting up sales outlets. Therefore, the information regarding the magnitude of the current and future demand for the product is indispensable. Theory of demand provides an insight into these problems. From the analysis of market demand, business executives can know:

   (*i*) the factors that determine the size of demand,

  (*ii*) elasticities of demand, i.e., how responsive or sensitive is the demand to the changes in its determinants,

 (*iii*) possibility of sales promotion through manipulation of prices,

  (*iv*) responsiveness of demand to advertisement expenditure, and

  (*v*) optimum levels of sales, inventories and advertisement cost, etc.

4.2.1 Law of Demand

The law of demand states the relationship between the quantity demanded and price of a commodity. Although quantity demanded of a commodity depends also on many other factors, e.g., consumer's income, price of the related goods, consumer's taste and preferences, advertisement, etc., price is the most important and the only determinant of demand in the short run. The law of demand is linked to price- quantity relationship.

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 decreases when its price increases.* This law implies that demand and price are inversely related. Marshall has stated 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, that is, *all other things remain unchanged*—other things include all other determinants of demand including consumer's income, price of the substitutes and complements, taste and preference, advertisement, etc. The law of demand can be illustrated through a *demand schedule* and a *demand curve.*

4.2.2 Movement along Vs. Shift in Demand Curve

The demand curve slopes downward to the right because of the law of the demand. It implies that factors that bring the law of demand into operation make the demand curve slope downward to the right. The **factors behind the law of demand** are following:

1. *Income Effect.* When price of a commodity falls, the purchasing power of its consumers increases since they are required to pay less for the same quantity. It means that with the fall in price, consumers' real income increases. According to another economic law, increase in real income (or purchasing power) increases demand for 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* (claiming a considerable proportion of the total consumption expenditure) falls substantially, consumers' real income increases. Consequently, they substitute superior goods for inferior ones. Therefore, *income effect* on the demand for inferior goods becomes *negative.*

2. *Substitution Effect.* When price of a commodity falls, it becomes relatively cheaper compared to its substitutes, their prices remaining constant. There is a natural tendency that consumers substitute cheaper goods for costlier ones, all other factors remaining the same. Consequently, rational consumers tend to substitute cheaper goods for costlier ones within the range of normal goods— goods whose demand increases with increase in consumer's income—other things remaining the same. Therefore, demand for the relatively cheaper commodity increases. The increase in demand on account of this factor is known as *substitution effect.*

3. *Diminishing Marginal Utility.* Marginal utility is the utility derived from the marginal unit consumed of a commodity. According to the law of diminishing marginal utility, the utility derived from the additional unit consumed goes over diminishing. Therefore, consumers consume more of a commodity only when its price decreases. Diminishing marginal utility is also responsible for increase in demand for a commodity when its price falls.

Thus, the income and substitution effects of fall in price and the law of diminishing marginal utility make the demand curve slope downward to the right.

Exceptions to the Law of Demand

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

1. ***Expectations Regarding Future Prices.*** When consumers expect a continuous increase in the price of a durable commodity, they buy more of it despite increase in its price, to avoid the pinch of a 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 serve as a 'status symbol', enhance social prestige or display wealth and richness, e.g., gold, precious stones, rare paintings and antiques, etc. Rich people buy such goods mainly because their prices are high. No body will buy diamonds, a prestigious good, if it sells at, say, ₹ 100 a kg.

3. ***Giffen Goods.*** Another exception to this law is the classic 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 essential commodity much cheaper than its substitutes, consumed mostly by the poor households and claiming a large part of their income. If 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 a kg and wheat ₹ 10 a 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.

Check Your Progress

1. What does the law of demand denote?
2. What is income effect?

4.3 CONCEPT OF MEASUREMENT OF ELASTICITY OF DEMAND

We have earlier discussed the *nature* of *relationship* between demand and its determinants. From managerial point of view, however, the knowledge of nature of relationship alone is not sufficient. What is more important is the extent of

relationship or the degree of responsiveness of demand to the changes in its determinants. The degree of responsiveness of demand to the change in its determinants is called *elasticity of demand*.

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 want to pass the rising cost on to the consumer by raising the price. Firms may decide to change the price even without any change in the cost of production. But whether raising price following the rise in cost or otherwise proves beneficial depends on:

(*a*) the price-elasticity of demand for the product, i.e., how high or low is the proportionate change in its demand in response to a certain percentage change in its price; and

(*b*) price-elasticity of demand for its substitutes because when the price of a product increases, the demand for its substitutes increases automatically even if their prices remain unchanged.

Raising price will be beneficial only if (*i*) demand for a product is less elastic; and (*ii*) demand for its substitute is much less elastic. Although most businessmen are intuitively aware of the elasticity of demand of the goods they make, the use of precise estimates of elasticity of demand will add precision to their business decisions.

In this section, we will discuss various methods of measuring elasticities of demand. The concepts of demand elasticities used in business decisions are:(*i*) Price elasticity, (*ii*) Cross-elasticity; (*iii*) Income elasticity; and (*iv*) Advertisement elasticity, and (*v*) Elasticity of price expectation.

Price Elasticity of Demand

Price elasticity of demand is generally defined as the responsiveness or sensitiveness of demand for a commodity to the changes in its price. More precisely, *elasticity of demand is the percentage change in demand as a result of one per cent change in the price of the commodity*. A formal definition of price elasticity of demand (e_p) is given as

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

A general formula for calculating coefficient of price elasticity, derived from this definition of elasticity, is given as follows:

$$e_p = \frac{\Delta Q}{Q} \div \frac{\Delta P}{P} = \frac{\Delta Q}{Q} \times \frac{P}{\Delta P}$$

$$= \frac{\Delta Q}{\Delta P} \times \frac{P}{Q} \qquad \qquad \dots(4.1)$$

where Q = original quantity demanded, P = original price, ΔQ = change in quantity demanded and ΔP = change in price.

It is important to note here that a minus sign (–) is generally inserted in the formula before the fraction in order to make the elasticity coefficient a non-negative value.

The elasticity can be measured between any two points on a demand curve (called *arc elasticity*) or at a point (called *point elasticity*).

Arc Elasticity

The measure of elasticity of demand between any two finite points on a demand curve is known as *arc elasticity*. For example, measure of elasticity between points J and K (Fig. 4.1) is the measure of arc elasticity. The movement from point J to K on the demand curve (D_x) shows a fall in the price from ₹20 to ₹10 so that $\Delta P = 20 - 10 = 10$. The fall in price causes an increase in demand from 43 units to 75 units so that $\Delta Q = 43 - 75 = -32$. The elasticity between points J and K (moving from J to K) can be calculated by substituting these values into the elasticity formula as follows:

$$e_p = -\frac{\Delta Q}{\Delta P} \cdot \frac{P}{Q} \text{ (with minus sign)}$$

$$= -\frac{-32}{10} \cdot \frac{20}{43} = 1.49 \qquad \qquad ...(4.2)$$
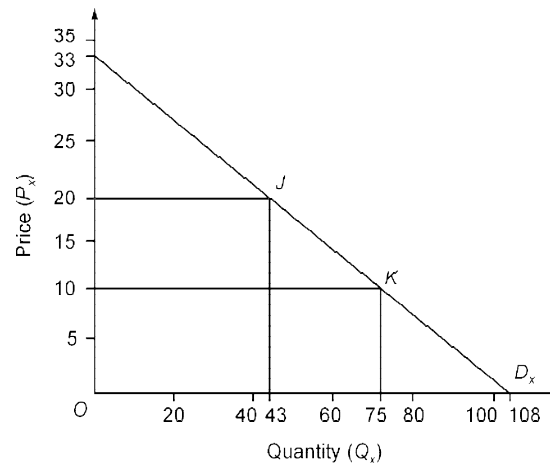
Fig. 4.1 Linear Demand Curve

This means that a one per cent decrease in price of commodity X results in a 1.49 per cent increase in demand for it.

Problem in Using Arc Elasticity: The arc elasticity should be measured and used carefully, otherwise it may lead to wrong decisions. Arc elasticity co-efficients

differ between the same two finite points on a demand curve if *direction* of change in price is reversed. For instance, as estimated in Eq. (4.2), the elasticity between points *J* and *K*—moving from *J* to *K* equals 1.49. It may be wrongly interpreted that the elasticity of demand for commodity *X* between points *J* and *K* equals 1.49 irrespective of the direction of price change. But it is not true. A reverse movement in the price, i.e., the movement from point *K* to *J* implies a different elasticity co-efficient (0.43). Movement from point *K* to *J* gives $P = 10$, $\Delta P = 10 - 20 = -10$, $Q = 75$ and $\Delta Q = 75 - 43 = 32$. By substituting these values into the elasticity formula, we get

$$e_p = -\frac{32}{-10}.\frac{10}{75} = 0.43 \qquad\qquad ...(4.3)$$

The measure of elasticity co-efficient in Eq. (4.3) for the reverse movement in price is obviously different from one given by Eq. (4.2). It means that *the elasticity depends also on the direction of change* in price. Therefore, while measuring price elasticity, the direction of price change should be carefully noted.

Some Modifications   Some modifications have been suggested in economic literature to resolve the problems associated with arc elasticity.

First, the problem arising due to the change in the direction of price change may be avoided by using the lower values of *P* and *Q* in the elasticity formula, so that

$$e_p = -\frac{\Delta Q}{\Delta P}.\frac{P_l}{Q_l}$$

where $P_l = 10$ (the lower of the two prices) and $Q_l = 43$ (the lower of the two quantities). Thus,

$$e_p = -\frac{32}{10}.\frac{10}{43} = 0.74 \qquad\qquad ...(4.4)$$

This method is however devoid of the logic of calculating percentage change because the choice of lower values of *P* and *Q* is arbitrary—it is not in accordance with the rule of calculating percentage change.

Second, another method suggested to resolve this problem is to use the average of upper and lower values of *P* and *Q* in fraction *P/Q*. In that case the formula is

$$e_p = -\frac{\Delta Q}{\Delta P}.\frac{(P_1 + P_2)/2}{(Q_1 + Q_2)/2}$$

or
$$e_p = -\frac{Q_1 - Q_2}{P_1 - P_2}.\frac{(P_1 + P_2)/2}{(Q_1 + Q_2)/2} \qquad\qquad ...(4.5)$$

where subscripts 1 and 2 denote lower and upper values of prices and quantitites.

Substituting the values from our example, we get,

$$e_p = - \frac{43 - 75}{20 - 10} \cdot \frac{(20 + 10)/2}{(43 + 75)/2} = 0.81$$

This method too has its own drawbacks as the elasticity co-efficient calculated through this formula refers to the elasticity mid-way between $P_1 P_2$ and $Q_1 Q_2$. The elasticity co-efficient (0.81) is not applicable for the whole range of price-quantity combinations at different points between *J* and *K* on the demand curve (Fig. 4.1)—it only gives a mean of the elasticities between the two points.

Point elasticity on a linear demand curve. Point elasticity is also a way to resolve the problem in measuring the elasticity. The concept of point elasticity is used for measuring price elasticity where change in price is infinitesimally small.

Point elasticity is the elasticity of demand at a finite point on a demand curve, e.g., at point *P* or *B* on the linear demand curve *MN* in Fig. 4.2. This is in contrast to the arc elasticity between points *P* and *B*. A movement from point *B* towards *P* implies change in price (ΔP) becoming smaller and smaller, such that point *P* is almost reached. Here the change in price is infinitesimally small. Measuring elasticity for an infinitesimally small change in price is the same as measuring elasticity at a point. The formula for measuring point elasticity is given below.

Point elasticity $(e_p) = \dfrac{P}{Q} \cdot \dfrac{\partial Q}{\partial P}$ ...(4.6)
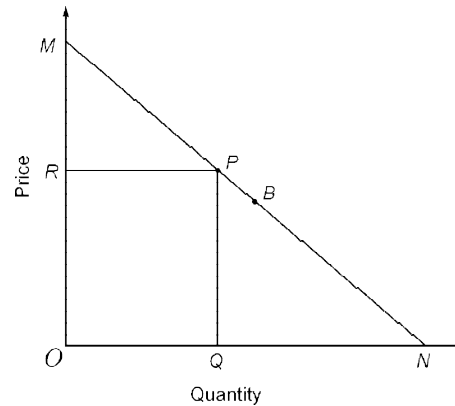
Fig. 4.2 *Point Elasticity*

Note that $\dfrac{\partial Q}{\partial P}$ has been substituted for $\dfrac{\Delta Q}{\Delta P}$ in the formula for arc elasticity.

The derivative $\dfrac{\partial Q}{\partial P}$ is reciprocal of the slope of the demand curve *MN*. Point elasticity is thus the product of price-quantity ratio at a particular point on the demand curve and the reciprocal of the slope of the demand line. The reciprocal

of the slope of the straight line *MN* at point *P* is geometrically given by *QN/PQ*. Therefore,

$$\frac{\Delta Q}{\Delta P} = \frac{QN}{PQ}$$

Note that at point *P*, price $P = PQ$ and $Q = OQ$. By substituting these values in Eq. (4.14), we get

$$e_p = \frac{PQ}{OQ} \cdot \frac{QN}{PQ} = \frac{QN}{OQ}$$

Given the numerical values for *QN* and *OQ*, elasticity at point *P* can be easily obtained. We may compare here the arc elasticity between points *J* and *K* and point elasticity at point *J* in Fig. 4.1. At point *J*,

$$e_p = \frac{QN}{OQ} = \frac{108 - 43}{43} = 1.51$$

Note that $e_p = 1.51$ is different from various measures of arc elasticities (i.e., $e_p = 1.49$, $e_p = 0.43$, $e_p = 0.7$, and $e_p = 0.81$).

As we will see below, geometrically, $QN/OQ = PN/PM$. Therefore, elasticity of demand at point *P* (Fig. 4.2) may be expressed as

$$e_p = \frac{PN}{PM}$$

Measuring Price Elasticity from a Demand Function

The price elasticity of demand for a product can be measured directly from the demand function. In this section, we will describe the method of measuring price elasticity of demand for a product from the demand function—both linear and non-linear. It may be noted here that if a demand function is given, arc elasticity can be measured simply by assuming two prices and working out ΔP and ΔQ. We will, therefore, confine ourselves here to point elasticity of demand with respect to price.

Price Elasticity from a Linear Demand Function

Suppose that a linear demand function is given as

$$Q = 100 - 5P$$

Given the demand function, point elasticity can be measured for any price. For example, suppose we want to measure elasticity at $P = 10$. We know that

$$e_p = \frac{\delta Q}{\delta P} \cdot \frac{P}{Q}$$

The term $\delta Q/\delta P$ in the elasticity formula is the slope of the demand curve. The slope of the demand curve can be found by differentiating the demand function. That is,

$$\frac{\delta Q}{\delta P} = \frac{\delta(100-5P)}{\delta P} = -5$$

Having obtained the slope of the demand curve as $\delta Q/\delta P = -5$, e_p at $P = 10$ can be calculated as follows. Since, $P = 10$, $Q = 100 - 5(10) = 50$. By substituting these values into the elasticity formula, we get,

$$e_p = (-5)\,\frac{10}{50} = -1$$

Similarly, at $\qquad\qquad P = 8, Q = 100 - 5(8) = 60 \qquad$ and

$$e_p = -5\,(8/60) = -40/60 = -0.67$$

And at $\qquad\qquad P = 15, Q = 100 - 5(15) = 25, \qquad$ and

$$e_p = -5(15/25) = -75/25 = -3$$

Price Elasticity from a Non-linear Demand Function

Suppose a non-linear demand function of multiplicative form is given as follows.

$$Q = aP^{-b}$$

and we want to compute the price elasticity of demand. The formula for computing the price elasticity is the same, i.e.,

$$e_p = \frac{\delta Q}{\delta P}\cdot\frac{P}{Q} \qquad\qquad\qquad …(4.7)$$

What we need to compute the price-elasticity coefficient is to find first the value of the first term, $\delta Q/\delta P$, i.e., the slope of the demand curve. The slope can be obtained by differentiating the demand function, Thus,

$$\frac{\delta Q}{\delta P} = -baP^{-b-1} \qquad\qquad\qquad …(4.8)$$

By substituting Eq. (4.8) in Eq. (4.7), e_p can be expressed as

$$e_p = -baP^{-b-1}\left(\frac{P}{Q}\right)$$

$$= \frac{-baP^{-b}}{Q} \qquad\qquad\qquad …(4.9)$$

Since $Q = aP^{-b}$, by substitution, we get

$$e_p = \frac{-baP^{-b}}{aP^{-b}} = -b \qquad\qquad …(4.10)$$

Equation (4.10) shows that when a demand function is of a multiplicative or power form, price elasticity coefficient equals the power of the variable *P*. This means that price elasticity in the case of a multiplicative demand function remains constant all along the demand curve regardless of a change in price.

Price Elasticity and Total Revenue

A firm aiming at enhancing its total revenue would like to know whether increasing or decreasing the price would achieve its goal. The price-elasticity coefficient of demand for its product at different levels of its price provides the answer to this question. The simple answer is that if $e_p > 1$, then decreasing the price will increase the total revenue and if $e_q < 1$, then increasing the price will increase the total revenue. To prove this point, we need to know the total revenue (*TR*) and the marginal revenue (*MR*) functions and measures of price-elasticity are required. Since $TR = Q.P$, we need to know *P* and *Q*. This information can be obtained through the demand function. Let us recall our demand function given as

$$Q = 100 – 5P$$

Price function (*P*) can be derived from the demand function as

$$P = 20 – 0.2Q \qquad\qquad …(4.11)$$

Given the price function, *TR* can be obtained as

$$TR = P . Q = (20 – 0.2Q)Q = 20Q – 0.2Q^2$$

From this *TR*-function, the *MR*-function can be derived as

$$MR = \frac{\partial TR}{\partial Q} = 20 – 0.4Q$$

The *TR*-function is graphed in panel (*a*) and the demand and *MR* functions are presented in panel (*b*) of Fig. 4.3. As the figure shows, at point *P* on the demand curve, $e = 1$ where output, $Q = 50$. Below point *P*, $e < 1$ and above point *P*, $e > 1$. It can be seen in panel (*a*) of Fig. 4.3 that *TR* increases so long as $e > 1$; *TR* reaches its maximum level where $e = 1$; and it decreases when $e < 1$.
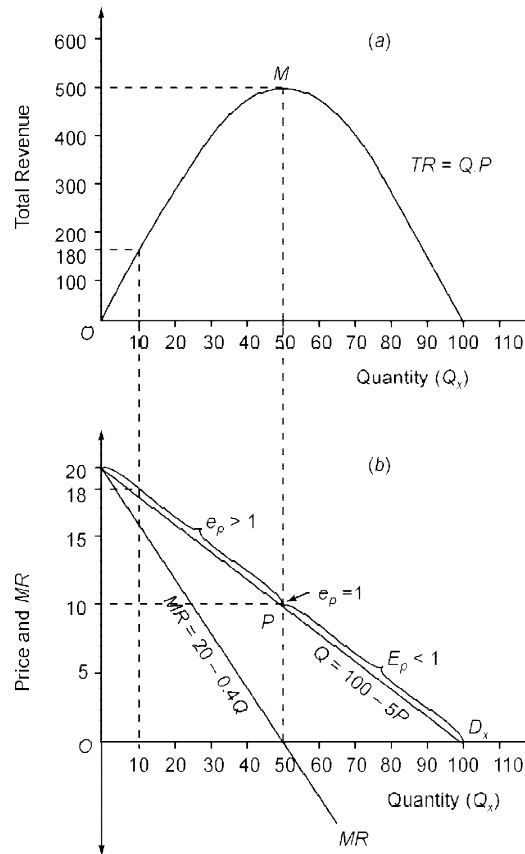
Fig. 4.3 *Price Elasticity and Total Revenue*

The relationship between price-elasticity and *TR* is summed up in Table 4.1. As the table shows, when demand is *perfectly inelastic* (i.e., $e_p = 0$ as is the case of a vertical demand line) there is no decrease in quantity demanded when price is raised and *vice versa*. Therefore, a rise in price increases the total revenue and *vice versa*.

In case of an *inelastic demand* (i.e., $e_p < 1$), quantity demanded increases by less than the proportionate decrease in price and hence the total revenue falls when price falls. The total revenue increases when price increases because quantity demanded decreases by less than the proportionate increase in price.

If demand for a product is *unit elastic* ($e_p = 1$) quantity demanded increases (or decreases) in the proportion of decrease (or increase) in the price. Therefore, total revenue remains unaffected.

If demand for a commodity has $e_p > 1$, change in quantity demanded is greater than the proportionate change in price. Therefore, the total revenue increases when price falls and *vice versa.*

The case of *infinitely elastic* demand represented by a horizontal straight line is rare. Such a demand line implies that a consumer has the opportunity to buy any quantity of a commodity and the seller can sell any quantity of a commodity, at a given price. It is the case of a commodity being bought and sold in a perfectly competitive market. A seller, therefore, cannot charge a higher or a lower price.

Table 4.1 *Elasticity, Price-change and Change in TR*

Elasticity Co-efficient	Change in Price	Change in TR
$e = 0$	Increase	Increase
	Decrease	Decrease
$e < 1$	Increase	Increase
	Decrease	Decrease
$e = 1$	Increase	No change
	Decrease	No change
$e > 1$	Increase	Decrease
	Decrease	Increase
$e = \infty$	Increase	Decrease to zero
	Decrease	Infinite increase*

*Subject to the size of the market.

Price Elasticity and Marginal Revenue

The relationship between price-elasticity and the total revenue (*TR*) can be known more precisely by finding the relationship between price-elasticity and marginal revenue (*MR*). *MR* is the first derivative of *TR*-function and $TR = P.Q$ (where $P =$ price, and $Q =$ quantity sold). The relationship between price-elasticity, *MR* and *TR* is shown below.

Since $TR = P.Q$,

$$MR = \frac{\partial(P \cdot Q)}{\partial Q} = P + Q\frac{\partial P}{\partial Q}$$

$$= P\left(1 + \frac{Q}{P} \cdot \frac{\partial P}{\partial Q}\right) \qquad \ldots(4.12)$$

Note that $\dfrac{Q}{P} \cdot \dfrac{\partial P}{\partial Q}$ in Eq. (4.12) is the reciprocal of elasticity. That is,

$$\frac{Q}{P} \cdot \frac{\partial P}{\partial Q} = -\frac{1}{e_p}$$

By substituting $-\dfrac{1}{e}$ for $\dfrac{Q}{P} \cdot \dfrac{\partial P}{\partial Q}$ in Eq. (4.12), we get

$$MR = P\left(1 - \frac{1}{e_p}\right) \qquad\qquad …(4.13)$$

Given this relationship between *MR* and price-elasticity of demand, the decision-makers can easily know whether it is beneficial to change the price. If $e = 1$, $MR = 0$. Therefore, change in price will not cause any change in *TR*. If $e < 1$, $MR < 0$, *TR* decreases when price decreases and *TR* increases when price increases. And, if $e > 1$, $MR > 0$, *TR* increases if price decreases and *vice versa*.

Price Elasticity, AR and MR: Given the Eq. (4.13), the formula for price elasticity (e_p) can be expressed in terms of *AR* and *MR*. We know that $P = AR$. So Eq. (4.13) can be written as

$$MR = AR\left(1 - \frac{1}{e_p}\right)$$

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

By rearranging the terms, we get

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

or

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

The reciprocal of this equation gives the measure of the price elasticity (e_p) of demand which can be expressed as

$$\frac{AR}{MR - AR} = -e_p \qquad \text{or} \qquad e_p = \frac{AR}{AR - MR}$$

Factors Affecting Price Elasticity of Demand

We have noted above that price-elasticity of a product may vary between zero and infinity. However, price-elasticity of demand, at a given price, varies from product to product depending on the following factors.

1. *Availability of Substitutes.* One of the most important determinants of elasticity of demand for a commodity is the availability of its close substitutes. The higher the degree of closeness of the substitutes, the greater the elasticity of demand for the commodity. For instance, coffee and tea may be considered as close substitutes for one another. If price of one of these goods increases, the other commodity becomes relatively cheaper. Therefore, consumers buy more of the relatively cheaper good and less of the costlier one, all other things remaining the same. The elasticity of demand for both these goods will be higher. Besides, the wider the range of the substitutes, the greater the elasticity. For instance, soaps, toothpastes, cigarettes, etc., are available in different brands, each brand being a close substitute for the other. Therefore, the price-elasticity of demand for each brand is much greater than that for the generic commodity. On the other hand, sugar and salt do not have close substitutes and hence their price-elasticity is lower.

2. *Nature of Commodity. The nature of a commodity* also affects the price-elasticity of its demand. Commodities can be grouped as luxuries, comforts and necessities. Demand for luxury goods (e.g., high-price refrigerators, TV sets, cars, decoration items, etc.) is more elastic than the demand for necessities and comforts because consumption of luxury goods can be dispensed with or postponed when their prices rise. On the other hand, consumption of necessary goods, (e.g., sugar, clothes, vegetables) cannot be postponed and hence their demand is inelastic. Comforts have more elastic demand than necessities and less elastic than luxuries. Commodities are also categorized as durable goods and perishable or non-durable goods. Demand for durable goods is more elastic than that for non-durable goods, 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. *Weightage in the Total Consumption.* Another factor that influences the elasticity of demand is the proportion of income which consumers spend on a particular commodity. If proportion of income spent on a commodity is large, its demand will be more elastic. On the contrary, if the proportion of income spent on a commodity is small, its demand is less price-elastic. Classic examples of such commodities are salt, matches, books, pens, toothpastes, etc. These goods claim a very small proportion of income. Demand for these goods is generally inelastic because increase in the price of such goods does not substantially affect the

consumer's budget. Therefore, people continue to purchase almost the same quantity even when their prices increase.

4. *Time Factor in Adjustment of Consumption Pattern.* Price-elasticity of demand depends also on *the time consumers need to adjust their consumption pattern to a new price*: the longer the time available, the greater the price-elasticity. The reason is that over a period of time, consumers are able to adjust their expenditure pattern to price changes. For instance, if the price of TV sets is decreased, demand will not increase immediately unless people possess excess purchasing power. But over time, people may be able to adjust their expenditure pattern so that they can buy a TV set at a lower (new) price. Consider another example. If price of petrol is reduced, the demand for petrol does not increase immediately and significantly. Over time, however, people get incentive from low petrol prices to buy automobiles resulting in a significant rise in demand for petrol.

5. *Range of Commodity Use.* *The range of uses of a commodity* also influences the price-elasticity of its demand. The wider the range of the uses of a product, the higher the elasticity of demand for the decrease in price. As the price of a multi-use commodity decreases, people extend their consumption to its other uses. Therefore, the demand for such a commodity generally increases more than the proportionate increase in its price. For instance, milk can be taken as it is and in the form of 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 has a greater elasticity. However, for the increase in price, such commodities have a lower price-elasticity because the consumption of a normal good cannot be cut down substantially beyond a point when the price of the commodity increases.

6. *Proportion of Market Supplied.* The elasticity of market demand also depends on the *proportion of the market supplied at the ruling price*. If less than half of the market is supplied at the ruling price, price-elasticity of demand will be higher than 1 and if more than half of the market is supplied, $e < 1$.

4.3.1 Cross-Elasticity of Demand

The 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, cross-elasticity of demand for tea is the percentage change in its quantity demanded with respect to the change in the price of its substitute, coffee. The formula for measuring cross-elasticity of demand for tea ($e_{t,c}$) and the same for coffee ($e_{c,t}$) is given below.

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

$$= \frac{P_c}{Q_t} \cdot \frac{\Delta Q_t}{\Delta P_c} \qquad \ldots(4.14)$$

and $$e_{c,\,t} = \frac{P_t}{Q_c} \cdot \frac{\Delta Q_c}{\Delta P_t} \qquad \ldots(4.15)$$

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

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

Uses of Cross-Elasticity

An *important use* of cross-elasticity is to define substitute goods. If cross-elasticity between any two goods is positive, the two goods may be considered as substitutes of one another. Also, the greater the cross-elasticity, the closer the substitute. Similarly, if cross-elasticity of demand for two related goods is negative, the two may be considered as complementary of one another: the higher the negative cross-elasticity, the higher the degree of complementarity.

The concept of cross-elasticity is of vital importance in changing prices of products having substitutes and complementary goods. If cross-elasticity in response to the price of substitutes is greater than one, it would be inadvisable to increase the price; rather, reducing the price may prove beneficial. In case of complementary goods also, reducing the price may be helpful in maintaining the demand in case the price of the complementary good is rising. Besides, if accurate measures of cross-elasticities are available, the firm can forecast the demand for its product and can adopt necessary safeguards against fluctuating prices of substitutes and complements.

4.3.2 Income-Elasticity of Demand

Apart from the price of a product and its substitutes, consumer's income is another basic determinant of demand for a product. As noted earlier, the relationship between quantity demanded and income is of positive nature, unlike the negative price-demand relationship. The demand for most goods and services increases with increase in consumer's income and *vice versa. The responsiveness of demand to the changes in income is known as income-elasticity of demand.*

Income-elasticity of demand for a product, say X, (i.e., e_y) may be defined as:

$$e_y = \frac{\dfrac{\Delta X_q}{X_q}}{\dfrac{\Delta Y}{Y}} = \frac{Y}{X_q} \cdot \frac{\Delta X_q}{\Delta Y} \qquad \ldots(4.16)$$

(where X_q = quantity of X demanded; Y = disposable income; ΔX_q = change in quantity of X demanded; and ΔY = change in income)

Obviously, the formula for measuring income-elasticity of demand is the same as that for measuring the price-elasticity. The only change in the formula is that the variable 'income' (Y) is substituted for the variable 'price' (P). Here, income refers to the disposable income, i.e., income net of taxes. All other formulae for measuring price-elasticities may by adopted to measure the income-elasticities, keeping in mind the difference between them and the purpose of measuring income-elasticity.

To estimate income-elasticity, suppose, for example, government announces a 10 per cent dearness allowance to its employees. As a result average monthly salary of government employees increases from ₹ 20,000 to ₹ 22,000. Following the pay-hike, monthly petrol consumption of government employees increases from 150 litre per month to 165 litre. The income-elasticity of petrol consumption can now be worked out as follows. In this case, ΔY = ₹ 22,000 – ₹ 20,000 = ₹ 2,000, and ΔQ (oil demand) = 165 litre – 150 litre = 15 litre. By substituting those values in Eq. (4.16), we get

$$e_y = \frac{20,000}{150} \times \frac{15}{2,000} = 1$$

It means that income elasticity of petrol consumption by government employees equals 1. It means that a one per cent increase in income results in a one per cent increase in petrol consumption.

Unlike price-elasticity of demand, which is always negative, income-elasticity of demand is always positive because of a positive relationship between income and quantity demanded of a product. But there is an *exception* to this rule. Income-elasticity of demand for an inferior good is negative, because of the inverse substitution effect. The demand for inferior goods decreases with increase in consumer's income. The reason is that when income increases, consumers switch over to the consumption of superior substitutes, i.e., 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 foodgrains; non-vegetarians buy more of meat and less of potato, and travellers travel more by plane and less by train.

Nature of Commodity and Income-Elasticity

For all normal goods, income-elasticity is positive though the degree of elasticity varies in accordance with the nature of commodities. Consumer goods of the three categories, viz., necessities, comforts and luxuries have different elasticities. The general pattern of income-elasticities of different goods for increase in income and their effect on sales are given in Table 4.2.

Table 4.2 *Income-Elasticities*

	Consumer goods	Co-efficient of income-elasticity	Effect on sales with change in income
1.	Essential goods	Less than one ($e_y < 1$)	Less than proportionate change in sale
2.	Comforts	Almost equal to unity ($e_y \cong 1$)	Almost proportionate change in sale
3.	Luxuries	Greater than unity ($e_y > 1$)	More than proportionate increase in sale

Income-elasticity of demand for different categories of goods may, however, vary from household to household and from time to time, depending on the choice and preference of the consumers, levels of 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 frequency of rise in income is high, income-elasticities will conform to the general pattern.

Uses of Income-Elasticity in Business Decisions

While price and cross elasticities of demand are of greater significance in the pricing of a product aimed at maximizing the total revenue in the short run, income-elasticity of a product is of a greater significance in production planning and management in the long run, particularly during the period of a business cycle. The concept of income-elasticity can be used in estimating future demand provided that the rate of increase in income and income-elasticity of demand for the products are known. The knowledge of income elasticity can thus be useful in forecasting demand, when a change in personal incomes is expected, other things remaining the same. It also helps in avoiding over-production or under-production.

In forecasting demand, however, only the relevant concept of income and data should be used. It is generally believed that the demand for goods and services

increases with increase in GNP, depending on the marginal propensity to consume. This may be true in the context of aggregate national demand, but not necessarily for a particular product. It is quite likely that increase in GNP flows to a section of consumers who do not consume the product in which a businessman is interested. For instance, if the major proportion of incremental GNP goes to those who can afford a car, the growth rate in GNP should not be used to calculate income-elasticity of demand for bicycles. Therefore, the income of only a relevant class or income-group should be used. Similarly, where the product is of a regional nature, or if there is a regional division of market between the producers, the income of only the relevant region should be used in forecasting the demand.

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

Check Your Progress

3. Define elasticity of demand?

4. Why is the concept of point elasticity used?

5. State one important use of cross-elasticity of demand.

4.4 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. The law of demand denotes the relationship between the quantity demanded and price of a commodity.

2. The increase in demand on account of increase in real income is called income effect.

3. The degree of responsiveness of demand to the change in its determinants is called elasticity of demand.

4. The concept of point elasticity is used for measuring price elasticity where change in price is infinitesimally small.

5. One important use of cross-elasticity of demand is to define substitute goods. If cross-elasticity between any two goods is positive, the two goods may be considered as substitutes of one another.

4.5 SUMMARY

- The analysis of total demand for a firm's product plays a crucial role in business decision-making.

- The market demand or the size of the market at a point in time at different prices gives the overall scope of business; it gives prospects for expanding business; and it plays a crucial role in planning for future production, inventories of raw materials, advertisement, and setting up sales outlets.

- 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 decreases when its price increases. This law implies that demand and price are inversely related.

- The demand curve slopes downward to the right because of the law of the demand. It implies that factors that bring the law of demand into operation make the demand curve slope downward to the right.

- The degree of responsiveness of demand to the change in its determinants is called elasticity of demand.

- Price elasticity of demand is generally defined as the responsiveness or sensitiveness of demand for a commodity to the changes in its price. More precisely, elasticity of demand is the percentage change in demand as a result of one per cent change in the price of the commodity.

- The relationship between price-elasticity and the total revenue (TR) can be known more precisely by finding the relationship between price-elasticity and marginal revenue (MR).

- The demand for most goods and services increases with increase in consumer's income and vice versa. The responsiveness of demand to the changes in income is known as income-elasticity of demand.

4.6 KEY WORDS

- **Price Elasticity of Demand:** It is generally defined as the responsiveness or sensitiveness of demand for a commodity to the changes in its price.

- **Point Elasticity:** It is the elasticity of demand at a finite point on a demand curve.

- **Income Elasticity of Demand:** It measures the responsiveness of the quantity demanded for a good or service to a change in the income of the people demanding the good.

- **Marginal Revenue:** It is the revenue gained by producing one additional unit of a product or service.

4.7 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short Answer Questions

1. State in your own words the law of demand.
2. Write a short note on the importance of elasticity concept in economics.
3. Mention the determinants of price elasticity of demand.

Long Answer Questions

1. Aanlyse the need for market demand.
2. What are the factors responsible for the movement /shift in the demand curve?
3. Explain the concept of cross-elasticity of demand.

4.8 FURTHER READINGS

Dwivedi, D. N. 2008. *Principles of Economics*, Seventh Edition. New Delhi: Vikas Publishing House.

Weil. David N. 2004. *Economic Growth*. London: Addison Wesley.

Thomas, Christopher R. and Maurice S. Charles. 2005. *Managerial Economics: Concepts and Applications*, Eighth Edition. New Delhi: Tata McGraw-Hill Publishing Company Limited.

Mankiw, Gregory N. 2002. *Principles of Economics*, Second Edition. India: Thomson Press.

<div style="text-align:center">

BLOCK - II
FEATURES OF ECONOMICS

</div>

UNIT 5   THEORY OF PRODUCTION

Structure

5.0   INTRODUCTION

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

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

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

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

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

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

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

This unit deals with the discussion on the theory of production. Production theory deals with quantitative relationships—technical and technological relations—between inputs (especially labour and capital) and output. Let us first discuss some basic concepts used in production analysis.

5.1 OBJECTIVES

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

- State the meaning and concept of production
- Identify the factors of production and production function
- Define the law of variable proportion
- Discuss the law of returns to a scale

5.2 THEORY OF PRODUCTION: AN OVERVIEW

Let us begin our discussion about the theory of production by beginning with the meaning and basic concepts of production.

5.2.1 Meaning and Basic Concepts of Production

This section presents a brief discussion on the basic concepts and terminology used in the analysis of the theory of production.

Meaning of Production

In general sense of the term 'production' means transforming inputs (labour, capital, raw materials, time, etc.) into an output. This concept of production is however limited to only 'manufacturing'. In economic sense, the term 'production' means a process by which resources (men, material, time, etc.) are transformed into a *different* and more useful commodity or service. In other words, a process by which men, material, capital and time are converted into value added products is called *production*.

In economic sense, *production process* may take a variety of forms other than manufacturing. For example, 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 the sand from the river bank to the construction site and a coal-miner does virtually nothing more than digging coal out of the ground and transporting to the factories. Similarly, fishermen only catch and transport fishes from sea, river and fisheries to the fish market. Their activities too are 'production'. Transporting men and materials from one place to another is also a productive activity: it produces *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 raw materials 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, both inputs and outputs are intangible; lawyers, doctors, social workers, consultants, musicians, orchestra players, 'bar girls', etc., are all engaged in producing intangible goods.

5.2.2 Factors of Production

Factors of production is an economic term that explains the inputs that are used in the production of goods or services to earn economic profit.

Input and Output

An *input* is a good or service that goes into 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.

The term 'inputs' needs some more explanations. Production process requires a wide variety of inputs, depending on the nature of product. But, economists classified inputs as (*i*) labour, (*ii*) capital, (*iii*) land, (*iv*) raw materials and (*v*) entrepreneurship. Technology and time are also treated as inputs in the modern concept of production.

Fixed and Variable Inputs

Inputs are classified as (*i*) *fixed inputs* and (*ii*) *variable inputs*. Fixed and variable inputs are defined in economic sense and also in technical sense. In **economic sense**, a fixed input is one whose supply is inelastic in the short-run. Therefore, all of its users together cannot buy more of it in the short-run. In **technical sense,** a fixed factor is one that remains fixed (or constant) for a certain level of output.

A *variable input* is defined as one whose supply in the short-run is elastic, e.g., labour and raw material, etc. All the users of such factors can employ a larger quantity in the short-run as well as in the long-run. Technically, a variable input is one that changes with the change in output. In the long-run, all inputs are variable.

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*. The **short-run** refers to a period of time in which the supply of certain inputs (e.g., plant, building, machinery, etc.) is fixed or is inelastic. In the short-run, therefore, production of a commodity can be increased by increasing the use of only variable inputs like labour and raw materials.

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

On the other hand, **long-run** refers to a period of time in which the supply of all the inputs is elastic, but not enough to permit a change in technology. That is, in the long-run, all the inputs are variable. Therefore, in the long-run, production of a commodity can be increased by employing more of both variable and fixed inputs.

The economists use another term, i.e., **very long-run** which refers to a period in which the technology of production is also subject to change or can be improved. In the very long-run, the production function also changes. The technological advances result in a larger output from a given quantity of inputs per unit of time.

5.3 PRODUCTION FUNCTION

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

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

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

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

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

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

The reasons for excluding other inputs are following.

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

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

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

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

$$Q_C = f(K, L) \qquad \qquad \text{…(5.1)}$$

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

K = capital,   and L = labour.

The production function (5.1) implies that quantity of coal produced depends on the quantity of capital (K) and labour (L) employed to produce coal. Increasing coal production will require increasing K and L. Whether the firm can increase both K and L or only L depends on the time period it takes into account for increasing production, i.e., whether the firm considers a *short-run* or a *long-run*.

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

(*i*) *Short-run production function;* and

(*ii*) *Long-run production function.*

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

$$Q = f\ (\bar{K}, L), \text{ where } K \text{ is a constant} \qquad …(5.2a)$$

For example, suppose a production function is expressed as

$$Q = bL$$

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

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

$$Q = f\ (K, L) \qquad\qquad …(5.2b)$$

As mentioned above, a production function can be expressed in the form of an equation, a graph or a table, though each of these forms can be converted into its other forms. We illustrate here how a production function in the form of an equation can be converted into its tabular form. Consider, for example, the Cobb-Douglas production function—the most famous and widely used production function—given in the form of an equation as

$$Q = AK^a L^b \qquad\qquad …(5.3)$$

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

Production function (5.3) gives the *general form* of Cobb-Douglas production function. The numerical values of parameters A, a and b, can be estimated by using actual factory data on production, capital and labour. Suppose numerical values of parameters are estimated as $A = 50$, $a = 0.5$ and $b = 0.5$. Once numerical values are known, the Cobb-Douglas production function can be expressed in its *specific form* as follows.

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

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

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

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

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

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

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

Similarly, by assigning different numerical values to K and L, the resulting output can be worked out for different combinations of K and L and a tabular form of production function can be prepared. Table 5.1 shows the maximum quantity of a commodity that can be produced by using different combinations of K and L, both varying between 1 and 10 units.

Table 5.1 *Production Function in Tabular Form*

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

Table 5.1 shows the units of output that can be produced with different combinations of capital and labour. The figures given in Table 5.1 can be graphed in a three-dimensional diagram.

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

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

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

The laws of production state the relationship between output and input. In the short-run, input-output relations are studied with one variable input (labour), other inputs (especially, capital) held constant. The laws of production under these conditions are called the 'Laws of Variable Proportions' or the 'Laws of Returns to a Variable Input'.

5.4 THE LAW OF DIMINISHING RETURNS TO A VARIABLE INPUT

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

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

 (*i*) labour is the only variable input, capital remaining constant;

 (*ii*) labour is homogeneous;

 (*iii*) the state of technology is given; and

 (*iv*) input prices are given.

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

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

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

$$Q_c = -L^3 + 15L^2 + 10L \qquad \qquad …(5.4)$$

Given the production function (5.4), we may substitute different numerical values for *L* in the function and work out a series of Q_c, i.e., the quantity of coal that can be produced with different number of workers. For example, if $L = 5$, then by substitution, we get

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

A tabular array of output levels associated with different number of workers from 1 to 12, in our hypothetical coal-production example, is given in Table 5.2 (Cols. 1 and 2).

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

Marginal Productivity of Labour (MP_L) can be obtained by differentiating the production function (5.4). Thus,

$$MP_L = \frac{\partial Q}{\partial L} = -3L^2 + 30L + 10 \qquad \dots(5.5)$$

By substituting numerical value for labour (L) in Eq. (5.5), MP_L can be obtained at different levels of labour employment. However, this method can be used only where labour is perfectly divisible and $\partial L \to 0$. Since, in our example, each unit of $L = 1$, calculus method cannot be used.

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

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

The MP_L worked out by this method is presented in Col. 3 of Table 5.2.

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

$$AP_L = \frac{-L^3 + 15L^2 + 10L}{L} = -L^2 + 15L + 10 \qquad ..(5.6)$$

Now AP_L can be obtained by substituting the numerical value for L in Eq. (5.6). AP_L obtained by this method is given in Col. 4 of Table 5.2.

Tables 5.2 *Three Stages of Production*

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

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

The information contained in Table 5.2 is presented graphically in panels (*a*) and (*b*) of Fig. 5.1. Panel (*a*) of Fig. 5.1 presents the total product curve (TP_L) and panel (*b*) presents marginal product (MP_L) and average product (AP_L) curves. The TP_L schedule demonstrates the law of diminishing returns. As the curve TP_L shows, the total output increases at an increasing rate till the employment of the 5th worker, as indicated by the increasing slope of the TP_L curve. (See also Col. 3 of the table). Employment of the 6th worker contributes as much as the 5th worker. Note that beyond the employment of the 6th worker, although TP_L continues to increase (until the 10th worker), the rate of increase in TP_L (i.e., MP_L) begins to fall. This shows the operation of the law of diminishing returns.

The three stages in production. Table 5.2 and Fig. 5.1 present the three usual stages in the application of the laws of diminishing returns.

In **Stage I**, TP_L increases at increasing rate. This is indicated by the rising MP_L till the employment of the 5th and 6th workers. Given the production function (5.4), the 6th worker produces as much as the 5th worker. The output from the 5th and the 6th workers represents an intermediate stage of constant returns to the variable factor, labour.

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

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

5.4.1   Factors Behind the Laws of Returns

As shown in Fig. 5.1, the marginal productivity of labour (MP_L) increases is Stage I, whereas it decreases in Stage II. In other words, in Stage I, Law of Increasing Returns is in operation and in Stage II, the law of Diminishing Returns is in application. The reasons which underly the application of the laws of returns in Stages I and II may be described as follows.

One of the important factors causing increasing returns to a variable factor is the ***indivisibility of fixed factor*** (capital). The minimum size of capital is given as it cannot be divided to suit the number of workers. Therefore, if labour is less than its optimum number, capital remains underutilized. Let us suppose that optimum capital-labour combination is 1:6. If capital is indivisible and less than 6 workers are employed, then capital would remain underutilized. When more and more workers are added, utilization of capital increases and also the productivity of additional worker. The second and the most important reason for increase in labour productivity is the ***division of labour*** that becomes possible with the employment of additional labour, until optimum capital-labour combination is reached.
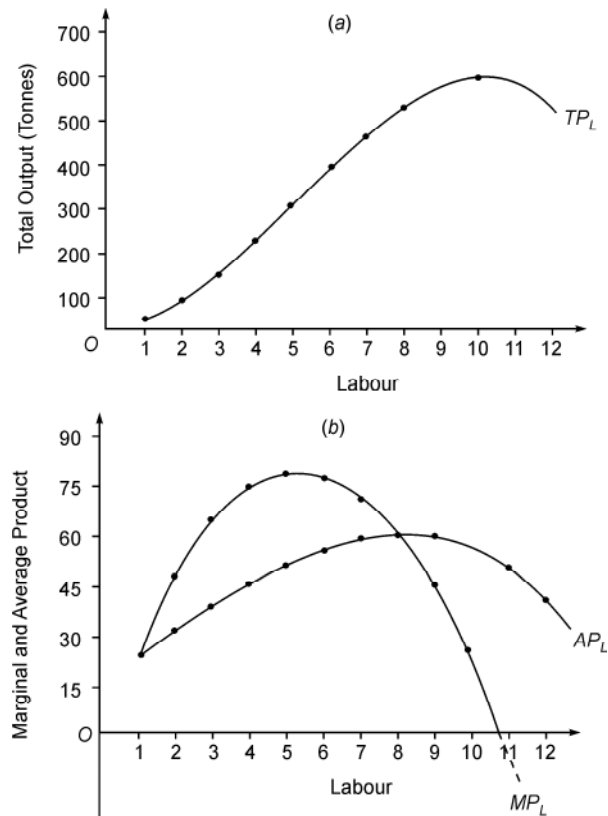
Fig. 5.1 *Total, Average and Marginal Products*

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

5.4.2 Application of the Law of Diminishing Returns

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

The Law of Diminishing Returns and Business Decisions. The law of diminishing returns as presented graphically has a relevance to the business decisions. The graph can help in identifying the rational and irrational stages of operations. It can also tell the business managers the number of workers (or other variable inputs) to apply to a given fixed input so that, given all other factors, output is maximum. As Fig. 5.1 exhibits, capital is presumably underutilized in Stage I. So a firm operating in Stage I is required to increase labour, and a firm operating in Stage II is required to reduce labour, with a view to maximizing its total production. From the firm's point of view, setting an output target in Stages I and II is irrational. The only meaningful and rational stage from the firm's point of view is Stage II in which the firm can find answer to the question 'how many workers to employ'. Figure 5.1 shows that the firm should employ a minimum of 7 workers and a maximum of 10 workers even if labour is available free of cost. This means that the firm has a limited choice—ranging from 7 to 10 workers. How many workers to employ against the fixed capital and how much to produce can be answered, only when the price of labour, i.e., wage rate, and that of the product are known. This question is answered below.

5.4.3 Determining Optimum Employment of Labour

It may be recalled from Fig. 5.1 that an output maximizing coal-mining firm would like to employ 10 workers since at this level of employment, the output is maximum. The firm can, however, employ 10 workers only if workers are available free of cost. But labour is not available free of cost—the firm is required to pay wages to the workers. Therefore, the question arises as to how many workers will the firm employ—10 or less or more than 10—to maximize its profit. A simple answer to this question is that the number of workers to be employed depends on the output that maximizes the firm's profit, given the product price and the wage rate. This point can be proved as follows.

We are aware that profit is maximum where

$$MC = MR$$

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

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

$$MRP = MP_L \times P$$

For example, suppose that the price (*P*) of coal is given at ₹ 10 per quintal. Now, *MRP* of a worker can be known by multiplying its MP_L (as given in Table

5.2) by ` 10. For example, *MRP* of the 3rd worker (see Table 5.2) equals $66 \times 10 = $ ` 660 and of the 4th worker, $78 \times 10 = $ ` 780. Likewise, if the entire column ($MP_L$) is multiplied by ` 10, it gives us a table showing marginal revenue productivity of workers. Let us suppose that wage rate (per time unit) is given at ` 660. Given the wage rate, the profit maximizing firm will employ only 8 workers because at this employment, *MRP* = wage rate = *MRP* of 8th worker; $66 \times 10 = $ ` 660. If the firm employs the 9th worker, his $MRP = 48 \times 10 = $ ` 480 < ` 660. Clearly, the firm loses ` 180 on the 9th worker. And, if the firm employees less than 8 workers, it will not maximize profit.

Graphic Illustration

The process of optimum employment of variable input (labour) is illustrated graphically in Fig. 5.2. When relevant series of *MRP* is graphed, it produces a *MRP* curve like one shown in Fig. 5.2. Similarly, the *MRP* curve for any input may be drawn and compared with *MC* (or *MW*) curve. Labour being the only variable input, in our example, let us suppose that wage rate in the labour market is given at *OW* (Fig. 5.2). When wage rate is constant, average wage (*AW*) equals the marginal wage (*MW*) i.e., *AW* = *MW*, for the entire range of employment in the short-run. When *AW* = *MW*, the supply of labour is shown by a straight horizontal line, as shown by the line *AW* = *MW*.
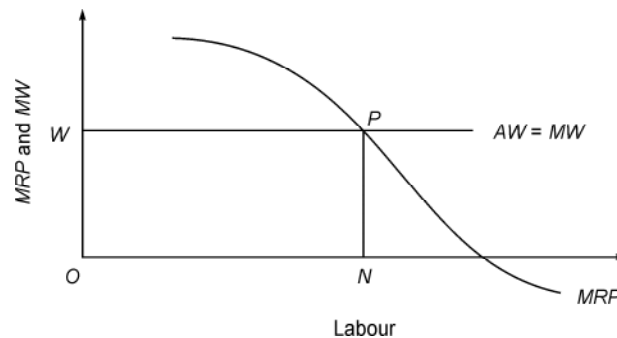
Fig. 5.2 *Determination of Labour Employment in the Short-Run*

With the introduction of *MRP* curve and *AW* = *MW* line (Fig. 5.2), a profit maximizing firm can easily find the maximum number of workers that can be optimally employed against a fixed quantity of capital. Once the maximum number of workers is determined, the optimum quantity of the product is automatically determined.

The marginality principle of profit maximization says that profit is maximum when *MR* = *MC*. This is a necessary condition of profit maximization. Fig. 5.2 shows that *MRP* = *MW* (= *MC*) are equal at point *P*, the point of intersection between *MRP* and *AW* = *MW*. The number of workers corresponding to this point is *ON*. A profit maximizing firm should therefore employ only *ON* workers. Given the number of workers, the total output can be known by multiplying *ON* with average labour productivity (*AP*).

5.5 LONG-TERM LAWS OF PRODUCTION: PRODUCTION WITH TWO VARIABLE INPUTS

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

5.5.1 Isoquant Curves

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

An idea of isoquant can be had from the curve connecting 158 units from 4 different combinations of capital and labour given in Table 5.1.

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

  (*i*) there are only two inputs, viz., labour (*L*) and capital (*K*), to produce a commodity *X*;

 (*ii*) both *L* and *K* and product *X* are perfectly divisible;

(*iii*) the two inputs—*L* and *K*—can substitute each other but at a diminishing rate as they are imperfect substitutes; and

(*iv*) the technology of production is given.
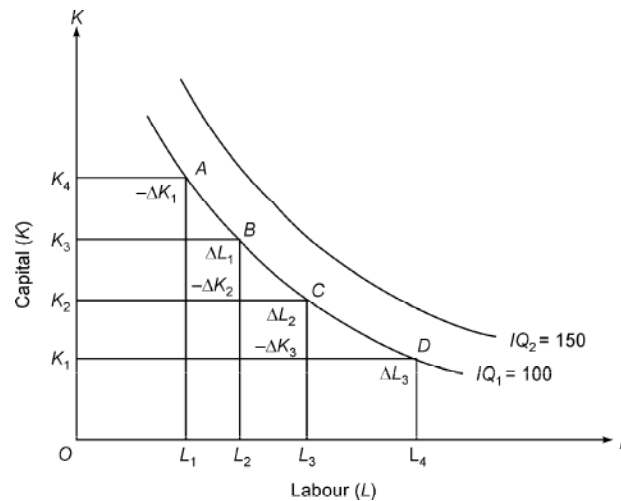
Fig. 5.3 Isoquant Curves

Given these assumptions, it is technically possible to produce a given quantity of commodity *X* with various combinations of capital and labour. The factor combinations are so formed that the substitution of one factor for the other leaves the output unaffected. This technological fact is presented through an isoquant curve ($IQ_1 = 100$) in Fig. 5.3. The curve IQ_1 all along its length represents a fixed quantity, 100 units of product *X*. This quantity of output can be produced with a number of labour-capital combinations. For example, points *A*, *B*, *C*, and *D* on the isoquant IQ_1 show four different combinations of inputs, *K* and *L*, as given in Table 5.3, all yielding the same output—100 units. Note that movement from *A* to *D* indicates decreasing quantity of *K* and increasing number of *L*. This implies substitution of labour for capital such that all the input combinations yield the same quantity of commodity *X*, i.e., $IQ_1 = 100$.

Table 5.3 Capital Labour Combinations and Output

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

5.5.2 Properties of Isoquant Curves

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

(*a*) *Isoquants have a negative slope.* An isoquant has a negative slope in the *economic region* and in the economic range of isoquant. The economic region is

NOTES

the region on the production plane and economic range of isoquant is the range in which substitution between inputs is technically feasible. Economic region is also known as the product maximizing region. The negative slope of the isoquant implies substitutability between the inputs. It means that if one of the inputs is reduced, the other input has to be so increased that the total output remains unaffected. For example, movement from *A* to *B* on IQ_1 (Fig. 5.3) means that if $K_4 K_3$ units of capital are removed from the production process, $L_1 L_2$ units of labour have to be brought in to maintain the same level of output.

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

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

In plain words, *MRTS* is the rate at which a marginal unit of labour can substitute a marginal unit of capital (moving downward on the isoquant) without affecting the total output. This rate is indicated by the slope of the isoquant. The *MRTS* decreases for two reasons: (*i*) no factor is a perfect substitute for another, and (*ii*) inputs are subject to diminishing marginal returns. Therefore, more and more units of an input are needed to replace each successive unit of the other input. For example, suppose various units of *K* (minus sign ignored) in Fig. 5.3 are equal, i.e.,

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

the corresponding units of *L* substituting *K* go (in Fig. 5.3) on increasing, i.e.,

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

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

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

(*c*) *Isoquants are non-intersecting and non-tangential.* The intersection or tangency between any two isoquants implies that a given quantity of a commodity can be produced with a smaller as well as a larger input-combination. This is untenable so long as marginal productivity of inputs is greater than zero. This point can be proved graphically. Note that in Fig. 5.4, two isoquants intersect each other at point *M*. Consider two other points—point *J* on isoquant marked $Q_1 = 100$ and point *K* on isoquant marked $Q_2 = 200$ such that points *K* and *J* fall on a vertical line KL_2, denoting the same amount of labour (OL_2) but different units of capital—KL_2 units of capital at point *K* and JL_2 units of capital at point *J*. Note that point *M* is common to both the isoquants. Given the definition of isoquant,

one can easily infer that a quantity that can be produced with the combination of K and L at point M can be produced also with factor combination at points J and K. On the isoquant $Q_1 = 100$, factor combinations at points M and J yield 100 units of output. And, on the isoquant $Q_2 = 200$, factor combinations at M and K yield 200 units of output. Since point M is common to both the isoquants, it follows that input combinations at J and K are equal in terms of output. This implies that in terms of output,

$$OL_2(L) + JL_2(K) = OL_2(L) + KL_2(K)$$
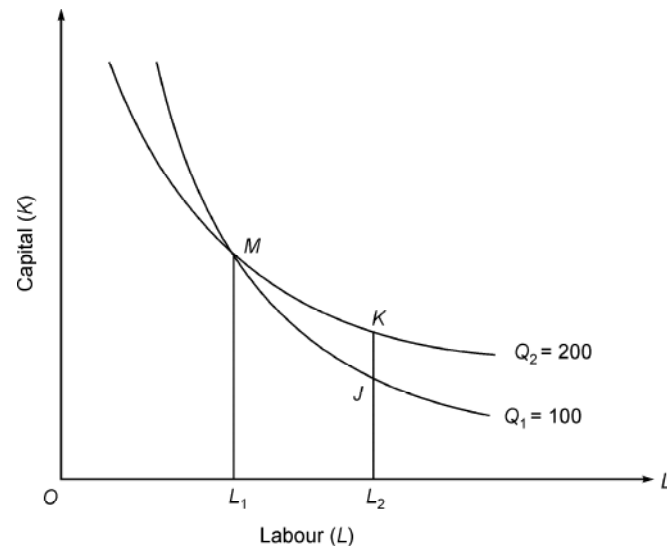
Fig. 5.4 *Intersecting Isoquants*

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

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

But it can be seen in Fig. 5.4 that

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

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

(***d***) ***Upper isoquants represent higher level of output.*** Between any two isoquants, the upper one represents a higher level of output than the lower one. The reason is, an upper isoquant has a larger input combination, which, in general, produces a larger output. Therefore, upper isoquant has a higher level of output.
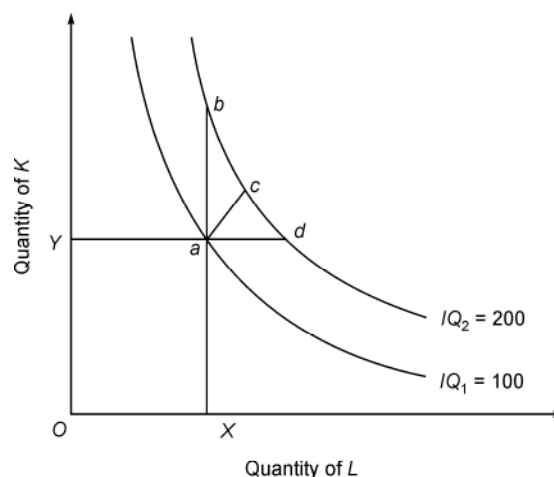
Fig. 5.5 Comparison of Output at Two Isoquants

For instance, IQ_2 in Fig. 5.5 will always indicate a higher level of output than IQ_1. For, any point at IQ_2 consists of more of either capital or labour or both. For example, consider point *a* on IQ_1 and compare it with any point at IQ_2. The point *b* on IQ_2 indicates more of capital (*ab*), point *d* more of labour (*ad*) and point *c* more of both, capital and labour. Therefore, IQ_2 represents a higher level of output (200 units) than IQ_1 indicating 100 units.

5.5.3   Isoquant Map and Economic Region of Production

Isoquant map. An *isoquant map* is a set of isoquants presented on a two-dimensional plane as shown by isoquants Q_1, Q_2, Q_3 and Q_4 in Fig. 5.6. Each isoquant shows various combinations of two inputs that can be used to produce a given quantity of output. An upper isoquant is formed by a greater quantity of one or both the inputs than the input combination indicated by the lower isoquants. For example, isoquant Q_2 indicates a greater input-combination than that shown by isoquant Q_1 and so on.

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

Economic region. Economic region is that area of production plane in which substitution between two inputs is technically feasible without affecting the output. This area is marked by locating the points on the isoquants at which $MRTS = 0$. *A zero MRTS implies that* further substitution between inputs is technically not feasible. It also determines the minimum quantity of an input that must be used to produce a given output. Beyond this point, an additional employment of one input will necessitate employing additional units of the other input. Such a point

on an isoquant may be obtained by drawing a tangent to the isoquant and parallel to the vertical and horizontal axes, as shown by dashed lines in Fig. 5.6. By joining the resulting points *a, b, c* and *d*, we get a line called the *upper ridge line*, *Od*. Similarly, by joining the points *e, f, g* and *h*, we get the *lower ridge line, Oh*. The ridge lines are locus of points on the isoquants where the marginal products (*MP*) of the inputs are equal to zero. The upper ridge line implies that *MP* of capital is zero along the line, *Od*. The *lower ridge line* implies that *MP* of labour is zero along the line, *Oh*.
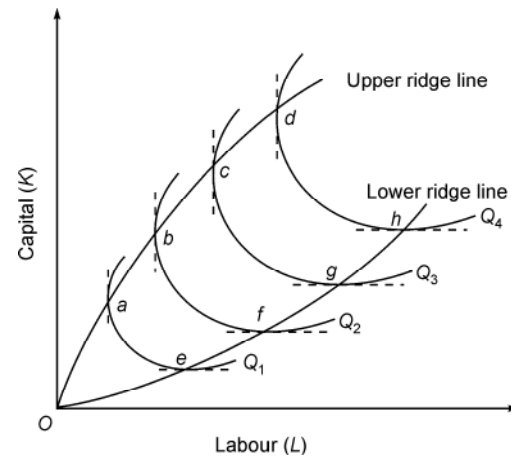
Fig. 5.6 *Isoquant Map*

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

5.5.4   Other Forms of Isoquants

We have introduced above a convex isoquant that is most widely used in traditional economic theory. The shape of an isoquant, however, depends on the degree of substitutability between the factors in the production function. The *convex isoquant* presented in Fig. 5.3 assumes a continuous substitutability between capital and labour but at a diminishing rate. The economists have, however, observed other degrees of substitutability between *K* and *L* and have demonstrated the existence of three other kinds of isoquants.

1.   Linear Isoquants. A linear isoquant is presented by the line *AB* in Fig. 5.7. A linear isoquant implies perfect substitutability between the two inputs, *K* and *L*. The isoquant *AB* indicates that a given quantity of a product can be produced by using only capital or only labour or by using both. This is possible only when the two factors, *K* and *L*, are perfect substitutes for one another. A linear isoquant also implies that the *MRTS* between *K* and *L* remains *constant* throughout.
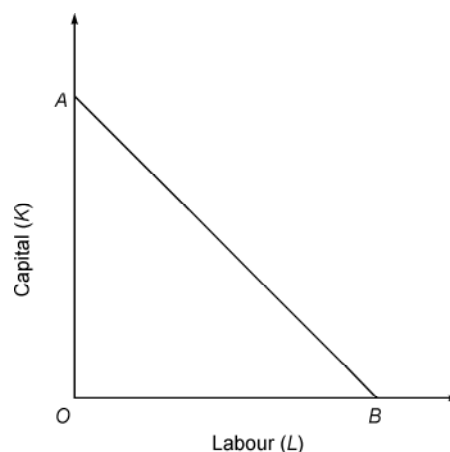
Fig. 5.7 *Linear Isoquant*

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

$$\text{If} \quad Q = f(K, L) \text{ then, } = aK + bL \qquad \qquad ...(5.7)$$

The production function (5.7) means that the total output, Q, is simply the weighted sum of K and L. The slope of the resulting isoquant from this production function is given by $-b/a$. This can be proved as shown below.

Given the production function (5.7),

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

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

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

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

2. Isoquants with Fixed Factor-Proportion or L-Shaped Isoquants. When a production function assumes a fixed proportion between K and L, the isoquant takes 'L' shape, as shown by isoquants Q_1 and Q_2 in Fig. 5.8. Such an isoquant implies zero substitutability between K and L. Instead, it assumes perfect complementarity between K and L. The *perfect complementarity* assumption implies that a given quantity of a commodity can be produced by one and only one combination of K and L and that the proportion of the inputs is fixed. It also implies that if the quantity of an input is increased and the quantity of the other input is held constant, there will be no change in output. The output can be increased only by increasing both the inputs proportionately.
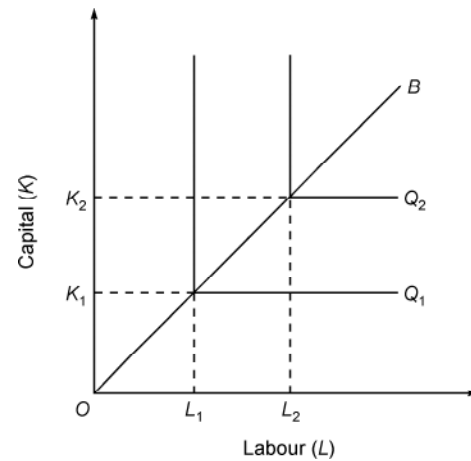
Fig. 5.8 *The L-Shaped Isoquant*

As shown in Fig. 5.8, to produce Q_1 quantity of a product, OK_1 units of K and OL_1 units of L are required. It means that if OK_1 units of K are being used, OL_1 units of labour must be used to produce Q_1 units of a commodity. Similarly, if OL_1 units of labour are employed, OK_1 units of capital must be used to produce Q_1. If units of only K or only L are increased, output will not increase. If output is to be increased to Q_2, K has to be increased by K_1K_2 and labour by L_1L_2. This kind of technological relationship between K and L gives a fixed proportion production function.

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

$$Q = f(K, L) = min \ (aK, bL) \qquad \qquad …(5.8)$$

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

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

3. Kinked Isoquants or Linear Programming Isoquants. The fixed proportion production function (Fig. 5.8) assumes that there is only one technique of production, and capital and labour can be combined only in a fixed proportion. It implies that to double the production, one would require doubling both the inputs, K and L. The line OB (Fig. 5.8) shows that there is only one factor combination for a given level of output. In real life, however, the businessmen and the production engineers find

in existence many, but not infinite, techniques of producing a given quantity of a commodity, each technique having a different fixed proportion of inputs. In fact, there is a wide range of machinery available to produce a commodity. Each machine requires a fixed number of workers to work it. This number varies from machine to machine. For example, 40 persons can be transported from one place to another by two methods: (*i*) by hiring 10 taxis and 10 drivers, or (*ii*) by hiring a bus and 1 driver. Each of these methods is a different process of production and has a different fixed proportion of capital and labour. Handlooms and power looms are other examples of two different factor proportions. One can similarly find many such processes of production in manufacturing industries, each process having a different fixed-factor proportion.

Let us suppose that for producing 10 units of a commodity, *X*, there are four different techniques of production available. Each techniques has a different fixed factor-proportion, as given in Table 5.4.

Table 5.4 *Alternative Techniques of Producing 100 Units of X*

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

The four hypothetical production techniques, as presented in Table 5.4, have been graphically presented in Fig. 5.9. The ray *OA* represents a production process having a fixed factor-proportion of 10*K*:2*L*. Similarly, the other three production processes having fixed capital-labour ratios 6:3, 4:6 and 3:10 have been shown by the rays *OB*, *OC* and *OD* respectively. Points *A*, *B*, *C* and *D* represent four different production techniques. By joining the points, *A*, *B*, *C* and *D*, we get a **kinked isoquant**, *ABCD*.
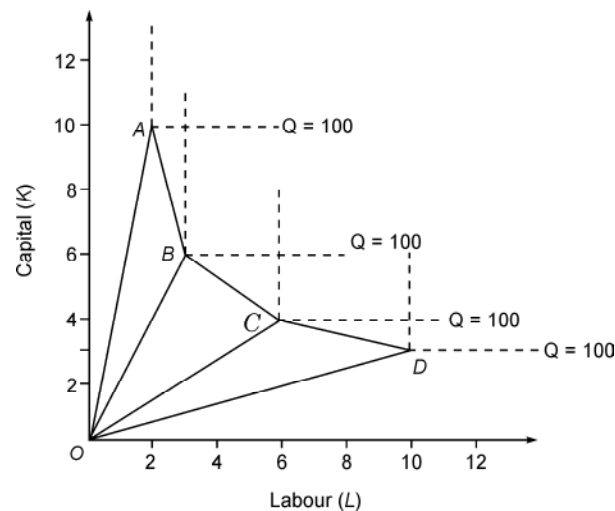
Fig. 5.9 *Fixed Proportion Techniques of Production*

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

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

5.6 ELASTICITY OF FACTOR SUBSTITUTION

We have noted that *MRTS* decreases along the isoquant. *MRTS* refers only to the slope of an isoquant, i.e., the ratio of marginal changes in inputs. It does not reveal the substitutability of one input for another—labour for capital—with changing combination of inputs.

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

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

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

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

The concept of elasticity of factor substitution is graphically presented in Fig. 5.10. The movement from point *A* to *B* on the isoquant *IQ*, gives the ratio of change in *MRTS*. The rays *OA* and *OB* represent two techniques of production with different factor intensities. While line *OA* indicates capital intensive technique, line *OB* indicates labour intensive technique. The shift from *OA* to *OB* gives the change in factor intensity. The ratio between the two factor intensities measures the *substitution elasticity*.
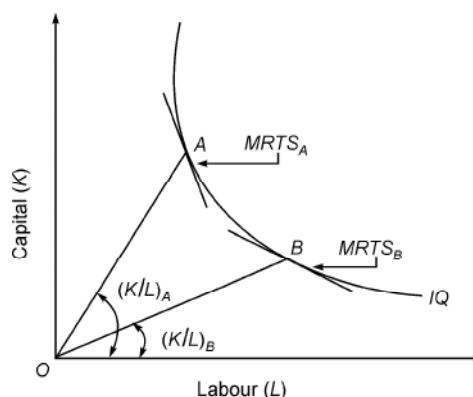
Fig. 5.10 *Graphic Derivation of Elasticity of Substitution*

The value of substitution elasticity depends on the curvature of the isoquants. It varies between 0 and ∞, depending on the nature of production function. It is, in fact, the production function that determines the curvature of the various kinds of isoquants. For example, in case of fixed-proportion production function [see Eq. (5.8)] yielding an *L*-shaped isoquant, $\sigma = 0$. If production function is such that the resulting isoquant is linear (see Fig. 5.7), $\sigma = \infty$. And, in case of a homogeneous production function of degree 1 of the Cobb-Douglas type, $\sigma = 1$.

5.7   LAWS OF RETURNS TO SCALE

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

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

  (*i*) total output may increase more than proportionately;

  (*ii*) total output may increase proportionately; and

  (*iii*) total output may increase less than proportionately.

Accordingly, there are three kinds of returns to scale:

  (*i*) Increasing returns to scale;

  (*ii*) Constant returns to scale, and

  (*iii*) Diminishing returns to scale.

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

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

1. *Increasing Returns to Scale*

When inputs, *K* and *L*, are increased at a certain proportion and output increases more than proportionately, it exhibits *increasing returns to scale.* For example, if quantities of both the inputs, *K* and *L*, are successively doubled and the resultant output is more than doubled, the returns to scale is said to be increasing. The increasing returns to scale is illustrated in Fig. 5.11. The movement from point *a* to *b* on the line *OB* means doubling the inputs. It can be seen in Fig. 5.11 that input-combination increases from $1K + 1L$ to $2K + 2L$. As a result of doubling the inputs, output is more than doubled: it increases from 10 to 25 units, i.e., an increase of 150%. Similarly, the movement from point *b* to point *c* indicates 50% increase in inputs as a result of which the output increases from 25 units to 50 units, i.e., by 100%. Clearly, output increases more than the proportionate increase in inputs. This kind of relationship between the inputs and output shows *increasing returns to scale*.
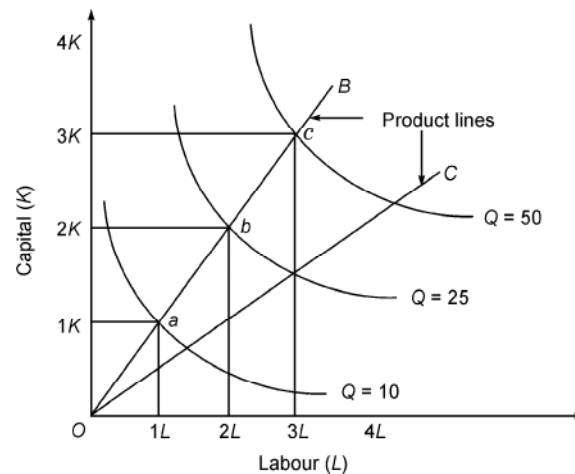
***Fig. 5.11** Increasing Returns to Scale*

5.7.1   The Factors Behind Increasing Returns to Scale

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

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

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

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

2. *Constant Returns to Scale*

When the increase in output is proportionate to the increase in inputs, it exhibits *constant returns to scale.* For example, if quantities of both the inputs, K and L, are doubled and output is also doubled, then the returns to scale are said to be constant. Constant returns to scale are illustrated in Fig. 5.12. The lines *OA* and *OB* are 'product lines' indicating two hypothetical techniques of production with optimum capital-labour ratio. The isoquants marked $Q = 10$, $Q = 20$ and $Q = 30$ indicate the three different levels of output. In the figure, the movement from points *a* to *b* indicates doubling both the inputs. When inputs are doubled, output is also doubled, i.e., output increases from 10 to 20. Similarly, the movement from *a* to *c* indicates trebling inputs—*K* increase to 3*K* and *L* to 3*L*. This leads to trebling the output—from 10 to 30.
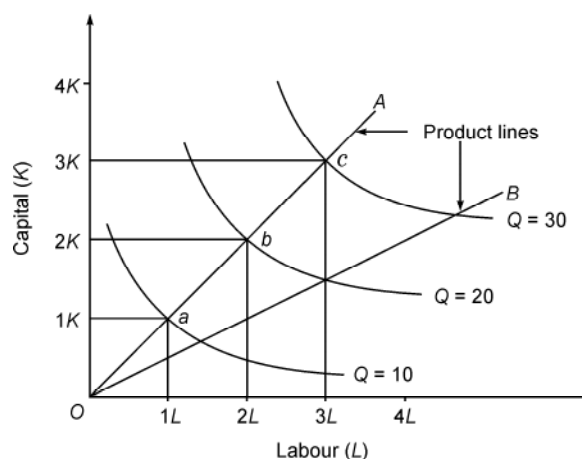
Fig. 5.12 Constant Returns to Scale

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

$$1K + 1L \Rightarrow 10$$
$$2K + 2L \Rightarrow 20$$
$$3K + 3L \Rightarrow 30$$

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

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

3. *Decreasing Returns to Scale*

The firms are faced with *decreasing returns to scale* when a certain proportionate increase in inputs, *K* and *L*, leads to a less than proportionate increase in output. For example, when inputs are doubled and output is less than doubled, then decreasing returns to scale is in operation. The decreasing returns to scale is illustrated in Fig. 5.13. As the figure shows, when the inputs *K* and *L* are doubled, i.e., when capital-labour combination is increased from $1K + 1L$ to $2K + 2L$, the output increases from 10 to 18 units. This means that when capital and labour are increased by 100 per cent, output increases by only 80 per cent. That is, increasing output is less that the proportionate increase in inputs. Similarly, movement from point *b* to *c* indicates a 50 per cent increase in the inputs. But, the output increases by only 33.3 per cent. This exhibits *decreasing returns to scale*.
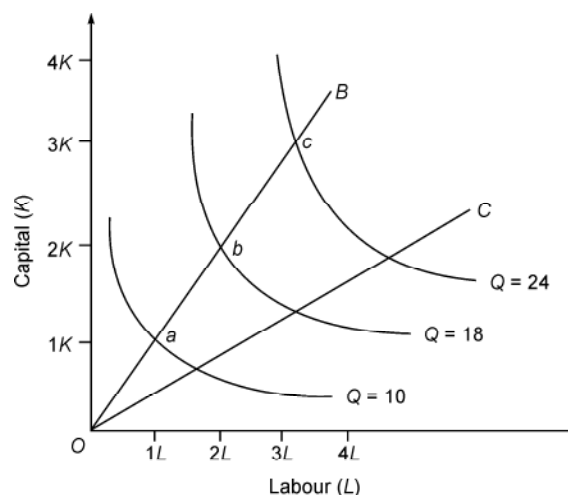
Fig. 5.13 Decreasing Return to Scale

Causes of Diminishing Return to Scale

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

Check Your Progress

1. What is a variable input?

2. Name the kinds of productions function.

3. What is an isoquant curve?

5.8 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

---

1. A variable input is defined as one whose supply in the short-run is elastic, for example, labour and raw material and so forth.

2. There are two kinds of production functions, namely:

    (i) Short-run production function

    (ii) Long-run production function

3. An isoquant curve can be defined as the locus of points representing various combinations of two inputs—capital and labour—yielding the same output.

5.9 SUMMARY

- In general sense of the term 'production' means transforming inputs (labour, capital, raw materials, time, etc.) into an output. This concept of production is however limited to only 'manufacturing'.

- Production process does not necessarily involve physical conversion of raw materials into tangible goods. Some kinds of production involve an intangible input to produce an intangible output.

- Inputs are classified as (*i*) *fixed inputs* and (*ii*) *variable inputs*. Fixed and variable inputs are defined in economic sense and also in technical sense.

- Production function is a mathematical presentation of input-output relationship. More specifically, a production function states the technological relationship.

- The laws of production state the relationship between output and input. In the short-run, input-output relations are studied with one variable input (labour), other inputs (especially, capital) held constant. The laws of production under these conditions are called the 'Laws of Variable Proportions' or the 'Laws of Returns.

- The law of diminishing returns is *an empirical law,* frequently observed in various production activities. This law, however, may not apply universally to all kinds of productive activities since it is not as true as the law of gravitation.

- The term 'isoquant' has been derived from the Greek word *iso* meaning 'equal' and Latin word *quantus* meaning 'quantity'. The 'isoquant curve' is, therefore, also known as 'Equal Product Curve' or 'Production Indifference Curve'.

- In the isoquant map, each upper isoquant indicates a larger input-combination than the lower ones, and each successive upper isoquant indicates a higher level of output than the lower ones.

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

5.10 KEY WORDS

- **Input:** An input is simply anything which the firm buys for use in its production or other processes.
- **Economic Region:** In economics, it is that area of production plane in which substitution between two inputs is technically feasible without affecting the output.

5.11 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short Answer Questions

1. Define the production function.
2. Identify the factors of production.
3. Write a short note on the elasticity of factor substitution.

Long Answer Questions

1. Describe the underlying assumptions of the production function.
2. Discuss the law of variable proportion.
3. Explain the law of returns to scale with diagrams.

5.12 FURTHER READINGS

Dwivedi, D. N. 2008. *Principles of Economics*, Seventh Edition. New Delhi: Vikas Publishing House.

Weil. David N. 2004. *Economic Growth*. London: Addison Wesley.

Thomas, Christopher R. and Maurice S. Charles. 2005. *Managerial Economics: Concepts and Applications*, Eighth Edition. New Delhi: Tata McGraw-Hill Publishing Company Limited.

Mankiw, Gregory N. 2002. *Principles of Economics*, Second Edition. India: Thomson Press.

UNIT 6   COST CONCEPTS - I

Structure

6.0   INTRODUCTION

Business decisions are generally taken on the basis of money value of the inputs and outputs. The total money spent on acquiring the inputs—labour, capital, materials and technology—gives the measure of the total *cost of production*. Money received from the sale of output gives the total revenue. In this unit, we are concerned with the theory of cost, i.e., the laws pertaining to cost-output relationship.

The cost of production is an important factor in almost all business analysis and decisions, specially those which pertain to the following aspects.

(*a*) locating the weak points in production management;

(*b*) minimizing the cost as far as possible, for a given output;

(*c*) finding the optimum level of output;

(*d*) determination of price and dealers margin; and

(*e*) estimating or projecting the cost of business operation.

Besides, cost analysis assumes a great significance in all major business decisions because the term 'cost' has different meaning under different settings and is subject to varying interpretations. It is, therefore, essential that only the relevant concept of costs is used in the business decisions. This unit deals with these aspects of cost analysis.

6.1 OBJECTIVES

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

- Define the cost function
- Describe the economies and diseconomies of scale
- Differentiate between explicit cost and implicit cost

6.2 CONCEPT OF COST

The cost concepts which are relevant to business analysis and decision-making can be grouped, on the basis of their nature and purpose, under two overlapping categories: (*i*) concepts used for accounting purposes, and (*ii*) analytical cost concepts used in economic analysis of business activities. We will discuss here some important concepts of the two categories. It is important to note here that this classification of cost concepts is only a matter of analytical convenience.

6.2.1 Accounting Cost Concepts

1. *Opportunity Cost and Actual Cost:* The concept of opportunity cost is related to the alternative uses of scarce resources. As noted earlier, resources, both natural and man-made, are *scarce*. They are scarce in relation to demand for them to satisfy the ever growing human needs. Resources, though scarce, have *alternative uses*. The *scarcity* and the *alternative uses* of the resources give rise to the concept of *opportunity cost*. Resources available to a business unit—be it an individual firm, a joint stock corporation or a multinational— are limited. But the limited resources available to a firm can be put to alternative uses.

For example, suppose a firm has ` 100 million at its disposal and the firm finds three risk-free alternative uses of the fund available to it: (*i*) to expand the size of the firm, (*ii*) to set up a new production unit in another city and (*iii*) to buy shares in another firm. Suppose also that the expected annual return from each of the three alternative uses of finance is given as follows.

Alternative 1 : Expansion of the size of the firm ` 20 million

Alternative 2 : Setting up a new production unit ` 18 million

Alternative 3 : Buying shares in another firm ` 16 million

All other things being the same, a rational decision for the firm would be to invest the money in alternative 1. This implies that the manager would have to sacrifice the annual return from the second best alternative, i.e., a return of ` 18 million expected from alternative 2. In economic jargon, ` 18 million is called an annual *opportunity cost* of an annual income of ` 20 million. Thus, *the opportunity cost of availing an opportunity is the foregone income expected from the second best opportunity of using the resources*. In our example, the opportunity cost of ` 20 million per annum is ` 18 million per annum.

The difference between *actual earning* and its *opportunity cost* is called *economic gain* or *economic profit*. The concept of opportunity cost assumes a great significance where economic gain is neither insignificant nor very large because then it requires a careful evaluation of the two options.

The applicability of the opportunity cost concept is not limited to decisions involving finances. The concept can be applied to all other kinds of resources involved in business decisions, especially where there are at least two alternative options involving costs and benefits. For example, suppose a firm has to take a decision on whether to fire an efficient labour officer (for treating labour unkindly) in settlement of a dispute with the labour union or to allow the matter to be taken to the labour court. If the firm decides to fire the labour officer, then the loss of an efficient, reliable labour officer is the *opportunity cost* of buying peace with the labour union. If the firm decides to retain the labour officer, come whatever may, then the cost of prolonged litigation, the cost arising out of a possible labour strike and the consequent reduction in output are the *opportunity costs* of retaining the labour officer. Given the two options, the firm will have to evaluate the cost and benefit of each option and take a decision accordingly.

2. *Business Costs and Full Costs:* *Business costs* include all the expenses which are incurred to carry out a business. The concept of business costs is similar to the actual or real costs. Business costs "include all the payments and contractual obligations made by the firm together with the book cost of depreciation on plant and equipment." These cost concepts are used for calculating business profits and losses and for filing returns for income-tax and also for other legal purposes.

The concept of *full cost,* includes business costs, opportunity cost and normal profit. The opportunity cost includes the expected earning from the second best use of the resources, or the market rate of interest on the total financial capital and also the value of an entrepreneur's own services which are not charged for in the current business. Normal profit is a necessary minimum earning in addition to the opportunity cost, which a firm must receive to remain in its present occupation.

3. *Explicit and Implicit or Imputed Costs:* *Explicit costs* are those which fall under actual or business costs entered in the books of accounts. The payments on account of wages and salaries, materials, license fee, insurance premium, depreciation charges are the examples of explicit costs. These costs involve cash payment and are recorded in normal accounting practices.

In contrast to explicit costs, there are certain other costs which do not take the form of cash outlays, nor do they appear in the accounting system. Such costs are known as *implicit* or *imputed costs*. Opportunity cost is an important example of implicit cost. For example, suppose an entrepreneur does not utilize his services in his own business and works as a manager in some other firm on a salary basis. If he sets up his own business, he foregoes his salary as manager. This loss of salary is the opportunity cost of income from his own business. This is an implicit cost of his own business. Thus, implicit wages, rent, and implicit interest are the

wages, rents and interest which an owner's labour, building and capital, respectively, can earn from their second best use.

Implicit costs are not taken into account while calculating the loss or gains of the business, but they do appear as an important consideration in whether or not to retain a factor in its present use. The explicit and implicit costs together make the *economic cost*.

4. *Out-of-Pocket and Book Costs:* The items of expenditure which involve cash payments including, both recurring and non-recurring expenses, are known as *out-of-pocket costs*. All the explicit costs (e.g., wages, rent, interest, cost of materials and maintenance, transport expenditure, etc.) fall in this category. On the contrary, there are certain actual business costs which do not involve cash payments, but a provision is therefore made in the books of account and they are taken into account while finalising the profit and loss accounts. For example, payments made by a firm to itself, depreciation allowances and unpaid interest on the owner's own fund are the example of *book costs*.

6.2.2 Analytical Cost Concepts

This covers the following costs:

1. *Fixed and Variable Costs:* These costs are covered under the category of analytical cost concept. It is to be noted that fixed and variable costs have been discussed in Unit 8 of this book.

2. *Total, Average and Marginal Costs: Total Cost* (*TC*) is the total expenditure incurred on the production of goods and service. It refers to the total outlays of money expenditure, both explicit and implicit, on the inputs used to produce a given level of output. It includes both fixed and variable costs. That is,

$$TC = TFC + TVC$$

Average Cost (*AC*) is of statistical nature—it is not actual cost. It is obtained by dividing the total cost (*TC*) by the total output (*Q*), i.e.,

$$AC = \frac{TC}{Q} \qquad \qquad \dots(6.1)$$

It is to be noted that the concept of marginal cost has been discussed in Unit 7 and Unit 8 of this book.

3. *Short-Run and Long-Run Costs:* Short-run and long-run cost concepts are related to variable and fixed costs, respectively, and often figure in economic analysis interchangeably.

Short-run costs are the costs which have short-run implication in production process and vary with the variation in output, the size of the firm remaining the same. *Short-run* costs are the same as *variable costs*. *Long-run* costs, on the other hand, are the costs which are incurred on the fixed assets like plant, building,

machinery, etc. It is important to note that the running cost and depreciation of the capital assets are included in the short-run or variable costs.

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 variables in the utilization of fixed plant or other facilities whereas long-run costs are associated with the changes in the size and kind of plant.'

4. *Incremental Costs and Sunk Costs:* Conceptually, *incremental costs* are closely related to the concept of marginal cost but with a relatively wider connotation. While marginal cost refers to the cost of the marginal unit of output, incremental cost refers to the total additional cost associated with the decisions to expand the output or to add a new variety of product, etc. The concept of incremental cost is based on the fact that in the real world, it is not practicable (for lack of perfect divisibility of inputs) to employ factors for each unit of output separately. Besides, in the long run, when firms expand their production, they hire more of men, materials, machinery and equipments. The expenditures of this nature are incremental costs and not the marginal cost (as defined earlier). Incremental costs arise also owing to the change in product lines, addition or introduction of a new product, replacement of worn out plant and machinery, replacement of old technique of production with a new one, etc.

The *sunk costs* are those which are incurred once for all. Such costs cannot be altered, increased or decreased, by varying the rate of output. For example, once it is decided to make incremental investment expenditure and the funds are allocated and spent, all the preceding costs are considered to be the sunk costs since they accord to the prior commitment and cannot be revised or reversed or recovered when there is a change in market conditions or change in business decisions.

5. *Historical and Replacement Costs:* *Historical cost* refers to the cost of an asset acquired in the past whereas *replacement cost* refers to the expenditure made for replacing an old asset. These concepts owe their significance to the unstable nature of price behaviour. Stable prices over time, other things given, keep historical and replacement costs on par with each other. Instability in asset prices makes the two costs differ from each other.

Historical *cost* of assets is used for accounting purposes, in the assessment of the net worth of the firm. The replacement cost figures in business decisions regarding the renovation of the plant.

6. *Private and Social Costs:* We have so far discussed the cost concepts that are related to the working of the firm and that are used in the cost-benefit analysis of business decisions. There are, however, certain other costs which arise due to the functioning of the firm but do not normally figure in the business decisions nor are such costs explicitly borne by the firms. The costs on this category are borne

by the society. Thus, the total cost generated by a firm's working may be divided into two categories: (*i*) costs paid out or provided for by the firms, and (*ii*) costs not paid or borne by the firms including the use of resources freely available plus the disutility created in the process of production. The costs of the former category are known as *private costs* and costs of the latter category are known as *external* or *social costs*. To mention a few examples of social cost: Mathura Oil Refinery discharging its wastage in the Yamuna river causes water pollution. Mills and factories located in a city cause air pollution, and so on. Let us now look at these concepts of cost in some detail.

Private costs are those which are actually incurred or provided for by an individual or a firm on the purchase of goods and services from the market. For a firm, all the actual costs both explicit and implicit are private costs. Private costs are internalized costs that are incorporated in the firm's total cost of production.

Social costs, on the other hand, refer to the total cost borne by the society due to production of a commodity. Social cost includes both private cost and the external cost. Social cost includes (*a*) the cost of resources for which the firm is not required to pay a price, i.e., atmosphere, rivers, lakes, and also for the use of public utility services like roadways, drainage system, etc., and (*b*) the cost in the form of 'disutility' created through air, water and noise pollution, etc. The costs of category (*c*) are generally assumed to equal the total private and public expenditure incurred to safeguard the individual and public interest against the various kinds of health hazards created by the production system. The private and public expenditure, however, serve only as an indicator of 'public disutility'—they do not give the exact measure of the public disutility or the social costs.

6.3 COST FUNCTION

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

$$TC = f(Q)$$

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

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

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

Kinds of Cost Functions and Change in TC

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

6.4 ECONOMIES AND DISECONOMIES OF SCALE

LAC decreases with the expansion of production scale upto OQ_2 and then it begins to rise. This behaviour of *LAC* is caused by the economies and diseconomies of scale. Economies of scale result in cost saving and diseconomies lead to rise in cost. *Economies and diseconomies of scale* determine also the returns to scale. Increasing returns to scale operate till economies of scale are greater than the diseconomies of scale, and returns to scale decrease when diseconomies are greater than the economies of scale. When economies and diseconomies are in balance, returns to scale are constant. In this section, we will briefly discuss the various kinds of economies and diseconomies of scale.

Economies of Scale

The economies of scale are classified as

- (*a*) Internal or Real Economies, and
- (*b*) External or Pecuniary Economies.

6.4.1 Internal Economies

Internal economies, also called 'real economies', are those which arise from the expansion of the plant-size of the firm and are internalized. This means that internal economies are exclusively available to the expanding firm.

Internal economies may be classified under the following categories.

   (*i*) Economies in production;

  (*ii*) Economies in marketing;

 (*iii*) Managerial economies; and

 (*iv*) Economies in transport and storage.

(*i*) **Economies in Production:** Economies in production arise from two sources: (*a*) technological advantages, and (*b*) advantages of division of labour and specialization.

Technological advantages: Large-scale production provides an opportunity to avail the advantages of technological advances. Modern technology is highly specialized. The advanced technology makes it possible to conceive the whole process of production of a commodity in one composite unit of production. For example, production of cloth in a textile mill may comprise such plants as (*i*) spinning; (*ii*) weaving; (*iii*) printing and pressing; and (*iv*) packing, etc. A composite dairy scheme may consist of plants like (*i*) chilling; (*ii*) milk processing; and (*iii*) bottling. Under small-scale production, the firm may not find it economical to have all the plants under one roof. It would, therefore, not be in a position to have the full advantage of a composite technology. But, when scale of production expands and firms hire more capital and labour, their total output increases more than proportionately till the optimum size of the firm is reached. It results in lower cost of production.

Advantages of division of labour and specialization: When a firm's scale of production expands, more and more workers of varying skills and qualifications are employed. With the employment of larger number of workers, it becomes increasingly possible to divide the labour according to their qualifications and skills and to place them in the process of production where they are best suited. This is known as division of labour. *Division of labour* leads to specialization. It increases productivity of labour and, thereby, reduces cost of production. Besides, specialized workers develop more efficient tools and techniques and gain speed of work. These advantages of division of labour improve productivity of labour– per unit of cost and time.

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

The economies in marketing the firm's own product are associated with (*a*) economies in advertisement cost; (*b*) economies in large-scale distribution through wholesalers, etc.; and (*c*) other large-sale economies. With the expansion of the firm, the total production increases. But the expenditure on advertising the

product does not increase proportionately. Similarly, selling through the wholesale dealers reduces the cost on distribution of the firm's production. The firm also gains on large scale distribution through better utilization of 'sales force, distribution of sample, etc.' This kind of economy however does not directly effect the production conditions. However, it proves to be a promotional factor even under rising cost.

(*iii*) **Managerial Economies:** Managerial economies arise from (*a*) specialization in management, and (*b*) mechanization of managerial functions. For a large-size firm, it becomes possible to divide its management into specialized departments under specialized personnel, such as production manager, sales manager, personnel or HR manager, labour officers, etc. This increases the efficiency of management at all the levels of management because of the decentralization of decision-making. Large-scale firms have the opportunity to use advanced techniques of communication, telephones and telex machines, inter-com, computers, and their own means of transport. All these lead to quick decision-making, help in saving valuable time of the management and, thereby, improve the managerial efficiency. For these reasons, managerial cost increases less than proportionately with the increase in production scale upto a certain level.

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

6.4.2 External or Pecuniary Economies of Scale

External or pecuniary economies accrue to the expanding firms from the advantages arising outside the firm, e.g., in the input markets. Pecuniary economies accrue to the large-size firms in the form of discounts and concessions on (*i*) large scale purchase of raw material, (*ii*) large scale acquisition of external finance, particularly from the commercial banks; (*iii*) massive advertisement campaigns; (*iv*) large scale hiring of means of transport and warehouses, etc. These benefits are available to all the firms of an industry—they are not specific to any one particular firm.

Besides, expansion of an industry invites and encourages the growth of ancillary industries which supply inputs. In the initial stages, such industries also enjoy the increasing returns to scale. In a competitive market, therefore, input

prices go down. This benefit accrues to the expanding firms in addition to discounts and concessions. For example, growth of the automobile industry helps the development of tyre industry and other motor parts. If Maruti Udyog Limited starts producing tyres for its own cars and ancillaries, its cost of production may be higher because of a small scale of production. Consider another example, growth of fishing industry encourages growth of firms that manufacture and supply fishing nets and boats. Competition between such firms and law of increasing returns at least in the initial stages, reduces the cost of inputs. Reduction in input costs is an important aspect of external economies.

Diseconomies of Scales

Diseconomies of scale are disadvantages that arise due to the expansion of production scale and lead to a rise in the cost of production. Like economies, diseconomies may be internal and external. Internal diseconomies are those which are exclusive and internal to a firm – they arise within the firm. External diseconomies arise outside the firms, mainly in the input markets. Let us describe the nature of internal and external diseconomies in some detail.

1. Internal Diseconomies

Internal diseconomies begin to arise when the advantages of division of labour and managerial staff have been fully exploited and excess capacity of plant, warehouses, transport and communication systems, etc. is fully used. Although some economies may still exist, diseconomies begin to overweigh the economies and the costs begin to rise. Internal diseconomies arise due to the following reasons.

(*a*) *Managerial inefficiency:* Diseconomies begin to appear first at the management level. Managerial inefficiencies arise, among other things, from the expansion of scale itself. With fast expansion of the production scale, personal contacts and communications between (*i*) owners and managers, (*ii*) managers and labour, and (iii) inter departmental linkages; get rapidly reduced. Close control and supervision is replaced by remote control management. With the increase in managerial personnel, decision-making becomes complex leading to lack of coordination and implementation delays. Besides, with the expansion of the scale of production, management is professionalised beyond a point. As a result, the owner's objective function of profit maximization is gradually replaced by managers' utility function, like job security and high salary, standard or reasonable profit target, satisfying functions. All these lead to laxity in management and, hence to a rise in the cost of production.

(*b*) *Labour Inefficiency***:** Another source of internal diseconomy is the overcrowding of labour leading to a loss of control over labour productivity. On the other hand, increase in the number of workers encourages formation of labour unions and is supported by the labour laws of the country. Formation of labour unions promotes labour union activities which take the form of labour strikes and

lock-out by the firms. Labour strikes and lock-outs simply means the loss of output per unit of time and hence, rise in the cost of production.

2. External Diseconomies

External diseconomies are the disadvantages that originate outside the firm especially in the input markets. External diseconomies arise also due to natural constraints, specially in agriculture and extractive industries. In case of manufacturing industries, the expansion of the firm, particularly when all the firms of the industry are expanding, the discounts and concessions that are available on bulk purchases of inputs and concessional finance come to an end. More than that, increasing demand for inputs puts pressure on the input markets and input prices begin to rise causing a rise in the cost of production. Such diseconomies are called *pecuniary diseconomies*.

On the production side, the law of diminishing returns to scale come into force due to excessive use of fixed factors, more so in agriculture and extractive industries. For example, excessive use of cultivable land turns it into barren land; pumping out water on a large scale for irrigation causes the water table to go down resulting in rise in cost of irrigation; extraction of minerals on a large scale soon exhausts the mineral deposits on upper levels and mining further deep causes rise in cost of production; extensive fishing reduces the availability of fish and the catch, even when fishing boats and nets are increased. These kinds of diseconomies make the *LAC* curve move upward.

Check Your Progress

1. What is short run cost?
2. What are internal economies?
3. Give one example of implicit cost.

6.5 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. The short-run cost is the cost which has short-term implications in the production process, i.e. these are used over a short range of output. These are the cost incurred once and cannot be used again and again, such as payment of wages, cost of raw materials and so forth.

2. Internal economies, also called 'real economies', are those which arise from the expansion of the plant-size of the firm and are internalized. This means that internal economies are exclusively available to the expanding firm.

3. Opportunity cost is one example of implicit cost.

6.6   SUMMARY

- The cost concepts which are relevant to business analysis and decision-making can be grouped, on the basis of their nature and purpose, under two overlapping categories: (*i*) concepts used for accounting purposes, and (*ii*) analytical cost concepts used in economic analysis of business activities.

- Business costs include all the expenses which are incurred to carry out a business. The concept of business costs is similar to the actual or real costs.

- Explicit costs are those which fall under actual or business costs entered in the books of accounts. The payments on account of wages and salaries, materials, license fee, insurance premium, depreciation charges are the examples of explicit costs.

- Conceptually, incremental costs are closely related to the concept of marginal cost but with a relatively wider connotation.

- Social cost includes (*a*) the cost of resources for which the firm is not required to pay a price, i.e., atmosphere, rivers, lakes, and also for the use of public utility services like roadways, drainage system, etc., and (*b*) the cost in the form of 'disutility' created through air, water and noise pollution, etc.

- Cost function is a symbolic statement of the technological relationship between cost and output. In its general form, it is expressed by an equation.

- Economies of scale result in cost saving and diseconomies lead to rise in cost. Economies and diseconomies of scale determine also the returns to scale. Increasing returns to scale operate till economies of scale are greater than the diseconomies of scale, and returns to scale decrease when diseconomies are greater than the economies of scale.

- Internal economies, also called 'real economies', are those which arise from the expansion of the plant-size of the firm and are internalized.

- Managerial economies arise from (*a*) specialization in management, and (*b*) mechanization of managerial functions.

- Diseconomies of scale are disadvantages that arise due to the expansion of production scale and lead to a rise in the cost of production.

6.7   KEY WORDS

- **Private Cost:** These are those which are actually incurred or provided for by an individual or a firm on the purchase of goods and services from the market.

- **Explicit Cost:** These are those which fall under actual or business costs entered in the books of accounts.

6.8 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short Answer Questions

1. Define the cost function.
2. What is social cost?
3. Write a short note on the significance of cost analysis.

Long Answer Questions

1. Discuss the various categories of internal economies.
2. Differentiate between explicit cost and implicit cost.
3. 'Reduction in input costs is an important aspect of external economies.' Explain the statement.

6.9 FURTHER READINGS

Dwivedi, D. N. 2008. *Principles of Economics*, Seventh Edition. New Delhi: Vikas Publishing House.

Weil. David N. 2004. *Economic Growth*. London: Addison Wesley.

Thomas, Christopher R. and Maurice S. Charles. 2005. *Managerial Economics: Concepts and Applications*, Eighth Edition. New Delhi: Tata McGraw-Hill Publishing Company Limited.

Mankiw, Gregory N. 2002. *Principles of Economics*, Second Edition. India: Thomson Press.

UNIT 7   COST CONCEPTS - II

Structure

7.0   INTRODUCTION

The previous unit introduced you to the various cost concepts. This unit will introduce you to the concept of marginal revenue and marginal cost, optimum firm, opportunity cost and real cost. Marginal revenue and marginal cost are vital calculations that assist companies in analyzing and maximize their profits. Put together, marginal revenue and marginal cost are used to determine how many units of a given product or service a company should produce, as well as the price per unit.

7.1   OBJECTIVES

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

- Discuss the relationship between marginal revenue and marginal cost
- Define optimum firm
- Differentiate between opportunity cost and real cost

7.2   MARGINAL REVENUE AND MARGINAL COST

Profit Maximizing Conditions

Total profit (Π) is defined as

$$\Pi = TR - TC \qquad \qquad \text{…(7.1)}$$

where TR = total revenue, and TC = total cost.

There are two conditions that must be fulfilled for $TR - TC$ to be maximum. These conditions are called (*i*) necessary or the first order condition, and

(*ii*) secondary or supplementary condition.

The necessary or the first-order condition: It requires that marginal revenue (*MR*) must be equal to marginal cost (*MC*). By definition, marginal revenue is the revenue obtained from the production and sale of one additional unit of output and marginal cost is the cost arising due to the production of one additional unit of output.

The secondary or the second-order condition: It requires that the necessary or first-order condition must be satisfied under the stipulation of decreasing *MR* and rising *MC*. The fulfilment of the two conditions makes it the *sufficient condition*.

The profit maximizing conditions can also be presented algebraically as follows.

We know that a profit maximizing firm seeks to maximize

$$\Pi = TR - TC$$

Let us suppose that the total revenue (*TR*) and total cost (*TC*) functions are, respectively, given as

$$TR = f(Q) \qquad \text{and} \quad TC = f(Q)$$

where Q = quantity produced and sold.

By substituting total revenue and total cost functions in Eq. (7.1), the profit function may be written as

$$\Pi = f(Q)_{TR} - f(Q)_{TC} \qquad \qquad ...(7.2)$$

Equation (7.2) can now be manipulated to illustrate the *first* and *second* order conditions of profit maximization as follows.

First-order condition: The first-order condition of maximizing a function is that its first derivative must be equal to zero. Thus, the first-order condition of profit maximization is that the first derivative of the profit function Eq. (7.2) must be equal to zero. Differentiating the total profit function and setting it equal to zero, we get

$$\frac{\partial \Pi}{\partial Q} = \frac{\partial TR}{\partial Q} - \frac{\partial TC}{\partial Q} = 0 \qquad \qquad ...(7.3)$$

This condition holds only when

$$\frac{\partial TR}{\partial Q} = \frac{\partial TC}{\partial Q}$$

In Eq. (7.3), the term $\partial TR/\partial Q$ gives the slope of the *TR* curve which in turn gives the marginal revenue (*MR*). Similarly, the term $\partial TC/\partial Q$ gives the slope of the total cost curve which is the same as marginal cost (*MC*). Thus, the first-order condition for profit maximization can be stated as

$$MR = MC$$

The first-order condition is generally known as *necessary condition*. A necessary condition is one that *must* be satisfied for an event to take place. In other words, the condition that *MR = MC* must be satisfied for profit to be maximum.

Second-order Condition: As already mentioned, in non-technical terms, the second-order condition of profit maximization requires that the first order condition is satisfied under rising *MC* and decreasing *MR*. This condition is illustrated in Fig. 7.1. The *MC* and *MR* curves are the usual marginal cost and marginal revenue curves respectively. Incidentally, *MC* and *MR* curves are derived from *TC* and *TR* functions respectively. *MC* and *MR* curves intersect at two points, P_1 and P_2. Thus, the first-order condition is satisfied at both the points, but the second order condition of profit maximization is satisfied only at point P_2. Technically, the second-order condition requires that the second de-rivative of the profit function is negative. The second derivative of the total profit function is given as

$$\frac{\partial^2 \Pi}{\partial Q^2} = \frac{\partial^2 TR}{\partial Q^2} - \frac{\partial^2 TC}{\partial Q^2} \qquad \text{...(7.4)}$$
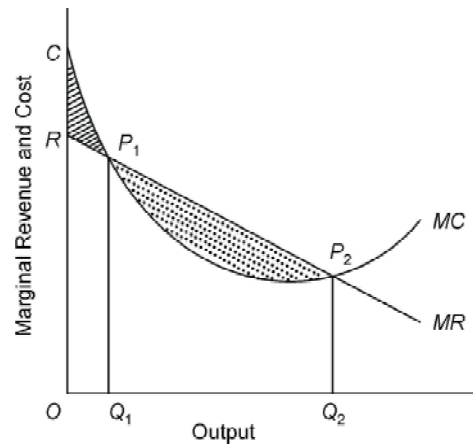
Fig. 7.1 Marginal Conditions of Profit Maximization

The second-order condition requires that

$$\frac{\partial^2 TR}{\partial Q^2} - \frac{\partial^2 TC}{\partial Q^2} < 0$$

or

$$\frac{\partial^2 TR}{\partial Q^2} < \frac{\partial^2 TC}{\partial Q^2} \qquad \text{...(7.5)}$$

Since $\partial^2 TR/\partial Q^2$ gives the slope of *MR* and $\partial^2 TC/\partial Q^2$ gives the slope of *MC*, the second-order condition may also be written as

Slope of *MR* < Slope of *MC*

It implies that *MC* must have a steeper slope than *MR* or *MC* must intersect the *MR* from below.

To conclude, profit is maximized where both the first and second order conditions are satisfied.

7.2.1 Optimum Firm and Representative Firm

The *long-run marginal cost curve* (*LMC*) is derived from the short-run marginal cost curves (*SMCs*). The derivation of *LMC* is illustrated in Fig. 7.3 in which *SACs*, *SMC$_s$* and *LAC* are the same as in Fig. 7.2 (*b*). To derive the *LMC*, consider the points of tangency between *SACs* and the *LAC*, i.e., points *A*, *B* and *C*. In the long-run production planning, these points determine the output at the different levels of production. Each of these outputs has an *SMC*. For example, if we draw a perpendicular from point *A*, it intersects *SMC$_1$* at point *M* determining *SMC* at *MQ$_1$*. The same process can be repeated for points *B* and *C* to find out *SMC* at outputs *Q$_2$* and *Q$_3$*. Note that points *B* and *C* determine *SMC* at *BQ$_2$* and *CQ$_3$* respectively. A curve drawn through points *M*, *B* and *N*, as shown by the *LMC*, represents the behaviour of the marginal cost in the long-run. This curve is known as the long-run marginal cost curve, *LMC*. It shows the trends in the marginal cost in response to the changes in the scale of production.
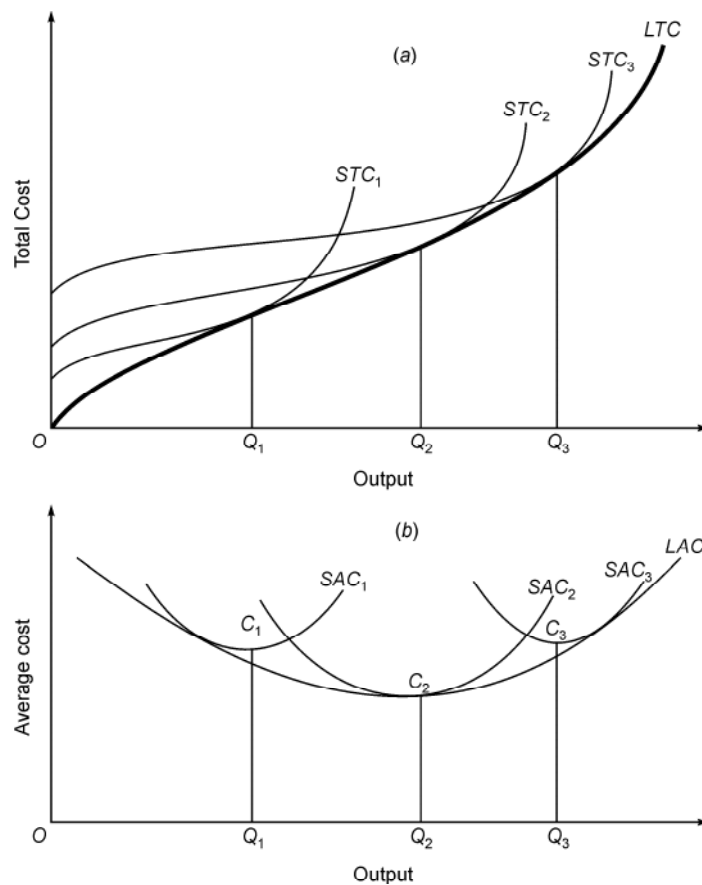
Fig. 7.2 *Long-run Total and Average Cost Curves*

Some important inferences may be drawn from Fig. 7.2. The *LMC* must be equal to *SMC* for the output at which the corresponding *SAC* is tangent to the *LAC*. At the point of tangency, *LAC* = *SAC*. Another important point to notice is that *LMC* intersects *LAC* when the latter is at its minimum, i.e., point *B*. There is one and only one short-run plant size whose minimum *SAC* coincides with the minimum *LAC*. This point is *B* where

$$SAC_2 = SMC_2 = LAC = LMC$$

The short-run cost curves are helpful in showing how a firm can decide on the *optimum utilization of the plant*—the fixed factor, or how it can determine the output level that minimizes cost. Long-run cost curves, on the other hand, can be used to show how a firm can decide on the *optimum size of the firm*.
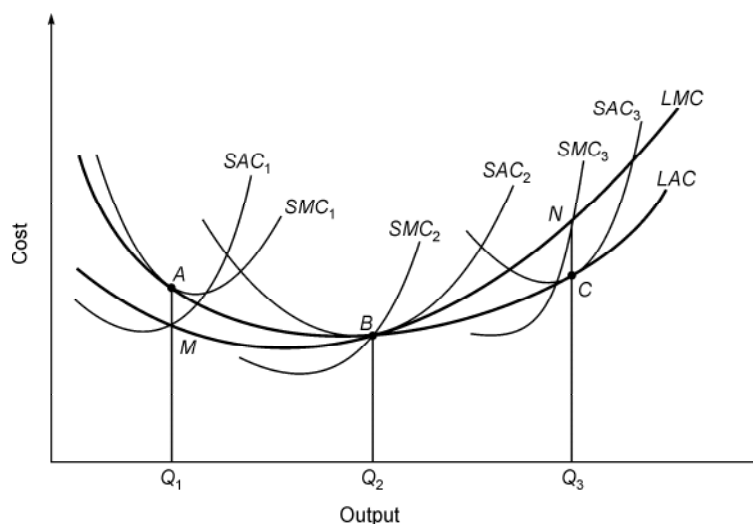
Fig. 7.3 Derivation of LMC

Conceptually, the optimum size of a firm is one which ensures the most efficient utilization of resources. Practically, the optimum size of the firm is one that minimizes the *LAC*. Given the state of technology over time, there is technically a unique size of the firm and level of output associated with the least-cost concept. In Fig. 7.2, the optimum size of the firm consists of two plants represented by *SAC*1 and *SAC*2. The two plants together produce *OQ*2 units of a product at minimum long-run average cost (*LAC*) of *BQ*2. The downtrend in the *LAC* indicates that until output reaches the level of *OQ*2, the firm is of less than optimal size. Similarly, expansion of the firm beyond production capacity *OQ*2, causes a rise in *SMC* and, therefore, in *LAC*. It follows that given the technology, a firm aiming to minimize its average cost over time must choose a plant that gives minimum *LAC* where *SAC* = *SMC* = *LAC* = *LMC*. This size of plant assures the most efficient utilization of the resources. Any change in output level—increase or decrease—will make the firm enter the area of inoptimality.

An optimum firm is that firm which fully utilizes its scale of operation and produces optimum output with the minimum cost per unit production.

7.2.2 Nature of Costs in Economics – Opportunity Cost Vs Real Cost

The term "cost" can be defined in terms of resources consumed to accomplish a specific goal. Classification based on what resources have been consumed in basic categories like wages, light and heat, advertisement and other expenses is termed as cost by nature. It is widely used in financial reporting for cost accounting.

Some of the accounting entries based on nature and found often in financial statements are wages, salaries, social costs, material, write-down of inventories, repair and maintenance costs, depreciation of fixed assets, amortization of intangible assets, administration expenses and marketing expenses.

In this context, let us study opportunity cost and real cost.

Opportunity Cost and Real Cost. Resources available to any person, firm or society are scarce but have alternative uses with different returns. Income maximizing resource owners put their scarce resources to their most productive use and thus, they forego the income expected from all other uses of the resources. The income foregone is called *opportunity cost*. While measuring the opportunity cost, the return from the second best use only is taken into account. The *opportunity cost* may be defined as the expected returns form the second best use of the resources were foregone due to the scarcity of resources. The opportunity cost is also called *alternative cost.* If resources available to a person, a firm or a society were unlimited there would be no opportunity cost.

For example, suppose that a firm has a sum of ` 100,000 for which it has only two alternative uses. It can buy either a printing machine or alternatively a lathe machine both having productive life of 10 years. From the printing machine, the firm expects an annual income of ` 20,000 and from the lathe, ` 15,000. A profit maximizing firm would invest its money in the printing machine and forego the expected income from the lathe. The opportunity cost of the income from printing machine is the expected income from the lathe, i.e., ` 15,000.

Associated with the concept of opportunity cost is the concept of *economic rent* or *economic profit.* In our example of expected earnings firm printing machine and economic rent of the printing machine is the excess of its earning over the income expected from the lathe. That is, economic rent equals ` 20,000 – ` 15,000 = ` 5,000. The implication of this concept for a business man is that investing in the printing machine is preferable so long as its economic rent is greater than zero. Also, if firms know the economic rent of the various alternative uses of their resources, it will be helpful in the choice of the best investment avenue.

In contrast to the concept of *opportunity cost*, **real costs** are those which are actually incurred by the firm in payment for labour, material, plant, building, machinery, equipment, travelling and transport, advertisement, etc. The total money expenses, recorded in the books of accounts are for all practical purposes, the *real costs*. In our example, the cost of printing machine, i.e., ` 100,000 is the actual cost. Real cost comes under the accounting cost concept.

Check Your Progress

1. Define marginal revenue.

2. What is an optimum firm?

7.3 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. Marginal revenue is revenue obtained from the production and sale of one additional unit of output.

2. An *optimum firm* is that *firm* which fully utilizes its scale of operation and produces *optimum* output with the minimum cost per unit production.

7.4 SUMMARY

- By definition, marginal revenue is the revenue obtained from the production and sale of one additional unit of output and marginal cost is the cost arising due to the production of one additional unit of output.

- A necessary condition is one that must be satisfied for an event to take place. In other words, the condition that *MR = MC* must be satisfied for profit to be maximum.

- The long-run marginal cost curve (LMC) is derived from the short-run marginal cost curves (SMCs).

- The short-run cost curves are helpful in showing how a firm can decide on the optimum utilization of the plant—the fixed factor, or how it can determine the output level that minimizes cost. Long-run cost curves, on the other hand, can be used to show how a firm can decide on the optimum size of the firm.

- Conceptually, the optimum size of a firm is one which ensures the most efficient utilization of resources.

- Resources available to any person, firm or society are scarce but have alternative uses with different returns. Income maximizing resource owners put their scarce resources to their most productive use and thus, they forego the income expected from all other uses of the resources.

- Associated with the concept of opportunity cost is the concept of economic rent or economic profit.

- In contrast to the concept of opportunity cost, actual costs are those which are actually incurred by the firm in payment for labour, material, plant, building, machinery, equipment, travelling and transport, advertisement, etc.

7.5 KEY WORDS

- **Real Cost:** These are those costs which are actually incurred by the firm in payment for labour, material, plant, building, machinery, equipment, travelling and transport and so forth.

- **Profit:** It is a financial benefit that is realized when the amount of *revenue* gained from a business activity surpasses the expenses, costs and taxes needed to sustain the activity

7.6 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short Answer Questions

1. Define marginal cost.
2. Why is the opportunity cost also known as alternative cost?
3. What do you understand by the term 'economic profit'?

Long Answer Questions

1. How is marginal revenue and marginal cost related to the profit maximization function of a firm?
2. Differentiate between opportunity cost and real cost.
3. Examine the nature of costs in economics.

7.7 FURTHER READINGS

Dwivedi, D. N. 2008. *Principles of Economics*, Seventh Edition. New Delhi: Vikas Publishing House.

Weil. David N. 2004. *Economic Growth*. London: Addison Wesley.

Thomas, Christopher R. and Maurice S. Charles. 2005. *Managerial Economics: Concepts and Applications*, Eighth Edition. New Delhi: Tata McGraw-Hill Publishing Company Limited.

Mankiw, Gregory N. 2002. *Principles of Economics*, Second Edition. India: Thomson Press.

UNIT 8   COST CONCEPTS - III

Structure

8.0 Introduction
8.1 Objectives
8.2 Fixed Costs Vs Variable Costs
    8.2.1   Notion of Marginal Cost
    8.2.2   Equilibrium of Industry
8.3 Conditions of Competitive Equilibrium
8.4 Answers to Check Your Progress Questions
8.5 Summary
8.6 Key Words
8.7 Self Assessment Questions and Exercises
8.8 Further Readings

8.0   INTRODUCTION

The total cost of a business is comprised of fixed costs and variable costs. Fixed costs and variable costs affect the marginal cost of production only if variable costs exist. The marginal cost of production is calculated by dividing the change in the total cost by a one-unit change in the production output level and determines the cost of production for one more unit of good. It is useful in measuring the point at which a business can achieve economies of scale.

8.1   OBJECTIVES

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

- Differentiate between fixed and variable costs

- Analyse the notion of marginal cost

- Describe the process of equilibrium in the industry

- Explain the conditions of competitive equilibrium

8.2   FIXED COSTS VS VARIABLE COSTS

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

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

8.2.1 Notion of Marginal Cost

Marginal cost (MC) is defined as the addition to the total cost on account of producing one additional unit of the product. Or, marginal cost is the cost of the marginal unit produced. Marginal cost is calculated as $TC_n - TC_{n-1}$ where n is the number of units produced. Using cost function, *MC* can be defined as

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

Total, average and marginal cost concepts are used in the economic analysis of firm's production activities.

8.2.2 Equilibrium of Industry

In physical sense, the term equilibrium means the 'state of rest'. In general sense, it means balance in forces working in opposite directions. In the context of market analysis, *equilibrium refers to a state of market in which quantity demanded of a commodity equals the quantity supplied of the commodity.* The equality of demand and supply produces an *equilibrium price.* The equilibrium price is the price at which quantity demanded of a commodity equals its quantity supplied. That is, at equilibrium price, demand and supply are in equilibrium. Equilibrium price is also called *market-clearing price.* Market is cleared in the sense that there is no unsold stock and no unsupplied demand.

Determination of Market Price

Equilibrium price of a commodity in a free market is determined by the market forces of demand and supply. In order to analyze 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 weekly market demand and supply schedules for shirts in Delhi are given as shown in Table 8.1.

Table 8.1 *Weekly Demand and Supply Schedules for Shirts*

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

As the table shows, there is only one price of shirts (₹300) at which quantity demanded per week equals the quantity supplied at 40 thousand shirts. It means that the shirt market in Delhi is in equilibrium at price ₹300. At all other prices, the shirt market is in *disequilibrium*—the state of 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, demand exceeds supply showing *shortage* of shirts in the market. Likewise, at all prices above ₹300, supply exceeds demand showing *excess supply.*

In a free market, disequilibrium itself creates the condition for equilibrium. When there is excess supply, it forces downward adjustments in the price and quantity supplied. When there is excess demand, it forces upward adjustments in the price and quantity demanded. The process of downward and upward adjustments in price and quantity continues till the price reaches ₹300 and quantities supplied and demanded are in balance at 40 thousand shirts. This process is automatic. Let us now look into the process of price and quantity adjustments called 'market mechanism'.

Market Mechanism: How Market Brings about Balance

Market mechanism is a process of interaction between the market forces of demand and supply to determine equilibrium price. To understand how it works, let the price of shirts be initially set at ₹100. At this price, the number of shirts demanded (80,000) exceeds the quantity supplied (10,000) by 70 thousand shirts. The shortage gives sellers an opportunity to raise the price and it prepares buyers to accept and pay a higher price. As a result, price goes up. Increase in price enhances the profit margin. This induces firms to produce and sell more in order to maximize their profits. This trend continues till price rises to ₹300. As Table 8.1 shows, at price ₹300, the buyers are willing to buy 40 thousand 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. 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 in order to clear their unsold stock. Some firms find low price unprofitable and go out of the 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, price at ₹300 per shirt is equilibrium price.

8.3 CONDITIONS OF COMPETITIVE EQUILIBRIUM

A competitive equilibrium can be defined as an equilibrium condition taking place between profit-maximizing producers and utility-maximizing consumers in competitive markets with freely determined prices with the objective of arriving at an equilibrium price. Hence, at this equilibrium price, the quantity supplied is equivalent to the quantity demanded.

It is to be noted that individual consumer and firm behaviour influence the basic supply. The competitive equilibrium model is constructed on the behaviour of aggregate consumers and organizations in competitive markets. It can be utilized to forecast the equilibrium price and total quantity in the market, along with the quantity consumed by each individual and output per organization.

Competitive equilibrium is a state of market, recognized by a set of prices and an allocation of commodities, such that at equilibrium prices, each agent make best use of his objective function subject to his technological constraints and resource limitations, and the market unblocks the aggregated supply and demand for the products in question.

Competitive equilibrium theory is recognized as a specialized branch of game theory that deals with making decisions in large markets. It is largely used to examine economic activities dealing with fiscal or tax policy, in finance for analysis of stock markets and commodity markets, to study interest, and exchange rates and other prices. It serves as a yardstick for efficiency in economic analysis. It depend on the assumption of competitive markets, where each trader decides upon a quantity that is so small compared to the total quantity traded in the market, such that their individual transactions have no influence on the prices. Competitive markets are a model, and a standard by which other market structure are assessed.

In a capitalist market, essential regulatory functions, like maintaining stability, proficiency and impartiality are left to the instruments of pricing. Thus, competitive equilibrium theory of equilibrium prices attained a noticeable position in mathematical economics.

Check Your Progress

1. What are fixed costs?
2. State the uses of total, average and marginal cost.
3. Define market mechanism.

8.4 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. Fixed costs are those which remain fixed in volume over a certain level of output.

2. Total, average and marginal cost concepts are used in the economic analysis of firm's production activities.

3. Market mechanism is a process of interaction between the market forces of demand and supply to determine equilibrium price.

8.5 SUMMARY

- Fixed costs are those which remain fixed in volume over a certain level of output. Fixed cost does not vary with variation in the output between zero and a certain level of output.

- The fixed costs include (*i*) costs of managerial and administrative staff, (*ii*) depreciation of machinery, building and other fixed assets, (*iii*) maintenance of land, etc. The concept of fixed cost is associated with the short-run. In the long run, not cost is fixed.

- Marginal cost (MC) is defined as the addition to the total cost on account of producing one additional unit of the product. Or, marginal cost is the cost of the marginal unit produced.

- In physical sense, the term equilibrium means the 'state of rest'. In general sense, it means balance in forces working in opposite directions. In the context of market analysis, *equilibrium refers to a state of market in which quantity demanded of a commodity equals the quantity supplied of the commodity.*

- Equilibrium price is also called *market-clearing price.* Market is cleared in the sense that there is no unsold stock and no unsupplied demand.

- 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.

- When market is in the state of disequilibrium, either demand exceeds supply or supply exceeds demand.

- *In a free market, disequilibrium itself creates the condition for equilibrium.* When there is excess supply, it forces downward adjustments in the price and quantity supplied.

- Market mechanism is a process of interaction between the market forces of demand and supply to determine equilibrium price.

- Competitive equilibrium theory is recognized as a specialized branch of game theory that deals with making decisions in large markets. It is largely used to examine economic activities dealing with fiscal or tax policy, in finance for analysis of stock markets and commodity markets, to study interest, and exchange rates and other prices.

8.6 KEY WORDS

- **Fixed Costs:** These are those which remain fixed in volume over a certain level of output. Fixed cost does not vary with variation in the output between zero and a certain level of output.

- **Marginal Cost (MC):** It is defined as the addition to the total cost on account of producing one additional unit of the product.

8.7 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short Answer Questions

1. State the notion of marginal cost.
2. How does the market bring about balance in the economy?

Long Answer Questions

1. Analyse the conditions of competitive equilibrium.
2. Describe the determinants of market price.

8.8 FURTHER READINGS

Dwivedi, D. N. 2008. *Principles of Economics*, Seventh Edition. New Delhi: Vikas Publishing House.

Weil. David N. 2004. *Economic Growth*. London: Addison Wesley.

Thomas, Christopher R. and Maurice S. Charles. 2005. *Managerial Economics: Concepts and Applications*, Eighth Edition. New Delhi: Tata McGraw-Hill Publishing Company Limited.

Mankiw, Gregory N. 2002. *Principles of Economics*, Second Edition. India: Thomson Press.

BLOCK - III

THEORY OF INTEREST AND MARKET BEHAVIOUR

UNIT 9   THEORIES OF INTEREST

Structure

9.0   INTRODUCTION

In this unit, we discuss how interest rate is determined in the capital market. Note that like wages are the price for labour and rent is the price for using a rented property, interest is the price for using borrowed funds or money. Like wage rate is determined by demand for and supply of labour, the interest rate is determined by demand for funds and supply of loanable funds. However, interest rate determination has been a much more complicated issue. Complications arise because borrowed money is first converted into productive assets (capital) like land, building and machinery. So the return of capital assets has also to be taken into account in analyzing how interest rate is determined in capital market.

9.1   OBJECTIVES

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

- Discuss the concept of interest and interest rate
- Describe liquidity preference theory
- Explain the meaning of profit, risk and uncertainty
- Understand the meaning of normal profits

9.2 INTEREST – INTEREST AS REWARD FOR WAITING

In general sense of the term, '***interest***' means the amount that a borrower pays to the lender in excess of the borrowed amount after a period of time. For example, if a person borrows Rs 100 today and pay Rs 110 after a year, the extra payment of Rs 10 in the interest.

Interest rate is expressed in terms of percentage of the extra payment to the borrowed amount calculated for a period of one year. For example, in our above example, extra payment of Rs 10 turns out to be 10 per cent. In means that interest rate is 10 per cent per annum.

Why is interest charged and why is interest paid? To answer this question, let us consider the interest of the lender in lending money and the interest of the borrower in borrowing money and his willingness to pay interest.

When a person lends his money, he bears the following costs:

 (*i*) If the lender does not have excess cash balance, he has to cut down his consumption to save and to lend. In that case, he foregoes his consumption and utility for which he wants a compensation.

 (*ii*) Even if he has excess cash balance, he loses liquidity and faces inconvenience when need arises to buy goods of his necessity. Therefore, he expects some compensation for his inconvenience after lending and parting with liquidity.

 (*iii*) By lending money, the lender has a feeling of financial insecurity – money may not be recoverable when required for some unavoidable reasons. For this insecurity he expects a compensatory reward.

 (*iv*) Money lent may not be recoverable at all for such reasons as (*a*) insolvency of the borrower; (*b*) dishonesty of the borrower, and (*c*) death of the borrower. There is therefore a risk. The lender needs a return on risk bearing.

 (*v*) He could put his money in some form of investment yielding income, which he foregoes. Therefore, he needs a compensatory return.

Similarly, the borrower gains from the borrowed money in different ways.

 (*i*) The borrower is able to meet his requirements from the borrowed money and, therefore, avoids the possible losses, if any. Therefore, his interest lies in borrowing and hence he is willing to pay interest.

 (*ii*) He can invest the borrowed money in some profitable business and earn a high rate of return.

 (*iii*) The borrower can avoid some uncertainties of life and feel more secure with funds available. Therefore, he is willing to pay for it.

Thus, both the lenders and the borrowers have interest in lending and borrowing. The interest arises out of this interest of lenders and borrowers. How rate of interest is determined is the matter of the theory of interest.

9.3 LIQUIDITY PREFERENCE THEORY

Having criticised the classical theories of interest, Keynes propounded his own Liquidity Preference Theory of Interest. The 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.

The Keynesian theory of interest states that the equilibrium rate of interest is determined by the aggregate supply of money and aggregate demand for money (or what he called, the *liquidity preference*). By demand for money, Keynes means liquidity preference or holding money in the form of idle cash balance. Let us look at Keynesian concept of aggregate supply and aggregate demand for money.

1. The Aggregate Supply of Money: As regards the *aggregate supply of money,* money supply is created by the central bank of the country on the basis of the country's need for money. The supply of money is determined as a matter of monetary policy of the country. In Keynesian theory of interest, money supply is assumed to be interest-inelastic and is assumed to remain constant at a point of time.

2. Keynesian Theory of Demand for Money: According to Keynes, people demand money or hold idle cash balance for three motives or purposes:

   (*i*) Transaction demand for money,

  (*ii*) Precautionary demand for money, and

 (*iii*) Speculative demand for money.

(*i*) *Transaction demand for money.* Transaction demand for money refers to the money held by the people to carry out their routine and planned transactions, e.g., for meeting routine consumption needs and planned business transactions. Transaction demand for *money* (M_t) is the function of current income, i.e.,

$$M_t = f(Y), \qquad \text{where } Y \text{ is the current regular income.}$$

Note that, according to Keynes, *transaction demand for money is interest-inelastic*, i.e., a certain amount of money has to be spent whatever the interest rate.

(*ii*) *Precautionary demand for money.* Precautionary demand for money refers to the money which households and business firms hold for precautionary purpose, i.e., the money held for meeting contingent expenses and expenses arising out of unpredictable events like theft, loss of job and medical expenditure on treatment

of illness, etc. Households hold some cash balance in excess of their routine expenses as a precaution for emergencies and business firms hold extra cash balance for taking advantage of changing market conditions. Precautionary demand for money (M_p) also is the function of current income, i.e.,

$$M_p = f(Y)$$

Like transaction demand for money, precautionary demand is also interest-inelastic.

Since both transaction and precautionary demand for *money* are function of current income, they can be jointly summed as M_T and can be written functionally as follows. Since

$$M_t = f(Y) \text{ and } M_p = f(Y),$$
$$M_t + M_p = M_T = f(Y)$$

(*iii*) *Speculative demand for money.* People hold money for speculative impose with a view to take the advantage of unpredictable change in the financial market, like share market. It is called speculative demand for money because there is no certainty or predictability of the changes in the financial market. Therefore, holding idle cash balance involves loss of interest. That is why it is called speculative demand for money. Speculative demand for money is interest-elastic. That is, speculative money holding increases when interest rate goes down and it decreases when interest rate goes up. Speculative money demand function is written as

$$M_s = f(i)$$

where M_s is speculative demand for money and i is the interest rate.

Aggregate demand for money. According to the Keynesian theory of demand for money, total money demand (M_D) can be expressed as follows:

$$M_D = M_T + M_S$$

In the theory of interest rate determination, Keynes links demand for money with interest rate. The relationship between the aggregate demand for money and interest rate is shown in Fig. 9.1. As can be seen in Fig. 9.1, the aggregate demand for money (M_D) is the horizontal sum of M_T and M_s curve. Since M_T is interest-inelastic and M_s is interest elastic, M_D curve turns out to be interest-elastic.
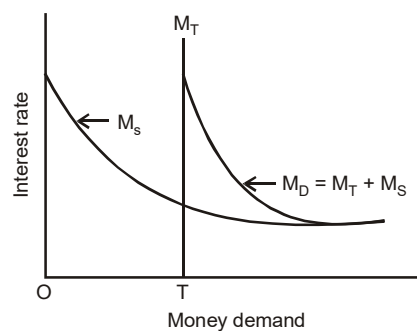
Fig. 9.1 *Demand for Money: Keynesian Approach*

3. Determination of Interest Rate: As mentioned above, according to Keynes, the equilibrium rate of interest is determined where aggregate demand for money equals the aggregate supply of money, i.e., at the level of money demand and supply where $M_D = M_S$. The Keynesian theory of interest rate determination is illustrated in Fig. 9.2. In this figure, M_{D1} schedule shows the aggregate demand for money in relation to the interest rate and vertical line marked M_S represents the aggregate money supply. As shown in Fig. 9.2, M_{D1} schedule intersects with M_S line at point E_1. At point E_1, therefore, the aggregate demand for money equals the aggregate supply of money. Point E_1 shows, therefore, the point of equilibrium in the money market. Money market being in equilibrium at point E_1 the equilibrium rate of interest is determined at Oi_1. The supply of money remaining constant, if demand for money increases for some reason and money-demand curve shifts upward as shown by the schedule M_{D2}, then equilibrium point shifts up to E_2 and equilibruim interest rate rises to Oi_2. Similarly, if money-demand schedule M_{D1} shifts downward, the interest rate will fall below Oi_1.
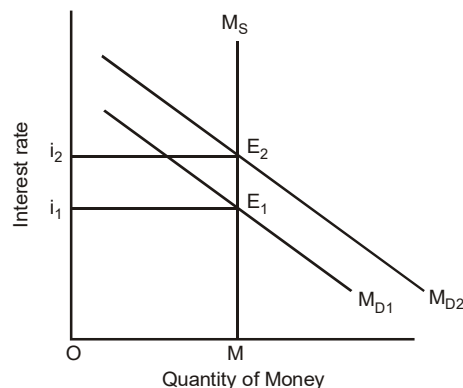
Fig. 9.2 *Determination of Interest Rate: Keynesian Approach*

Criticism of Keynesian Theory of Interest: In brief, Keynes had criticized classical theory of interest and concluded that interest rate is indeterminate under classical system. According to Alvin Hansen, the same criticism applies to Keynesian theory also. In his own words, "exactly the same criticism applies to Keynesian theory in its simple form." It means that the Keynesian theory is also indeterminate. His argument runs as follows. Given the money supply, the transaction demand for money depends on the level of income; income level depends on the investment; investment depends on the interest rate. Therefore, unless interest rate is determined, investment can not be determined; unless investment is determined, income cannot be determined; unless income is determined, money demand cannot be determined; unless money demand is determined, interest rate cannot be determined.

Furthermore, suppose money market is in equilibrium and interest rate is determined. Now let money-demand curve shift upward for some reason. As a result, interest will go up causing investment to decline. When investment goes down, income level goes down. When income level declines, money demand declines too. As a result, interest rate falls down causing investment to go up.

Increase in investment leads to rise in income and therefore an upward shift in money demand curve. Consequently, interest rate goes up. Thus, interest rate keeps fluctuating between an upper and a lower limit. The interest rate is thus indeterminate also under Keynesian system.

Check Your Progress

1. How is interest rate expressed?
2. What does the Keynesian theory of interest state?

9.4 PROFIT, RISK AND UNCERTAINTY

In common parlance, risk means a low probability of an expected outcome. From business decision-making point of view, risk refers to a situation in which a business decision is expected to yield more than one outcome and the probability of each outcome is known to the decision makers or it can be reliably estimated. For example, if a company doubles its advertisement expenditure, there are four probable outcomes: (*i*) its sales may more-than-double, (*ii*) they may just double, (*iii*) increase in sales may be less than double and (*iv*) sales do not increase at all. The company has the knowledge of these probabilities or has estimated the probabilities of the four outcomes on the basis of its past experience as (*i*) more-than double – 20% (or 0.2), (*ii*) almost double – 40% (or 0.4), (*iii*) less-than double – 50% (or 0.5) and (*iv*) no increase – 10% (or 0.1). It means that there is 80% risk in expecting more-than-doubling of sales, and 60% risk in expecting doubling of sale, and so on.

There are two approaches to estimating probabilities of outcomes of a business decision, viz., (*i*) *a priori approach*, i.e., the approach based on deductive logic or intuition and (*ii*) *posteriori approach*, i.e., estimating the probability statistically on the basis of the past data. In case of *a priori probability*, we know that when a coin is tossed, the probabilities of 'head' or 'tail' are 50:50, and when a dice is thrown, each side has 1/6 chance to be on the top. The *posteriori* assumes that the probability of an event in the past will hold in future also. The probability of outcomes of a decision can be estimated statistically by way of 'standard deviation' and 'coefficient of variation'.

Meaning of Uncertainty

Uncertainty refers to a situation in which there is more than one outcome of a business decision and the probability of no outcome is known nor can it be meaningfully estimated. The unpredictability of outcome may be due to lack of reliable market information, inadequate past experience, and high volatility of the market conditions. For example, if an Indian firm, highly concerned with population burden on the country, invents an irreversible sterility drug, the outcome regarding its success is completely unpredictable. Consider the case of insurance companies.

It is possible for them to predict fairly accurately the probability of death rate of insured people, accident rate of cars and other automobiles, rate of buildings catching fire, and so on, but it is not possible to predict the death of a particular insured individual, a particular car meeting an accident or a particular house catching fire, etc.

The long-term investment decisions involve a great deal of uncertainty with unpredictable outcomes. But, in reality, investment decisions involving uncertainty have to be taken on the basis of whatever information can be collected, generated and 'guesstimated'. For the purpose of decision-making, the uncertainty is classified as:

(*a*) complete ignorance and

(*b*) partial ignorance.

In case of *complete ignorance*, investment decisions are taken by the investor using their own judgement or using any of the rational criteria. What criterion he chooses depends on his attitude towards risk. The investor's attitude towards risk may be that of

(*i*) a risk averter,

(*ii*) a risk neutral or

(*iii*) a risk seeker or risk lover.

In simple words, a risk averter avoids investment in high-risk business. A risk-neutral investor takes the best possible decision on the basis of his judgement, understanding of the situation and his past experience. He does his best and leaves the rest to the market. A risk lover is one who goes by the dictum that 'the higher the risk, the higher the gain'. Unlike other categories of investors, he prefers investment in risky business with high expected gains.

In case of *partial ignorance*, on the other hand, there is some knowledge about the future market conditions; some information can be obtained from the experts in the field, and some probability estimates can be made. The available information may be incomplete and unreliable. Under this condition, the decision-makers use their subjective judgement to assign an *a priori probability* to the outcome or the pay-off of each possible action such that *the sum of such probability distribution is always equal to one*. This is called *subjective probability distribution*. The investment decisions are taken in this case on the basis of the *subjective probability distribution*. Decision-making process under partial ignorance is described in the following section beginning with the pay-off matrix.

9.4.1 Normal Profits

Generally two different concepts of profits are used in economic literature, viz., *normal profit* and *pure profit.* Before we answer the question, let us look into these concepts of profit.

Normal profit is the minimum rate of return that a firm must earn to remain in the industry. In other words, normal profit equals the *transfer earning.* Normal profit is also referred to as the wages of management. Marshall calls it the supply price of average business ability. The concept of normal profit is related to the concept of long run. It refers to the long-term earning of the entrepreneurs under competitive conditions. Under competitive conditions, in the long-run, the earnings of all the entrepreneurs of an industry tends to equalise. Besides, the concept of normal profit is also related to the state of equilibrium in which there is no risk or uncertainty involved, nor is there any tendency of firms to enter or to leave the industry. That is, in a static equilibrium all firms earn only normal profit, or what Knight calls, the wages of management.

Let us now return to the question whether profits enter the cost of production. When reference is made to normal profit, undoubtedly, it enters the cost of production, in the same way as rent, interest and wages. For, normal profit is treated simply as the wages of management. But, when reference is made to pure profit, it does not enter the cost of production. Pure profit is rather a surplus over and above the cost of production.

9.4.2 Marginal Productivity and Profits

Marginal profit is the additional profit from selling one extra unit. A profit per unit will be achieved when marginal revenue (MR) is greater than marginal cost (MC). At profit maximisation, marginal profit is zero because MC = MR.

Profit maximisation

Firms achieve maximum profits when marginal revenue (MR) is equal to marginal cost (MC), that is when the cost of producing one more unit of a good or service is exactly equal to the revenue derived from selling one extra unit.
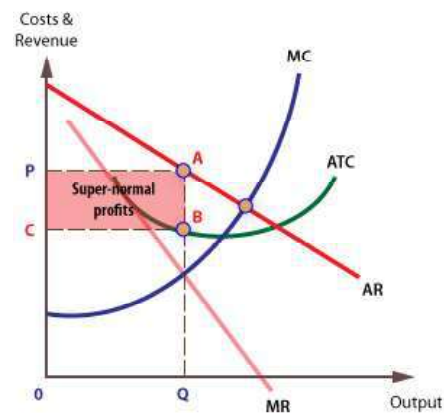
If marginal profit is greater than zero

If the firm stops short of producing Q, (at Q1 below) then MR is greater than MC, and marginal profit is still greater than zero. Hence, the firm should increase output.
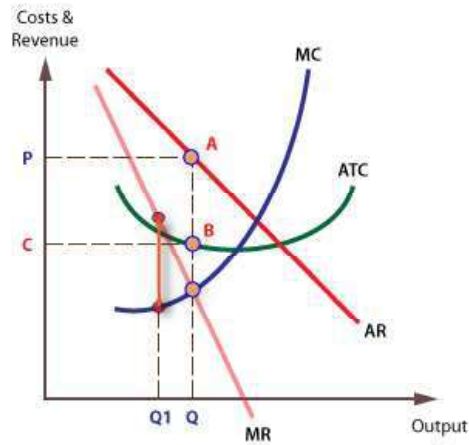
If marginal profit is less than zero

If the firm produces greater than Q, (at Q2 below) MC is greater than MR and marginal profit is negative. Hence, the firm should reduce its output. Only when MR = MC, at Q, will total profits be maximised.
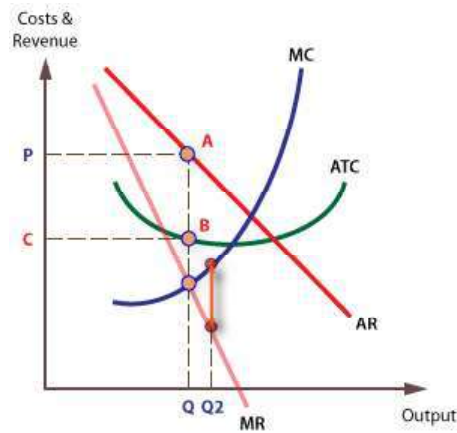
Check Your Progress
3. How is uncertainty classified for the purpose of decision making?
4. What is the other name for normal profit?

9.5 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. Interest rate is expressed in terms of percentage of the extra payment to the borrowed amount calculated for a period of one year.

2. The Keynesian theory of interest states that the equilibrium rate of interest is determined by the aggregate supply of money and aggregate demand for money (or what he called, the *liquidity preference*).

3. For the purpose of decision-making, the uncertainty is classified as:
   - Complete ignorance and
   - Partial ignorance.
4. Normal profit is also known as the wages of management.

9.6 SUMMARY

- In general sense of the term, '*interest*' means the amount that a borrower pays to the lender in excess of the borrowed amount after a period of time.

- *Interest rate* is expressed in terms of percentage of the extra payment to the borrowed amount calculated for a period of one year.

- Both the lenders and the borrowers have interest in lending and borrowing. The interest arises out of this interest of lenders and borrowers. How rate of interest is determined is the matter of the theory of interest.

- Keynes propounded his own Liquidity Preference Theory of Interest. The 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.

- The Keynesian theory of interest states that the equilibrium rate of interest is determined by the aggregate supply of money and aggregate demand for money (or what he called, the *liquidity preference*).

- In common parlance, risk means a low probability of an expected outcome. From business decision-making point of view, risk refers to a situation in which a business decision is expected to yield more than one outcome and the probability of each outcome is known to the decision makers or it can be reliably estimated.

- There are two approaches to estimating probabilities of outcomes of a business decision, viz., (*i*) *a priori approach*, i.e., the approach based on deductive logic or intuition and (*ii*) *posteriori approach*, i.e., estimating the probability statistically on the basis of the past data.

- Uncertainty refers to a situation in which there is more than one outcome of a business decision and the probability of no outcome is known nor can it be meaningfully estimated.

- The long-term investment decisions involve a great deal of uncertainty with unpredictable outcomes. But, in reality, investment decisions involving uncertainty have to be taken on the basis of whatever information can be collected, generated and 'guesstimated'.

- The concept of normal profit is related to the concept of long run. It refers to the long-term earning of the entrepreneurs under competitive conditions.

Under competitive conditions, in the long-run, the earnings of all the entrepreneurs of an industry tends to equalise.

9.7   KEY WORDS

- **Liquidity Preference Theory:** In macroeconomic theory, liquidity preference is the demand for money, considered as liquidity.

- **Priori Approach:** The term usually describes lines of reasoning or arguments that proceed from the general to the particular, or from causes to effects.

9.8   SELF ASSESSMENT QUESTIONS AND EXERCISES

Short Answer Questions

1. Write a short note on speculative and precautionary demand for money.
2. State the criticism regarding Keynesian theory of interest.
3. Write a short note on marginal productivity and profits.

Long Answer Questions

1. Analyse the arguments regarding why is interest charged and paid.
2. The Keynesian theory of interest is a purely monetary theory of interest. Discuss.
3. Discuss the concept of normal profits.

9.9   FURTHER READINGS

Dwivedi, D. N. 2008. *Principles of Economics*, Seventh Edition. New Delhi: Vikas Publishing House.

Weil. David N. 2004. *Economic Growth*. London: Addison Wesley.

Thomas, Christopher R. and Maurice S. Charles. 2005. *Managerial Economics: Concepts and Applications*, Eighth Edition. New Delhi: Tata McGraw-Hill Publishing Company Limited.

Mankiw, Gregory N. 2002. *Principles of Economics*, Second Edition. India: Thomson Press.

UNIT 10 TYPES OF MARKET STRUCTURES

Structure

10.0 Introduction
10.1 Objectives
10.2 Market Structures: Nature of Competition
     10.2.1 Market Structure: Meaning, Importance and Implication
10.3 Answers to Check Your Progress Questions
10.4 Summary
10.5 Key Words
10.6 Self Assessment Questions and Exercises
10.7 Further Readings

10.0 INTRODUCTION

Maximization of output or minimization of cost or optimization of resource allocation is, however, only one aspect of the profit maximizing behaviour of the firm. Another and equally important aspect of profit maximization is to find the price from the set of prices revealed by the demand schedule that is in agreement with profit maximization objective of the firm. It must be noted that there is only one price for each product commensurate with profit maximization, under the given conditions. The profit maximizing price does not necessarily coincide with minimum cost of production. Besides, the level of profit-maximizing price is generally different in different kinds of markets, depending on the degree of competition between the sellers. Therefore, while determining the price of its product, a firm has to take into account the nature of the market. In this unit, we will discuss about the types of market structures.

10.1 OBJECTIVES

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

- Analyse the nature of market competition

- Differentiate between different market structures

- Describe the meaning, importance and implications of different market structures

10.2 MARKET STRUCTURES: NATURE OF COMPETITION

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

We are concerned in this section with the question: How is the price of a commodity determined in different kinds of markets? The determination of price of a commodity depends on the number of sellers and the number of buyers. Barring a few cases, e.g., occasional phases in share and property markets, the number of buyers is larger than the number of sellers. The number of sellers of a product in a market determines the *nature and degree of competition* in the market. The nature and degree of competition make the *structure of the market*. Depending on the number of sellers and the degree of competition, the market structure is broadly classified as given in Table 10.1.

Table 10.1 *Types of Market Structures*

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

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

10.2.1 Market Structure: Meaning, Importance and Implication

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

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

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

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

In case of a **monopoly**, the degree of competition is close to nil. An uncontrolled monopoly firm has full control over the price of its product. A monopoly,

in the true sense of the term, is free to fix any price for its product, of course, under certain constraints, viz., (*i*) the objective of the firm, and (*ii*) demand conditions.

The theory of pricing explains pricing decisions and pricing behaviour of the firms in different kinds of market structures.

Check Your Progress

1. What is a market for a commodity?

2. What is a market structure?

3. What happens under perfect competition?

10.3 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. Market does not necessarily mean a place. The market for a commodity may be local, regional, national or international. What makes a market is a set of buyers, a set of sellers and a commodity. Buyers are willing to buy and sellers are willing to sell, and there is a price for the commodity.

2. The market structure determines a firm's power to fix the price of its product a great deal. The degree of competition determines a firm's degree of freedom in determining the price of its product.

3. Under perfect competition, a large number of firms compete against each other for selling their product. Therefore, the degree of competition under perfect competition is close to one, i.e., the market is highly competitive.

10.4 SUMMARY

- In economic sense, a market is a system by which buyers and sellers bargain for the price of a product, settle the price and transact their business—buy and sell a product.

- Personal contact between the buyers and sellers is not necessary. In some cases, e.g., forward sale and purchase, even immediate transfer of ownership of goods is not necessary.

- The determination of price of a commodity depends on the number of sellers and the number of buyers. Barring a few cases, e.g., occasional phases in share and property markets, the number of buyers is larger than the number of sellers.

- The market structure determines a firm's power to fix the price of its product a great deal. The degree of competition determines a firm's degree of freedom in determining the price of its product.

- The degree of freedom implies the extent to which a firm is free or independent of the rival firms in taking its own pricing decisions. Depending on the market structure, the degree of competition varies between zero and one.

- In case of a ***monopoly***, the degree of competition is close to nil. An uncontrolled monopoly firm has full control over the price of its product.

- A monopoly, in the true sense of the term, is free to fix any price for its product, of course, under certain constraints, viz., (*i*) the objective of the firm, and (*ii*) demand conditions.

10.5 KEY WORDS

- **Monopolistic competition**: It is a type of imperfect competition such that many producers sell products that are differentiated from one another (e.g. by branding or quality) and hence are not perfect substitutes.

- **Oligopoly:** A state of limited competition, in which a market is shared by a small number of producers or sellers.

10.6 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short Answer Questions

1. 'A market does not necessarily mean a place.' Do you agree? Give reasons for your answer.

2. What does the number of sellers in a market determine?

Long Answer Questions

1. Differentiate between perfect and imperfect market structures.

2. Analyse the meaning, importance and implication of market structures.

10.7 FURTHER READINGS

Dwivedi, D. N. 2008. *Principles of Economics*, Seventh Edition. New Delhi: Vikas Publishing House.

Weil. David N. 2004. *Economic Growth*. London: Addison Wesley.

Thomas, Christopher R. and Maurice S. Charles. 2005. *Managerial Economics: Concepts and Applications*, Eighth Edition. New Delhi: Tata McGraw-Hill Publishing Company Limited.

Mankiw, Gregory N. 2002. *Principles of Economics*, Second Edition. India: Thomson Press.

UNIT 11 PRICING UNDER PERFECT COMPETITION

Structure

11.0 INTRODUCTION

In the previous unit, you studied about the types of market structures. In this unit, you will study in detail about the theory of price and output determination under different kinds of market conditions. Two basic points need to be understood here. One, the main consideration behind the determination of price and output is to achieve the objectives of the firm. Two, although there can be various business objectives, traditional theory of price and output determination is based on the assumption that all firms have only one and the same objective to achieve, that is, profit maximization.

11.1 OBJECTIVES

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

- Discuss price determination under perfect competition
- Explain price determination under pure monopoly
- Identify the causes and kinds of monopolies
- Describe price discrimination under monopoly
- Examine pricing and output decisions under monopolistic competition
- Evaluate pricing and output decisions under oligopoly
- Prepare an overview of the oligopoly models

11.2 PRICE DETERMINATION UNDER PERFECT COMPETITION

The term perfect competition refers to a set of conditions prevailing in the market. A perfectly competitive market is one which has the following characteristics.

1. *A large number of sellers and buyers.* Under perfect competition, the number of sellers and buyers is very large. The number of sellers is so large that the share of each seller in total supply of a product is too small for a single seller to affect the market price by changing his supply. Likewise, the number of buyers is so large that the share of each buyer in total demand is too small for a single buyer to influence the market price by changing his/her demand.

2. *Homogeneous products.* Products supplied by all firms are almost homogeneous. Homogeneity of products means that products supplied by various firms are so identical in appearance and use that buyers do not distinguish between them nor do they prefer the product of one firm to that of another. Product of each firm is regarded as a perfect substitute for the product of other firms. Hence, no firm can gain any competitive advantage over the other firms. Nor do the firms distinguish between the buyers. For example, wheat and vegetables produced by all the farmers, other things given, are treated as homogeneous.

3. *Perfect mobility of factors of production.* For a market to be perfectly competitive, there should be perfect mobility of resources. This means that the factors of production must be in a position to move freely into or out of an industry and from one firm to another. This is however a purely theoretical assumption.

4. *Free entry and free exit of firms.* There is no barrier, legal or market-related, on the entry of new firms into or exit of existing ones from the industry. Firms are free to enter the industry and quit it at their free will.

5. *Perfect knowledge.* There is perfect dissemination of the information about the market conditions. Both buyers and sellers are fully aware of the nature of the product, its availability or saleability and of the price prevailing in the market.

6. *Absence of collusion or artificial restraint.* There is no sellers' union or other kinds of collusions between the sellers such as cartels or guilds, nor is there any kind of collusion between the buyers, e.g., consumers' associations or consumer forum. Each seller and buyer acts independently. The firms enjoy the freedom of independent decisions.

7. *No government intervention.* In a perfectly competitive market, there is no government intervention with the working of the market system. There is no licencing system regulating the entry of firms to the industry, no regulation of market prices, i.e., fixation of lower or upper limits of prices, no control over the supply of inputs, no fixation of quota on production, and no rationing of consumer demand, no subsidy to producers or to consumers, etc.

Perfect competition, as characterized above, is **an uncommon phenomenon** in the real business world. However, the actual markets that approximate to the conditions of perfectly competitive model include the share markets, securities and bond markets, and agricultural product markets, e.g., local vegetable markets. Although perfectly competitive markets are uncommon phenomena, perfect competition model has been the most popular model used in economic theories due to its analytical value as it provides a starting point and analytical framework for pricing theory.

11.2.1 Perfect Competition and Pure Competition

Sometimes a distinction is made between perfect competition and pure competition. The difference between the two is only a matter of degree. *Perfect competition less perfect mobility of factors and perfect knowledge is regarded as pure competition.* In this book, however, we shall use the two terms interchangeably.

Price and Output

As noted above, perfect competition is a market setting in which there are a large number of sellers of a homogeneous product. Each seller supplies a very small fraction of the total supply. No single seller is powerful enough to influence the market price. Nor can a single buyer influence the market price. Market price in a perfectly competitive market is determined by the market forces—market demand and market supply. *Market demand* refers to the demand for the industry as a whole: it is the sum of the quantity demanded by each individual consumer or user at different prices. Similarly, market supply is the sum of quantity supplied by the individual firms in the industry. The market price is,

therefore, determined for the industry, and is given for each individual firm and for each buyer. Thus, a seller in a perfectly competitive market is a 'price-taker, not a 'price-maker'.

In a perfectly competitive market, therefore, the main problem for a profit maximizing firm is not to determine the price of its product but to adjust its output to the market price so that profit is maximum.

The mode of price determination—price level and its variation—depends on the time taken by the supply position to adjust itself to the changing demand conditions. Therefore, price determination under perfect competition is analyzed under three different time periods:

(*i*) market period or very short-run,

(*ii*) short-run and

(*iii*) long-run.

The short-run and long-run have already been defined. As regards the market period or very short-run, it refers to a time period during which quantity supplied is absolutely fixed or, in other words, supply response to price is nil, i.e., supply of the product is inelastic. Price determination in the three types of time periods is described below.

(*i*) *Price Determination in Market Period.* In the market period, the total output of a product is fixed. Each firm has a stock of commodity to be sold. The stock of goods with all the firms makes the total supply. Since the stock is fixed, the supply curve is perfectly inelastic, as shown by the line *SQ* in Fig. 11.1(*a*). In this situation, price is determined solely by the demand condition. Supply remains an inactive factor. For instance, suppose that the number of marriage houses (or tents) in a city in a marriage season is given at *OQ* [(Fig. 11.1(*a*)] and the supply curve takes the shape of a straight vertical line, as shown by the line *SQ*. Suppose also that the demand curve for marriage houses (or tents) during an average marriage season is given by D_1. Demand curve and supply line intersect at point *M*, determining the rent for each marriage house at $MQ = OP_1$. But, suppose during a marriage season, demand for marriage houses (or tents) increases suddenly because a larger number of parents decide to celebrate the marriage of their daughters and sons, because auspicious dates for marriage are not available in the next few years. In that case, the demand curve D_1 shifts upward to D_2. The equilibrium point—the point of intersection between demand and supply curves—shifts from point *M* to *P*, and marriage house rentals rise to $PQ = OP_2$. This price becomes a parametric price for all the buyers.

NOTES
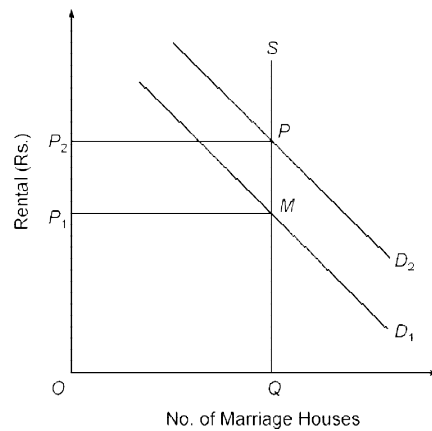
Fig. 11.1 (a) Demand Determined   *Fig. 11.1 (b) Supply Determined*
Price in Market Period   *Price in Market Period*

Similarly, given the demand for a product, if its supply decreases suddenly for such reasons as droughts, floods (in case of agricultural products) and sudden increase in export of a product, prices of such products shoot up. For example, price of onions had shot up in Delhi from ₹ 12 per kg to ₹ 36 kg. in 1998 due to export of onion. In case of supply determined price, supply curve shifts leftward causing rise in price of the goods in short supply. This phenomenon is illustrated in Fig. 11.1(*b*). Given the demand curve (*D*) and supply curve (S_2), the price is determined at OP_1. Demand curve remaining the same, the fall in supply makes the supply curve shift leftward to S_1. As a result price increases from OP_1 to OP_2.

The other examples of very short-run markets may be daily fish market, stock markets, daily milk market, coffin markets during a period of natural calamities, certain essential medicines during epidemics, etc.

(***ii***) ***Price in the Short-Run.*** A short-run is, by definition, a period in which firms can neither change their scale of production or quit, nor can new firms enter the industry. While in the 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.

The determination of market price in the short-run is illustrated in Fig. 11.2(*a*) and adjustment of output by the firms to the market price and firm's equilibrium are shown in Fig. 11.2(*b*). Fig. 11.2(*a*) shows the price determination for the industry by the demand curve *DD* and supply curve *SS*, at price OP_1 or *PQ*. This price is fixed for all the firms in the industry.
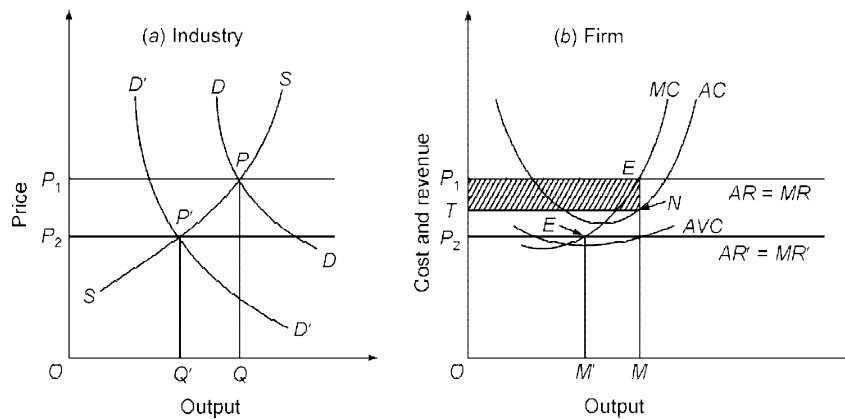
Fig. 11.2 *Pricing under Perfect Competition in the Short-run*

Given the price $PQ\ (= OP_1)$, an individual firm can produce and sell any quantity at this price. But any quantity will not yield maximum profit. Given their cost curves, the firms are required to adjust their output to the price PQ so that they maximize their profit.

The process of firm's output determination and its equilibrium are shown in Fig. 11.2(*b*). Profit is maximum at the level of output where $MR = MC$. Since price is fixed at PQ, firm's $AR = PQ$. If AR is constant, $MR = AR$. The firm's MR is shown by $AR = MR$ line. Firm's upward sloping MC curve intersects $AR = MR$ at point E. At point E, $MR = MC$. Point E is, therefore, the firm's equilibrium point. An ordinate drawn from point E to the output axis, as shown by the line EM, determines the profit-maxmizing output at OM. At this output the firm's $MR = MC$. This satisfies the necessary condition of maximum profit. The total maximum profit has been shown by the area P_1TNE.

The total profit is calculated as Profit $= (AR - AC)\,Q$. In Fig. 11.2(*b*), $AR = EM$; $AC = NM$; and $Q = OM$. Substituting these values into the profit equation, we get Profit $= (EM - NM)\,OM$. Since $EM - NM = EN$, Profit $= EN \times OM = P_1TNE$. This is the maximum supernormal profit, given the price and cost curves, in the short run.

Firms may make losses in the short-run: While firms may make supernormal profit, there may be conditions under which firms make losses in the short-run. For instance, this may happen if market price decreases to $P'Q'$ due to downward shift in the demand curve from DD to $D'D'$ [Fig. 11.2(*a*)]. This will force a process of output adjustments till firms reach a new equilibrium at point E'. Here again firm's $AR' = MR' = MC$. But, as Fig. 11.2(*b*) shows, $AR < AC$. Therefore, the firms incur a loss. But, since in the short-run, it may not be desirable to close down the production, the firms try to minimize the loss, by adjusting their output downward to OM' where it covers only its MC, i.e., $E'M'$. The firms survive in the short-run so long as they cover their MC.

It is important to note here that in the short-run, a firm in a perfectly competitive market may be in a position to earn ***economic profit***. It may as well be forced to make losses. Once market price for the product is determined, it is given for all the firms. No firm is large enough to influence the prices. If a firm fixes the price of its product lower than the market price, it may lose a part of its total profit, or may even incur losses. If it raises the price of its product above the market price, it may not be in a position to sell its produce in a competitive market. The only option for a firm is to produce as much as it can sell at the given price.

(*iii*) ***Pricing in the Long-Run.*** In contrast to the short-run conditions, in the long-run, the firms can adjust their size or quit the industry and new firms can enter the industry. If market price in the long run is such that $AR > AC$, then the firms make economic or super normal profit. As a result, new firms get attracted towards the industry causing increase in market supply at the given price. Increase in market supply causes rightward shift in the supply curve. Similarly, if $AR < AC$, then firms make losses. Therefore, marginal firms quit the industry causing decrease in market supply. This causes a leftward shift in the supply curve. The rightward shift in the supply curve pulls down the price and its leftward shift pushes it up. This process continues until price is so determined that $AR = AC$, and firms earn only normal profit.

The price determination in the long-run and output adjustment by individual firms are illustrated graphically in Fig. 11.3(*a*) and (*b*).

Let us suppose that the long-run demand curve is given by the curve DD'; the short-run supply curve is given by the curve SS_1 and price is determined at OP_1. Let us suppose also that all the firms of the industry face identical LAC and LMC curves as shown in Fig. 11.3(*b*). At market price OP_1, all the firms find their equilibrium at point *M* in panel (*b*) of the figure. At equilibrium point *M*, $OP_1 = AR' = MR' = LMC$. Given the price and cost, firms make an economic profit of *MS* per unit. The supernormal profit lures other firms into the industry. Consequently, industry's supply curve shifts rightward to SS_2 causing a fall in price to OP_2. At this price, firms are in a position to cover only $LMC (= NQ_2)$ at output OQ_2 and are making losses because $AR < LAC$. Firms incurring losses cannot survive in the long-run. Such firms, therefore, quit the industry. As a result, the total production in the industry decreases causing a leftward shift in the supply curve, say, to the position of *SS* curve. Price is determined at OP_0. The existing firms adjust their output to the new market price OP_0 and reach a new equilibrium at point *E* where equilibrium output is *OQ*. At the output *OQ*, firms are in a position to make only normal profit, since at this output, $OP_0 = AR = MR = LMC = LAC (= EQ)$. No firm is in a position to make economic profit, nor does any firm make losses. Therefore, there is no tendency of new firms entering the industry or the existing ones going out. At this price and output, individual firms and the industry are both in long-run equilibrium.
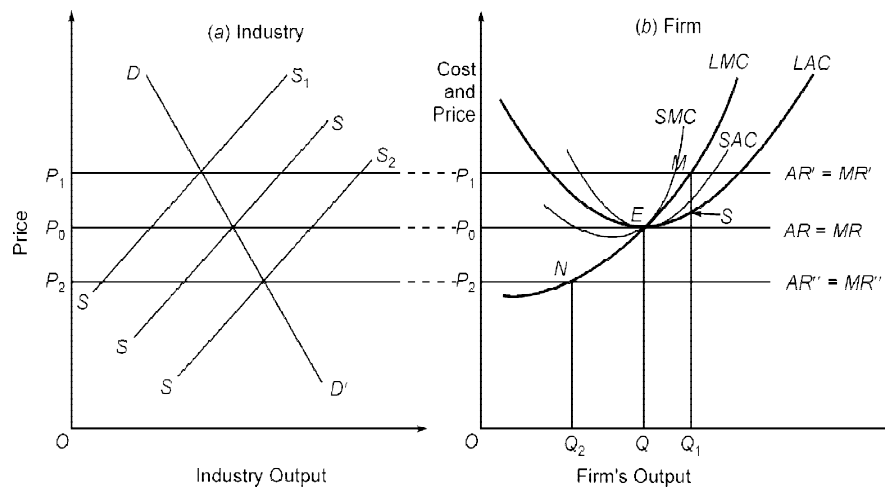
Fig. 11.3 *Pricing under Perfect Competition in the Long-run*

11.3   PRICE DETERMINATION UNDER PURE MONOPOLY

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

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

11.3.1 Causes and Kinds of Monopolies

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

(*i*) legal restrictions or barriers to entry of new firms

(*ii*) sole control over the supply of scarce and key raw materials

(*iii*) efficiency in production and

(*iv*) economies of scale.

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

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

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

(iv) *Economies of Scale.* The economies of scale are a primary and technical reason for the emergence and existence of monopolies in an unregulated market. If a firm's long-run minimum cost of production or its most efficient scale of production almost coincides with the size of the market, then the large-size firm finds it profitable

in the long-run to eliminate competition through price cutting in the short-run. Once its monopoly is established, it becomes almost impossible for the new firms to enter the industry and survive. Monopolies created on account of this factor are known as ***natural monopolies***. A natural monopoly may emerge out of the technical conditions of efficiency or may be created by law on efficiency grounds.

11.3.2 Pricing and Output Decision: Short-Run Analysis

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

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

The short-run revenue and cost conditions faced by a monopoly firm are presented in Fig. 11.4. Firm's average and marginal revenue curves are shown by the *AR* and *MR* curves, respectively, and its short-run average and marginal cost curves are shown by *SAC* and *SMC* curves, respectively. The price and output decision rule for profit maximizing monopoly is the same as for a firm in the competitive industry.
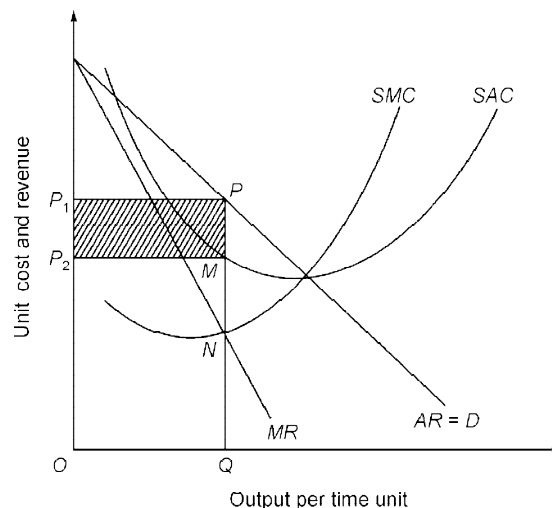
Fig. 11.4 *Price Determination under Monopoly: Short-run*

As noted earlier, profit is maximized at the level of output at which $MC = MR$. Given the profit maximization condition, a profit maximizing monopoly firm chooses a price-output combination at which $MR = SMC$. Given the firm's cost and revenue curves in Fig. 11.4, its MR and SMC intersect at point N. An ordinate drawn from point N to X-axis, determines the profit maximizing output for the firm at OQ. At this output, firm's $MR = SMC$. The ordinate NQ extended to the demand curve ($AR = D$) gives the profit maximizing price at PQ. It means that given the demand curve, the output OQ can be sold per time unit at only one price, i.e., PQ ($= OP_1$). Thus, the determination of output simultaneously determines the price for the monopoly firm. Once price is fixed, the unit and total profits are also simultaneously determined. Hence, the monopoly firm is in a state of equilibrium.

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

Does a Monopoly Firm Always Earn Economic Profit?

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

   (*i*) its cost and revenue conditions;

  (*ii*) threat from potential competitors; and

 (*iii*) government policy in respect of monopoly.

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

   (*i*) if $AR > AC$, there is economic profit for the firm,

  (*ii*) if $AR = AC$, the firm earns only normal profit, and

 (*iii*) if $AR < AC$, though only a theoretical possibility, the firm makes losses.

11.3.3 Monopoly Pricing and Output Decision in the Long-Run

The decision rules regarding optimal output and pricing in the long-run are the same as in the short-run. In the long-run, however, a monopolist gets an opportunity to expand the size of its firm with a view to enhance its long-run profits. The expansion of the plant size may, however, be subject to such conditions as (*a*) size of the market, (*b*) expected economic profit and (*c*) risk of inviting legal restrictions.

Let us assume, for the time being, that none of these conditions limits the expansion of a monopoly firm and discuss the price and output determination in the long-run.

The equilibrium of monopoly firm and its price and output determination in the long-run is shown in Fig. 11.5. The *AR* and *MR* curves show the market demand and marginal revenue conditions faced by the monopoly firm. The *LAC* and *LMC* show the long-run cost conditions. It can be seen in Fig. 11.5, that monopoly's *LMC* and *MR* intersect at point *P* determining profit maximizing output at OQ_2. Given the *AR* curve, the price at which the total output OQ_2 can be sold is P_2Q_2. Thus, in the long-run, equilibrium output will be OQ_2 and price P_2Q_2. This output-price combination maximizes monopolist's long-run profit. The total long-run monopoly profit is shown by the rectangle $LMSP_2$.
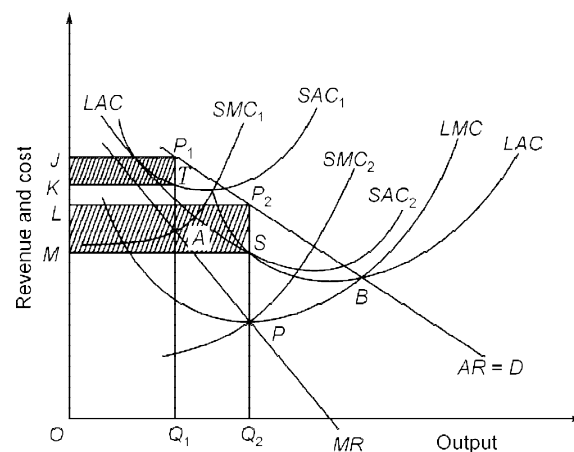
Fig. 11.5 *Equilibrium of Monopoly in the Long-run*

It can be seen in Fig 11.5 that compared to short-run equilibrium, the monopolist produces a larger output and charges a lower price and makes a larger monopoly profit in the long-run. In the short-run, monopoly's equilbirium is determined at point *A*, the point at which SMC_1 intersects the *MR* curve. Thus, monopoly's short-run equilibrium output is OQ_1 which is less than long-run output OQ_2. But the short-run equilibrium price P_1Q_1 is higher than the long-run equilibrium price P_2Q_2. The total short-run monopoly profit is shown by the rectangle JP_1TK which is much smaller than the total long-run profit LP_2SM. This, however, is not necessary: it all depends on the short-run and long-run cost and revenue conditions.

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

11.4 PRICE DISCRIMINATION UNDER MONOPOLY

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

(*i*) physicians and hospitals, lawyers, consultants, etc., charge their customers at different rates mostly on the basis of the latter's ability to pay;

(*ii*) merchandise sellers sell goods to relatives, friends, old customers, etc., at lower prices than to others and offer off-season discounts to the same set of customers;

(*iii*) railways and airlines charge lower fares from the children and students, and for different class of travellers;

(*iv*) cinema houses and auditoria charge differential rates for cinema shows, musical concerts, etc.,

(*iv*) some multinationals charge higher prices in domestic and lower prices in foreign markets, called 'dumping', and

(*v*) lower rates for the first few telephone calls, lower rates for the evening and night trunk-calls; higher electricity rates for commercial use and lower for domestic consumption, etc. are some other examples of price discrimination.

Necessary Conditions

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

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

elasticities of demand in different markets are the same, price discrimination would reduce the profit by reducing demand in the high price markets.

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

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

11.4.1 Price Discrimination by Degrees

The degree of price discrimination refers to the extent to which a seller can divide the market or the consumers and can take advantage of it in extracting the consumer's surplus. The economic literature presents three degrees of price discrimination.

First degree: The *first degree price discrimination* is the limit of discriminatory pricing. First degree or perfect price discrimination is feasible when the market size of the product is small and the monopolist is in a position to know the price each consumer or each group of consumers is willing to pay, (i.e., he knows his buyer's demand curve for his product), then he sets the price accordingly and tries to extract the entire *consumer surplus*. What the seller does is that he sets the price at its highest level—the level at which all those who are willing to buy the commodity buy at least one unit each. After extracting the consumer surplus of this segment of consumers for the first unit of commodity, the monopolist gradually lowers down the price, so that the consumer surplus of the users of the second unit is extracted. This procedure is continued until the entire consumers' surplus available at the equilibrium price, i.e., at the price at which $MC = MR$, is extracted. Consider, for example, the case of medical services of exclusive use. 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 patient.
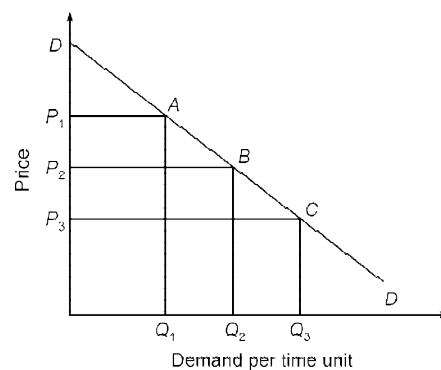
Fig. 11.6 *Second Degree Price Discrimination*

Second degree: Where market size is very large, perfect discrimination is neither feasible nor desirable. In that case, a monopolist uses second degree discrimination or the 'block pricing method'. A monopolist adopting the *second degree price discrimination* intends to siphon off only the major part of the consumer's surplus, rather than the whole of it. The monopolist divides the potential buyers into blocks, e.g., rich, middle class and poor, and sells the commodity in blocks. The monopolist sells its product first to the rich customers at the highest possible price. Once this part of the market is supplied, the firm lowers down the price for middle class buyers. Finally, bottom price is used for the poor class of buyers.

The second degree price discrimination is feasible where (*i*) the number of consumers is large and price rationing can be done, as in case of utility services like telephones, supply of water, etc.; (*ii*) demand curve for all the consumers is identical; (*iii*) a single rate is applicable for a large number of buyers. As shown in Fig. 11.6, a monopolist practicing second degree price discrimination, charges the highest price OP_1 for OQ_1 units and a lower price OP_2 for the next Q_1Q_2 units, and the lowest price OP_3 for the next Q_2Q_3 units. Thus, by adopting a block pricing system, the monopolist maximizes his total revenue (*TR*) at

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

Third degree: When a profit maximizing monopolist sets different prices in different markets having demand curves with different elasticities, he is practising the *third degree price discrimination*. It happens quite often that a monopolist has to sell his goods in two or more markets, completely separated from one another, each having a demand curve with different elasticity. A uniform price cannot be set for all the markets without losing profits. The monopolist is, therefore, required to find different price-quantity combinations that can maximize his profit in each market. For this purpose, he divides his total output between the market segments so that his *MC* = *MR* in each market, and fixes price accordingly.

For example, suppose that a monopolist has only two markets, *A* and *B*. The demand curve (D_a) and marginal revenue curve (MR_a) given in Fig. 11.7(*a*), represent the *AR* and *MR* curves in market *A*. D_b and MR_b in Fig. 11.7(*b*) represent the *AR* and *MR* curves in market *B*. The horizontal summation of D_a and D_b gives the total demand curve for the two markets, a shown by *AR* = *D* in Fig. 11.7(*c*) and the horizontal summation of MR_a and MR_b gives the aggregated *MR* [(Fig. 11.7(*c*)]. The firm's marginal cost is shown by *MC* that intersects *MR* at point *T*. Thus, the optimum level of output for the firm is determined at *OQ* at which *MR* = *MC*.
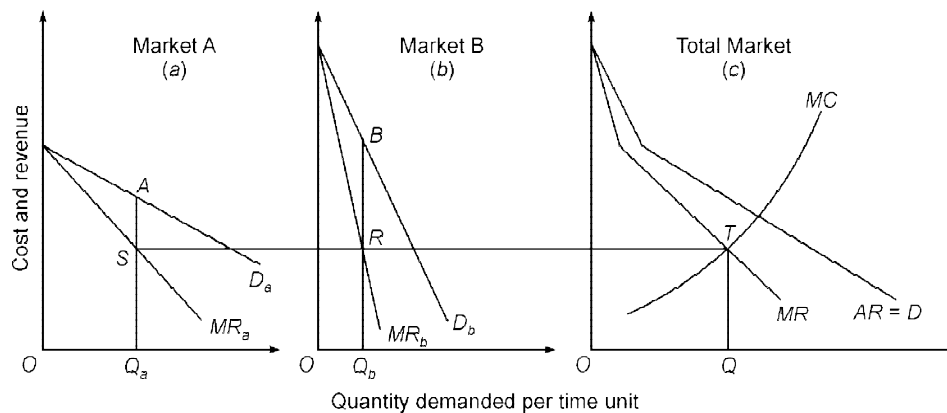
Fig. 11.7 Third Degree Price Discrimination

The problem that a monopolist faces is that the whole of his output *OQ* cannot be sold in any one of the markets at a profit maximizing price. Therefore, the monopolist has to allocate output *OQ* between the two markets in such proportions that the necessary condition of profit maximization is satisfied in both the markets, i.e., *MC* must be equal to *MR* in both the markets. This is accomplished by drawing a line from point *T* parallel to *X*-axis, through MR_b and MR_a. The points of intersection, *S* and *R* on curves MR_a and MR_b, respectively, determine the optimum share for markets *A* and *B*. As shown in the Fig. 11.7, the monopolist maximizes his profit in market *A* by selling OQ_a units at price AQ_a and in market *B*, by selling OQ_b units at price BQ_b. Note that $OQ_a + OQ_b = OQ$.

The third degree price discrimination may be suitably practised between any two or more markets separated from each other by geographical distance, transport barriers, cost of transportation and legal restrictions on the inter-regional or inter-state transportation of commodities by individuals.

Price Determination

The profit maximizing prices can now be obtained by substituting Q_a and Q_b with their estimated values (4 and 3, respectively) in price functions (11.4) and (11.5), respectively. The price for market *A* can be obtained as

$$P_a = 32 - 2\ Q_a = 32 - 2\ (4) = 24$$

and price for market *B* as

$$P_b = 22 - Q_b = 22 - 3 = 19$$

Thus, in market *A*, price = ₹ 24 and in market *B*, price = ₹ 19.

Profit Determination

Now that prices and sales for the two markets are known, total profit can be obtained by substituting numerical values for Q_a and Q_b in profit function (11.11). The profit function (11.11) is reproduced below.

$$\pi = 30\ Q_a + 20\ Q_b - 3Q_a{}^2 - 2Q_b{}^2 - 2Q_aQ_b - 10$$

By substituting 4 for Q_a and 3 for Q_b, we get

$$\pi = 30 \ (4) + 20(3) - 3(4)(4) - 2 \ (3)(3) - 2(4)(3) - 10$$
$$= 120 + 60 - 48 - 18 - 24 - 10 = 80$$

The total profit is ₹ 80. This profit satisfies the conditions of the maximum profit. It is, therefore, maximum.

11.5 PRICING AND OUTPUT DECISIONS UNDER MONOPOLISTIC COMPETITION

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

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

   (*i*) large number of sellers

  (*ii*) free entry and free exit

 (*iii*) perfect factor mobility

 (*iv*) complete dissemination of market information

  (*v*) differentiated product.

11.5.1 Monopolistic vs. Perfect Competition

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

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

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

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

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

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

11.5.2 Price and Output Decisions in the Short-Run

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

The short-term pricing and output determination under monopolistic competition is illustrated in Fig. 11.8. It gives short-run revenue and cost curves faced by the monopolistic firm.
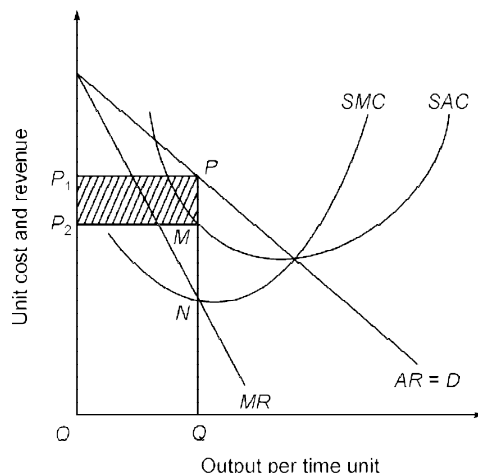
Fig. 11.8 *Price-Output Determination under Monopolistic Competition*

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

11.5.3 Price and Output Determination in the Long-Run

The mechanism of price and output determination in the long-run under monopolistic competition is illustrated graphically in Fig. 11.9. To begin the analysis, let us suppose that, at some point of time in the long-run, firm's revenue curves are given as AR_1 and MR_1 and long-run cost curves as *LAC* and *LMC*. As the figure shows, MR_1 and *LMC* intersect at point *M* determining the equilibrium output at OQ_2 and price at P_2Q_2. At price P_2Q_2, the firms make a supernormal or economic profit of P_2T per unit of output. This situation is similar to short-run equilibrium.
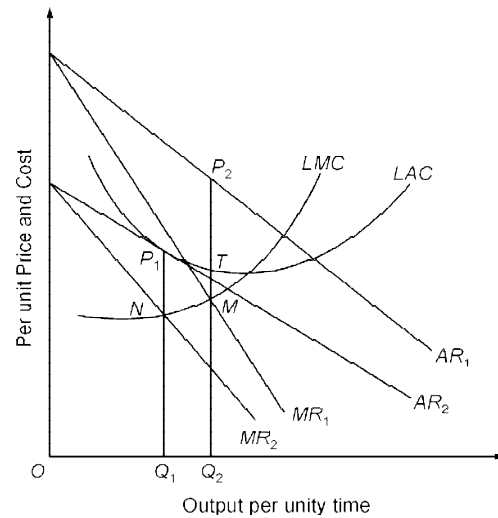
Fig. 11.9 *The Long-Run Price and Output Determination*
under Monopolistic Competition

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

First, the supernormal profit attracts new firms to the industry. As a result, the existing firms lose a part of their market share to new firms. Consequently, their demand curve shifts downward to the left until AR is tangent to LAC. This kind of change in the demand curve is shown is Fig. 11.9 by the shift in AR curve from AR_1 to AR_2 and the MR curve from MR_1 to MR_2.

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

The ultimate picture of price and output determination under monopolistic competition is shown at point P_1 in Fig. 11.9. As the figure shows, LMC intersects MR_2 at point N where firm's long-run equilibrium output is determined at OQ_1 and price at P_1Q_1. Note that price at P_1Q_1 equals the LAC at the point of tangency. It means that under monopolistic competition, firms make only normal profit in the long-run. Once all the firms reach this stage, there is no attraction (i.e., super normal profit) for the new firms to enter the industry, nor is there any reason for the existing firms to quit the industry. This signifies the long-run equilibrium of the industry.

Numerical Illustration

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

$$Q_1 = 100 - 0.5P_1$$

or $$P_1 = 200 - 2Q_1$$ …(11.1)

Given the price function (11.1), firms' TR_1 function can be worked out as

$$TR_1 = P_1 \cdot Q_1 = (200 - 2Q_1)Q_1$$
$$= 200Q_1 - 2Q_1{}^2$$ …(11.2)

The marginal revenue function (MR_1) can be obtained by differentiating the TR_1 function (11.2). Thus,

$$MR_1 = 200 - 4Q_1$$ …(11.3)

Suppose also that firms' TC function is given as

$$TC = 1562.50 + 5Q - Q^2 + 0.05Q^3$$ …(11.4)

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

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

$$= \frac{1562.50}{Q} + 5 - Q + 0.05Q^2$$ …(11.5)

We get firms' LMC function by differentiating its TC function (11.4). Thus,

$$LMC = 5 - 2Q + 0.15Q^2$$ …(11.6)

Let us now work out the *short-run equilibrium* levels of output and price that maximize firms' profit. The profit maximizing output can be obtained by equating MR_1 and LMC functions given in Eqs. (11.3) and (11.6), respectively, and solving for Q_1. That is,

$$MR_1 = LMC$$
$$200 - 4Q_1 = 5 - 2Q + 0.15Q^2$$ …(11.7)

For uniformity sake, let us replace Q in MC function as Q_1 and solve the Eq. (11.7) for Q_1.

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

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

Let us now find firms' equilibrium price (P_1), *LAC* and supernormal profit. Price P_1 can be obtained by substituting 30 for Q_1 in the price function (11.1).

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

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

Firms' *LAC* can be obtained by substituting equilibrium output 30 for *Q* in function (11.5). Thus,

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

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

Equilibrium output = 30

$$P_1 = 140$$
$$LAC = 72.08$$

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

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

$$Q_2 = 98.75 - P_2$$
and $$P_2 = 98.75 - Q_2 \qquad …(11.8)$$

To work out the long-run equilibrium, we need to find the new *TR* function (TR_2) and the new *MR* function (MR_2) corresponding to the new price function (11.8). For this, we need to first work out the new *TR* function (TR_2).

$$TR_2 = P_2 \cdot Q_2 = (98 \cdot 75 - Q_2) \ Q_2$$
$$= 98 \cdot 75Q_2 - Q_2^2$$
$$…(11.9)$$

We get MR_2 by *differentiating TR* function (11.9). Thus,

$$MR_2 = 98 \cdot 75 - 2Q_2 \qquad …(11.10)$$

The long-run equilibrium output can now be obtained by equating MR_2 with the *LMC* function (11.6). For the sake of uniformity, we designate *Q* in the *LMC* function as Q_2. The long-run equilibrium output is then determined where

$$MR_2 = LMC$$

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

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

$$625 = Q_2^2$$

$$Q_2 = 25$$

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

$$P_2 = AR_2 = LAC$$

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

By substitution, we get

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

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

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

11.6 NON-PRICE COMPETITION: SELLING COST AND EQUILIBRIUM

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

Incurring selling cost increases sales, but with varying degrees. Generally, sales increase initially at increasing rates, but eventually at decreasing rates. Consequently, the average cost of selling (*ASC*) initially decreases but ultimately it increases. The *ASC* curve is, therefore, *U*-shaped, similar to the conventional *AC* curve. This implies that total sales are subject to diminishing returns to increasing

selling costs. Non-price competition through selling cost leads all the firms to an almost similar equilibrium. Chamberlin calls it "Group Equilibrium." We discuss here Chamberlin analysis of firm's group equilibrium.

11.6.1 Selling Cost and Group Equilibrium

To analyze group equilibrium of firms with selling costs, let us recall that the main objective of all firms is to maximize their total profit. When they incure selling costs, they do so with the same objective in mind. All earlier assumptions regarding cost and revenue curves remain the same. The analysis of group equilibrium is presented in Fig. 11.10. Suppose *APC* represents firms' average production cost and competitive price is given at OP_3. None of the firms incurs any selling cost. Also, let all the firms be in equilibrium at point *E* where they make only normal profits.
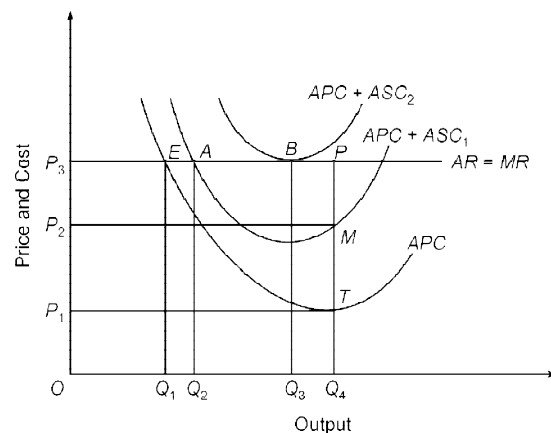
Fig. 11.10 *Selling Costs and Group Equilibrium*

Now suppose that one of the firms incurs selling cost so that its *APC* added with average selling costs (*ASC*) rises to the position of the curve $APC + ASC_1$ and its total sale increases to OQ_4. At output OQ_4, the firm makes supernormal profits of P_3PMP_2. This profit is, however, possible only so long as other firms do not incur selling cost on their products. If other firms do advertise their products competitively and incure the same amount of selling cost, the initial advantage to the firm advertising first disappears and its output falls to OQ_2. In fact, all the firms reach equilibrium at point *A* and produce OQ_2 units. But their short-sightedness compels them to increase their selling cost because they expect to reduce their *APC* by expanding their output. With increased selling cost, their $APC + ASC$ curve shifts further upward. This process continues until $APC + ASC$ rises to $APC + ASC_2$ which is tangent to the $AR = MR$ line. This position is shown by point *B*. Beyond point *B*, advertising is of no avail to any firm. The equilibrium will be stable at point *B* where each firm produces OQ_3 and makes only normal profit.

11.6.2 Critical Appraisal of Chamberlin's Theory

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

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

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

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

Fourthly, 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.

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

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

Despite the above criticism, Chamberlin's contribution to the theory of price cannot be denied. Chamberlin was the first to introduce the concept of *differentiated*

product and *selling costs* as a decision variable and to offer a systematic analysis of these factors. Another important contribution of Chamberlin is the introduction of the concept of demand curve based on market share as a tool of analyzing behaviour of firms, which later became the basis of the *kinked-demand* curve analysis.

11.7 PRICING AND OUTPUT DECISIONS UNDER OLIGOPOLY

In this section, we will discuss price and output determination under *oligopoly*. Let us first look at the market organization characterized by oligopoly.

11.7.1 Oligopoly: Definition, Sources and Characteristics

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

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

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

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

11.7.2 Sources of Oligopoly

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

1. *Huge capital investment.* Some industries are by nature capital-intensive, e.g., manufacturing automobiles, aircraft, ships, TV sets, computers, mobile phones, refrigerators, steel and aluminium goods, etc. Such industries require huge initial investment. Therefore, only those firms which can make huge investment can enter these kinds of industries. In fact, a huge investment requirement works as a natural barrier to entry to the oligopolistic industries.

2. *Economies of scale.* By virtue of huge investment and large scale production, the large units enjoy *absolute cost advantage* due to economies of scale in production, purchase of industrial inputs, market financing, and sales organization. This gives the existing firms a comparative advantage over new firms in price competition. This also works as a deterrent for the entry of new firms.

3. *Patent rights.* In case of *differentiated oligopoly,* firms get their differentiated product patented which gives them an exclusive right to produce and market the patented commodity. This prevents other firms from producing the patented commodity. Therefore, unless new firms have something new to offer and can match the existing products in respect of quality and cost, they cannot enter the industry. This keeps the number of firms limited.

4. *Control over certain raw materials.* Where a few firms acquire control over almost the entire supply of important inputs required to produce a certain commodity, new firms find it extremely difficult to enter the industry. For example, if a few firms acquire the right from the government to import certain raw materials, they control the entire input supply.

5. *Merger and takeover***.** Merger of rival firms or takeover of rival firms by the bigger ones with a view to protecting their joint market share or to put an end to waste of competition is working, in modern times, as an important factor that gives rise to oligopolies and strengthens the oligopolistic tendency in modern industries. Mergers and takeovers have been one of the main features of recent trend in Indian industries.

11.7.3 Features of Oligopoly

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

1. *Small number of sellers.* As already mentioned, there is a small number of *sellers* under oligopoly. How small is the number of sellers in oligopoly markets is

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

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

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

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

An illuminating example of strategic manoeuvering is cited by Robert A. Meyer. To quote the example, one of the US car manufacturing companies announced in one year in the month of September an *increase* of $ 180 in the price list of its car model. Following it, a few days later a second company announced an increase of $ 80 only and a third announced an increase of $ 91. The first company made a

counter move: it announced a *reduction* in the enhancement in the list price from $ 180 to $ 71. This is a pertinent example of interdependence of firms in business decisions under oligopolistic market structure. In India, when Maruti Udyog Limited (MUL), announced a price cut of ₹ 24,000 to ₹ 36,000 in early 2005 on its passenger cars, other companies followed suit. However, *price competition* is not the major form of competition among the oligopoly firms as price war destroys the profits. A more common form of competition is *non-price competition* on the basis of product differentiation, vigorous advertising and provision of survive.

3. *Barriers to entry.* Barriers to entry to an oligopolistic industry arise due to such market conditions as (*i*) huge investment requirement to match the production capacity of the existing ones, (*ii*) economies of scale and absolute cost advantage enjoyed by the existing firms, (*iii*) strong consumer loyalty to the products of the established firms based on their quality and service and (*iv*) preventing entry of new firms by the established firms through price cutting. However, the new entrants that can cross these barriers can and do enter the industry, though only a few, that too mostly the branches of *MNCs* survive.

4. *Indeterminate price and output.* Another important feature, though a controversial one, of the oligopolistic market structure is the indeterminateness of price and output. The characteristic fewness and interdependence of oligopoly firms makes derivation of the demand curve a difficult proposition. Therefore, price and output are said to be indeterminate. However, price and output are said to be determinate under collusive oligopoly. But, there too, collusion may last or it may break down. *An opposite view is that price under oligopoly is sticky*, i.e., if price is once determined, it tends to stabilize.

11.7.4 The Oligopoly Models: An Overview

As already mentioned, under oligopolistic conditions, rival firms indulge in an intricate pattern of actions, reactions and counter-actions showing a variety of behavioural patterns. As Baumol puts it, "Under [these] circumstances, a very wide variety of behaviour pattern becomes possible. Rivals may decide to get together and cooperate in the pursuit of their objectives,… or, at the other extreme, may try to fight each other to the death. Even if they enter an agreement, it may last or it may break down." The economists have, therefore, found it extremely difficult to make a systematic analysis of price and output determination under oligopoly. This has, however, not deterred the economists from their efforts to find an agreeable solution to the problem.

In accordance with the a wide variety of behavioural patterns, the economists have developed a variety of analytical models based on different behavioural

assumptions. The widely quoted oligopoly models include Cournot's duoply model (1838), Bertrand's leadership model (1880), Edgeworth's duoply model (1897), Stackelberg's model (1933), Sweezy's kinked demand curve model (1939), Neumann and Margenstern Game Theory model (1944) and Baumol's sales maximization model (1959). None of these models, however, provides a universally acceptable analysis of oligopoly, though these models do provide an insight into oligopolistic behaviour.

In this section, we have mentioned some selected oligopoly models with the purpose of showing the behaviour of oligopoly firms and working of the oligopolistic markets. These analytical models are selected on the basis of how price and output are determined under price competition, cartel system and the dilemma that oligopoly firms face in their price and output decisions. The famous oligopoly models are the following:

(*i*) Cournot's duopoly model,

(*ii*) Sweezy's kinked demand curve model,

(*iii*) Price leadership models:

(*a*) Price leadership by low-cost firm, (*b*) Price leadership by dominant firm and (*c*) Price leadership by barometric firm,

(*iv*) Collusive model: The Cartel Arrangement,

(*v*) The Game Theory model of oligopoly, and

(*vi*) Prisoner's Dilemma

11.7.5 A Classical Model of Duopoly: Cournot's Model

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

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

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

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

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

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

Cournot's duopoly model is presented in Fig. 11.11. The demand curve for mineral water is given by the *AR* curve and firm's *MR* by the *MR* curve. To begin with, let us suppose that there are only two sellers *A* and *B,* but initially, *A* is the only seller of mineral water in the market. By assumption, his *MC* = 0. Following the profit maximizing rule, he sells quantity *OQ* where his *MC* = 0 = *MR,* at price OP_2. His total profit is OP_2PQ.
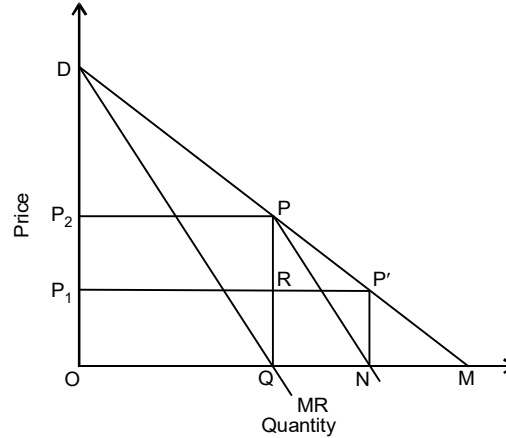
Fig. 11.11 *Price and Output Determination under Duopoly: Cournot's Model*

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

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

Now it is *B*′s turn to react. Following Cournot's assumption, *B* assumes that *A* will continue to supply only 3/8 of the market and the market open to him equals 1 – 3/8 = 5/8. To maximise his profit under the new conditions, *B* supplies

$1/2 \times 5/8 = 5/16$ of the market. It is now for *A* to reappraise the situation and adjust his price and output accordingly.

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

The actions and reactions and equilibrium of the sellers *A* and B, according to Cournot's model, are presented in Table 11.1.

Table 11.1 *Determination of Market Share*

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

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

Cournot's equilibrium solution is stable. For, given the action and reaction, it is not possible for any of the two sellers to increase their market share as shown in the last row of the table.

Cournot's model of duopoly can be extended to a general oligopoly model. For example, if there are three sellers in the industry, each one of them will be in equilibrium when each firm supplies 1/4 of the market. The three sellers together supply 3/4 of the total market, 1/4 of the market remaining unsupplied. Similarly, when there are four firms each one of them supply 1/5th of the market and 1/5th of the market remains unsupplied. The formula for determining the share of each seller in an oligopolistic market is: $Q \div (n + 1)$ where Q = market size, and n = number of sellers.

11.7.6 Bilateral Monopoly

Bilateral monopoly is said to exist in the factor market when there is a single buyer and a single seller of labour. This section analyses factor pricing and employment under the conditions that *labour is supplied by monopolist* (*a labour union*) *and demanded by a monopsonist*, a monopolist in the product market. We assume also that all buyers of a factor (labour) an Employers' Union which acts as a monopsonist in the labour market. On the supply side, labour is organised in a Labour Union which acts as a monopolist. Wage determination in this kind of market situation is generally analysed under **collective bargaining**. It is to be noted that bilateral monopoly has been discussed in detail in Unit 12.

Check Your Progress

1. List two characteristics of a perfect competitive market.

2. Mention the salient features of an oligopoly market.

3. Define price discrimination.

4. Name the two common forms of non-price competition.

11.8 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. Two characteristics of a perfect competitive market are the following:

   - Homogeneous products

   - Absence of collusion or artificial restraint

2. The salient features of an oligopoly market are the following:

   - Small number of sellers

   - Interdependence of decision-making

   - Barriers to entry

   - Indeterminate price and output

3. Price discrimination means selling the same or slightly differentiated product to different sections of consumers at different prices, not commensurate with the cost of differentiation.

4. The two common forms of non-price competition are product innovation and advertisement.

11.9 SUMMARY

- Under perfect competition, the number of sellers and buyers is very large. The number of sellers is so large that the share of each seller in total supply of a product is too small for a single seller to affect the market price by changing his supply.

- Sometimes a distinction is made between perfect competition and pure competition. The difference between the two is only a matter of degree. Perfect competition less perfect mobility of factors and perfect knowledge is regarded as pure competition.

- The term pure monopoly means an absolute power of a firm to produce and sell a product that has no close substitute. In other words, a monopolized market is one in which there is only one seller of a product having no close substitute.

- The emergence and survival of a monopoly firm is attributed to the factors which prevent the entry of other firms into the industry and eliminate the existing ones.

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

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

- Oligopoly is defined as a market structure in which there are a few sellers selling homogeneous or differentiated products. Where oligopoly firms sell a homogeneous product, it is called pure or homogeneous oligopoly.

- In accordance with the wide variety of behavioural patterns, the economists have developed a variety of analytical models based on different behavioural assumptions. The widely quoted oligopoly models include Cournot's duoply model (1838), Bertrand's leadership model (1880), Edgeworth's duoply model (1897), Stackelberg's model (1933), Sweezy's kinked demand curve model (1939), Neumann and Margenstern Game Theory model (1944) and Baumol's sales maximization model (1959).

11.10 KEY WORDS

- **Pure monopoly:** This implies an absolute power of a firm to produce and sell a product that has no close substitute.

- **Monopolistic competition:** It is defined as market setting in which a large number of sellers sell differentiated products.

- **Oligopoly:** It is defined as a market structure in which there are a few sellers selling homogeneous or differentiated products.

11.11 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short Answer Questions

1. What is the difference between perfect competition and pure competition?
2. Give examples of price discrimination.
3. Write a short note on selling cost and group equilibrium.
4. What are the main sources of monopoly?
5. Prepare an overview of the oligopoly models.
6. Briefly mention Cournot's Duopoly model.

Long Answer Questions

1. 'The economies of scale are a primary and technical reason for the emergence and existence of monopolies in an unregulated market.' Analyse the statement.
2. Graphically represent the determination of price and output by a monopoly firm in the short-run.
3. Discuss price discrimination by degrees.
4. Distinguish between monopolistic competition and perfect competition.
5. Critically analyse Chamberlin's theory of monopolistic competition.

11.12 FURTHER READINGS

Dwivedi, D. N. 2008. *Principles of Economics*, Seventh Edition. New Delhi: Vikas Publishing House.

Weil. David N. 2004. *Economic Growth*. London: Addison Wesley.

Thomas, Christopher R. and Maurice S. Charles. 2005. *Managerial Economics: Concepts and Applications*, Eighth Edition. New Delhi: Tata McGraw-Hill Publishing Company Limited.

Mankiw, Gregory N. 2002. *Principles of Economics*, Second Edition. India: Thomson Press.

NOTES

BLOCK - IV
THEORY OF WAGES AND ECONOMIC ANALYSIS

UNIT 12 THEORY OF WAGE DETERMINATION

Structure

12.0 Introduction
12.1 Objectives
12.2 Theory of Wages: An Overview
     12.2.1 Static and Dynamic Wage Differentials
     12.2.2 Labour Demand Curve with Increasing Marginal Productvity
     12.2.3 Collective Bargaining
12.3 Wage Policy and Regulation
     12.3.1 Objectives of National Wage Policy
     12.3.2 Regulations Adopted by the State
12.4 Answers to Check Your Progress Questions
12.5 Summary
12.6 Key Words
12.7 Self Assessment Questions and Exercises
12.8 Further Readings

12.0 INTRODUCTION

In this unit, 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.

The theory of wage determination in these kinds of market structure has been discussed in the subsequent section of this unit.

12.1 OBJECTIVES

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

- Discuss the theory of wages
- Analyse the concept of marginal productivity with reference to wage determination
- Define collective bargaining
- Explain wage differentials

12.2 THEORY OF WAGES: AN OVERVIEW

Let us now study the theory of wages under various market conditions.

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. The labour demand and supply curves can be used straightaway to explain and illustrate wage determination under perfect competition in both product market and labour market. Wage determination in this kind of market structure is illustrated in Fig. 12.1.
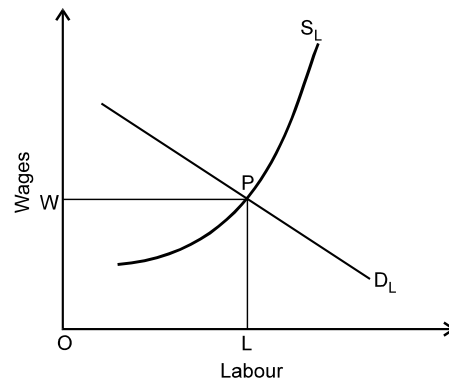
Fig. 12.1 *Determination of Wages in a Perfectly Competitive Market*

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.

12.2.1 Static and Dynamic Wage Differentials

The theory of wage determination discussed above assumes that the units of a factor (labour) are homogeneous. If all the units of labour are identical and if non-monetary advantages are the same in all uses, then the wage rate for each labour tends to be the same in a perfectly competitive market. In the real world, however, the labours are neither identical nor homogeneous. Nor are the different workers paid the same wage rate. In this section, we discuss the reasons for wage differentials.

The reasons for wage differentials may be classified under two groups:

   (i) **Dynamic (or disequilibrium) causes**; and

  (ii) **Static (or equilibrium) causes**.

(a) Dynamic Wage Differentials

Dynamic wage differentials are those which arise due to disequilibrium in commodity and labour markets. The dynamic wage differentials act on the demand and supply conditions of labour to restore the equilibrium and, thereby, remove the wage differences. Such differentials are therefore temporary and exist only till the disequilibrium persists. For example, suppose that there are only two industries, *A* and *B,* in a country, which use only one factor of production, i.e., labour (*L*). Both industries are in equilibrium and pay the same wage rate (*W*) to the labour. Suppose, for some reason, the demand for product of industry *B* decreases. As a result, price of *B*'s product falls and VMP_L curve shifts leftward, causing a fall in the wages paid by industry *B,* while industry *A* continues to pay the same wage rate (*W*). Wage rate in industry *B* will be lower than that in industry *A*. **This is called dynamic wage differential**.

The dynamic wage differentials are self-adjusting, especially in case of homogeneous labour. Due to lower wage rate in industry *B,* labour tends to move to industry *A* in which wages are relatively higher. Shift of labour to industry *A* continues until wages in industry *A* fall to the level in industry *B*. Thus, in course of time, a new equilibrium is reached and factor price (wage) differential disappears.

Whether a new equilibrium is reached and wage differences disappear depends on the *factor mobility.* If factors are immobile, the factor-price differentials persist. If factors are mobile, the factor-price differences disappear over time. How long does it take for wage differentials to disappear? It depends on the degree of factor mobility. The greater the factor mobility, the quicker the factor price equalisation in various industries, and *vice versa.*

(b) Static Wage Differentials

The static wage differentials are those that persist in the state of labour market equilibrium. Such differences are not removed by the competitive forces of the market. Wage differentials of this type arise mainly due to the following reasons:

   (i) qualitative differences in labour, i.e., non-homogeneity of labour;

(ii) difference in the nature of occupations in which labour is employed;

(iii) differences in the prices of product produced by labour; and

(iv) market imperfections.

We will explain first the static wage differentials caused by **qualitative differences in labour** in some detail and then discuss briefly the wage differentials caused by **other factors.**

Wage Differentials Due to Non-homogeneity of Labour

Labour is the most non-homogeneous factor of production. Therefore, wage differentials are far more common than the other factor price differentials. Wage differentials between the two broad groups of labour—*skilled* and *unskilled*–are more obvious than those within the groups. The nature of wage differentials between the skilled and unskilled labour is illustrated in Fig. 12.2(a) and (b). The total supply of both unskilled and skilled labour is assumed to be perfectly *inelastic.* The initial supply curve of unskilled labour is shown by S_u line in Fig. 12.2(a) and that of skilled labour by S_s line in Fig. 12.2(b). The market demand curve for unskilled and skilled labour are, respectively, shown by D_u and D_s curves. Consider first Fig. 12.2(a). Given the D_u and S_u curves, the unskilled labour market is in equilibrium at point E_u and wage rate for unskilled labour is determined at OW_u. At this wage rate (OW_u), the total employment of unskilled labour is ON_u.
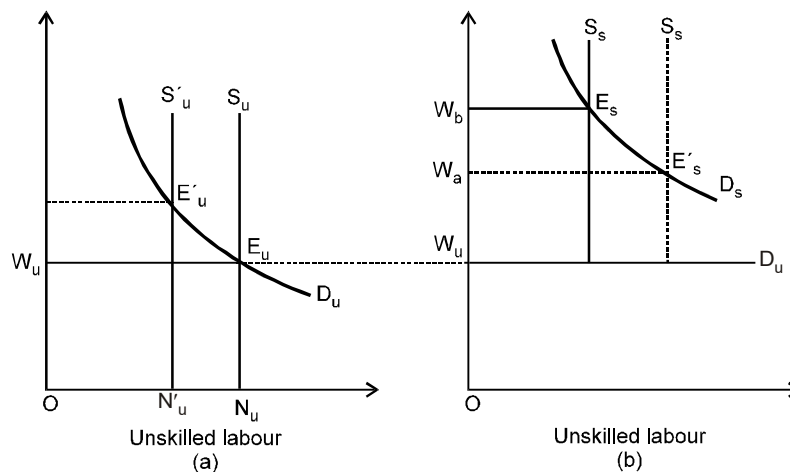
Fig. 12.2 *Wage Differential: (a) Unskilled; and (b) Skilled Labour*

Note that at wage rate OW_u [Fig. 12.2(b)], the supply of skilled labour is zero as no skilled labour offers to work at this wage rate. The wage rate for skilled labour will be determined at a much higher level. Given the demand curve D_s and supply curve S_s, the market wage rate for skilled labour is determined at OW_s, which is greater than the wage rate (OW_u) for unskilled labour. This wage differentials between the skilled and unskilled labour are of *static* nature.

The wage differentials may narrow down or widen over time. If higher wage rates for skilled labour encourage people to acquire skill through training and

education and the supply of skilled labour increases. As a result, skilled labour supply curve will shift rightward to S_s' [Fig. 12.2(b)], and wage rate will tend to settle at point E_s', provided there is no barrier to entry into the skilled labour market. This will reduce the wage differential. If supply of skilled labour increases because workers of unskilled group acquire training and skill, the labour supply curve of unskilled labour [Fig. 12.2(a)] will shift leftward to S_u' causing a rise in the unskilled labour rate wage point E_u'. This will further reduce the wage differentials. Note that the difference between unskilled-labour-wage rate OW_u and skilled-labour-wage rate OW_s, narrows down as shown in Fig. 12.2(b). Note that wage difference decreases from W_sW_u to W'_sW_u.

Monopoly in Commodity Market and Perfect Competition in labour Market

In this section, we discuss wage rate determination in a situation in which a monopolist in the commodity market hires labour from a perfectly competitive labour market.

Let us assume that the monopolist uses a single variable factor, labour, to produce a commodity, X. He hires labour from perfectly competitive labour market in which labour supply is perfectly elastic and wage rate is given. However, being a monopolist in the commodity market, the firm faces a downward sloping AR and MR curves. Since a monopolist can sell additional quantity only at decreasing prices, VMP_L curve derived under perfectly competitive conditions does not represent the labour demand curve of a monopoly seller in the commodity market. For a monopolist, the basis of demand for labour is not the value of marginal product curve (VMP_L), but the marginal revenue product curve (MRP_L). The MRP_L is defined as $MRP_L = MP_L \cdot MR_x$. Note that, as shown in Fig. 12.3, monopolist MRP_L is not the same as the VMP_L. The difference between MRP_L and VMP_L is illustrated numerically in Table 12.1.

Table 12.1 *The MRP_L and VMP_L Schedules*

Units of variable input (labour)	Total output (Q_x)	Marginal Product (MP_L)	Selling price (P_x)	Total revenue ($TR = Q_x \cdot P_x$) (₹)	Marginal revenue ($MR = \Delta TR \div MP_L$)	$MRP_L = (MP_L \times MR)$	$VMP_l = (MP_L \cdot P_x)$
1	7.00	10	70	70	10	70	70
2	13.50	6.50	9	122	8	52	58
3	19.00	5.50	8	152	6	33	44
4	23.00	4.00	7	161	2	8	28
5	24.00	1.00	6	144	–17	–17	–06
6	24.00	0.00	5	120	∞	—	0.00

*Note***:** Figures in the fifth column onwards are rounded off.

It can be seen from the last two columns of the table that, except at the initial stage, MRP_L is not the same as VMP_L. While both MRP_L and VMP_L tend to decrease, MRP_L falls at a much faster rate than the VMP_L. The difference between MRP_L and VMP_L is more clearly brought out in Fig. 12.3. It can be seen that VMP_L curve lies above the MRP_L curve at all the levels of employment. The reason is that $P_x > MR_x$ at all levels of employment and output. The common feature of VMP_L and MRP_L is that both have a negative slope though the slope of VMP_L is less than that of MRP_L, i.e.,

$$\frac{\Delta VMP_L}{\Delta L} < \frac{\Delta MRP_L}{\Delta L}$$
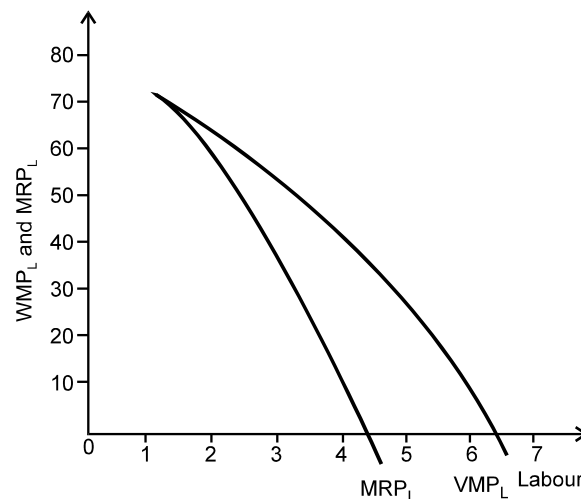
Fig. 12.3 MRL_L and VMP_L *Schedules of Variable Factor (Labour)*

Monopoly Demand for Labour—A Single Variable Factor

Let us recall the assumption that the monopoly firm hires labour from a perfectly competitive labour market. Since labour market is perfectly competitive, a monopoly firm faces a perfectly elastic labour supply curve, as shown by the curve S_L in Fig. 12.4. Recall that given wage rate (W), $MC_L = \overline{w}$. As regards the monopoly firm's demand for labour, **the curve MRP_L represents its demand curve for labour.** A profit maximising monopoly firm will hire labour up to the point where its MRP_L equals marginal cost of labour (MC_L). Since wage rate $(\overline{w})$ remains constant in a perfectly competitive labour market, the monopolist will employ labour up to the point where $MRP_L = MC_L = \overline{w}$.
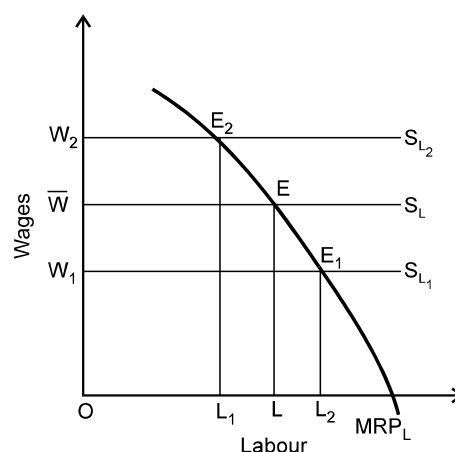
Fig. 12.4 *Monopoly Demand for Labour*

As Fig. 12.4 shows, MRP_L and S_L curves intersect at point E. At this point, $MRP_L = \overline{W} = MC_L$. Point E, therefore, determines the equilibrium demand for labour, i.e., OL. In other words, the profit maximising monopoly will employ OL units of labour because at this level of employment, its $MRP_L = MC_L = \overline{W}$. By employing OL units of labour, the monopolist reaches equilibrium at point E. When wage rate decreases to OW_1, the monopolist employs OL_2 units of labour. Similarly, when wage rate rises to OW_2, the firm will employ OL_1 units of labour. The conclusion that can be drawn from this relationship is that as the wage rate decreases (increases) the demand for labour increases (decreases) along the MRP_L. Therefore, the MRP_L curve is monopoly's demand curve for labour, if labour is the only variable factor used by the monopoly firm.

Monopoly Demand for Labour With More than One Variable Factor

When more than one variable input or factor is used in the process of production, the VMP_L curve is not the demand curve for labour. The same logic applies to the monopoly firm's demand for labour too. That is, **when more than one variable input is used by a monopolist, the marginal revenue product curve (MRP_L) is not the monopoly's demand curve for labour**. The process of deriving monopoly demand curve for labour when the monopolist uses more than one variable input is also the same as discussed. We briefly repeat the process below.

Let us suppose that monopolist's MRP_L curve and labour supply curve (SL_2) are given as in Fig. 12.5. Given the wage rate OW_2, monopolist is in equilibrium at point E_1 where he employs OL_1 units of labour. Now let the wage rate fall to OW_1. All other things remaining the same, the monopolist should have moved down along the MRP_{L_1} to point F. But other things do not remain the same because of interdependence of factor demand. As noted earlier, a fall in the wage rate from OW_2 to OW_l causes three kinds of effects: **Substitution effect, output effect** and **profit maximisation effect**.
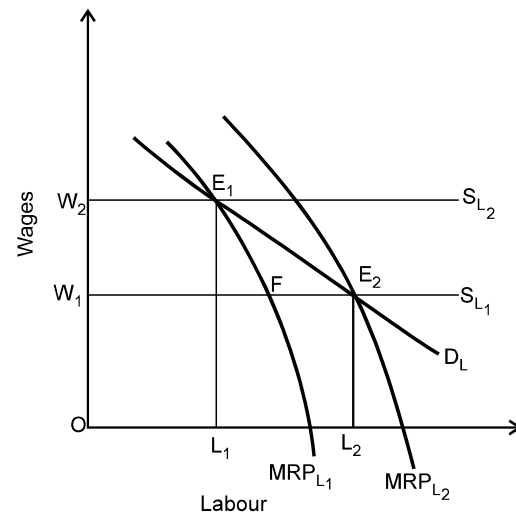
Fig. 12.5 *Derivation of Monopoly Demand for Labour when other Variable Inputs are Used*

The output and profit maximisation effects generally cause a rightward shift in the MRP_L, for example, from the position of MRP_{L_1} to that of MRP_{L_2} (Fig. 12.5). The MRP_{L_2} intersects SL_1 at point E_2. Therefore, the profit maximising monopolist moves to a new equilibrium point E_2. By joining the equilibrium points E_1 and E_2, we get the monopolist's demand curve for labour when it uses several inputs. The curve D_L represents the monopoly demand curve for labour when more than one variable input is used.

Comparison of Market Demand for Labour Under Monopoly and Monopolistic Competition

Recall that market demand curve for a variable factor (labour) is the horizontal sum of the individual demand curves of various competitive firms. Since in case of monopoly there is a single firm, there would be no external or industry effect resulting from a decrease in the price of labour. Therefore, in case of a monopoly firm hiring labour in a competitive labour market, the market demand curve for labour is not the horizontal summation of individual labour demand curves.

But in case of oligopolistic and monopolistic competitors, the market demand curve for a variable input (labour) is the summation of the individual demand curves for it. The reason is that when all oligopolistic or monopolistic firms expand their output, the market price of the output falls. As a result, the individual demand curve and marginal revenue curve for the commodity shift to the left. Due to leftward shift in the demand (AR) and marginal revenue curves, the individual demand curves for the variable input shifts leftward. As such, the procedure of deriving market demand curve for a variable factor is exactly the same as in case of perfectly competitive industry.

Determination of Factor Price (Wages) and Employment

When product market is characterised by monopoly or monopolistic competition and the variable factor (labour) market is perfectly competitive, the *factor price*

(wage rate) and employment are determined in the same fashion as they are determined under perfectly competitive market. That is, market equilibrium *price and employment of the variable factor labour are determined by the intersection of market demand curve for and market supply curve of the factor.* Thus, whether product market is monopolistically or perfectly competitive, the analysis of equilibrium price of a variable factor and its employment is the same. There is, however, one important difference between the two. In case of perfectly competitive product-market, the market demand curve for labour is based on its VMP_L, but in case of monopoly and monopolistic competition, the market demand curve for labour is based on its MRP_L.

Monopolistic Exploitation of Variable Factor (Labour)

As noted above, market demand curve for the variable factor, labour, in case of monopoly in the product market, is based on its MRP_L rather than on its VMP_L. This leads to **monopolistic exploitation of labour**. There is exploitation because labour is paid a price equal to its *MRP* which is less than its *VMP*. According to Joan Robinson, a productive factor (labour) is exploited if it is paid a price less than the value of its marginal product (*VMP*). Robinson's analysis of monopolistic exploitation of labour is presented below.
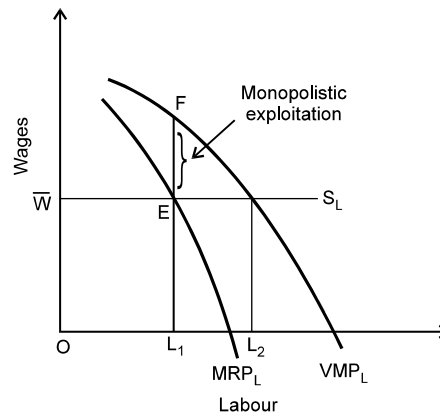
Fig. 12.6 *Exploitation of Labour by a Monopoly Firm*

The exploitation of labour (a variable factor) by an individual monopoly firm is illustrated in Fig. 12.6 and by the monopolistic firms as a group in Fig. 12.7. As shown in Fig. 12.6, given the MRP_L and S_L curves, a profit maximising monopolist will be in equilibrium at point E, employ OL_1 units of labour and pay wage $O\overline{W}$ $(= EL_1)$. But, under prefect competition in the product market, OL_1 units of labour would be demanded at wages FL_1, VMP_L being the relevant labour demand curve. Or else, the employment will be OL_2 at wage rate $O\overline{W}$ $= EL_1$. Thus, the difference between monopoly wage rate (EL_1) and competitive wage

rate (FL_1), i.e., $FL_1 - EL_1 = EF$) is the extent of **monopolistic exploitation** of labour. Besides, the monopolist restricts employment of labour to OL_1 units whereas the perfectly competitive firm would have employed OL_2 units of labour. The lower level of employment by a monopolist also results in loss of output.

Figure 12.7 illustrates the case of *monopolistic* exploitation of a variable factor (labour) at the market level. Let us suppose: (i) that curve D_m represents the market demand curve for labour by the monopolistic firms; (ii) that curve D_c represents the market demand curve for labour by the perfectly competitive firms; and (iii) that curve S_L represent the market supply curve of labour. In case of monopoly and monopolistic competition, the labour market will be in equilibrium at point E_M, the wage rate will be OW_M and employment will be OL_M units of labour. But, in case product market is perfectly competitive, the labour market will be in equilibrium at point E_C which determines wage rate at OW_C and employment at OL_C units of labour. Note that wage rate (OW_M) under monopolistic competition in the product market is much less than wage rate OW_C under a perfectly competitive product market. Thus, according to Joan Robinson, $OM_C - OW_M = W_C W_M$ is the extent of **monopolistic exploitation.**
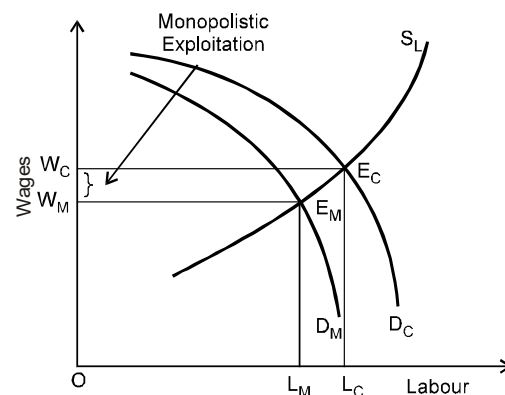
Fig. 12.7 *Exploitation of Labour under Monopolistic Competition*

This view of Joan Robinson on monopolistic exploitation has however been questioned on the ground that lower wage payment is inevitable because of divergence between MRP_L and VMP_L. The MRP_L is lower than VMP_L (at all levels of employment) not because of monopoly powers of the monopolistic sellers but because of **product differentiation**. Product differentiation creates brand loyalty which makes the demand curve slope downward to the right. In case of a downward sloping demand curve, there is bound to be a diversion between the price (AR) and the marginal revenue (MR), marginal revenue being lower than the price. Since all firms, whether in perfect or in imperfect market, attempt at profit maximisation, monopoly firms pay labour a wage rate that equals MRP_L. Therefore, the difference between OW_C and OW_M cannot be considered as exploitation. The difference arises due to the market conditions. However, if product differentiation is excessive and commodities are imposed on the consumers by the monopolistic sellers, then the argument of monopolistic exploitation may be acceptable.

Wage Determination Under Monopoly In product and Monopsony in Labour Market

In this section, we discuss wage determination in a market setting in which there is monopoly in the commodity market and monopsony in the labour market. When there is a single buyer of labour, there exists **monopsony** in the labour market. Thus, in the present model, the monopolist is also a monopsonist in the factor market. It may be **noted** here that the analysis of determination of factor-price and employment under the conditions of monopoly in the commodity-market and monopsony in the factor market is relevant also for monopolistic competition and oligopoly in the commodity-market, and monopsony and oligopsony in the factor market.

Factor price and employment determination in the market setting described above is discussed under two different assumptions: (i) that only a single variable factor (labour) is used; and (ii) that more than one variable factor is used. Before we analyse the factor price and employment determination, let us explain the **concept of marginal cost of factor input** which is used in the analysis of factor price determination.

Marginal Cost of a Variable Factor

Factor price and employment under product monopoly and factor monopsony are determined by intersection of labour demand curve and the marginal-cost-of-labour or *MW* curve. The monopsonist's demand curve for labour is given by the MRP_L curve. Let us now explain the **concept of marginal cost of variable factor**.

A monopsonist in labour market faces a positively sloping labour supply curve, i.e., more labour is supplied at increasing wage rate. Due to positive slope of the labour supply curve, there is a divergence between the average and marginal costs of labour. The monopsonist must therefore consider his **marginal cost of labour**, i.e., marginal wage (*MW*) to decide on the units of labour to be employed. The marginal cost of labour may be computed as shown in Table 12.2.

Table 12.2 Total and Marginal Cost of Labour

Units of Labour	Average cost of labour (AW) (₹)	Total expen- diture on labour (TW) (₹)	Marginal cost of labour (MW) = ΔTW)(₹)
1	5	5	—
2	6	12	7
3	7	21	9
4	8	32	11
5	9	45	13
6	10	60	15
7	11	77	17

It can be observed from the table that marginal cost of labour is greater than the average wage rate at all levels of employment. Figures in the 2nd and 4th columns when graphed, as in Fig. 12.8, give the normal labour supply curve (*AW*) and marginal cost of labour (*MW*) curve. The *AW* curve shows the labour supply for the monopsonist and *MW* curve shows its marginal cost at different levels of labour employment.
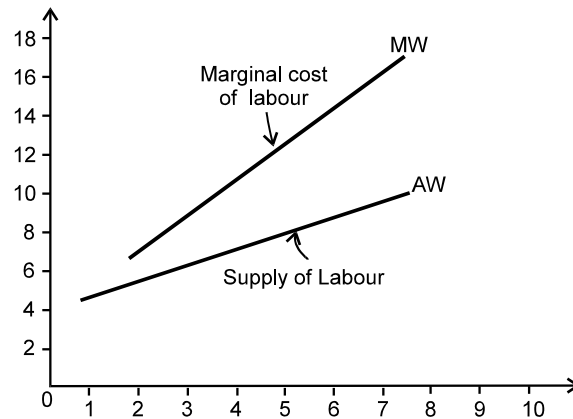
Fig. 12.8 *Marginal Cost of Labour*

12.2.2 Labour Demand Curve with Increasing Marginal Productvity

The derivation of demand curve for a variable factor (labour) when capital is also a variable factor is illustrated in Fig. 12.9. 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, substitution and profit maximisation effects. The VMP_{L_2} intersects the new labour to supply curve SL_1 at point E_2. The point E_2 is therefore be the new 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).
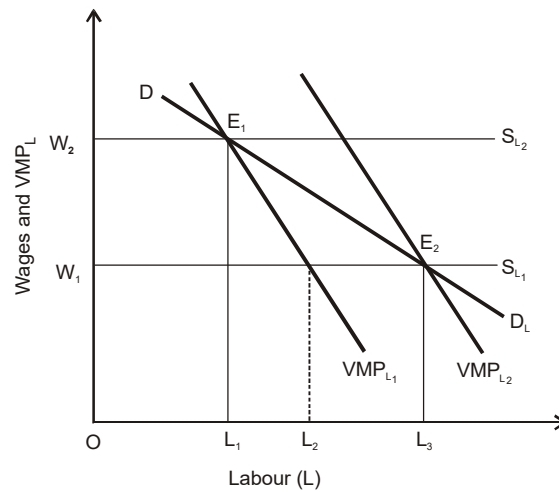
Fig. 12.9 *Labour Demand Curve when both L and K are Variable*

Wages and Employment Under Monopsony

Having derived the *AW* and *MW* curves, we can now explain wage and employment determination under monopsony. As already mentioned, wage rate and employment under monopsony are determined by the intersection of monopsonist's demand curve for labour and marginal cost curve of labour. When a monopsonist uses only one variable factor (labour), MRP_L curve is its demand curve for labour and *MW* curve is its labour supply curve which is the same as marginal cost curve of labour, as shown in Fig. 12.8.
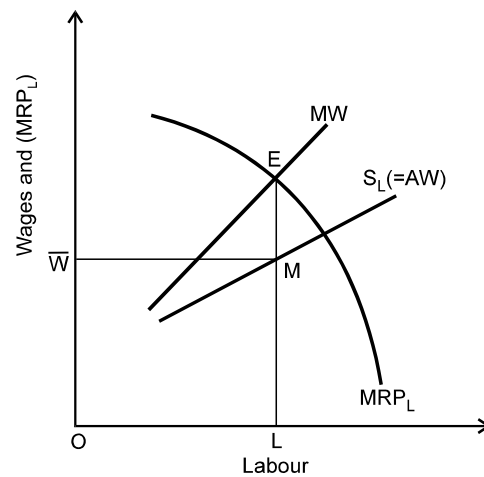
Fig. 12.10 *Wages and Employment Determination under Monopsony*

The determination of wage rate and employment under monopsony with a single variable factor (labour) is illustrated in Fig. 12.10. The MRP_L curve represents monopsonist's demand curve for labour and *MW* curve represents its marginal cost. A monopsonist employs labour up to the point at which the marginal revenue product of labour (MRP_L) equals its marginal cost of labour (*MW*), i.e., where

$MRP_L = MW$. In Fig. 12.10, the MRP_L and MW curves intersect each other at point E at which $MRP_L = MW$. The monopsonist firm is therefore in equilibrium at point E where it employs OL units of labour. Thus, OL is the equilibrium level of labour employment. Given the labour supply curve (AW), the equilibrium wage rate (corresponding to the equilibrium level of employment (OL) is $O\overline{W}$ ($= ML$).

Monopsonist Exploitation of Labour

Determination of wage rate and monopsonistic exploitation of labour are both analysed simultaneously in this section. Recall that monopolistic exploitation results from the fact that commodity demand curve for a monopolist firm has a negative slope, and hence, its $MR < P$. The profit maximising monopolist firm must pay for a factor service according to their MRP ($= MPP . MR$) which is less than their VMP ($= MP.P$). The level of monopolistic exploitation equals the difference between VMP and MRP. **Monopsonistic exploitation** arises for reasons similar to monopolistic competition. But monopsonistic exploitation is greater than the monopolistic exploitation. A monopsonist pays a price to a factor which is less than not only its VMP but also less than its MRP. This gives rise to the *monopsonistic exploitation* which results from the monopsonistic power of the firm. The extent of monopsonistic exploitation of labour may be measured by comparing the wage rate in perfectly competitive and labour markets with the wage rate under monopolistic product market and monopsonistic product labour market conditions. A comparative analysis of the two wage rates is presented in Fig. 12.11.

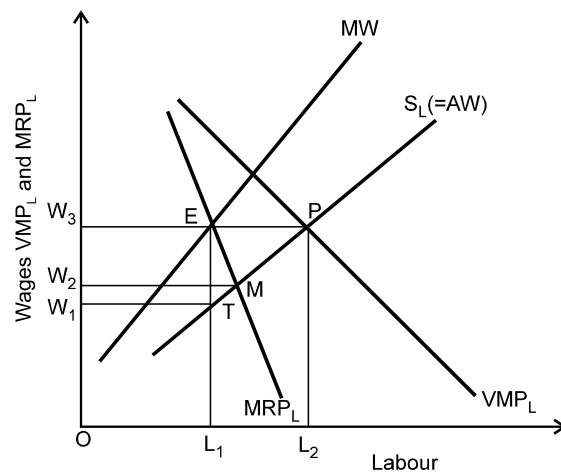**Fig. 12.11** *Monopsonistic and Monopolistic Exploitation of Labour*

When both product and labour markets are perfectly competitive, the curve VMP_L represents the industry or market *demand curve* for labour, and the curve S_L ($= AW$) represents the market *supply of labour*. Labour demand and supply curves intersect at point P determining the wage rate at OW_3—which equals VMP_L, i.e., $w = VMP_L$.

Now, let commodity market be monopolistic while labour market remains perfectly competitive. The market demand curve for labour now is MRP_L curve which is the sum of individual demand curves of the monopolists. The MRP_L curve intersects the labour supply curves, S_L at point *M,* determining wage rate at OW_2. Thus, labour market reaches a new equilibrium point (*M*) where wage rate is determined at OW_2. Note that monopoly wage rate, OW_2, is less than the competitive wage rate, OW_3. The difference between OW_3 and OW_2, (i.e., $OW_3 - OW_2 = W_2W_3$) is the **monopolistic exploitation** of labour. Besides, when there is monopoly in the commodity market, employment of labour decreases from OL_2 to OL_1. Thus, the effect of monopolistic exploitation is a lower level of employment at a lower wage rate.

Let us now introduce monopsony in the labour market while commodity market remains monopolised. This is the category of market organisation with which we are mainly concerned in this section. The monopsonist must employ labour until $MRP_L = MW$, the marginal cost of labour. As Fig. 12.11 shows, MRP_L and *MW* curves intersect at point *E* which determines the equilibrium level of employment at OL_1. The ordinate EL_1 intersects the labour supply curve S_L at point *T,* which determines the equilibrium wage rate for the monopsonist at OW_1. Thus, the wage rate under monopsony in labour market goes further below the competitive wage rate. The difference between the competitive wage rate, OW_3, and the monopsony wage rate, OW_1 measures the **monopsonistic exploitation** of labour. That is, $OW_3 - OW_1 = W_1W_3$ is the monopsonistic exploitation.

Monopsonistic exploitation W_1W_3 may be split into two parts W_2W_3 and W_1W_2. The exploitation W_2W_3 is attributable to monopoly power in the commodity market. This part of factor exploitation is not unique to the monopsonist. But remaining part, W_1W_2, is unique to the monopsonist. Thus, the main feature of the monopsonistic exploitation is that each factor is paid a price less than even its *MRP.*

Perfect Competition in Commodity Market and Monopoly in Labour Market (Labour Union)

Wage determination has so far been analysed assuming perfect competition in the labour market. The labour markets are however generally imperfect due mainly to growth of **labour unions**. In this section, we introduce **labour union** to the labour market and assume that the union has control over the total labour supply of an area or to an industry. Here, we explain the wage determination in a model assuming monopoly in the factor market and perfect competition in the commodity market.

The effect of monopoly in the labour market and factor price (wage) determination are illustrated in Fig. 12.12. The monopoly labour supply curve is shown by curve $S_L = MC_L$. Note that the curve $S_L = MC_L$ shown in this model represents the marginal cost of labour, not the average cost of labour. The curve $D_L = VMP_L = AR_L$ represents the demand curve for labour by the competitive

firms. Individual labour demand curve is assumed to be horizontal. This curve also represents the average revenue curve for the labour union. Under competitive conditions, the labour market equilibrium is determined at point E with equilibrium wage rate determined at OW_c and equilibrium employment OL_c.
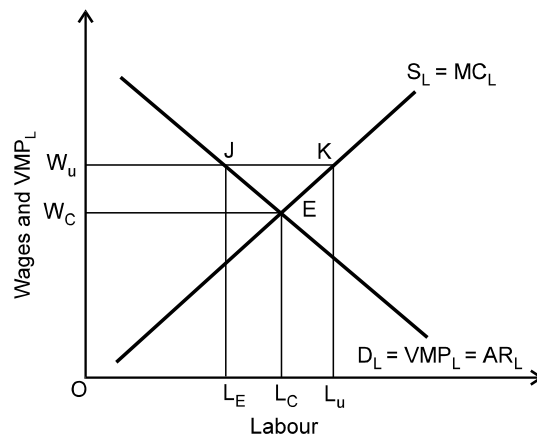
Fig. 12.12 *Wage Rate under Monopoly in Labour Market*

Now suppose that labour market is unionised and monopolised by the union. Suppose also that union bargains for higher wage rate and succeeds in getting wage rate raised to OW_u. The supply of labour at this wage rate is OL_u. Thus, the rise in wage rate creates a perfectly elastic labour supply curve up to OL_u. Therefore, the new labour supply curve is W_uKS_L. With the emergence of the new supply curve, the new equilibrium takes place at point J and equilibrium level of employment is OL_E. As is clear from Fig. 12.12, the union has succeeded in raising the wage rate above its competitive level, OW_c, but at the same time has caused a fall in employment from OL_c to OL_E. Since union creates excess supply of labour to the extent of $JK = L_EL_u$, and if unemployed workers are prepared to work at a wage rate lower than the union wage rate (W_u), there is a possibility of wage cutting, unless union wage rate is strictly enforced.

12.2.3 Collective Bargaining

It may be noted at the outset that economic theory does not predict a precise outcome of collective bargaining. The outcome of collective bargaining is not certain. That is, *solution to a bilateral monopoly situation is indeterminate*. The economic analysis of bilateral monopoly (or collective bargaining) does, however, bring out the upper and lower limits within which factor price (e.g., wage rate) can be determined through the process of collective bargaining. The determination of wage rate, ultimately depends on a number of factors, e.g., bargaining powers and skills, economic and political power of labour unions and of employer's association, the effect of government intervention, etc. We will first show the **indeterminateness** of equilibrium under bilateral monopoly and then analyse the effects of labour unions in respect of **monopsonistic exploitation**.

Indeterminate Wages Under Bilateral Monopoly

The indeterminateness of equilibrium under bilateral monopoly is illustrated in Fig. 12.13. The monopsonist's demand curve for labour is given by $D_L = MRP_L$. The monopolist (i.e., the labour union) considers $MRP_L = AR_u$ curve as its average revenue (AR_u). Given the AR_u curve, the marginal revenue curve of the labour union may be derived as MR_L—it must bisect the labour market midway. The labour supply curve from the monopsonist's point of view is given by $S_L = AW_m = MC_u$ and marginal cost of labour curve by MW.

Let us now consider the situation from the viewpoint of the monopsonist, the employers' association. The monopsonist must employ labour where $MRP_L = MW$ in order to maximise profit. Thus, given the cost and revenue curves, the monopsonist finds his equilibrium at point E where he would employ OL_u units of labour and pay wages OW_m.

On the other hand, the monopolist, the labour union, maximises its gains (earning per unit of time) where its marginal cost equals its marginal revenue (MR_L). The marginal cost of labour from labour union's point of view is the same as S_L curve. For the monopolist, the marginal cost of labour equals the marginal revenue (OW_u) at employment OL_m. Given the labour demand curve D_L, the labour union would press for wage rate OW_u.
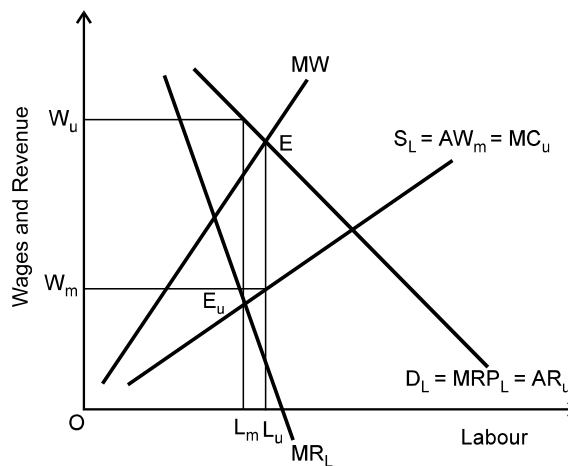
Fig. 12.13 *Indeterminate Bilateral Monopoly*

Thus, there are two possible wage rates —OW_m offered by the monopsonist and OW_u demanded by the labour union. If monopsonist can force the monopolist to behave like a competitive seller, the wage rate will be fixed at OW_m. And, if the monopolist (labour union) can force the monopsonist to behave like a competitive buyer, wage rate will be OW_u. Whether they succeed in doing so is *uncertain*. The wage rate therefore remains indeterminate. Whether or not a compromise wage rate is settled is a matter of bargaining power of the two parties.

12.3 WAGE POLICY AND REGULATION

One of the objectives of economic planning is to raise the standard living of the people. This means that the benefits of planned economic development should be distributed among the different sections of society. Therefore, in achieving a socialistic pattern of society the needs for proper rewards to the working class of the country can never be overemphasized. A national wage policy thus aims at establishing wages at the highest possible level, which the economic conditions of the country permit and ensuring that the wage earner gets a fair share of the increased prosperity of the country as a whole resulting from the economic development.

The term 'wage policy' here refers to legislation or government action calculated to affect the level or structure of wages or both, for the purpose of attaining specific objectives of social and economic policy.

12.3.1 Objectives of National Wage Policy

The objectives of the National Wage Policy are as follows:

(a) To eliminate malpractices in the payment of wages.

(b) To set minimum wages for workers, whose bargaining position is weak due to the fact that they are either unorganized or inefficiently organized.

(c) To rationalize inter-occupational, inter-industrial and inter-regional wage differentials in such a way that disparities are reduced in a phased manner.

(d) To ensure reduction of disparities of wages and salaries between the private and public sectors in a phased manner.

(e) To compensate workers for the raise in the cost of living in such a manner that in the process the ratio of disparity between the highest paid and the lowest paid worker is reduced.

(f) To provide for the promotion and growth of trade unions and collective bargaining.

(g) To obtain for the worker's a just share in the fruits of economic development.

(h) To avoid following a policy of high wages to such an extent that it results in substitution of capital for labour thereby reducing employment.

(i) To prevent high profitability units with better capacity to pay a level of wages in excess to the prevailing level of wages in other sectors.

(j) To permit bilateral collective bargaining within national framework so that high wage islands are not created.

(k) To encourage the development of incentive systems of payment with a view to raising productivity and the real wages of workers.

(l) To bring about a more efficient allocation, utilization of man-power through wage differentials and appropriate systems of payments.

12.3.2 Regulations Adopted by the State

In order to achieve the above objectives under the National Wage Policy, the following regulations have been adopted by the State:

- Prescribing minimum rates of wages
- Compulsory conciliation and arbitration
- Wage boards

  1. **Minimum wages:** In order to prescribe the minimum rate of wages, the Minimum Wages Act, 1948, was passed. The act empowers the government to fix minimum rates of wages in respect of certain sweated and unorganized employments. It also provides for the review of these wages at intervals not exceeding five years.

  2. **Compulsory conciliation and arbitration:** With the object of providing for conciliation and arbitration, the Industrial Disputes Act, 1947, was passed. It provides for the appointment of Industrial Tribunals and National Industrial Tribunals for settlement of industrial disputes including those relating to wages.

  3. **Wage boards:** A wage board is a tripartite body with representatives of management and workers, presided over by a government nominated chairman who can act as an umpire in the event of disagreement among the parties. Technically, a wage board can make only recommendations, since there is no legal sanction for it, but for all practical purposes, they are awards which if made unanimously are considered binding upon employers.

Wage policy in a developing economy

A suitable wage policy for a developing economy must ensure economic growth with stability. If the wage level is too high it will hamper industrial growth. If the wage level is too low, it will adversely affect the workers. Therefore, a proper wage level is necessary to sustain a study growth of the economy. There are two main considerations in wage fixation. They are:

1. To adjust wages to cost of living (need based wages),
2. To link wages with productivity.

1. *Need-based wages:* The meaning of the term 'need-based wage' is that a worker and his family are provided with not only bare necessities of food, clothing and shelter but also children's education, proper health service requirements of essential social needs and a measure of insurance against misfortunes and old age. The Indian Labour Conference held in 1957 accepted the following norms of determining the need based wage:

   (i) The standard working family should consist of three consumption units.

(ii) The minimum requirements of food should be calculated on the basis of net intake of calories as recommended by Dr Aykroyd.

(iii) The clothing requirements should be taken as 18 yards per head per annum.

(iv) As for housing, the rent corresponding to minimum provided under the Government Industrial Housing Scheme.

(v) Fuel, lighting, and other miscellaneous items should constitute 20 per cent of the total minimum wage.

However need-based wage has many practical difficulties. If wages are raised to the need-based wage level and there is no corresponding increase in productivity, there is bound to be inflationary rise in prices. Further the capacity of the industry to pay is relevant. This capacity of industry to pay will depend on the productivity of labour.

2. ***Linking wages with productivity:*** Improvement in wages can result mainly from increased productivity. However, no attention is being paid to productivity, and wages are being either increased on an ad hoc basis or on the basis of cost of living. The Third Plan observed that 'for workers no real advance in their standard of living was possible without steady increase in productivity, because any increase in wages generally beyond certain narrow limits, would otherwise be nullified by a rise in prices'. However, linking wages with productivity arises on account of the following difficulties. They are:

(i) Productivity in India is low. Since productivity is low, wages will have to be low. This position is totally unacceptable to the workers.

(ii) Employers are opposed to the linking of wages with productivity because they are not interested in productivity but profitability.

(iii) Even employees are opposed to the linking of wages with productivity because they feel that low productivity is due to poor management.

(iv) Employers argue that the raise in output is not due to the worker's effort but because of improvement in technology, plant and machinery.

(v) There is the difficulty of assessing productivity especially in industries where the output does not consist of standardized units.

A suitable wage policy

A suitable wage policy in a developing economy should aim at:

1. Containing the rise in prices which can be achieved through a suitable monetary and fiscal policy

2. Linking wage increases to increase in productivity

In view of the above factors, the Indian government enacted the following laws with respect to wage regulation.

- The Payment of Wages Act, 1936
- The Minimum Wages Act, 1948

Let us briefly study about the main objective behind the enactment of these laws.

1. **The Payment of Wages Act, 1936:** The Payment of Wages Act, 1936 was enacted with a view to ensuring that wages payable to employed persons covered by the Act were disbursed by the employers within the prescribed time limit and that no deductions other than those authorized by law were made by them. The last amendment was made in 1982 and several provisions of the Act have become obsolete over the years. Many proposals have been received by the government for amending various provisions which are creating practical difficulties in enforcement of this Act. This Act has been amended in 2005 and latest in 2017.

2. **The Minimum Wages Act, 1948:** The *Minimum Wages Act* came into existence to safeguard the interests of the workers engaged in the unorganized sector. This *Act* provides for fixation and revision of *minimum wages* of the workers engaged in employment.

Check Your Progress

1. What are dynamic wage differentials?
2. What are the reasons responsible for the rise of static wage differentials?

12.4 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. Dynamic wage differentials are those which arise due to disequilibrium in commodity and labour markets. The dynamic wage differentials act on the demand and supply conditions of labour to restore the equilibrium and, thereby, remove the wage differences.

2. Static Wage differentials arise mainly due to the following reasons:
   (i) qualitative differences in labour, i.e., non-homogeneity of labour;
   (ii) difference in the nature of occupations in which labour is employed;
   (iii) differences in the prices of product produced by labour; and
   (iv) market imperfections.

12.5 SUMMARY

- Under the condition of a perfectly labour market and a perfectly product market, wages are determined by the demand for and supply of labour. The labour demand and supply curves can be used straightaway to explain and illustrate wage determination under perfect competition in both product market and labour market.

- The theory of wage determination discussed above assumes that the units of a factor (labour) are homogeneous. If all the units of labour are identical and if non-monetary advantages are the same in all uses, then the wage rate for each labour tends to be the same in a perfectly competitive market.

- The static wage differentials are those that persist in the state of labour market equilibrium. Such differences are not removed by the competitive forces of the market.

- Labour is the most non-homogeneous factor of production. Therefore, wage differentials are far more common than the other factor price differentials.

- When more than one variable input is used by a monopolist, the marginal revenue product curve (*MRPL*) is not the monopoly's demand curve for labour.

- When product market is characterized by monopoly or monopolistic competition and the variable factor (labour) market is perfectly competitive, the factor price (wage rate) and employment are determined in the same fashion as they are determined under perfectly competitive market.

- Factor price and employment determination in the market setting described above is discussed under two different assumptions: (i) that only a single variable factor (labour) is used; and (ii) that more than one variable factor is used.

12.6 KEY WORDS

- **Monopsony:** It is a market structure in which a single buyer substantially controls the market as the major purchaser of goods and services offered by many would-be sellers.

- **Bilateral monopoly:** It is a market structure consisting of both a *monopoly* (a single seller) and a monopsony (a single buyer).

12.7 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short Answer Questions

1. What is the derivation of a demand curve for a commodity different from that for a factor of production?

2. What role does MRP play in the derivation of demand curve for a factor of production?

3. What is meant by the interdependence of factor demand? What problems does it cause in deriving a factor demand curve?

4. Write a short note on the concept of collective bargaining.

Long Answer Questions

1. Explain the derivation of demand for a factor of production assuming that only one factor is demanded.

2. Derive a wage offer curve and explain its role in deriving the labour supply curve? Also, derive labour supply curve from wage-offer curve.

3. How is marginal productivity related with the concept of wage determination?

12.8 FURTHER READINGS

Dwivedi, D. N. 2008. *Principles of Economics*, Seventh Edition. New Delhi: Vikas Publishing House.

Weil. David N. 2004. *Economic Growth*. London: Addison Wesley.

Thomas, Christopher R. and Maurice S. Charles. 2005. *Managerial Economics: Concepts and Applications*, Eighth Edition. New Delhi: Tata McGraw-Hill Publishing Company Limited.

Mankiw, Gregory N. 2002. *Principles of Economics*, Second Edition. India: Thomson Press.

UNIT 13 THEORY OF RENT AND QUASI-RENT

Structure

13.0 INTRODUCTION

The doctrine of quasi rent was introduced by Marshall and extended by Ricardo's theory of land rent to other factors fixed in supply during the short period. It refers in Marshall's words to "income derived from machines and other appliances made by man."

There are certain durable factors whose supply cannot be increased or decreased during the short period. Machines, ships, houses and even human ability like land are fixed in supply but only in the short run. When the demand for them increases, their supply being fixed they earn a surplus which is not rent but is like rent as their supply can be increased in the long run. Marshall preferred to call it quasi rent.

13.1 OBJECTIVES

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

- Differentiate between scarcity and differential rent
- Describe the antecedents of rent theory
- Discuss the concept of quasi rent
- Explain rent as an economic surplus

13.2 SCARCITY VS DIFFERENTIAL RENTS

Scarcity rent

Scarcity rent refers to the price paid for the use of homogeneous land when its supply is limited in relation to demand. If all units of land are homogeneous but demand exceeds supply, all land will earn economic rent by virtue of its scarcity.

Differential rent

Differential rent refers to the rent that arises owing to differences in fertility of land. The surplus that arises due to difference between the marginal and intra-marginal land is the differential rent. It is generally accrued under conditions of extensive land cultivation. The term was first proposed by David Ricardo.

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 the House of Lords and the House of Commons appointed 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.

13.2.1 Quasi-Rent

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 in their current use 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. 13.1. Suppose, given the *AVC, AC* and *MC* curves, price is *OP,* and the firm is in equilibrium at point *E.*
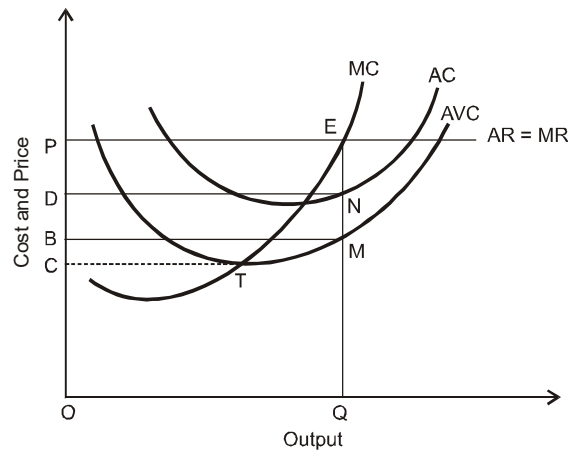
Fig. 13.1 *Determination of Quasi-Rent*

At equilibrium, 'firms' total revenue is

$$OP \times OQ = OPEQ$$

and $$TVC = OB \times OQ = OBMQ$$

The firm must pay a total sum of *OBMQ* to retain the variable factors. Under perfectly competitive conditions, this sum equals their transfer earnings, i.e., the earning that a factor expects from its second best use. The quasi-rent may be obtained as

$$\text{Quasi-Rent} = OPEQ – OBMQ = PBME$$

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.,

$$TR = OC \times CT \text{ and } TVC = OC \times CT$$

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

13.3 RENT AS SURPLUS OVER TRANSFER EARNINGS

Transfer earnings are the minimum payment required to keep a factor of production in its present use.

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. 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 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."[1] 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 cost of production being the same. Suppose there are three grades of land—*A*, *B* and *C* and an equal amount of capital and labour is used to cultivate the same area of each grade of land. However, the respective yields from the three grades of land 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. But *C* grade land yields only 70 quintals of wheat. This gives rise

to rent on *B* grade land and it 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 observed that it happens often that before land *B* is brought under cultivation, more of capital and labour can be employed to increase productivity of land *A.* But, it is quite likely that doubling the capital and labour 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., cultivation, and hence, there is no transfer earning. So all that is paid in the name of rent becomes economic rent. There are, however, alternative uses of land. There are, therefore, transfer earnings, and the total rent cannot be economic rent.

Finally, Ricardo considered land supply to be fixed because he considered land of the country 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.

13.3.1 Rent as Economic Surplus

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., $e_s = \infty$) 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 is illustrated in Fig. 13.2. 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.
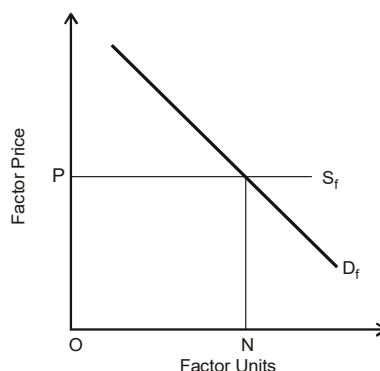
Fig. 13.2 *Zero Economic Rent*

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. 13.3. The market price is fixed at *OM.* The whole of which is economic rent. When factor supply curve has a positive slope, economic rent equals factor price.
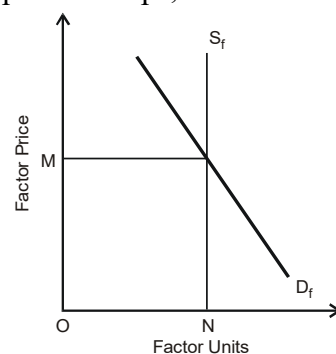
Fig. 13.3 *Zero Transfer Earning*

Check Your Progress

1. Why was Ricardo criticised as being anti-landed aristocracy?

2. What is perfectly elastic factor supply?

13.4 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. 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.

2. Perfectly elastic factor supply (i.e., $es = \infty$) 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.*

13.5 SUMMARY

- 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.

- 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.

- 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.

- In the short-run, variable factors can be transferred to their alternative uses if they are paid in their current use 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.

- 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. By niggardliness of nature, Ricardo meant 'fixed supply' of land and its limited productivity.

- Ricardian theory of rent is based on the principle of demand and supply. If, in a country, the supply of land exceeds the total demand for land, no rent will be paid, like nothing is paid for the use of air.

- In case of intensive cultivation, Ricardo observed that it happens often that before land *B* is brought under cultivation, more of capital and labour can be employed to increase productivity of land *A*.

- 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.

- 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.

13.6 KEY WORDS

- **Scarcity rent:** Scarcity rent refers to the price paid for the use of homogeneous land when its supply is limited in relation to demand.

- **Differential rent:** Differential rent refers to the rent that arises owing to differences in fertility of land. The surplus that arises due to difference between the marginal and intra-marginal land is the differential rent.

13.7 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short Answer Questions

1. Differentiate between scarcity and differential rents.
2. Define quasi rent with the help of the formula.

Long Answer Questions

1. Describe the antecedents of rent theory.
2. Analyse the concept of quasi rent in the short and long run.
3. Discuss rent as surplus to transfer earnings.

13.8 FURTHER READINGS

Dwivedi, D. N. 2008. *Principles of Economics*, Seventh Edition. New Delhi: Vikas Publishing House.

Weil. David N. 2004. *Economic Growth*. London: Addison Wesley.

Thomas, Christopher R. and Maurice S. Charles. 2005. *Managerial Economics: Concepts and Applications*, Eighth Edition. New Delhi: Tata McGraw-Hill Publishing Company Limited.

Mankiw, Gregory N. 2002. *Principles of Economics*, Second Edition. India: Thomson Press.

UNIT 14 MACROECONOMIC ANALYSIS

Structure

14.0 INTRODUCTION

Macroeconomics is the study of economy as a whole. The study of the economy as a whole is carried out by analyzing the behaviour of and interaction between macroeconomic variables including national output (*GDP* and *GNP*), aggregate employment, the general price level, aggregate consumption, savings and investment, price level and economic transactions with the rest of the world. Precisely, macroeconomics studies the relationship and interaction between the macroeconomic variables and other internal and external 'factors or forces' that determine the level and growth of national output and employment, the general price level and the balance of payments position of an economy.

14.1 OBJECTIVES

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

- Analyse the classical and modern approach to the theory of income and employment

- Explain the determination of national income through graphical presentation

- Discuss the concept of circular flow of income

- Analyse the various measures of national income

14.2 THEORY OF INCOME AND EMPLOYMENT

Let us now study the theory of income and employment.

14.2.1 Classical Approach

Following the publication of Adam Smith's classic entitled *An Inquiry into the Nature and Causes of the Wealth of Nations* in 1776, a body of economic theory was gradually developed during the following century and a half. The chief architects of this theory, known as the classical economic theory, were David Ricardo, John Stuart Mill, Jean Baptiste Say and Alfred Marshall. The problem of unemployment was not the primary concern of this theory. Assuming that full employment exists in the economy in the long run, the classical economic theory was mainly concerned with the discussion of those factors which determined:

- What goods and services would be produced in the economy with its given resources;

- The allocation of the economy's given resources between their different rival uses;

- The relative prices of different goods and services and of the factors of production; and

- The distribution of income earned from production between the different co-operating factors of production.

The classical economists assumed that full employment was a normal feature in the economy. According to them, in a *laissez-faire* economy market forces operated in the system which maintained full employment and consequently kept the aggregate output at the level producible under conditions of full employment. In the classical economic theory, full employment was a rule in the long period. Deviations from it were viewed only as temporary exceptions. Full employment did not, however, rule out the existence of some unemployment in the economy. Even at the 'full employment' level, there would be some people in the economy who could be either frictionally or voluntarily unemployed. The frictional unemployment was temporary unemployment between job changes or on entry into the labour force while searching for jobs due to the want of adequate knowledge about the available job opportunities in the economy on the part of workers. Voluntary unemployment was due to the reluctance or refusal on the part of workers to work at the going wage. Workers agitating for higher wages were an example of the voluntarily unemployed workers. While frictional unemployment would disappear with the workers getting acquainted with the available job opportunities in due course of time, voluntary unemployment was due to the workers' refusal to work at the current wage and did not worry the classical economists. In short, full employment only implied that involuntary unemployment—a state of being unemployed in spite of the workers' willingness to work at the going wage rate—did not exist in the economy.

But what would happen if there were workers who were involuntarily unemployed in the economy? According to the classical economic theory, if there is unemployment in the economy, forces of correction will soon eliminate it and

will restore full employment in the economy. The basic classical tenet was that in a free market economy, the aggregate demand for goods and services could not, except temporarily, fall short of the aggregate supply of goods and services. As long as the aggregate demand equalled the aggregate supply, there was no barrier to the production of goods and services corresponding to full employment in the economy. In the classical view, lapses from full employment were infrequent and short-lived. Depressions were, therefore, considered infrequent and short-lived occurrences. This conclusion is, however, puzzling to any serious student of economic history who knows about the severe and prolonged depressions of the 1870s, 1930s and other periods.

Although the classical theory of employment, output and price level was attacked by a few dissenters in the 19th century—Thomas Robert Malthus, Jean Charles Leonard de Sismondi, Karl Marx, J A Hobson, Silvio Gesell and others—the attack was unsuccessful because no alternative theory was constructed to replace the classical theory. 'Since Malthus was unable to explain clearly (apart from an appeal to the facts of common observation) how and why effective demand could be deficient or excessive, he failed to furnish an alternative construction, and Ricardo conquered England as completely as the Holy Inquisition conquered Spain. Not only was his theory accepted by the city, by statesmen and by the academic world. But controversy ceased; the other point of view completely disappeared; it ceased to be discussed. The great puzzle of Effective Demand with which Malthus had wrestled vanished from economic literature.'

John Maynard Keynes successfully attacked the classical explanation of the determination of aggregate employment, output and general price level. It was the assumption of a given volume of total output, rather than its composition and technique of production, which was severely attacked by Keynes. The great depression of the 30s gave a severe blow to the naïve classical economic theory.

The essential feature of classical macroeconomic analysis is that it presents a model of full employment in the economy in the long period. Underlying the analysis, are the assumptions of perfect competition in the factor and product markets and profit-maximization on the part of firms. There are three markets to study. *First,* there is the labour market which deals with the supply of and the demand for labour. The equilibrium condition for full employment in the labour market requires that the wage should be one corresponding to which the demand for and the supply of labour in the market are in equilibrium, i.e., there is neither an excess supply of nor an excess demand for labour in the market. In the labour market we are concerned with the analysis of the form of the aggregate demand and the aggregate supply functions of labour.

Secondly, there is the product market with its equilibrium flow condition which is equivalent in macroeconomic equilibrium to an equality between saving and investment. The equilibrium condition in the capital–bonds–market requires the equilibrium between the *ex ante* investment and *ex ante* saving. *Thirdly,* there

is the money market which is concerned with the demand for and the supply of money. The first two markets deal with the equilibrium of the real sector of the economy while the money market is concerned with the equilibrium of the monetary sector of the economy. The equilibrium in the monetary sector determines the absolute price level which does not influence the relative prices, aggregate employment and output which are determined in the real sector of the economy. In short, there is a dichotomy or separation between the real and monetary sectors of the economy in the classical economic system. This dichotomy arises from the argument of the classicists that 'money is a veil' (neutral).

In the classical economic theory, money does not matter and its function in the economy is merely to facilitate the real transactions by serving as a medium of exchange. It is neutral and does not interfere with the real processes of production and distribution in the economy; it only facilitates production, i.e., lubricates the wheels of the economic system. According to the classicists, changes in the money supply cause proportionate changes only in the equilibrium values of the *nominal* variables, leaving the equilibrium values of the *real* variables (output, employment, real wage, interest rate, etc.,) unchanged. The equilibrium values of these real variables are exclusively and solely determined in the real sector—in the labour, capital and commodity markets. In the classical macroeconomics, the economy's real sector can, therefore, be dichotomized from its monetary sector.

Money, however, does something more than merely act as a medium of exchange in the economy. In a dynamic world with uncertain future, money is also demanded for asset purposes. Consequently, it influences both the production and distribution in the economic system. In other words, changes which take place in the monetary sector also influence the real sector of the economy.

14.2.2 Modern (Keynesian), Approach

The two major questions with which we shall be concerned here are: (*i*) What factors determine the level of national income and (*ii*) How is the equilibrium level of national income determined? These questions were first answered by J.M. Keynes, in 1936, in his book *The General Theory of Employment, Interest and Money*. We will outline here the Keynesian theory of income determination.

Keynesian Theory of National Income Determination

To explain the Keynesian theory of income determination, the entire economy is divided into four sectors, viz.,

1. Household sector,

2. Firms or the business sector,

3. Government sector and

4. Foreign sector.

The Keynesian theory of income determination is present in the following three models: (*i*) Two-sector model including only the household and the business

sectors; (*ii*) Three-sector model including household, business and government sectors and (*iii*) Four-sector model including foreign sector with the three-sector model.

For the sake of simplicity and systematic exposition of the Keynesian theory of income determination, we will first discuss in this section income determination in a two-sector model involving only the household and firm sectors.

Determination of National Income: Two-Sector Model

Assumptions

The following simplifying assumptions are made to specify the two-sector model of a hypothetical simple economy.

First, the hypothetical simple economy has only two sectors: households and firms. The households own the factors of production and they sell factor services to the firms to earn their living in the form of factor payments—wages, rent, interest and profits. Also, households are the consumers of all final goods and services. The firms, on the other hand, hire factor services from the households and produce goods and services which they sell to the households.

Secondly, there is no government. Or, if government is there, it does not perform any economic function; it does not tax, it does not spend and it does not consume.

Thirdly, the economy is a closed one: there is no foreign trade. It implies that there is no outflow or inflow of goods and services to and from foreign countries.

Fourthly, there are no corporate savings or undistributed (or retained) corporate profits, i.e., the total corporate profit is distributed as dividends.

Finally, prices of all goods and services, supply of labour and capital, and the state of production technology remain constant.

Aggregate Supply and Aggregate Demand

According to Keyens, national income of a country is determined by two factors: (*i*) aggregate demand (*AD*) and (*ii*) aggregate supply (*AS*) of goods and services. And, the equilibrium level of national income is determined where *AD* equals *AS*. Before we illustrate graphically the determination of national income, let us explain the concepts of aggregate demand and aggregate supply.

(i) Aggregate Supply. The *aggregate supply* (*AS*) refers to the total value of goods and services produced and supplied in an economy per unit of time. Aggregate supply includes both consumer goods and producer goods. The goods and services produced per time unit multiplied by their respective (constant) prices give the total value of the national output. This is the aggregate supply in terms of money value.

Aggregate Supply Schedule. The Keynesian aggregate supply schedule or aggregate supply curve is drawn on the assumption that total income is always spent – no part of it is retained or withheld. That is, total income always equals total expenditure. This relationship between income and expenditure is shown by

a 45° line in Fig. 14.1. This line is also called *aggregate supply schedule*. In the Keynesian theory of income determination, aggregate income equals consumption (*C*) plus savings (*S*). Therefore, *AS* schedule is generally named as *C + S* schedule. The aggregate supply (*AS*) curve is also sometimes called 'aggregate expenditure' (*AE*) curve.
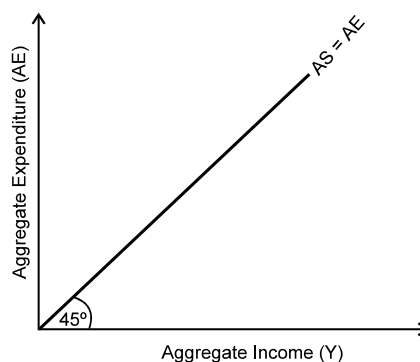

Fig. 14.1 *The Aggregate Supply Curve*

(ii) Aggregate Demand. The aggregate demand is an *ex-post* concept. It implies effective demand which equals actual expenditure. The aggregate effective demand means the aggregate expenditure made by the society per unit of time, usually, one year. Aggregate demand (*AD*) consists of two components:

   (*i*) aggregate demand for consumer goods (*C*) and

   (*ii*) aggregate demand for capital goods (*I*).

Thus, $AD = C + I$ ...(14.1)

Aggregate Demand Schedule

The aggregate demand *AD* schedule is also called *C + I* schedule. In the Keynesian framework, investment (*I*) is assumed to remain constant in the short-run. But, consumption (*C*) is treated to be a function of income (Y). Pending detailed discussion on the consumption function till the next section, let us assume that the consumption function is given as

$$C = a + bY$$ ...(14.2)

where *a* is a constant denoting *C* when *Y* = 0 and *b* is the proportion of income consumed, i.e., $b = \Delta C / \Delta Y.$

By substituting Eq. (14.2) in Eq. (14.1), *AD* function can be expressed as

$$AD = a + bY + I$$ ...(14.3)

Let us now illustrate the construction of the *C + I* schedule by assuming:

   (*i*) $C = 50 + 0.5Y$, and   (*ii*) $I = ₹ 50$ billion

The *AD* function given in Eq. (14.3) can now be written as

$$AD = 50 + 0.5Y + 50 = 100 + 0.5Y$$

An aggregate demand schedule based on the above assumptions is given in Table 14.1. The *C + I* schedule is plotted in Fig. 14.2.

Table 14.1 *Aggregate Demand Schedule*

(₹ *in billion*)

Income (Y)	*C = 50 + 0.5Y*				*I = 50*	*C + I Schedule*
0	50 +	0	=	50	50	100
50	50 +	25	=	75	50	125
100	50 +	50	=	100	50	150
150	50 +	75	=	125	50	175
200	**50 +**	**100**	**=**	**150**	**50**	**200**
250	50 +	125	=	175	50	225
300	50 +	150	=	200	50	250
350	50 +	175	=	225	50	275
400	50 +	200	=	250	50	300

National Income Determination: Graphical Presentation

Having explained the concept and derivation of aggregate supply and aggregate demand curves, we now turn to the question of income determination. In the above table, the last column represents the aggregate demand and the first column represents the aggregate supply. It can be seen in the table that *AS* and *AD* are equal only at one level of income and expenditure, i.e., at ₹ 200 billion. The equilibrium level of the national income is therefore determined at ₹ 200 billion. If for some reason, *AD* exceeds *AS* or *AS* exceeds *AD*, an adjustment process will bring them back in balance at ₹ 200 billion. The information contained in Table 14.1 can be used to illustrate income determination graphically.
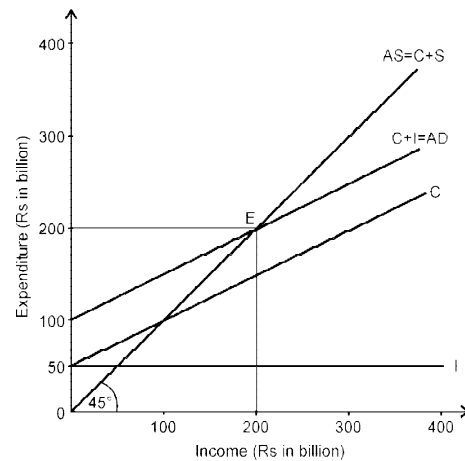
Fig. 14.2 *National Income Determination*

The data contained in Table 14.1 is presented graphically in Fig. 14.2. The *AS* schedule is drawn on the assumption that total income (*Y*) is always equal to total expenditure (*E*). The *AS* schedule has, therefore, a constant slope of 1. The *C + I* schedule is the vertical summation of the *C* and *I* schedules.

As Fig. 14.3 shows, *C + I* and *C + S* schedules intersect at point *E* determining the equilibrium level of income at ₹ 200 billion. Note that at point *E*,

$$AD = AS$$
$$C + I = C + S$$
$$150 + 50 = 200$$

Fig. 14.3 *National Income Determination*

Thus, the equilibrium level of national income is determined at ₹ 200 billion.

Why Not Equilibrium at Any Other Point?

Note that beyond the equilibrium point, *E*, *AD* < *AS* or (*C* + *I*) < (*C* + *S*). It means that at any point on the *AS* schedule beyond point *E*, the firms would be producing more than what households demand. If firms produce goods and services worth more than ₹ 200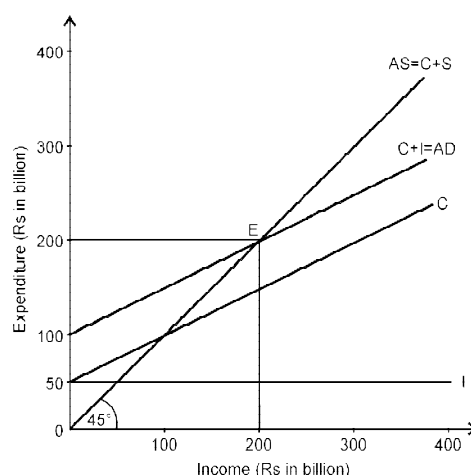 billion, they will find that they have produced in excess of aggregate demand and their unsold stocks are piling up. For example, suppose firms produce goods and services worth ₹ 250 billion. As Table 14.1 shows, this level of output (*AS*) exceeds the aggregate demand (*AD*) by ₹ 25 billion. Note that at output or *Y* = ₹ 250 billion, *AD* equals ₹ 225 billion (see Table 14.1). Therefore, firms' unsold stock equals goods and services worth ₹ 25 billion. Hence, they reduce their production and cut down their expenditure on inputs. As a result, the demand for factors of production decreases. This reduces household incomes and, thererfore, their expenditure on goods and services. This process continues until the equilibrium level of income reaches ₹ 200 billion.

Similarly, below ₹ 200 billion level of national income, aggregate demand exceeds aggregate supply. The firms, therefore, find that their output is less than what the society is willing to consume. They realize that they could make a greater income by producing and selling a larger output. For example, if firms produce goods worth only ₹ 150 billion, they find *AD* exceeding *AS* by ₹ 25 billion. That is, demand worth ₹ 25 billion remains unsupplied. They are, therefore, encouraged to produce more and generate more income to the society. The society, in its turn, spends more as its income increases. The process continues until the equilibrium level of national income is reached. Once the equilibrium level of national income is determined, it is supposed to remain stable.

Check Your Progress

1. Who were the chief architects of the classical economic theory?

2. State the four sectors of the economy.

14.3 CIRCULAR FLOW OF INCOME

We begin our study of macroeconomics by having a look at how an economy works. An economy is a system of interrelated economic activities and economic transactions. Basic economic activities include production, exchange and consumption. The economic activities are carried out in an integrated manner that creates a continuous process of economic transactions – buying and selling. Economic transactions generate two kinds of flows:

(*i*) product or goods flow and

(*ii*) money flow.

The product flow consists of factor and product flows, i.e., flow of factors of production and of goods and services. In a monetized economy, factor and product flows generate money flows in the form of factor payments and payments for goods and services. The two kinds of flows go in opposite direction in a circular manner and make two kinds of *circular flows.* An economy keeps working so long as the two flows go on uninterrupted. The working of an economy can, therefore, be viewed as circular flows of product and money and the size of the economy as the volume of goods flow.

This section presents a brief description of how goods and factor flows are generated and how an economy works in a systematic manner. To begin with, we will first give a description of circular flows in a simple economy consisting of only two sectors: (*i*) **households** and (*ii*) **firms**. The **households** have two characteristics: (*a*) they are owners of all factors of production and (*b*) they are consumers of all final goods and services. **Firms**, on the other hand, have two characteristics too: (*a*) they hire factors of production from the households and (*b*) they produce and sell their final products to the households. This model is then extended to include the **government sector** making it a 3-sector model. Finally, the model is extended further to include also the **foreign sector** (comprising only exports and imports goods and services) to make it a complete circular flow model consisting of households, firms, government and foreign sector.

Circular Flows in a Simple Economy Model

We begin with the description of circular flows in a simple economy consisting of only two sectors, viz., households and firms, and there is no government and no foreign trade. In our simple economy model, *households* are assumed (*i*) to own all the factors of production, (*ii*) to consume all final goods and services and (*iii*) their income consists of wages, rent, interest and profits. The *business firms,* on the other hand, are assumed (*i*) to hire factors of production from the households; (*ii*) they produce and sell goods and services to the households; and (*iii*) they do not save, i.e., there is no corporate saving.

The working of and circular flows of incomes and expenditure in the two-sector model are illustrated in Fig. 14.4. As the figure shows, factors of production

flow from the households to the factor market and from the factor market to the firms. As shown in the lower half of the figure, goods and services produced by the firms flow from the firms to the households. The arrows showing factor and product flows make the *product flows* or *real flows.* Note that real flows take a circular path.
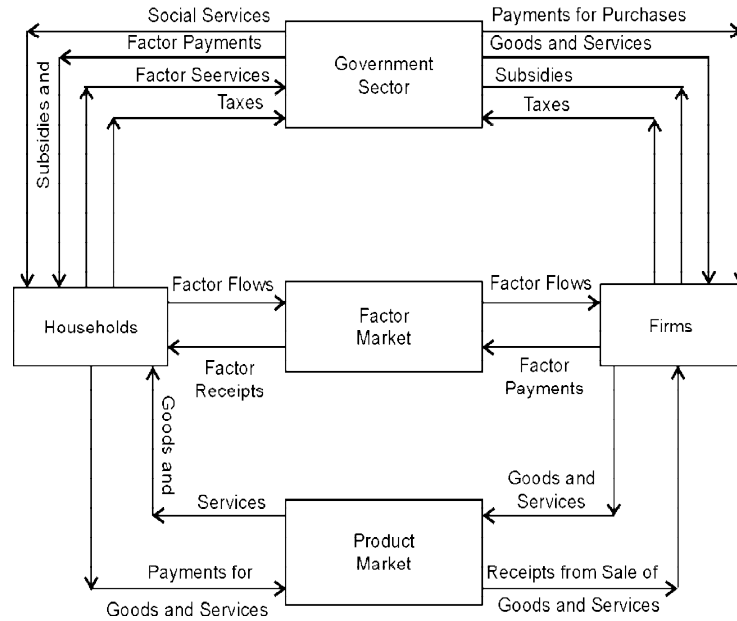

Fig. 14.4 *The Circular Flows of Goods and Money in a Simple Economy*

Product flow generates *money flow.* As the figure shows, money flows from the firms to the households in the form of factor payments as wages, interest, rent and dividends. Factor payments take the form of household incomes. Households spend their incomes on goods and services they consume. As a result, money incomes flow from the households to the firms in the form of payments for goods and services. Thus, money paid by the firms as factor payments flows back to the firms as payments made by the households for goods and services. This makes the circular flow of money.

Note that product and money flows make the circular flows in the economy and that products and money flow in opposite directions. These flows represent the working of the simple economy. An important feature of product and money flows is that the *value* of real flow equals the money flow. This equality results from the fact that *factor payments* are equal to *household incomes* and since households spend their total income on consumer goods and services, *household expenditure* equals the total receipts of the firms, which equals *the value of the output.* These equalities can be summarized in the form of *identities* as follows. In the final analysis, household incomes ≡ factor payments ≡ the money value of output. This identity holds so long as households spend their total income, i.e., households do not hoard any part of their income, and firms spend their total receipts on hiring factors of production from the households.

The Effect of Withdrawals and Injections

The product and money flows shown in Fig. 14.4 assume that there are no *withdrawals* from and *injections* into the economy. *Withdrawals* means withholding money incomes as idle cash balance. This withholding is not 'saving' for savings are returned to the circular flows in the form of purchase of capital goods (investment). Withdrawals are also called ***leakages***. *Injection,* on the other hand, means money expenditure in addition to factor incomes. In reality, however, there are *withdrawals from* and *additions to* the circular flows.

Let us look at the forms and nature of withdrawals and injections. In our two-sector model, a *withdrawal is an amount set aside by the households and/or by the firms, not to be spent on the goods and services over a period of time.* For example, if households set aside a part of their current income as a provision for old age or as security against the loss of job, etc., it is called a *withdrawal.* It is important to note that *a withdrawal is not a saving.* For, savings are ultimately returned to the circular flows in the form of investment expenditure. Likewise, firms may withhold a part of their sales revenue and not return it to the circular flows if they anticipate depression. *Withdrawals reduce the volume of the circular flow.*

Injections, on the other hand, are *the amount that is spent by the households and/or firms in addition to their current incomes generated within the regular economy.* Injections may be made by the households in the form of spending past savings or hoardings. Injection by the firms may take the form of spending their accumulated savings. Firms may inject money into the economy by borrowing from households. *Injections increase the size of the flow.*

14.4 NATIONAL INCOME CONCEPTS, DEFINITION AND ITS MEASUREMENT

National income is the final outcome of all economic activities of a nation valued in terms of money. National income is the most important macroeconomic variable and determinant of the business level and economic status of a country. The level of national income determines the level of aggregate demand for goods and services. Its distribution pattern determines the pattern of demand for goods and services, i.e., how much of what kinds of goods and services are demanded and produced. The trend in national income determines the trends in aggregate demand and also the economic prospects. Therefore, business decision-makers and economic analysts need to keep in mind these aspects of national income, especially those having long-run implications. National income or a relevant component of it is an indispensable variable considered in economic forecasting.

In this section, we will discuss basic concepts of national income used in business analysis and business decisions, methods of measuring national income and the trend and the growth rates in India's national income.

Definition

Conceptually, national income is the money value of all final goods and services produced in a country during a period of one year. Economic activities generate a large number of goods and services and make net addition to the national stock of capital. These together constitute the national income of a 'closed economy'—an economy which has no economic transactions with the rest of the world. In an 'open economy', national income also includes the net results of its transactions with the rest of the world (i.e., exports *less* imports).

Economic activities should be distinguished from the non-economic activities from a national point of view. Broadly speaking, ***economic activities*** include all human activities which create goods and services that can be valued at market price. Economic activities include production by farmers (whether for household consumption or for market), production by firms in the industrial sector, production of goods and services by the government enterprises, and services produced by business intermediaries (wholesalers and retailers), banks and other financial organizations, universities, colleges and hospitals, and professionals like medical practitioners, lawyers, consultants, etc. On the other hand, ***non-economic activities*** are those which produce goods and services that do not have any market value. Non-economic activities include spiritual, psychological, social and political services. The non-economic category of activities also includes hobbies, service to self, services of housewives, services of members of family to other members and exchange of mutual services between neighbours.

National Income as Money Flow

We have defined national income from the angle of *product flows*. The same can be defined in terms of *money flows*. While economic activities generate flow of goods and services, on the one hand, they generate money flows, on the other, in the form of factor payments—wages, interest, rent, profits, and earnings of self-employed. Thus, national income may also be estimated by adding the factor earnings and adjusting the sum for indirect taxes and subsidies. The national income thus obtained is known as *national income at factor cost*. It is related to money income flows.

The concept of national income is linked to the society as a whole. It differs fundamentally from the concept of *private income*. Conceptually, national income refers to the money value of the entire volume of final goods and services resulting from all economic activities of the country. This is not true of private income. Also from the calculation point of view, there are certain receipts of money or of services and goods that are not ordinarily included in private incomes but are included in the national incomes, and *vice versa*. National income includes, for example, employer's contribution to the social security and welfare funds for the benefit of employees, profits of public enterprises and services of owner occupied houses. But it excludes the interest on warloans, social security benefits and pensions. These items are, however, included in the private incomes. The national income is, therefore, not merely an aggregation of the private incomes.

Measures of National Income

Gross National Product (GNP)

Of the various measures of national income used in national income analysis, *GNP* is the most important and widely used measure of national income. It is the most comprehensive measure of the national productive activities in an open economy. *The GNP is defined as the value of all final goods and services produced during a specific period, usually one year, plus incomes earned abroad by the nationals minus incomes earned locally by the foreigners*. The *GNP* so defined is identical to the concept of gross national income (*GNI*). Thus, *GNP* = *GNI*. The difference between the two is only of procedural nature. While *GNP* is estimated on the basis of product-flows, *GNI* is estimated on the basis of money income flows, (i.e., wages, profits, rent, interest, etc.).

Gross Domestic Product (GDP)

The Gross Domestic Product (*GDP*) is defined as the market value of all final goods and services produced in the domestic economy during a period of one year, *plus* income earned locally by the foreigners *minus* incomes earned abroad by the nationals. The concept of *GDP* is similar to that of *GNP* with a significant procedural difference. In case of *GNP*, incomes earned by the nationals in foreign countries are added and incomes earned locally by the foreigners are deducted from the market value of domestically produced goods and services. But, in case of *GDP*, the process is reversed—incomes earned locally by foreigners are added and incomes earned abroad by the nationals are deducted from the total value of domestically produced goods and services.

Net National Product (NNP)

NNP is defined as *GNP less* depreciation, i.e.,

NNP = *GNP* – Depreciation

Depreciation is that part of total productive assets which is used to replace the capital worn out in the process of creating *GNP*. Briefly speaking, in the process of producing goods and services (including capital goods), a part of total stock of capital is used up. 'Depreciation' is the term used to denote the worn out or used up capital in the process of production. An estimated value of depreciation is deducted from the *GNP* to arrive at *NNP*.

The *NNP*, as defined above, gives the measure of net output available for consumption by the society (including consumers, producers and the government). *NNP* is the real measure of the national income. *NNP* = *NNI* (net national income). In other words, *NNP* is the same as the national income at *factor cost*. It should be noted that *NNP* is measured at market prices including direct taxes. Indirect taxes are, however, not a part of actual cost of production. Therefore, to obtain real national income, indirect taxes are deducted from the NNP. Thus, NNP *less* indirect taxes = National Income.

Some Accounting Definitions

(*a*) Accounting Identities at Market Price

$GNP \equiv GNI$ (Gross National Income)

$GDP \equiv GNP\ less$ net income from abroad

$NNP \equiv GNP\ less$ depreciation

NDP (Net Domestic Product) $\equiv NNP\ less$ net income from abroad

(*b*) Some Accounting Indentities at Factor Cost

GNP at factor cost $\equiv GNP$ at market price *less* net indirect taxes

NNP at factor cost $\equiv NNP$ at market price *less* net indirect taxes

NDP at factor cost $\equiv NNP$ at market price *less* net income from abroad

NDP at factor cost $\equiv NDP$ at market price *less* net indirect taxes

NDP at factor cost $\equiv GDP$ at market price *less* depreciation

Methods of Measuring National Income

National income of a country is generated by its people participating in different kinds of economic activities and producing goods and services. For measuring national income, an economy is viewed from three different angles.

1. The national economy is considered as an *aggregate of productive units* of different sectors such as agriculture, mining, manufacturing, trade and commerce, services, etc.

2. The whole national economy is viewed as a *combination of individuals and households owning different kinds of factors of production* which they use themselves or sell factor-services to make their livelihood.

3. The national economy may also be viewed as a collection of consuming, saving and investing units (individuals, households, firms and government).

Following these notions of a national economy, national income may be measured by three different corresponding methods:

1. *Net product method*—when the entire national economy is considered as an aggregate of producing units;

2. *Factor-income method*—when national economy is considered as combination of factor-owners and users;

3. *Expenditure method*—when national economy is viewed as a collection of spending units.

The procedures which are followed in measuring the national income in a *closed economy*—an economy which has no economic transactions with the rest of the world—are briefly described here.

Net Output or Value Added Method

The net output method is also called the *value added method*. In its standard form, this method consists of three stages: "(*i*) estimating the gross value of domestic

output in the various branches of production; (*ii*) determining the cost of material and services used and also the depreciation of physical assets; and (*iii*) deducting these costs and depreciation from gross value to obtain the net value of domestic output...". The net value of domestic product thus obtained is often called the *value added* or *income product* which is equal to the sum of wages, salaries, supplementary labour incomes, interest, profits, and net rent paid or accrued. Let us now describe the stages (*i*) and (*ii*) in some detail.

(a) Measuring Gross Value. For measuring the gross value of domestic product, output is classified under various categories on the basis of the nature of activities from which they originate. The output classification varies from country to country depending on (*i*) the nature of domestic activities; (*ii*) their significance in aggregate economic activities and (*iii*) availability of requisite data. For example, in the US, about seventy-one divisions and subdivisions are used to classify the national output; in Canada and the Netherlands, classification ranges from a dozen to a score; and in Russia, only half a dozen divisions are used. According to the CSO publication, fifteen sub-categories are currently used in India.

After the output is classified under the various categories, the value of gross output is computed in two alternative ways: (*i*) by multiplying the output of each category of sector by its respective market price and adding them together or (*ii*) by collective data about the gross sales and changes in inventories from the account of the manufacturing enterprises and computing the value of GDP on the basis thereof. If there are gaps in data, some estimates are made thereof and gaps are filled.

(b) Estimating Cost of Production. The next step in calculating the net national product is to estimate the cost of production including depreciation. Estimating cost of production is, however, a relatively more complicated and difficult task because of non-availability of adequate and requisite data. Much more difficult is the task of estimating depreciation since it involves both conceptual and statistical problems. For this reason, many countries adopt factor-income method for estimating their national income.

However, countries adopting net-product method find some ways and means to calculate the deductible cost. The costs are estimated either in absolute terms (where input data are adequately available) or as an overall ratio of input to the total output. The general practice in estimating depreciation is to follow the usual business practice of depreciation accounting. Traditionally, depreciation is calculated at some percentage of capital, permissible under the tax-laws. In some estimates of national income, the estimators deviate from the traditional practice and estimate depreciation as some ratio of the current output of final goods.

Following a suitable method, deductible costs including depreciation are estimated for each sector. The cost estimates are then deducted from the sectoral gross output to obtain the net sectoral products. The net sectoral products are then added together. The total thus obtained is taken to be the measure of net national product or national income by net product method.

Factor-Income Method

This method is also known as *income method* and *factor-share method*. Under this method, the national income is calculated by adding up all the "incomes accruing to the basic factors of production used in producing the national product". Factors of production are conventionally classified as land, labour, capital and organization. Accordingly, the national income equals the sum of the corresponding factor earnings. Thus,

National income = Rent + Wages + Interest + Profit

However, in a modern economy, it is conceptually very difficult to make a distinction between earnings from land and capital, on the one hand, and between the earnings from ordinary labour and entrepreneurial functions, on the other. For the purpose of estimating national income, therefore, factors of production are broadly grouped as labour and capital. Accordingly, national income is supposed to originate from two primary factors, viz., labour and capital. In some activities, however, labour and capital are jointly supplied and it is difficult to separate the labour and capital contents from the total earnings of the supplier. Such incomes are termed as *mixed incomes*. Thus, the total factor-incomes are grouped under three categories; (*i*) labour incomes, (*ii*) capital incomes and (*iii*) mixed incomes.

(a) Labour Incomes. Labour incomes included in the national income have three components: (*a*) wages and salaries paid to the residents of the country including bonus and commission, and social security payments; (*b*) supplementary labour incomes including employer's contribution to social security and employee's welfare funds, and direct pension payments to retired employees; (*c*) supplementary labour incomes in kind, e.g., free health and education, food and clothing, and accommodation, etc. Compensations in kind (food and clothes) to domestic servants and such other free-of-cost services provided to the employees are included in labour income. War bonuses, pensions, service grants are not included in labour income as they are regarded as 'transfer payments'. Certain other categories of income, e.g., incomes from incidental jobs, gratuities, tips, etc., are ignored for lack of data.

(b) Capital Incomes. According to Studenski, capital incomes include the following kinds of earnings:

(*a*) dividends excluding inter-corporate dividends;

(*b*) undistributed before-tax-profits of corporations;

(*c*) interest on bonds, mortgages, and saving deposits (excluding interests on war bonds, and on consumer-credit);

(*d*) interest earned by insurance companies and credited to the insurance policy reserves;

(*e*) net interest paid out by commercial banks;

(*f*) net rents from land, buildings, etc., including imputed net rents on owner-occupied dwellings;

(*g*) royalties and

(*h*) profits of government enterprises.

The data for the first two items are obtained mostly from the firms' books of accounts submitted for taxation purposes. But the definition of profit for national accounting purposes differs from that employed by taxation authorities. Some adjustments in income tax data therefore, become, necessary. The data adjustments generally pertain to (*i*) excessive allowance of depreciation made by the firms; (*ii*) elimination of capital gains and losses since these items do not reflect the changes in current income and (*iii*) elimination of under or over-valuation of inventories on book-value.

(c) Mixed Income. Mixed incomes include earnings from (*a*) farming enterprises, (*b*) sole proprietorship (not included under profit or capital income) and (*c*) other professions, e.g., legal and medical practices, consultancy services, trading and transporting, etc. This category also includes the incomes of those who earn their living through various sources as wages, rent on own property, interest on own capital, etc.

All the three kinds of incomes, viz., labour incomes, capital incomes and mixed incomes, added together give the measure of national income by *factor-income method*.

Expenditure Method

The expenditure method, also known as **final product method,** measures national income at the final expenditure stages. In estimating the total national expenditure, any one of the two following methods are used: **first,** all the money expenditures at market price are computed and added up together and **second,** the value of all the products finally disposed off are computed and added up, to arrive at the total national expenditure. The items of expenditure which are taken into account under the *first method* are (*a*) private consumption expenditure; (*b*) direct tax payments; (*c*) payments to the non-profit-making institutions and charitable organizations like schools, hospitals, orphanages, etc. and (*d*) private savings. Under the *second method*, the following items are considered: (*a*) private consumer goods and services, (*b*) private investment goods, (*c*) public goods and services and (*d*) net investment abroad. The second method is more extensively used because the data required in this method can be collected with greater ease and accuracy.

Treatment of Net Income from Abroad. We have so far discussed methods of measuring national income of a 'closed economy'. But most economies are *open* in the sense that they carry out foreign trade in goods and services and financial transactions with the rest of the world. In the process, some nations make net income through foreign trade while some lose their income to foreigners. The net earnings or losses from foreign trade change the national income. In measuring the national income, therefore, the net result of external transactions is adjusted to the total. Net incomes from abroad are added to, and net losses from the foreign transactions are deducted from the total national income arrived at through any of the above three methods.

Briefly speaking, all exports of merchandise and of services like shipping, insurance, banking, tourism, and gifts are added to the national income. And, all the imports of the corresponding items are deducted from the value of national output to arrive at the approximate measure of national income. To this is added the net income from foreign investment. These adjustments for international transactions are based on the international balance of payments of the nations.

Check Your Progress

3. What are the two kinds of flows generated by economic transactions?

4. What do you mean by injections in economics?

5. What is a Gross Domestic Product (GDP)?

14.5 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. The chief architects of this theory, known as the classical economic theory, were David Ricardo, John Stuart Mill, Jean Baptiste Say and Alfred Marshall.

2. The entire economy is divided into four sectors:

   - Household sector

   - Firms or the business sector

   - Government sector

   - Foreign sector

3. Economic transactions generate two kinds of flows:

   - Product or goods flow

   - Money flow

4. Injections are the amount that is spent by the households and/or firms in addition to their current incomes generated within the regular economy. Injections may be made by the households in the form of spending past savings or hoardings.

5. The Gross Domestic Product (GDP) is defined as the market value of all final goods and services produced in the domestic economy during a period of one year, *plus* income earned locally by the foreigners *minus* incomes earned abroad by the nationals.

14.6 SUMMARY

- Macroeconomics is the study of economy as a whole. The study of the economy as a whole is carried out by analyzing the behaviour of and

interaction between macroeconomic variables including national output (*GDP* and *GNP*), aggregate employment, the general price level, aggregate consumption, savings and investment, price level and economic transactions with the rest of the world.

- Following the publication of Adam Smith's classic entitled *An Inquiry into the Nature and Causes of the Wealth of Nations* in 1776, a body of economic theory was gradually developed during the following century and a half.

- The chief architects of this theory, known as the classical economic theory, were David Ricardo, John Stuart Mill, Jean Baptiste Say and Alfred Marshall.

- The classical economists assumed that full employment was a normal feature in the economy. According to them, in a *laissez-faire* economy market forces operated in the system which maintained full employment and consequently kept the aggregate output at the level producible under conditions of full employment.

- Although the classical theory of employment, output and price level was attacked by a few dissenters in the 19th century—Thomas Robert Malthus, Jean Charles Leonard de Sismondi, Karl Marx, J A Hobson, Silvio Gesell and others—the attack was unsuccessful because no alternative theory was constructed to replace the classical theory.

- John Maynard Keynes successfully attacked the classical explanation of the determination of aggregate employment, output and general price level. It was the assumption of a given volume of total output, rather than its composition and technique of production, which was severely attacked by Keynes.

- In the classical economic theory, money does not matter and its function in the economy is merely to facilitate the real transactions by serving as a medium of exchange.

- The Keynesian theory of income determination is present in the following three models: (*i*) Two-sector model including only the household and the business sectors; (*ii*) Three-sector model including household, business and government sectors and (*iii*) Four-sector model including foreign sector with the three-sector model.

- In our simple economy model, *households* are assumed (*i*) to own all the factors of production, (*ii*) to consume all final goods and services and (*iii*) their income consists of wages, rent, interest and profits.

- National income is the final outcome of all economic activities of a nation valued in terms of money. National income is the most important macroeconomic variable and determinant of the business level and economic status of a country.

14.7 KEY WORDS

- **Gross National Product:** The GNP is defined as the value of all final goods and services produced during a specific period, usually one year, plus incomes earned abroad by the nationals minus incomes earned locally by the foreigners.

- **Gross Domestic Product (GDP):** It is defined as the market value of all final goods and services produced in the domestic economy during a period of one year, *plus* income earned locally by the foreigners *minus* incomes earned abroad by the nationals.

14.8 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short Answer Questions

1. What are the factors which determined most of the discussion of classical economic theory?

2. Write a short note on the determination of national income in the two sector model.

3. Comment on the net output or value added method.

Long Answer Questions

1. Analyse the classical and modern approach of income and employment.

2. Explain the determination of national income through graphical presentation.

3. Discuss the concept of circular flow of income.

4. Analyse the various measures of national income.

14.9 FURTHER READINGS

Dwivedi, D. N. 2008. *Principles of Economics*, Seventh Edition. New Delhi: Vikas Publishing House.

Weil. David N. 2004. *Economic Growth*. London: Addison Wesley.

Thomas, Christopher R. and Maurice S. Charles. 2005. *Managerial Economics: Concepts and Applications*, Eighth Edition. New Delhi: Tata McGraw-Hill Publishing Company Limited.

Mankiw, Gregory N. 2002. *Principles of Economics*, Second Edition. India: Thomson Press.