ALAGAPPA UNIVERSITY

[Accredited with 'A+' Grade by NAAC (CGPA:3.64) in the Third Cycle
and Graded as Category–I University by MHRD-UGC]

(A State University Established by the Government of Tamil Nadu)

KARAIKUDI – 630 003

Directorate of Distance Education

B.A. [History]

I - Semester

108 14

PRINCIPLES OF ECONOMICS

Reviewer	
Dr. K. Krishnamoorthy	Assistant Professor in Education, Directorate of Distance Education, Alagappa University, Karaikudi

Authors

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

Units (1-5, 9-10, 11.3-11.10, 12-14)

Dr Shalini Singh, *Senior Consultant, Poshan Abhiyan, Ministry of Women and Child Development, New Delhi*

Units (6, 7.0-7.2, 7.4-7.8)

Vikas® Publishing House: Units (7.3, 8, 11.0-11.2)

VIKAS® PUBLISHING HOUSE PVT. LTD.
E-28, Sector-8, Noida - 201301 (UP)
Phone: 0120-4078900 • Fax: 0120-4078999
Regd. Office: 7361, Ravindra Mansion, Ram Nagar, New Delhi 110 055
• Website: www.vikaspublishing.com • Email: helpline@vikaspublishing.com

Work Order No. AU/DDE/DE1-238/Preparation and Printing of Course Materials/2018 Dated 30.08.2018 Copies - 500

SYLLABI-BOOK MAPPING TABLE

Principles of Economics

Syllabi	Mapping in Book
BLOCK 1: Nature, Definition and Utility of Economics **UNIT-I:** Introduction – Nature and scope of economics **UNIT-II:** Definition of economic - Economics laws - Methods of economic analysis. **UNIT-III:** Consumption – Utility – Law of diminishing marginal utility – Equi-marginal utility.	**Unit 1:** Nature and Scope of Economics **(Pages 1-14)** **Unit 2:** Economic Laws **(Pages 15-24)** **Unit 3:** Consumption **(Pages 25-38)**
BLOCK 2: Demand, Production and Population Theories **UNIT- IV:** Demand – law of demand Elasticity of demand – Indifference curve analysis - Consumer's surplus. **UNIT-V:** Production – Factors of production – Laws of returns. **UNIT-VI:** Theories of population **UNIT-VII:** Malthusian and optimum theories.	**Unit 4:** Consumer's Demand **(Pages 39-77)** **Unit 5:** Theory of Production **(Pages 78-97)** **Unit 6:** Theories of Population **(Pages 98-109)** **Unit 7:** Malthusian and Optimum Theories **(Pages 110-120)**
BLOCK 3: Optimum Theories Law of Supply **UNIT-VIII:** Division of labour – Capital formation-functions of an entrepreneur. **UNIT-IX:** Internal and External economics – Cost of production – Average and marginal cost.	**Unit 8:** Division of Labour **(Pages 121-130)** **Unit 9:** Internal and External Economies **(Pages 131-143)**
BLOCK 4: Time Management and Theories of Wages and Distribution **UNIT-X:** Exchange (Theory of pricing)-Market average revenue and marginal revenue – Law of supply. **UNIT-XI:** Marshall's time analysis – perfect competition – Price determination –monopoly. **UNIT-XII:** Discriminating monopoly – Monopolistic competition – Selling cost – Oligopoly. **UNIT-XIII:** Distribution (Theory of Factor Pricing) – Marginal productivity theory of distribution. **UNIT-XIV:** Rircadian theory of rent – Quasi rent – Theories of wages – Trade union and wages - Theories of interest – Theories of profit.	**Unit 10:** Exchange (Theory of Pricing) **(Pages 144-161)** **Unit 11:** Perfect Competition and Price Determination **(Pages 162-173)** **Unit 12:** Monopoly, Monopolistic Competition and Oligopoly **(Pages 174-202)** **Unit 13:** Distribution: Theory of Factor Pricing **(Pages 203-215)** **Unit 14:** Theories of Rent and Quasi-Rent, Wages, Interest and Profit **(Pages 216-242)**

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
NATURE, DEFINITION AND UTILITY OF ECONOMICS

NOTES

UNIT 1 NATURE AND SCOPE OF ECONOMICS

Structure

1.0 INTRODUCTION

Economics is basically a study of the utilization of resources under specific conditions, all bound with a strong hope that the subject under scrutiny is a rational entity which seeks to improve its overall well-being. Two branches within the subject have evolved thus, namely, microeconomics (individual choices) which deals with entities and the interaction between those entities, while macroeconomics (aggregate outcomes) deals with the entire economy as a whole. This unit will introduce you to the nature and scope of economics.

1.1 OBJECTIVES

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

- Define economics
- State the nature of economics

- Analyse the scope of economics
- List the various branches of economics

NOTES

1.2 INTRODUCTION TO 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 is '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 fact is that economics has not yet been defined precisely and appropriately. The reason is, as Zeuthen has observed, 'Economics is an unfinished science' and as Schultz has remarked, 'Economics is still a very young science and many problems in it are almost untouched'. These observations made half-a-century ago hold still true. It seems that, after Robbins, no serious attempt was made to define economics. Defining economics has been so fruitless effort that some modern authors of economics text, including those by reputed economists like Samuelson, Baumol, and Stiglitz avoid the issue of defining economics. For example, William J. Baumol (a Nobel laureate) and Allen S. Blinder write in their own text, 'Many definitions of economics have been proposed, but we prefer to avoid any attempt to define the discipline in a single sentence or paragraph', and let 'the subject matter speak for itself.'

However, 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.

1.2.1 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, economic 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.*

1.2.2 Why Does the Problem of Making Choice Arise?

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

NOTES

1. ***Human Wants, Desires and Aspirations are Limitless:*** The history of human civilization bears evidence to the fact that human desire to consume more and more of better and better goods and services has ever since been increasing. For example, housing need has risen from a hut to luxury palace, and if possible, a house in space; the need for means of transportation has gone up from mule and camel to supersonic jet planes; demand for means of communication has risen 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) increase in 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 and pressing, 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. Also, while satisfying some wants gives a greater satisfaction than others. Given their intensity and urgency, human wants can be arranged in the order of their priority. The priority of wants, however, varies from person to person, and from time to time for the same person. Therefore, the question arises as to 'which want to satisfy first' and 'which the last'. Thus, the consumers has to make choice 'what to consume' and 'how much to consume'. Economics studies how consumers (individuals and household) make choice between their wants and how they allocate their expenditure between different kinds of goods and services they choose to consume.

2. ***Resources are Scarce:*** The need for making choice between the various goods that people want to produce and consume arises mainly because *resources that are available to the people at any point of time for satisfying their wants are scarce and limited.* What are the resources? Conceptually, 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, equipment, 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 other categories of resources, viz., *time, technology* and *information.* All these *resources are scarce.* Resource scarcity is a relative term. It implies that resources are scarce in relation to the demand for resources. The *scarcity of resources is the mother of all economic problems.* If resources were unlimited, like human wants, there would be no economic problem and, perhaps, no economics as a subject of study. *It is the scarcity of resources in relation to human wants that forces people to make choices*.

Furthermore, the problem of making choice arises also because resources have *alternative uses* and alternative uses have *different returns or earnings.* For example, a building can be used to set up a shopping center, business office, a 'public school', a hospital or for residential purpose. But the return on building varies from use to use of the building. Therefore, a return maximizing building 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. This is called **opportunity cost**. Economics as a social science analyses how people (individuals and society) make their choices between the economic goals they want to achieve, between the goods and services they want to produce, and between the alternative uses of their resources with the objective of maximizing their gains. The gain maximizers 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. This is what constitutes, according to Robbins, the subject matter of economics.

3. ***People are Gain Maximizers.*** Yet another important aspect of human nature that leads to the choice-making behaviour is that most people aim at maximizing their gains from the use of their limited resources. 'Why people want to maximize their gains' is no concern of economics? Traditional economics assumes maximizing behaviour of the people as a part of their rational economic behaviour. This assumption is based on observed facts. As *consumers,* they want to maximize their utility or satisfaction; as *producers,* they want to maximize their output or profit; and as *factor owners,* they want to maximize their earnings. *People's desire to maximize their gains* is a very important aspect of economic behaviour of the people giving rise to economics. If 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.

1.2.3 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. Robbins' definition of economics confines subject matter of economics to the study of economic behaviour of the people at its micro level. It may thus be said that Robbins' definition confines economics to what is now called *microeconomics*. But economics goes far beyond the scope of microeconomics. If economics is confined to the study of choice-making behaviour of the individual economic man, many other and more important economic issues that constitute a major part of modern economic science will have to be left out. 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 determine the overall economic growth of a country?
- What causes fluctuations in the national output, employment and the general price level?
- How do international trade and international capital flows affect the domestic economy?
- What causes inflation 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 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?

One can point out many other issues which do not fall within the purview of microeconomics. Analysis of and finding answer to such economic problems constitute now a major and also a more important subject matter of economics than the choice-making aspect of it. The study of the issues mentioned above has created a relatively new branch of economics, called *macroeconomics*. As noted above, the *scope of economics* continues to grow and expand in its scope, size and analytical rigour. Boundaries of economic science are not yet precisely marked though economics is claimed to be 'the oldest and best developed of the social sciences'. Let us now have a glance at the *scope of economics* as it is known today and its major branches and specialized areas of economics.

1.3 NATURE AND SCOPE OF ECONOMICS

NOTES

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 now divided into two major branches: *Microeconomics* and *Macroeconomics.* A brief description of the subject matter and approach of microeconomics and macroeconomics follows.

1.3.1 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, microeconomics 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*.

1.3.2 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 studies also the impact of public revenue and public expenditure, government's economic activities and policies on the economy. An important aspect of macroeconomics studies is the consequences of international trade and other economic relations between the nations. The study of these aspects of economic phenomena constitutes the major themes of *macroeconomics*.

1.3.3 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, international 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.

NOTES

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'.

1.3.4 Is Economics a Positive or a Normative Science?

What is a Positive and a Normative Science? A *positive science* studies the phenomena as they actually are or as they actually happen. It does not involve any

NOTES

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.

1.3.5 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.

1.3.6 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,

NOTES

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.

NOTES

Check Your Progress

1. Define economic behaviour.
2. State any two reasons due to which the problem of making choices arises.
3. Name the two main branches of modern economics.
4. What does monetary economics deal with?

1.4 ANSWERS TO 'CHECK YOUR PROGRESS' QUESTIONS

1. 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.
2. Two reasons due to which the problem of making choices arises are the following:
   - Human Wants, Desires and Aspirations are Limitless
   - Resources are Scarce
3. The two main branches of modern economics are Microeconomics and Macroeconomics.
4. 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.

1.5 SUMMARY

- Adam Smith (1776) defined economics is 'an inquiry into the nature and causes of the wealth of the nations'.
- 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'.
- *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.
- 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.*

NOTES

- *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 between the various goods that people want to produce and consume arises mainly because *resources that are available to the people at any point of time for satisfying their wants are scarce and limited.*
- Microeconomics 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.
- *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.
- 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.
- 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.
- 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.

1.6 KEY WORDS

- **Macroeconomics:** It 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.
- **Industrial Economics:** It 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.

- **Microeconomics:** It is concerned with microscopic study of the various elements of the economic system and not with the system as a whole.

NOTES

1.7 SELF-ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. Define economics.
2. Write a short note on the nature of economics.
3. Briefly mention the specialized branches of economic studies.
4. What is the main function of economics?

Long-Answer Questions

1. Analyse economics as a social science.
2. 'Economics is both a positive and normative science.' Discuss.
3. What is opportunity cost?

1.8 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 2 ECONOMIC LAWS

NOTES

Structure

2.0 INTRODUCTION

In the previous unit you studied about the nature and scope of economics. Let us now study the nature and limitations of economic laws. The pertinent questions which may be asked here are: How valid are the 'economic laws'? Are the economic laws as exact as the laws of natural sciences? To what extent can economic laws guide the formulation of economic policies? Let us now see how the economists have answered these questions.

2.1 OBJECTIVES

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

- Define economics
- Examine the nature of economic laws
- Discuss the methods of economic analysis

2.2 DEFINTION OF ECONOMICS

Managerial economics essentially constitutes of economic theories and analytical tools that are widely applied to business decision-making. It is therefore useful to know, ***'what is economics'***. Economics is a social science. Its basic function is to study how people—individuals, households, firms and nations—maximize their

NOTES

gains from their limited resources and opportunities. In economic terminology, this is called *maximizing behaviour* or, more appropriately, *optimizing behaviour. Optimizing behaviour is, selecting the best out of available options with the objective of maximizing gains from the given resources*. Economics is thus *a social science, which studies human behaviour in relation to optimizing allocation of available resources to achieve the given ends*. For example, economics studies how households allocate their limited resources (income) between the various goods and services they consume so that they are able to maximize their total satisfaction. It analyzes how households with limited income decide 'what to consume' and 'how much to consume' with the aim of maximizing total utility.

Consider the case of firms, the producers of goods and services. Economics studies how producers—the firms—make the choice of the commodity to produce, the production technology, location of the firm, market or market segment to cater to, price of the product, the amount to spend on advertizing (if necessary) and the strategy for facing competition, etc.

At macro level, economics studies how nations allocate their resources, men and material, between competing needs of the society so that economic welfare of the society can be maximized. Also, economics studies how government formulates its economic policies—taxation policy, expenditure policy, price policy, fiscal policy, monetary policy, employment policy, foreign trade (export-import policy), tariff policy, etc. – and, the effects of these policies.

Economics is obviously a study of the choice-making behaviour of the people. In reality, however, choice-making is not as simple as it looks because the economic world is very complex and most economic decisions have to be taken under the condition of imperfect knowledge, risk and uncertainty. Therefore, taking an appropriate decision or making an appropriate choice in an extremely complex situation is a very difficult task. In their endeavour to study the complex decision-making process, economists have developed a large kit of *analytical tools and techniques* with the aid of mathematics and statistics and have developed a large corpus of *economic theories* with a fairly high predictive power. *Analytical tools and techniques, economic laws and theories constitute the body of economics.*

2.2.1 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.

NOTES

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 due mainly 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 help to recognize the structure of managerial problems, eliminate the minor details that might obstruct decision-making, and help to concentrate on the main issue.

NOTES

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.

2.3 ECONOMIC LAWS

Let us now study the nature of economic laws.

2.3.1 Economic Laws are Statement of Tendencies

As Marshall defined it, *economic laws are statement of economic tendencies*. They bring out the relationships between the economic variables. The existence of such relationships is subject to the conditions specified by underlying assumptions. Validity of the economic laws therefore depends on (a) the reliability of assumptions, and

(b) logical consistency between assumptions and the hypotheses. Given these conditions, if one examines the validity of economic laws within the framework of its model, one may find economic laws as valid as the laws of natural science. For, the laws of all sciences are statements of tendencies. For example, even the time-honoured Newton's law of gravitation is a statement of tendency. As the law says, gravitation makes things fall to the ground *if nothing occurs to prevent it*. But when a balloon filled with gas (which is lighter than air) is dropped from a height, the pressure of air makes it go further up in the air despite the existence of gravitational force. Thus, the law of gravitation works only within the specified conditions. "The law of gravitation is, therefore, a statement of tendencies." To this extent economic laws are comparable to natural laws.

2.3.2 Economic Laws are Not as Exact and Precise as Natural Laws

In spite of economic laws being comparable, to an extent, to natural laws, economic laws are not as exact and precise as the laws of natural sciences. Also, economic laws do not operate as steadily as the laws of natural sciences. The reason is natural sciences deal with the elements of nature which move or act in a stable and predictable manner. On the contrary, economics deals with human behaviour. Human beings differ in their likings, dislikings, choices and preferences. Their behaviour is governed by impulses, reactions, imitative habits, social custom and values. Therefore, human beings act and react differently to the same economic event. For example, if price of a commodity falls, all its consumers or users do not increase their purchase uniformly—some buy more; some buy the same quantity as they used to before the fall in price; and some may not buy the commodity at all. Thus, as Marshall has observed, "…there are no economic tendencies which act as steadily and can be measured as exactly as gravitation can: and consequently

NOTES

there are no laws of economics which can be compared for precision with the laws of gravitation." But, there are sciences other than astronomy, whose laws are inexact, e.g., the 'science of the tide.' Marshall suggested, "The laws of economics are to be compared with the laws of tides, rather than with the simple and exact law of gravitation. For the actions of men are so various and uncertain that the best statement of tendencies which we can make in a science of human conduct, must… be inexact and faulty."

2.3.3 Economic Laws are Hypothetical

It has also been alleged that economic laws are hypothetical as they are based on *ceteris paribus* assumptions, i.e., "all other things remain the same". But, all scientific doctrines assume, tacitly or implicitly, certain conditions and are in this sense hypothetical. It cannot however be denied that economic laws are more hypothetical than the laws of natural sciences. Natural laws possess greater predictive power than the economic laws. The reason for economic laws being less exact is that the nature and behaviour of variables with which economists deal are extremely complex and lack uniformity. The conditions under which they carry out their studies are very uncertain to hold. While in natural sciences, scientists can subject their object of study to laboratory tests and experiments, economists cannot experiment the behaviour of a household or a firm or an industry in a laboratory and determine their behavioural pattern. The interrelatedness and inseparability of economic variables make the experiments of economic laws impossible. Economists have therefore to rely to a great extent on logic and intuition which are subject to variation from economist to economist. These problems hamper immensely the formulation of exact and precise economic laws.

2.3.4 Economic Laws are Based on Unrealistic Assumptions

Another criticism against economic theories is that they are based on "unrealistic assumptions" and, therefore, they are unrealistic. For example, the whole conventional theory of firm is based on the assumption that "all business firms maximise their profits." As we shall see later, many economists claim that not all firms maximise profits. Profit maximisation assumption is therefore unrealistic and so is the theory of firm.

Milton Friedman has however argued that what is important in formulation of a theory is not whether its assumptions are "realistic" or "unrealistic". What is important is the predictive power of the theory notwithstanding the lack of "realism" in their assumptions. In his own words, "Insofar as a theory can be said to have 'assumptions' at all, and insofar as their 'realism' can be judged independently of the validity of predictions, the relation between the significance of a theory and the 'realism' of its 'assumptions' is almost the opposite of that suggested by the view under criticism. Truly important and significant hypotheses will be found to have 'assumptions' that are wildly inaccurate descriptive representations of reality, and in general, the more significant the theory, the more unrealistic the assumptions (in

NOTES

this sense). The reason is simple. A hypothesis is important if it 'explains' much by little, that is, if it subtracts the common and crucial elements from the mass of complex and detailed circumstances surrounding the phenomena to be explained and permits valid predictions on the basis of them alone. To be important, therefore, a hypothesis must be descriptively false in its assumptions; it takes account of, and accounts for, none of the many other attendant circumstances, since its very success owes them to be irrelevant for the phenomena to be explained. To put this point less paradoxically, the relevant question to ask about the assumption of a theory is not whether they are descriptively 'realistic' for they never are, but whether they are sufficiently good approximations for the purpose in hand. And this question can be answered only by seeing whether it yields sufficiently accurate predictions." Although, many authors have attempted to refute Friedman's argument, it still holds its ground.

2.3.5 Why Economic Laws?

The question that arises now is: If economic laws are hypothetical, inexact and imprecise, why do economists then formulate such laws? There are two basic purposes behind the formulation of economic laws, even if they are inexact and imprecise. First, to understand the economic conduct of the society, and second, to predict the future course of social conduct and of economic variable under stipulated conditions. Besides, the experience shows that, in spite of their inexactness, economic laws serve a great deal in formulating appropriate economic policies to control and regulate economic activities of the people to achieve certain predetermined social goals. In the words of Marshall, "By the fundamental impulses of our nature we all—high and low, learned and unlearned—are in our several degrees constantly striving to understand the course of human action, and to shape them for our purposes, whether selfish or unselfish, whether noble or ignoble."

According to Keynes, "Theory of economics does not furnish a body of settled conclusions immediately applicable to policy. It is a method rather than a doctrine, an apparatus of mind, a technique of thinking which helps its processor draw correct conclusions."

Besides, as economists suggests, *the objective of economic analysis is not to formulate exact and precise economic laws but to develop the understanding of the economic system. The basic function of economic theories is to provide a framework for logical economic thinking.* The need for such a framework arises because economic world is too complex to permit considering every bit of economic facts that affect economic decisions. It therefore becomes inevitable to abstract a particular phenomenon from complex reality and analyse it, by assuming away the complexities which may arise due to 'spill-out' and 'feed-back' effect. In the words of Keynes himself, "The object of our analysis is, not to provide a machine, or method of blind manipulation, which will furnish an infallible answer, but to provide ourselves with an organised and orderly method of thinking out particular problems." As Boulding has pointed out, the objective of economic analysis is to present the 'map' of *reality* rather than a perfect picture of it. And, "just as we do not expect a map to show every tree, every house, and every blade of grass in the landscape, so we should not expect economic analysis to take into account every detail and quirk of real economic behaviour."

Apart from satisfying the human urge to possess knowledge of the economic world in which they live and mapping the realities of the economic world, economic laws serve many useful purposes in our practical life.

First, economic laws indicate the consequences of various alternative actions and thus provide "an intelligent basis for choice among the possible actions."

Secondly, economic laws not only explain the behaviour of various economic units, as consumers, firms, resource owners, etc., but also guide them to a rational and logical behaviour, given their goals of economic activities.

Thirdly, economic principles, despite their limitations, provide not only the rules for optimum allocation of available resources but also the test for examining the efficiency of existing pattern of resource allocation.

Fourthly, economic principles provide guidelines for formulating appropriate economic policies of the government to achieve the desirable social ends.

Finally, another very important function of the economic theories is to predict the future. Each economic unit consciously or unconsciously plans the course of its future action and speculates about the future conditions. Economics principles provide tools for predicting the future course of economic events under stipulated conditions. Knowledge about the future economic conditions facilitates planning the future actions.

To conclude, in spite of the fact that economic laws are inexact and hypothetical in nature, they serve many useful purposes in practical life as they guide human actions so as to achieve their economic goals.

2.4 METHODS OF ECONOMIC ANALYSIS

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

NOTES

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

1. Mention the two factors on which the validity of economic laws is dependent.
2. What is the basic function of economics?
3. State the basic steps involved in the process of model building and theorization in economics.

2.5 ANSWERS TO 'CHECK YOUR PROGRESS' QUESTIONS

1. The two factors on which the validity of economic laws is dependent are:
   (a) The reliability of assumptions and (b) logical consistency between the assumptions and the hypotheses.
2. Economics is a social science. Its basic function is to study how people—individuals, households, firms and nations—maximize their gains from their limited resources and opportunities.

3. The basic steps involved in the process of model building and theorization in economics are the following:
   (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.

2.6 SUMMARY

- *Managerial economics* essentially constitutes of economic theories and analytical tools that are widely applied to business decision-making.
- At macro level, economics studies how nations allocate their resources, men and material, between competing needs of the society so that economic welfare of the society can be maximized.
- 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.
- As Marshall defined it, *economic laws are statement of economic tendencies*. They bring out the relationships between the economic variables.
- In spite of economic laws being comparable, to an extent, to natural laws, economic laws are not as exact and precise as the laws of natural sciences.
- Another criticism against economic theories is that they are based on "unrealistic assumptions" and, therefore, they are unrealistic. For example, the whole conventional theory of firm is based on the assumption that "all business firms maximise their profits."
- According to Keynes, "Theory of economics does not furnish a body of settled conclusions immediately applicable to policy. It is a method rather than a doctrine, an apparatus of mind, a technique of thinking which helps its processor draw correct conclusions."
- 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*.
- 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.
- *Economic variables* are measurable quantities, e.g., consumer goods, output, inputs, money, income, etc. The economic variables assumed to remain constant are called *parameters*.

2.7 KEY WORDS

NOTES

- **Economic variables:** These are measurable qualities for example, consumer goods, outputs, money and so forth.
- **Economic theory:** It is the statement of cause-and-effect relationship between two or more observed facts of real economic life.
- **Model:** It is an abstraction from reality. It represents reality in a simplified form.

2.8 SELF-ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. Define economic laws.
2. Why do managers need to know economics?
3. Mention the three main contributions of economic theory to business economics as discussed by Baumol.

Long-Answer Questions

1. Analyse the need of having economic laws.
2. Discuss the methods of economic analysis.
3. ‘Economic laws are hypothetical.’ Explain the statement.

2.9 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 3 CONSUMPTION

NOTES

Structure

3.0 INTRODUCTION

The consumers demand a commodity because they derive or expect to derive *utility* from 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. In this unit, you will study about the concept of utility, Law of Diminishing Marginal Utility and Equi-marginal Utility.

3.1 OBJECTIVES

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

- Define the concept of utility
- Discuss the law of diminishing marginal utility
- Explain equi-marginal utility

3.2 UTILITY-LAW OF DIMINISHING MARGINAL UTILITY

The concept of utility can be looked upon from two angles—from the commodity angle and from the consumer's angle. From the commodity 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*.

NOTES

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 or a profession like prostitution.

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 a 'relative' concept. In the subjective sense, utility is a matter of one's own feeling of satisfaction. In the relative sense (*i*) a commodity need not be useful for all, for example, 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 at different moods of a consumer. In consumer analysis, only the 'subjective' concept of utility is used.

3.2.1 Measurability of Utility

Utility is a psychological phenomenon. It is a feeling of satisfaction, pleasure or happiness. Is utility measurable quantitatively? Measurability of utility has, however, been a contentious issue. The classical economists, viz., Jeremy Bentham, Leon Walrus, Carl Menger, etc. and the neo-classical economist, 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. The modern economists, most notably J.R. Hicks and R.G.D. Allen, however, hold the view that utility is not quantitatively measurable—it is not measurable in absolute terms. Utility can be expressed only ordinally comparatively 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. 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.

(i) Cardinal Measurement of Utility. Some early psychological experiments on an individual's responses to various stimuli led classical and 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. The neo-classical economists, especially Marshall, devised a method of measuring utility. According to Marshall, utility of a commodity for a person equals the amount of money he/she is willing to pay for a unit of the commodity. In other words, price one is prepared to pay for a unit of a commodity equals the utility he expects to derive from the commodity. They formulated 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 economic analysis, they assumed (*i*) that one 'util' equals one unit of money, and (*ii*) that utility of money remains constant.

NOTES

It has, however, been realised 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., the utility. Numerous factors affect the state of consumer's mood, which are impossible to determine and quantify. *Utility is therefore immeasurable in cardinal terms*.

(ii) Ordinal Measurement of Utility. The modern economists have discarded the concept of *cardinal utility* and have instead employed the concept of *ordinal utility* for analysing 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 numerical terms, but it is always possible for them to tell introspectively whether a commodity is more or less or equally useful as compared to another. For example, a consumer may not be able to tell that a bottle of pepsi gives 5 utils and a glass of fruit juice gives 10 utils. But he or she can always tell whether a glass of fruit juice gives more or less utility than a bottle of pepsi. This assumption forms the basis of the ordinal theory of consumer behaviour.

To sum up, the neo-classical economists maintained that cardinal measurement of utility is practically possible and is meaningful in consumer analysis. The modern economists, on the other hand, maintain that utility being a psychological phenomenon is inherently immeasurable quantitatively. They also maintain that the concept of ordinal utility is a feasible concept and it meets the conceptual requirement of analysing the consumer behaviour. However, both the concepts of utility are used in analysing consumer behaviour.

3.2.2 Two Approaches to 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 or Marshallian approach.

(*ii*) ***Ordinal Utility Approach***, pioneered by J.R. Hicks, a Nobel laureate and R.G.D. Allen, is also called Hicks-Allen approach or 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 managerial decisions 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 neo-classical demand theory based on cardinal utility has retained its appeal and applicability to the analysis of

NOTES

market behaviour. Besides, the study of neo-classical demand theory serves as a foundation for understanding the advanced theories of consumer behaviour. The study of neo-classical theory of demand is of particular importance and contributes a great deal in managerial decisions.

3.3 CARDINAL UTILITY APPROACH TO CONSUMER DEMAND

The central theme of the consumption theory is the analysis of *utility maximizing behaviour* of the consumer. The fundamental postulate of the consumption theory is that all the consumers—individuals and households—aim at *utility maximisation* 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?

(*ii*) how does he or she allocate his/her total consumption expenditure on various commodities he/she consumes so that his/her total utility is maximized?

As mentioned above, the theory of consumer behaviour postulates that consumers seek to maximize their total utility or satisfaction. On the basis of this postulate, consumption theory explains how a consumer attains the level of maximum satisfaction under the following assumptions.

Assumptions: The cardinal utility approach to consumer analysis makes the following assumptions.

(*i*) **Rationality**. It is assumed that the consumer is a rational being in the sense that he/she satisfies his/her wants in the order of their preference. That is, he/she buys that commodity first which yields the highest utility and that 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 assumed that utility is cardinally measurable and that utility of one unit of a commodity equals the money a consumer is prepared to pay for it or 1 util = 1 unit of money.

(*v*) **Diminishing marginal utility.** Consumption of a commodity is subject to the law of diminishing marginal utility, i.e., the utility derived from the successive units of a commodity goes on decreasing as a consumer consumes more and more units of the commodity. 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 by a consumer can be added together to obtain the total utility. For example, suppose a person consumes X number of goods. His total utility can be expressed as:

$$TU = UX_1 + UX_2 + UX_3 + ... + UX_n$$

where X_1, X_2, ... X_n denote the total quantities of the various goods consumed.

NOTES

3.3.1 Total and Marginal Utility

Before we proceed to explain and illustrate the law of diminishing marginal utility, let us explain the concept of *total* and *marginal utility* used in the explanation of the law of diminishing marginal utility.

Total Utility: Assuming that utility is measurable and additive, total utility may be defined as the sum of the utilities derived by a consumer from the various units of goods and services he consumes. Suppose a consumer consumes four units of a commodity, X, at a time and derives utility as u_1, u_2, u_3 and u_4. His total utility (TU_x) from commodity X can be measured as follows.

$$TU_x = u_1 + u_2 + u_3 + u_4$$

If a consumer consumes n number of commodities, his total utility, TU_n, will be the sum of total utilities derived form 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: The marginal utility is another most 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 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. It may be expressed as

$$MU = \frac{\Delta TU}{\Delta Q}$$

where TU = total utility, and ΔQ = change in quantity consumed by one unit.

NOTES

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

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

Having explained the concept of total utility (*TU*) and marginal utility (*MU*), let us now discuss the law of diminishing marginal utility.

3.3.2 The Law of Diminishing Marginal Utility

Let us begin our study of consumer demand with the law of 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 increases, the utility derived from each successive unit decreases, 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., *rasgullas*, keeping the consumption of all other commodities constant, the utility which he derives from the successive rasgullas he consumes goes on diminishing. This law applies to all kinds of consumer goods—durable and non-durable sooner or later.

To explain the law of diminishing marginal utility, let us suppose that a consumer consumes 6 units of a commodity *X* and his/her total and marginal utility derived from various units of *X* are as given in Table 3.1.

Table 3.1 *Total and Marginal Utility Schedules of X*

No. of units consumed	*Total utility*	*Marginal utility*
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 consumed per unit of time, the *TU* increases but at a diminishing rate. The diminishing rate of increase in the total utility gives the measure of marginal utility. The diminishing *MU* is shown in the last column of the table. Fig. 3.1 illustrates graphically the law of diminishing *MU*. The rate of increase in *TU* as the result of increase in the number of units consumed is shown by the *MU* curve in Fig. 3.1. The downward sloping *MU* curve shows that marginal utility goes on decreasing as consumption increases. At 4 units consumed, the *TU* reaches its maximum level, i.e., 65 utils. Beyond this, *MU* becomes negative and *TU* begins to decline. The downward sloping *MU* curve illustrates the law of diminishing marginal utility.

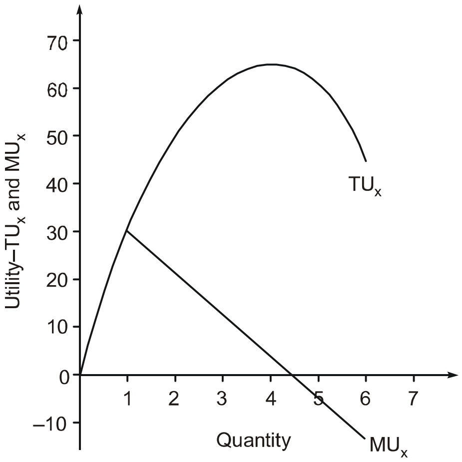


Fig. 3.1 *Diminishing Marginal Utility*

NOTES

Why the *MU* Decreases: 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 and the intensity of his need goes on decreasing. Therefore, the utility obtained from each successive unit goes on decreasing.

Necessary Conditions: 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. Otherwise, the law of diminishing *MU* may not apply.

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, but eventually it does decrease. As a matter of fact, the law of marginal utility generally operates universally.

3.3.3 Consumer Equilibrium - Equi-marginal utility

From economic analysis point of view, a consumer is a utility maximizing entity. From theoretical point of view, therefore, a consumer is said to have reached his equilibrium position when he has maximized the level of his satisfaction, given his

NOTES

resources and other conditions. Technically, a utility-maximizing consumer reaches his equilibrium position when allocation of his consumption expenditure is such that the last penny spent on each commodity yields the same utility. How does a consumer reach this position?

Given the assumptions, a rational and utility-maximising consumer consumes commodities in the order of their utilities. He picks up first 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 another 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 the state of consumer's equilibrium.

Consumer's equilibrium is analysed under two conditions:

(*i*) a consumer consuming only one commodity, and

(*ii*) a consumer consuming many commodities.

Let us first explain and illustrate consumer's equilibrium in a simple case assuming that the consumer spends his total income on only one commodity.

(i) Consumer's Equilibrium: One-Commodity Case. We explain and illustrate here consumer's equilibrium in a simple one-commodity model. Suppose that a consumer with certain 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) as asset, 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) as an asset remains constant (assumption 6). Therefore, the consumer will exchange 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 point 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 $P_x\,(MU_m)$ 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 above

NOTES

point E, $MU_x > P_x$ (MU_x). Therefore, the utility maximizing consumer would exchange his money for commodity X, and will increase his total satisfaction because his gain in terms of MU_x is greater than his loss in terms of MU_m. This conditions exists till he reaches point E. And, 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.

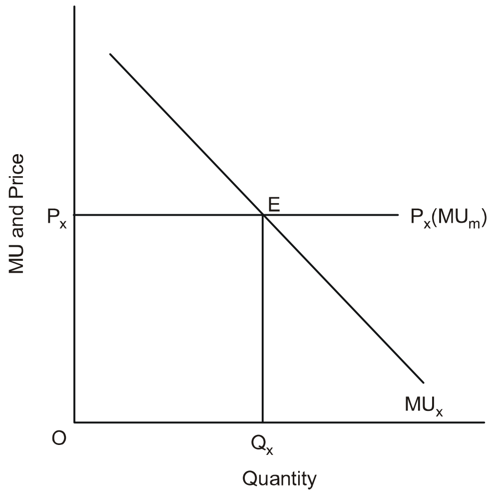


***Fig. 3.2** Consumer's Equilibrium*

(ii) Consumer's Equilibrium in Multicommodity Case: The Law of Equi-Marginal Utility. In the previous section, we have explained consumer's equilibrium assuming that the consumer consumes a single commodity. In real life, however, a consumer consumes multiple 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.

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 on 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 simplicity of, however, we will 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 distribute his income between commodities X and Y, so that

$$MU_x = P_x\ (MU_m)$$

and $$MU_y = P_y\,(MU_m)$$

Given these conditions, the consumer is in equilibrium where

NOTES

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

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

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.

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

$$\frac{MU_A}{P_A} = \frac{MU_B}{P_B} = \cdots = \frac{MU_Z}{P_Z} = MU_m \qquad ...(3.3)$$

Equation (3.3) 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.

3.3.4 Derivation of Individual Demand Curve for a Commodity

We have explained, in the preceding sections, the consumer's equilibrium using one-commodity and multi-commodity models. The theory 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. Marshall gave the equilibrium condition for the consumption of a commodity, say *X*, as $MU_x = P_x\,(MU_m)$. 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 curve for the commodity *X* is illustrated in Fig. 3.3 (*a*) and 3.3 (*b*). Let us first look at Fig. 3.3 (*a*). Suppose price of *X* is initially given at P_3 and the consumer is in equilibrium at point E_1 where $MU_x = P_3$ (MU_m). Here, equilibrium quantity is OQ_1. Now if price of the commodity *X* falls to P_2, the consumer reaches a new equilibrium position at point E_2 where $MU_x =$

P_2 (MU_m). Similarly, if price falls further, he/she buys and consumes more to maximize his/her satisfaction. This behaviour of the consumer can be used to derive his/her demand curve for commodity *X*.

NOTES

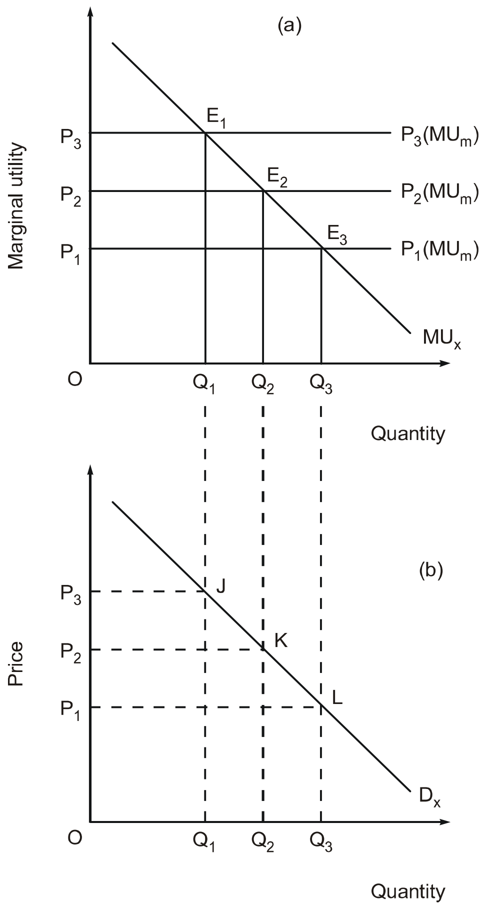


Fig. 3.3 *Derivation of Demand Curve*

Fig. 3.3 (*a*) reveals that when price is P_3, equilibrium quantity is OQ_1. When price decreases to P_2, equilibrium point shifts downward to point E_2 at which 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 when price goes on decreasing, the corresponding quantity goes on decreasing. This means that as price decreases, the equilibrium quantity increases. This inverse price-quantity relationship is the basis of the law of demand.

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

NOTES

Demand under Variable MU_m. We have explained above the consumer's equilibrium and derived the demand curve under the assumption that MU_m remains constant. This analysis holds even if MU_m is assumed to be variable. This aspect is explained below.

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.

Check Your Progress

1. Mention the two approaches to consumer demand analysis.
2. Define total utility.

3.4 ANSWERS TO 'CHECK YOUR PROGRESS' QUESTIONS

1. The two approaches to consumer demand analysis are the following:
   - Cardinal utility approach
   - Ordinal utility approach
2. Total utility may be defined as the sum of the utilities derived by a consumer from the various units of goods and services he consumes.

3.5 SUMMARY

- The concept of utility can be looked upon from two angles—from the commodity angle and from the consumer's angle. From the commodity 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*.
- Some early psychological experiments on an individual's responses to various stimuli led classical and neo-classical economists to believe that utility is measurable and cardinally quantifiable.

NOTES

- 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 numerical terms, but it is always possible for them to tell introspectively whether a commodity is more or less or equally useful as compared to another.
- The central theme of the consumption theory is the analysis of *utility maximizing behaviour* of the consumer.
- Assuming that utility is measurable and additive, total utility may be defined as the sum of the utilities derived by a consumer from the various units of goods and services he consumes.
- The marginal utility is another most important concept used in economic analysis. Marginal utility may be defined in a number of ways.
- 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 and the intensity of his need goes on decreasing. Therefore, the utility obtained from each successive unit goes on decreasing.
- From economic analysis point of view, a consumer is a utility maximizing entity. From theoretical point of view, therefore, a consumer is said to have reached his equilibrium position when he has maximized the level of his satisfaction, given his resources and other conditions.

3.6 KEY WORDS

- **Utility:** It is an economic term alluding to the total satisfaction received from consuming a good or service.
- **Marginal utility:** It is the addition to the total utility resulting from the consumption of one additional unit of the commodity.
- **Ordinal utility:** 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 numerical terms, but it is always possible for them to tell introspectively whether a commodity is more or less or equally useful as compared to another.

3.7 SELF-ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. State the Law of Diminishing Marginal Utility.
2. Mention the assumptions of the Law of Diminishing Marginal Utility.

3. Give one example to illustrate the analysis of consumer equilibrium in one-commodity case.

Long-Answer Questions

NOTES

1. Discuss the cardinal utility approach to consumer demand.
2. Explain the law of Equi-marginal Utility.
3. Differentiate between cardinal measurement of utility and ordinal measurement of utility.

3.8 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 - II

DEMAND, PRODUCTION AND POPULATION THEORIES

NOTES

UNIT 4 CONSUMER’S DEMAND

Structure

4.0 INTRODUCTION

In the previous unit, you studied about the concept of consumption, utility, law of diminishing marginal utility and equi-marginal utility. In this unit, you will study about the concept of demand in detail. The unit will introduce you to the law of demand, demand curve, demand schedule, concept of market demand, determinants of market demand, elasticity of demand and the measurement of various kinds of demand elasticities, indifference curve analysis and finally the concept of consumer equilibrium.

4.1 OBJECTIVES

NOTES

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

- Define the law of demand
- Discuss the concept of elasticity of demand
- Explain indifference curve analysis
- Define the concept of consumer equilibrium

4.2 DEMAND: LAW OF DEMAND

Conceptually, demand can be defined as the desire for a good for which one has sufficient purchasing power and willingness to pay for the good. In simple words, *demand is a desire for a good, backed by ability and willingness to pay.* A desire without sufficient resources (money income) is merely a wish. A desire with resources but without willingness to spend is only a potential demand. Before we proceed, let us know the distinction between individual and market demand.

Individual and Market Demand. *Individual demand* can be defined as the quantity of a commodity that an individual is willing to buy at a given price over a specified period of time, say per day, per week, per month, etc. *Market demand* refers to the total quantity that all the users of a commodity are willing to buy at a given price over a specific period of time. In fact, market demand is the sum of individual demands. Both individual and market demands are governed by the law of demand. The law of demand is discussed below in detail.

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 Demand Schedule

A *demand schedule* is a tabular presentation of different prices of a commodity and its corresponding quantity demanded per unit of time. A hypothetical market demand schedule is given in Table 4.1. This table presents price of mobiles and the corresponding number of mobiles demanded per month.

Table 4.1 illustrates the law of demand. As data given in the table shows, the demand for mobiles increases as its price decreases. For instance, at price ` 8000 per phone, only 10 thousand mobiles are demanded per month. When price decreases to ` 4000, the demand for mobiles increases to 30 thousand and when price falls further to ` 1000, the demand rises to 80 thousand phones. This relationship between price and quantity demanded gives the law of demand.

Table 4.1 *Monthly Demand Schedule for Mobiles*

(Price in `)	*(No. of Phones in '000)*
8000	10
6000	18
4000	30
3000	40
2000	60
1000	80

4.2.3 Demand Curve

A *demand curve* is a graphical presentation of the demand schedule. A demand curve is obtained by plotting a demand schedule. For example, when the data given in the demand schedule (Table 4.1) is plotted as in Fig. 4.1, it yields points *B*, *J*, *K*, *L*, *M* and *N*. By joining these points by a line, we get the demand curve *DD*′. The curve *DD*′ in Fig. 4.1 depicts the law of demand. It slopes downward to the right. It has a negative slope. The negative slope of the demand curve *DD*′ shows the *inverse relationship* between the price of mobiles and its quantity demanded. It shows that demand for mobiles increases with the decrease in its price and decreases with rise in its price. As can be seen in Fig. 4.1, downward movement on the demand curve *DD*′ from point *D* towards *D*′ shows fall in price and rise in demand. Similarly, an upward movement from point *D*′ towards *D* reads rise in price and fall in demand.

NOTES

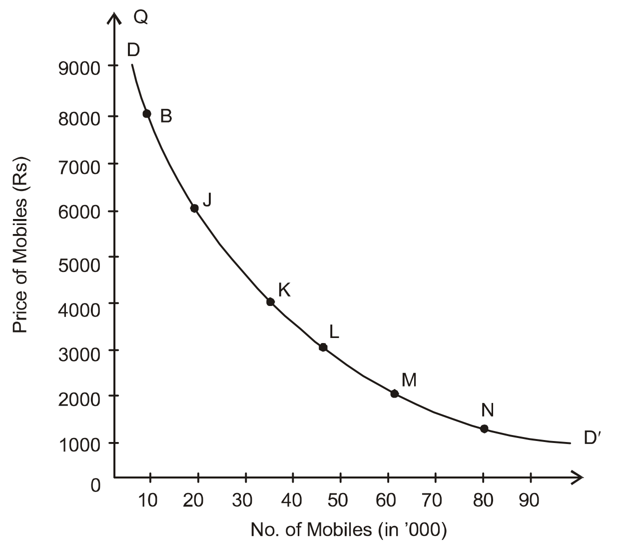


Fig. 4.1 *The Demand Curve*

The law of demand is based on an empirical fact. For example, when prices of mobiles and personal computers (PCs), specially of the latter, were astronomically high, only a few rich persons and big firms could afford them. Now with the revolution in computer and mobile phone technology and the consequent fall in their prices, demand for these goods has shot up in India. Currently, India has the second largest market for mobiles—second only to China. Every fourth Indian has a mobile phone.

4.2.4 Why the Demand Curve Slopes Downward to the Right

Figure 4.1 shows that *demand curve slopes downward to the right.* Why does it happen? 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 costier 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.

4.2.5 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 prstigeous 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).

NOTES

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[2] 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.

4.2.6 The Concept of Market Demand

Market demand for a commodity is the sum of all individual demands for the commodity at a given price, per unit of time. Suppose, there are only three consumers (*A*, B and C) of Pepsi and their weekly individual demand for Pepsi at its different prices is given as in Table 4.2. The last column of the table shows the market demand, i. e., the aggregate of individual demands for Pepsi.

Table 4.2 *Individual and Market Demand for the Pepsi Cans*

Price (`)	*No. of Pepsi Cans demanded by* *A*	*B*	*C*	*Market demand =A+B+C*
12	0	0	0	0
10	0	0	4	4
8	0	4	8	12
6	3	8	12	23
4	5	12	16	33
2	8	16	20	44
0	11	20	24	55

Derivation of Market Demand Curve. The last column of Table 4.2 shows weekly market demand for Pepsi. The *market demand curve* can be obtained by plotting the data in the last column of the table. Alternatively, market demand curve can be derived graphically by horizontal summation of the individual demand curves at each price of Pepsi. Graphical derivation of the market demand curve is illustrated in Fig. 4.2. The individual demand curves of buyers *A, B* and *C* are shown by the demand curves D_A, D_B and D_C, respectively. Horizontal summation of these demand curves produces weekly *market demand curve* for Pepsi as shown by the curve D_M. Thus, a *market curve is horizontal summation of individual demand curves at different prices.*

It is important to note here that there is a significant difference between the individual demand curves and the market demand curve. The individual demand curves may not slope downward in case of many consumer goods, e.g., a book by an author, umbrella, cinema ticket for a show, or a travel ticket, etc. But market demand for all such goods, slopes downward following the decrease in their prices, due to increase in the number of consumers.

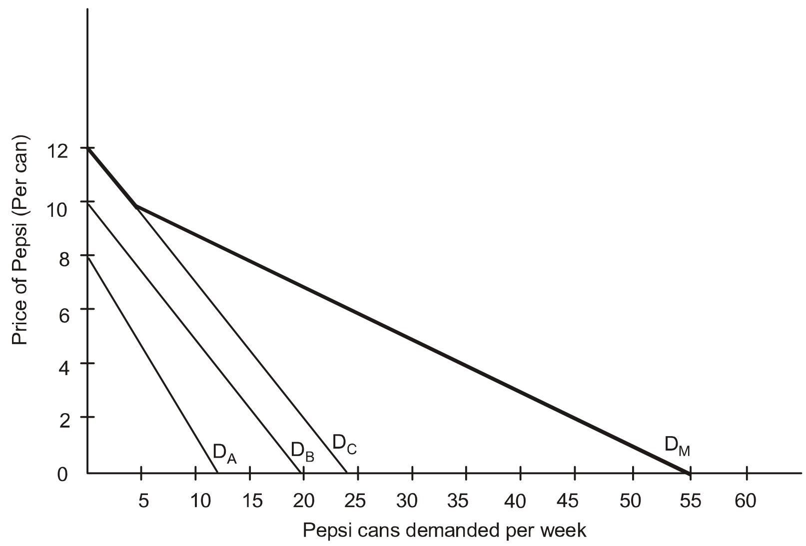


Fig. 4.2 *Derivation of Market Demand Curve*

4.2.7 Determinants of Market Demand

(i) Price of the Commodity. In the short run, price of a product is the main determinant of its market demand. How price determines the demand for a product has already been discussed in detail under the law of demand. In the long run, however, market demand for a product is determined by a number of factors other than price. We discuss here, some important quantifiable and non-quantifiable determinants of demand for a product.

(ii) Price of Substitutes and Complementary Goods. The demand for a commodity depends also on the prices of its substitutes and complementary goods. In general sense of the term, two goods are substitutes for one another if consumer use them for the same purpose and derive more or less the same utility. From economic analysis point of view, two commodities are deemed to be *substitutes* for one another, if change in price of one affects the demand for the other in the same direction. For instance, commodities *X* and *Y* are, in economic sense, substitutes for one another if a rise in the price of *X* increases the demand for *Y*, and *vice versa*. Tea and coffee, hamburger and hot-dog, wheat and rice, alcohol and drugs are some common examples of common substitutes. By definition, the relation between demand for a product and price of its substitute is of **positive** nature. When price of a product (say, tea) falls (or increases), then demand for its substitute (coffee) falls (or increases). The relationship of this nature is given in Fig. 4.3 (a).

NOTES

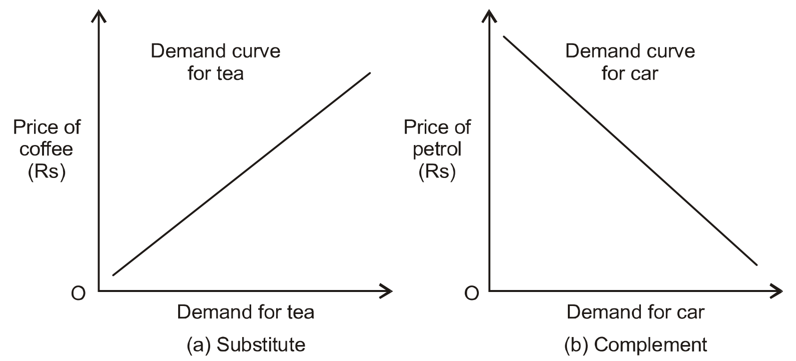


Fig. 4.3 *Demand for Substitute and Complement*

A commodity is deemed to be a *complement* of another when it complements the use of the other. For example, petrol is a complement to motor vehicles; butter and jam are complements to bread; milk and sugar are complement to tea and coffee and so on. In economic sense, two goods are treated as complements for one another, if an increase in the price of one causes a decrease in the demand for another. By definition, *there is an inverse relationship between the demand for a good and the price of its complement.* For instance, an increase (or a decrease) in the price of petrol causes a decrease (or an increase) in the demand for car, other things remaining the same. The nature of relationship between the demand for a product and the price of its complement is given in Fig. 4.3 (b).

(iii) Consumers' Income. Consumer's income is the basic determinant of the quantity demanded of a product. It is a common knowledge that the people with higher disposable income spend a larger amount on the normal goods and services than those with lower income. Income-demand relationship is of a more varied nature than that between demand and its other determinants.

For the purpose of income-demand analysis, goods and services may be grouped under four broad categories, *viz*. (*a*) essential consumer goods; (*b*) inferior goods; (*c*) normal goods; and (*d*) prestige and luxury goods. The relationship between income and the different kinds of goods is presented through the **Engel Curves**.

(a) Essential consumer goods (ECG). The goods and services which fall in this category are essentially consumed by almost all persons of a society, e.g., food grains, clothes, vegetable oils, sugar, matches, cooking fuel, and housing, etc. The quantity demanded of such goods increases with increase in consumer's income but only upto a certain limit, other factors remaining the same. The relationship between demand for goods and services of this category and consumer's income is shown by curve *ECG* in Fig. 4.4. As the curve shows, consumer's demand for essential goods increases until his income rises to OY_2 and beyond this level of income, it does not.

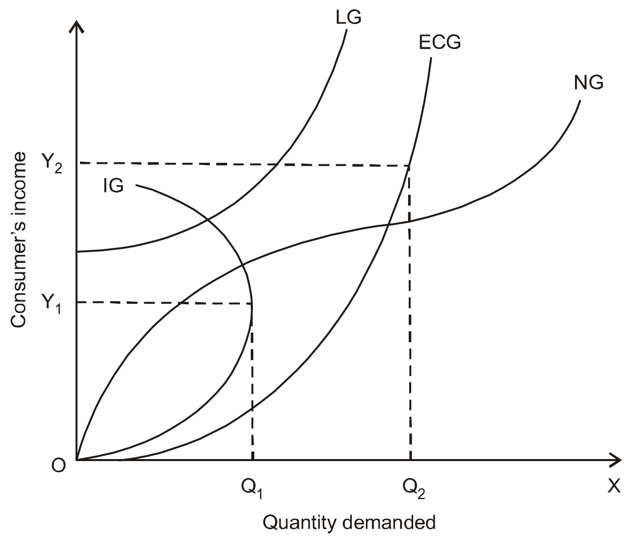


Fig. 4.4 *Income-Demand Curves*

NOTES

(*b*) *Inferior goods.* Inferior and superior goods are generally known to both consumers and sellers. For instance, every consumer knows that *bajra* is inferior to wheat and rice; *bidi* (an indigenous cigarette) is inferior to cigarette, coarse textiles are inferior to refined ones, kerosene stove is inferior to gas-stove; travelling by bus is inferior to travelling by taxi, and so on. In economic terminology, however, *a commodity* is *deemed to be inferior if its demand decreases with the increase in consumers'* income. The relation between income and demand for an inferior good is shown by curve *IG* in Fig. 4.4 assuming that other determinants of demand remain the same. Demand for such goods may initially increase with increase in income (say up to Y_1) but it decreases when income increases beyond this level.

(*c*) *Normal goods.* In economic sense, normal goods are the goods which are demanded in increasing quantities as consumers' income rises. Clothing is the most important example of this category of goods. Household furniture, electricity, telephones, household gadgets, etc. are other examples of normal goods. The nature of relationship between income and demand for normal goods is shown by the curve *NG* in Fig. 4.4. As the curve shows, demand for such goods increases with increase in income of the consumer, but at different rates at different levels of income. Demand for normal goods initially increases rapidly with the increase in income and later, at a lower rate.

It may be noted from Fig. 4.4 that upto a certain level of income (Y_1) the relation between income and demand for all types of goods is similar. The difference is of degree only. The relation between income and different kind of goods becomes distinctly different only beyond a certain level of income.

NOTES

(*d*) *Prestige and luxury goods.* Prestige goods are the goods which are consumed mostly by the rich section of the society, e.g., precious stones, diamond studded jewellery, costly cosmetics, luxury cars, airconditioners, costly decoration items (e.g., antiques), etc. Demand for such goods arises only beyond a certain level of consumer's income. The income-demand relationship of this category of goods is shown by the curve LG in Figure 4.4.

(iv) Consumer's Taste and Preference. Consumer's taste and preferences play an important role in determining the demand for a product. Taste and preferences depend, generally, on the social customs, religious values attached to a commodity, (e.g., alcohol, non-vegfood items, etc.), habits of the people, the general life-style of the social group and also the age and sex of the consumers. Change in these factors changes consumers' taste and preferences. When there is a change in consumers' liking, tastes and preferences for certain goods and services following the change in fashion, people switch their consumption pattern from cheaper and old fashioned goods over to costlier 'mod' goods, so long as price differentials commensurate with their preference. For example, preference for 'junk food' in the younger generation has increased as compared to normal home-made nutritious food. Consumers are prepared to pay higher prices for 'mod' goods even if their virtual utility is the same or even lower than of old-fashioned goods. This fact reveals that tastes and preferences also influence demand for goods and services.

(v) Consumers' Expectations. Consumers' expectations regarding the future course of changes in prices, income, and supply position of goods, play an important role in determining the demand for goods and services in the short run. If consumers expect a rise in the price of a commodity, they would buy more of it at its current price, with a view to avoiding the pinch of price-rise in future. On the contrary, if consumers expect prices of certain goods to fall, they postpone their purchases of such goods with a view to taking advantage of lower prices in future, mainly in case of non-essential goods. This behaviour of consumers reduces (or increases) the current demand for the goods whose prices are expected to decrease (or increase) in future. Similarly, an expected increase in income on account of the announcement of revision of pay-scales, dearness allowance, bonus, etc., induces increase in curent purchase, and *vice versa.*

(vi) Demonstration Effect. When new commodities or new models of existing ones appear in the market, rich people buy them first. Some people buy new goods or new model of goods because they have genuine need for them while others buy because they want to exhibit their affluence. But once new commodities come in vogue, many households buy them, not because they have a genuine need for them but because others or neighbours

NOTES

have bought these goods. The purchase by the latter category of buyers are made out of such feelings as jealousy, competition, equality in the peer group, social inferiority and the desire to raise social status. Purchases made on account of these factors are the result of ***'Demonstration Effect'*** or the '***Bandwagon Effect***'. These effects have a positive effect on the demand. On the contrary, when a commodity becomes the thing of common use, some people, mostly rich, decrease or give up the consumptions of such goods. This is known as '***Snob Effect***'. It has a negative effect on the demand for the related goods.

(vii) Consumer-credit Facility. Availability of credit to the consumers from the sellers, banks, relations and friends or from any other source encourages the consumers to buy more than what they would buy in the absence of credit facility. That is why the consumers who can borrow more consume more than those who can borrow less or cannot borrow at all. Credit facility affects mostly the demand for consumer durables, particularly those which require bulk payment at the time of purchase. For example, demand for cars and residential flats have had an unprecedented increase in demand in India due mainly to availability of bank loans.

(viii) Population of the Country. The total domestic demand for a product depends also on the size of population. Given the price, per capita income, taste and preferences etc., the larger the population, the larger the demand for a product of common use. With an increase (or decrease) in the size of population, employment percentage remaining the same, demand for the product increases (or decreases). The relation between market demand for a product (normal) and the size of population can be understood well by the fact that in 2008 US President, George Bush, blamed increasing demand for food in India and China for world food crises of 2008 because of their large population and economic growth to the income-demand relationship.

(ix) Distribution of National Income. The distribution pattern of national income also affects the demand for a commodity. If national income is evenly distributed, market demand for ***normal goods*** will be the largest. If national income is unevenly distributed, i.e., if majority of population belongs to the lower income groups, market demand for essential goods will be the largest whereas the same for other kinds of goods will be relatively low.

4.3 ELASTICITY OF DEMAND

Market demand for a product depends on its determinants and it changes with changes in its determinants. The change in demand in response to change its determinants is called *elasticity of demand.* If the demand for a product changes with change in any or all of its determinants, then it is called elastic demand. For example, if demand for computers changes due to change in computer prices, demand for computer is said to be price elastic. If demand does not respond to change in any or many of its determinants, it is called *inelastic demand. The elasticity of demand is defined as the sensitiveness of demand to the change in its determinants.* Specifically, *elasticity of demand can be defined as the percentage change in quantity demanded of a product due to 1 per cent change in its determinant.*

Given the definition of demand elasticity, we discuss now the measurement of various kinds of demand elasticities. Let us discuss first the method of measuring price elasticity of demand.

4.3.1 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*.

1. Measuring Price Elasticity of Demand

A simple formula for measuring 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 the definition of elasticity, is given as follows:

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

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

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

The price elasticity of demand (e_p) so measured is called ***elasticity coefficient***.

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

2. Arc and Point Elasticity

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

NOTES

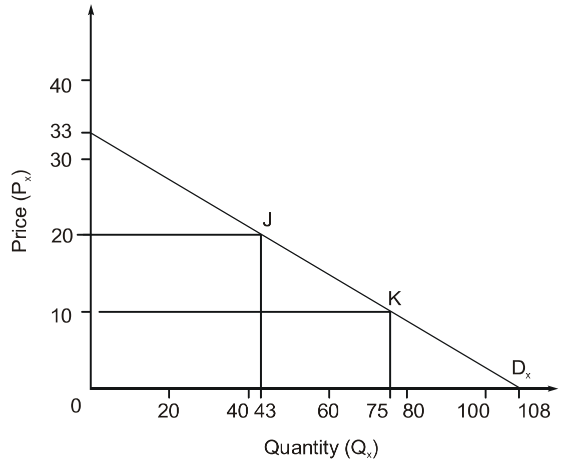


Fig. 4.5 *Linear Demand Curve*

Arc Elasticity

The measure of elasticity of demand between any two finite points on a demand curve is known as *arc elasticity*. The measure of arc elasticity is used when change in price is fairly large. For example, measure of elasticity between points *J* and *K* (Fig. 4.5) 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:

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

$$= -\frac{-32}{10} \cdot \frac{20}{43} = 1.49 \qquad \text{...(4.2)}$$

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. It means that elasticity co-efficient is 1.49.

Problem in using arc elasticity: The arc elasticity should be measured, interpreted 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 revsersed. 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 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

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

NOTES

The measure of elasticity co-efficient (0.43) in Eq. (4.3) for the reverse movement in price is significantly different from (1.49) 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

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

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

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

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

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

or

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

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 has its own drawbacks as the elasticity co-efficient calculated through this formula, refers to the elasticity mid-way between P_1 and P_2 and between Q_1 and Q_2. The elasticity co-efficient (0.81) is not applicable to the whole range of price-quantity combinations at different points between *J* and *K* on the demand curve (Fig. 4.5)—it only gives a mean of the elasticities between the two points.

None of the two modifications provide an appropriate solution to the problem arising out of change in the direction of the price-change. Therefore, it is always better to raise the original formula keeping in mind the direction of change in price – whether the price rises or falls.

Point Elasticity

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. Given the nature of the demand function, one may be required to measure point elasticity on a linear demand curve and on a non-linear demand curve. Let us first discuss the method of measuring point elasticity of demand on a linear demand curve.

Point Elasticity on a Linear Demand Curve. *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* (Fig. 4.6). This is in contrast to the arc elasticity between points *P* and *B*. A movement form 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.

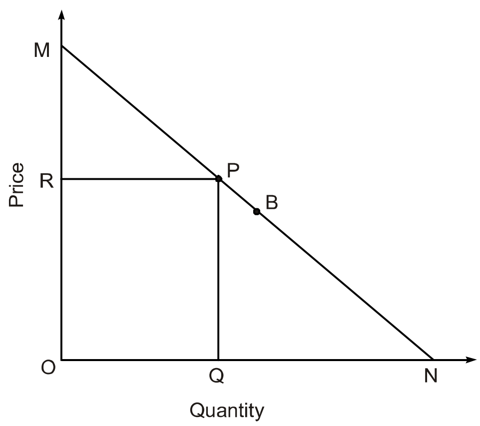


Fig. 4.6 *Point Elasticity*

$$\text{Point elasticity } (e_p) = \frac{\partial Q}{\partial P} \cdot \frac{P}{Q} \qquad ...(4.6)$$

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

The derivative $\frac{\partial Q}{\partial P}$ 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 $\frac{QN}{PQ}$.

Therefore,

NOTES

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

Note that at point P, price $P = PQ$ and $Q = OQ$. By substituting these values in Eq. (4.6), 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. 4.5. 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., 1.49, 0.43, 0.74, and 0.81).

As we will see below, geometrically, $QN/OQ = PN/PM$. Therefore, elasticity of demand at point P (Fig. 4.6) may be expressed as

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

Proof. The fact that $e_p = PN/PM$ can be proved as follows. Note that in Fig. 4.6, there are three triangles—ΔMON, ΔMRP and ΔPQN—and $\angle MON$, $\angle MRP$ and $\angle PQN$ are right angles. Therefore, the other corresponding angles of the three triangles will always be equal and hence, ΔMON, ΔMRP and ΔPQN are similar.

According to geometrical properties of similar triangles, the ratio of any two sides of a triangles are always equal to the ratio of the corresponding sides of the other triangles. By this rule, between ΔPQN and ΔMRP,

$$\frac{QN}{PN} = \frac{RP}{PM}$$

Since $RP = OQ$, by substituting OQ for RP in the above equation, we get

$$\frac{QN}{PN} = \frac{OQ}{PM}$$

It follows that $$\frac{QN}{OQ} = \frac{PN}{PM}$$

It means that price elasticity of demand at point P (Fig. 4.6) is given by

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

It may thus be concluded that the price elasticity of demand at any point on a linear demand curve is equal to the ratio of the lower segment to the upper segments of the demand line, i.e.,

$$e_p = \frac{\text{Lower segment}}{\text{Upper segment}}$$

Point Elasticity on a Non-linear Demand Curve. The ratio $\Delta D/\Delta P$, i.e., the slope of a non-linear demand curve is different at each point. Therefore, the method used to measure point elasticity on a linear demand curve cannot be applied straightaway to measure point elasticity at a non-linear demand curve.

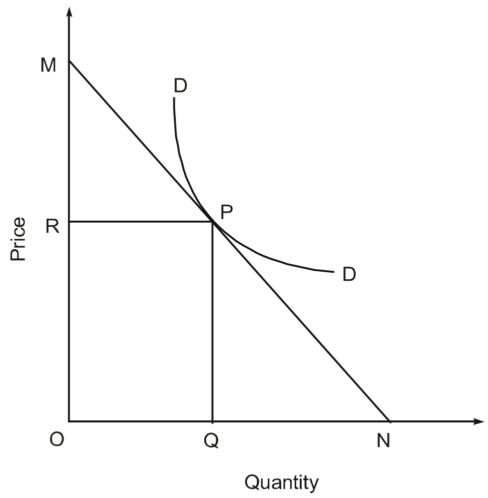


Fig. 4.7 *Price and Demand*

A simple technique is used to measure point elasticity on a non-linear demand curve. The technique is to draw a tangent through the chosen point. For example, suppose we want to measure elasticity on a non-linear demand curve, *DD* (Fig. 4.7) at point *P*. For this purpose, a tangent *MN* is drawn through point *P*. Since demand curve *DD* and the line *MN* pass through the same point (*P*), the slope of the *DD* curve and of the demand line *MN* at point *P* is the same. Therefore, the elasticity of demand curve at point *P* will be equal to that of the demand line at this point. Elasticity of the demand line at point *P* can be measured as

$$e_p = \frac{P}{Q} \cdot \frac{\partial P}{\partial P} = \frac{PQ}{OQ} \cdot \frac{QN}{PQ} = \frac{QN}{OQ}$$

As proved above, geometrically, $\frac{QN}{OQ} = \frac{PN}{PM}$. Thus, the price elasticity of demand curve *DD* at point *P* is given by *PN/PM*. The same process is adopted to measure price elasticity on any point on the demand curve.

The Range of Point Elasticity on a Linear Demand Curve. Given the measure of point elasticity, at midpoint of a linear demand curve, $e_p = 1$, as shown at point *P* in Fig. 4.8. It follows that at any point above the mid-point *P*, $e_p > 1$, and at any point below the mid-point P, $e_p < 1$. According to this formula, at the extreme point N, $e_p = 0$, and at extreme point M, e_p is undefined because division by zero is undefined. It must be noted here that these results are relevant between points M and N and that the elasticities at the extreme points M and N are, in effect, undefined.

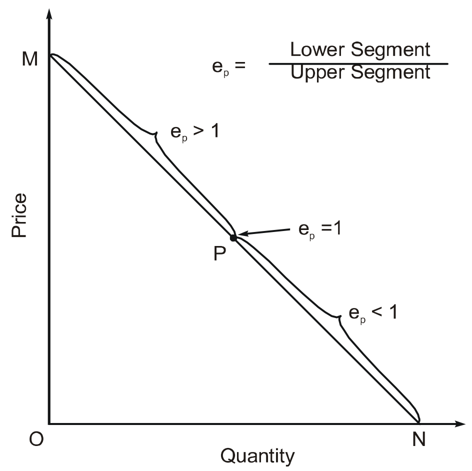


Fig. 4.8 *Point Elasticities of Demand*

4.3.2 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 describe the method of measuring price elasticity of demand for a product from the demand function—both linear and nonlinear. 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-demand combination. 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, which can be measured by differentiating the demand function. That is,

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

Note that $\delta Q/\delta P = -5$, is the same value as given in the demand function. 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 $P = 8$, $Q = 100 - 5(8) = 60$ and

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

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

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

Price Elasticity from a Non-linear Demand Function. Suppose a nonlinear demand function of multiplicative form is given as follows.

$$Q = aP^{-b} \qquad \ldots (4.7)$$

For nonlinear demand function aslo, the formula for computing the price elasticity is the same, i.e.,

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

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,

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

By substitution, e_p can be expressed as

$$e_p = -\ baP^{-b-1}\frac{P}{Q}$$

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

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

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

Eq. (4.10) shows that when a nonlienar 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 regardless of a change in price.

4.3.3 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

NOTES

measuring cross-elasticity of demand for tea ($e_{t,c}$) and the same for coffee ($e_{c,t}$) is given below.

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

$$= \frac{P_c}{Q_t} \cdot \frac{\Delta Q_t}{\Delta P_c} \qquad \ldots(4.11)$$

and

$$e_{c,\,t} = \frac{P_t}{Q_c} \cdot \frac{\Delta Q_c}{\Delta P_t} \qquad \ldots(4.12)$$

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 automobile, 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 its use in defining substitute goods. If cross-elasticity between two goods is positive, the two goods are considered to be 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 goods may be considered as complements: the higher the negative cross-elasticity, the higher the degree of complementarity.

The concept of cross-elasticity is of vital importance in pricing decision in respect 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 price of substitutes and complements.

4.3.4 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 and income-elasticity coefficient bears a positive sign. The demand for 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.*

The measure of income-elasticity of demand for a product, say X, (i.e., e_y) is given below in Eq. (4.13).

NOTES

$$e_y = \frac{\dfrac{\Delta X_q}{X_q}}{\dfrac{\Delta Y}{Y}} = \frac{Y}{X_q} \cdot \frac{\Delta X_q}{\Delta Y} \qquad \ldots(4.13)$$

(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.

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 and *vise-versa*. The reason is that when income increases, consumers switch over to the consumption of superior substitutes. For instance, when income rises, people prefer to buy more of rice and wheat and less of inferior foodgrains; buy more of meat and less of potato, and 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 4.3.

Table 4.3 *Income-Elasticities*

Consumer goods	*Co-efficient of income-elasticity*	*Effect on Sale*
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

NOTES

The income-elasticity of demand for different categories of goods may, however, vary from consumer to consumer 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 income increases frequently, income-elasticities will conform to the general pattern.

Uses of Income-Elasticity in Business Decisions. While price and cross elasticities are of greater significance in the pricing of a product aimed at maximizing the total revenue in the short period, 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, or are not in a position to, 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'.

4.3.5 Advertisement Elasticity of Sales

The expenditure on advertisement and on other sales-promotion activities does help in promoting sales, but *not in the same degree at all levels of the total sales*. The concept of advertisement elasticity is useful in determining the optimum level of advertisement expenditure. The concept of *advertisement elasticity* assumes a greater significance in deciding on advertisement expenditure, particularly when the government imposes restriction on advertisement cost or there is competitive advertising by the rival firms. Advertisement elasticity (e_A) of sales is measured as follows:

$$e_A = \frac{\Delta S/S}{\Delta A/A} = \frac{\Delta S}{\Delta A} \cdot \frac{A}{S}$$

NOTES

where S = sales; ΔS = increase in sales; A = initial advertisement cost, and ΔA = additional expenditure on advertisement.

Interpretation of Advertisement-elasticity. The advertisement elasticity of sales varies between $e_A = 0$ and $e_A = \propto$. Interpretation of some measures of advertising elasticity is given below.

Elasticities	*Interpretation*
$e_A = 0$	Sales do not respond to the advertisement expenditure.
$e_A > 0$ but < 1	Increase in total sales is less than proportionate to the increase in advertisement expenditure.
$e_A = 1$	Sales increase in proportion to the increase in expenditure on advertisement.
$e_A >$	Sales increase at a higher rate than the rate of increase of advertisement expenditure.

Determinants of Advertisement Elasticity. Some of the important factors which determine advertisement elasticity are the following:

(*i*) ***The level of total sales.*** In the initial stages of sale of a product, particularly of one which is newly introduced in the market, the advertisement elasticity is greater than unity. As sales increase, the elasticity decreases. For instance, after the potential market is supplied, the function of advertisement is to create additional demand by attracting more consumers to the product, particularly those who are slow in adjusting their consumption expenditure to provide for new commodities. Therefore, demand increases at a rate lower than the rate of increase in advertisement expenditure.

(*ii*) ***Advertisement by rival firms.*** In a highly competitive market, the effectiveness of advertisement by a firm is also determined by the money spent on and effectiveness of advertisement by the rival firms.

(*iii*) ***Cumulative effect of past advertisement.*** In case expenditure incurred on advertisement in the initial stages is not adequate enough to be effective, elasticity may be very low. But over time, additional doses of advertisement expenditure may have a cumulative effect on the promotion of sales and advertising elasticity may increase considerably.

(*iv*) ***Other factors.*** Advertisement elasticity is also affected by other factors affecting the demand for a product, e.g., change in products' price, consumers' income, growth of substitutes and their prices.

4.3.6 Price Expectation-Elasticity of Demand

NOTES

Sometimes, mainly during the period of price fluctuations, consumer's price expectations play a much more important role in determining demand for a commodity than any other factor. The concept of price-expectation-elasticity was devised and popularised by J.R. Hicks in 1939. The price-expectation-elasticity refers to the expected change in future price as a result of change in current prices of a product. The elasticity of price-expectation is defined and measured by the formula given below. As given in Eq. (4.14)

$$e_x = \frac{\Delta P_f / P_f}{\Delta P_c / P_c} = \frac{\Delta P_f}{\Delta P_c} \cdot \frac{P_c}{P_f} \qquad \ldots(4.14)$$

where P_c and P_f are current and future prices, respectively.

The coefficient e_x gives the measure of expected percentage change in future price as a result of 1 per cent change in present price. If $e_x > 1$, it indicates that future change in price will be greater than the present change in price, and *vice versa*. If $e_x = 1$, it indicates that the future change in price will be equal to the change in the current price.

The concept of elasticity of price-expectation is very useful in formulating future pricing policy. For example, if $e_x > 1$, it indicates that sellers will be able to sell more in the future at higher prices. Thus, businessmen may accordingly determine their future pricing policy.

4.4 INDIFFERENCE CURVE ANALYSIS: CONSUMER'S SURPLUS

Unlike Marshall, the modern economists—Hicks in particular—have used the ordinal utility concept to analyse consumer's behaviour. This is called 'ordinal utility approach'. Hicks has used a different tool of analysis called 'indifference curve' or equal utility curve to analyse consumer behaviour. In this section, we will first explain the 'indifference curve' and then explain consumer's behaviour through the indifference curve technique.

4.4.1 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 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. It gives him an opportunity to substitute one commodity for another, if need arises and to make various combinations of two substitutable goods which give him the same level of satisfaction. If a consumer is faced with such combinations, he would be indifferent between the combinations. When such combinations are plotted graphically, it produces a curve called *indifference curve*. An indifference curve is also called *Isoutility Curve* or *Equal Utility Curve*.

For example, let us suppose that a consumer makes five combinations *a*, *b*, *c*, *d* and *e* of two substitute commodities, *X* and *Y*, as presented in Table 4.4. All these combinations yield the same level of satisfaction indicated by *U*.

Table 4.4 *Indifference Schedule of Commodities X and Y*

Combination		*Units of Commodity Y*	+	*Units of Commodity X*	=	*Total Utility*
a	=	25	+	3	=	*U*
b	=	15	+	6	=	*U*
c	=	8	+	10	=	*U*
d	=	4	+	17	=	*U*
e	=	2	+	30	=	*U*

Table 4.4. 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 4.4 are plotted and joined by a smooth curve (as shown in Fig. 4.9). The resulting curve is known as an *indifference curve*. On this curve, one can locate many other 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.

Indifference Map. We have drawn a single indifference curve in Fig. 4.9 on the basis of the indifference schedule given in Table 4.4. 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. 4.9. As such, an indifference curve below the one given in Fig. 4.9 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 satisfac-tion but greater than the satisfaction indicated by *IC*. Thus, another indiffe-rence curve can be drawn above *IC*, say, through points *j*, *k* and *l* as shown in Fig. 4.9. This exercise may be repeated as many times as one wants, each time generating a new indifference curve.

NOTES

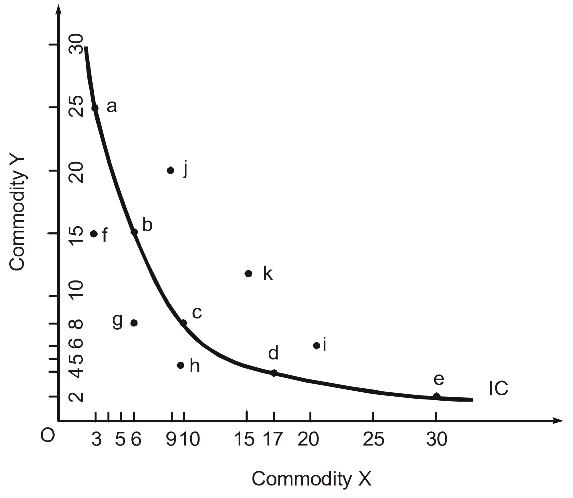


Fig. 4.9 *Indifference Curve*

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 indicating different combinations of goods X and Y yielding the same satisfaction. It is thus possible to draw a number of indifference curves without intersecting or touching the other, as shown in Fig. 4.10. The set of indifference curves IC_1, IC_2, IC_3 and IC_4 drawn in this manner make the *indifference map*. It is important to note here that utility represented by each upper IC is higher than that on the lower ones. For example, the utility represented by IC_2 is greater than utility represented by IC_1. In terms of utility, $IC_1 < IC_2 < IC_3 < IC_4$.

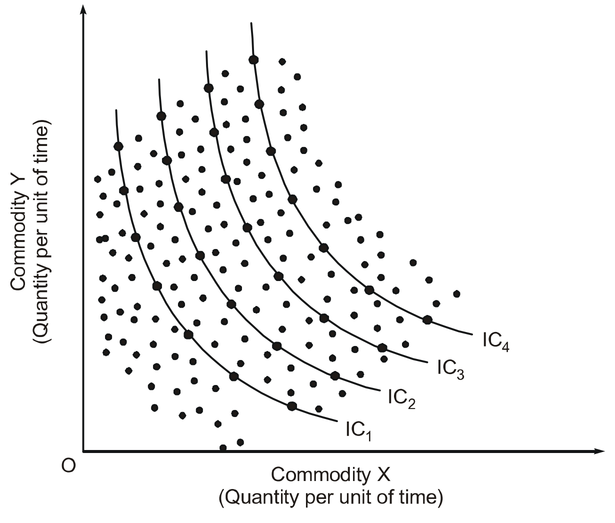


Fig. 4.10 *The Indifference Map*

4.4.2 Marginal Rate of Substitution (MRS)

NOTES

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*). The *MRS* is expressed as $\Delta Y/\Delta X$, moving down the curve.

The Diminishing *MRS*. 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 different combinations of *X* and *Y* given in Table 4.4 are given in Table 4.5.

Table 4.5 *The Diminishing MRS between Commodities X and Y*

Indifference Points	*Combinations* $Y + X$	*Change in Y* $(-\Delta Y)$	*Change in X* (ΔX)	$MRS_{y,x}$ $(\Delta Y/\Delta X)$
a	25 + 3	–	–	–
b	15 + 6	– 10	3	– 3.33
c	8 + 10	– 7	4	– 1.75
d	4 + 17	– 4	7	– 0.60
e	2 + 30	– 2	13	– 0.15

As Table 4.5 shows, when the consumer moves from point *a* to *b* on his indifference curve (Fig. 4.9) he gives up 10 units of commodity *Y* and gets only 3 units of commodity *X*, so that

$$MRS_{y,\,x} = \frac{-\Delta Y}{\Delta X} = \frac{-10}{3} = -3.33$$

As he moves down from point *b* to *c*, he gives up 7 units of *Y* for 4 units of *X*, giving

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

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

Why Does the *MRS* Diminish?

(i) Diminishing subjective marginal utility. The *MRS* decreases along the *IC* curve because, in most cases, no two goods are perfect substitutes for

NOTES

one another. In case any two goods are perfect substitutes, the indifference curve will be a straight line with a negative slope and constant *MRS*. Since most 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. That is the reason why *MRS* decreases.

(ii) Decreasing ability to sacrifice a good. When combination of two goods at a point on 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 4.4), the quantity of *Y* (25 units) is much larger than that of *X* (3 units). That is why the consumer is willing to sacrifice 10 units of *Y* for 3 unit of *X*. 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. Besides, as mentioned above, the *MRS* decreases also because of the law of the diminishing *MU*. The *MU* of a commodity available in larger quantity is lower than that of a commodity available on smaller quantity. Therefore, the consumer has to sacrifice a large quantity of *Y* for a small quantity of *X* in order to maintain total utility at the same level. These are the reasons why *MRS* between the two substitute goods decreases all along the indifference curve.

4.4.3 Properties of Indifference Curves

Indifference curves drawn for two normal substitute goods have the following four basic properties:

1. Indifference curves have a negative slope;
2. Indifference curves 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 Figure 4.11.

NOTES

Figure 4.11 shows two *IC* curves:

(*i*) a curvilinear *IC*, and

(*ii*) a straight line *IC* as shown by the line *PS*.

The curvilinear *IC* represents *IC* for two imperfect susbtitute goods whereas straight line *PS* represents *IC* for two perfect substitute goods. In both the cases, the *IC* has a downward or a negative slope. 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.

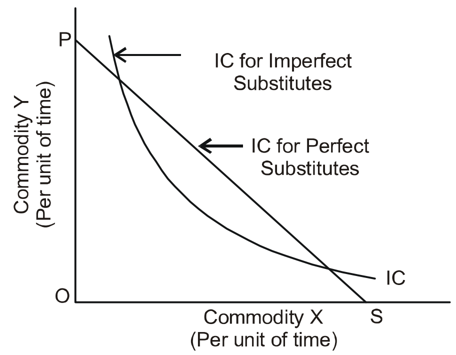


Fig. 4.11 *Normal Indifference Curves*

2. **Indifference Curves are Convex to Origin.** Indifference curves are not only negatively sloped, but are also *convex to the origin*. 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.

   The *MRS* decreases because of 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 reflects two rather 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

NOTES

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.

Let us now see what happens when two indifference curves, IC_1 and IC_2, intersect each other at point *A* (Fig. 4.12). Point *A* falls on both the indifference curves, IC_1 and IC_2. 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.

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_2 and point *C* on indifference curve IC_1 both being on a vertical line.

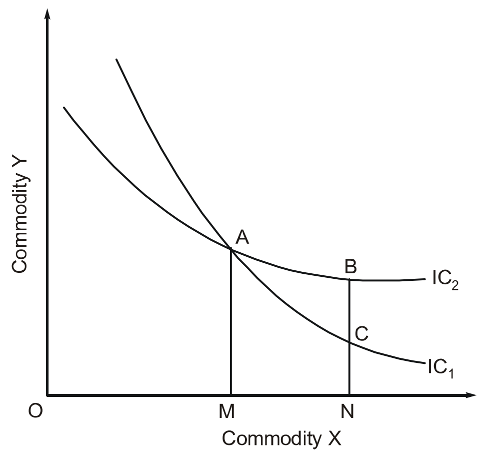


Fig. 4.12 *Intersecting indifference Curves*

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

$$A = B$$

and $$A = C$$

$\therefore$ $$B = C$$

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

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

Since '*ON* of *X*' is common to both the sides, it would mean that

$$BN \text{ of } Y = CN \text{ of } Y$$

But as Fig. 4.12 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.

NOTES

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. 4.13, indifference curve IC_2 is placed above the curve IC_1. It represents, therefore, 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 $MU > 0$.

For example, consider the indifference curves IC_1 and IC_2 in Fig. 4.13. Let us begin at point *a*. 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).

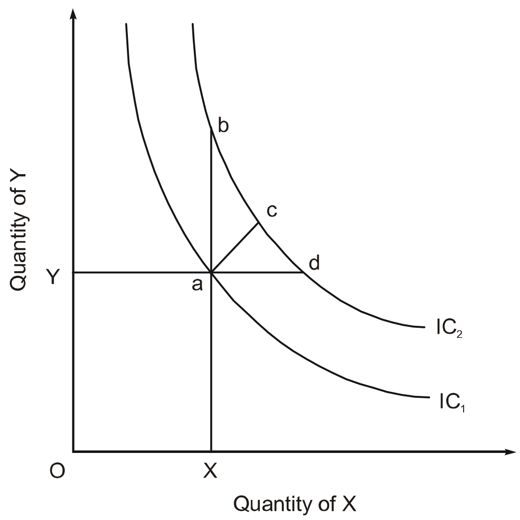


Fig. 4.13 *Comparison between Lower and Upper Indifference Curves*

4.4.4 Consumer Equilibrium

In this section onwards, we take up the main theme of the theory of consumer behaviour as developed under the ordinal utility approach. The main issue is how a consumer attains his equilibrium. 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:

NOTES

(*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*.

Consumer's equilibrium is illustrated in Fig. 4.14. 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, MU_y of $\Delta Y = MU_x$ of ΔX and, therefore, the slope of an indifference curve is given by

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

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

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

As shown in Fig. 4.14, at point *E*, $MRS_{y,\,x} = P_y / P_x$. Therefore, the consumer is in equilibrium at point *E*. The tangency of IC_2 with the budget line *AB*, indicates that IC_2 is the highest possible indifference curve which the consumer can reach, given his budgetary constraint and the prices. At equilibrium point *E*, the consumer consumes OQ_x of *X* and OQ_y of *Y*, which yield him the maximum satisfaction.

NOTES

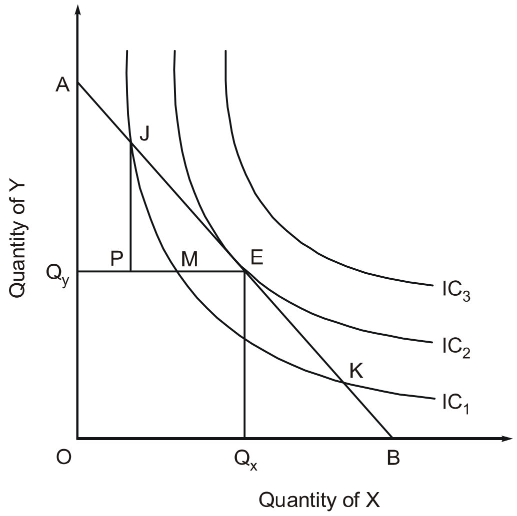


Fig. 4.14 *Equilibrium of the Consumer*

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 a lower 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, at any point on IC_1, the level of satisfaction is lower than the level of satisfaction indicated by IC_2. So long as the utility maximizing consumer has an opportunity to reach the curve IC_2, he would not like to settle on a lower indifference curve.

From the information contained in Fig. 4.14, 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 would 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* of *X* gives him extra utility, the consumer moves to point *E*. Since point *E* falls on a higher *IC*, it represents 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* on a higher IC_2 from any other point on the curve IC_1, all other things (taste, preference and prices of goods) remaining the same.

Another fact which is obvious from Fig. 4.14 is that, due to budget constraint, the consumer cannot move to an indifference curve placed above and to the right of IC_2. For example, his income would be insufficient to buy any combination of two goods at the curve IC_3. Note that the indifference curve IC_3 falls in the infeasibility area.

4.4.5 Effects of Change in Income on Consumer Demand

NOTES

In this section, we examine the effects of changes in consumer's income on his consumption behaviour, assuming that prices of all goods and consumer's tastes and preferences remain constant.

When a consumer's income changes, his capacity to buy goods and services changes too, other things remaining the same. These changes are shown by a parallel upward or downward shift in the consumer's budget line. As shown in Fig. 4.15, when a consumer's income decreases, his budget line shifts downward and when his income increases, the budget line shifts upward. With the changes in his income, the consumer moves from one equilibrium point to another. Such movements show the rise and fall in the consumption basket. This is called, "income effect". The direction and magnitude of income effect depends on the nature of the goods. For the purpose of measuring income effect, the goods are classified as (*i*) normal goods, and (*ii*) inferior goods. We explain here the income effect on the consumption of the normal and inferior goods.

Income Effect on Normal Goods

The effect of change in income on the consumption of normal goods is that their consumption income increases. This fact is illustrated in Fig. 4.15. The indifference curves IC_1, IC_2, IC_3 and IC_4 represent the consumer's indifference map. To analyse the effect of change in income on consumption, let us suppose that the consumer has a given income and prices of goods *X* and *Y* are given. Given the consumer's income and prices of *X* and *Y*, his budget line is given by *AJ*, and the consumer is initially in equilibrium at E_1 on the IC_1. Now let the consumer's income increase so that his budget line shifts from position *AJ* to *BK* and the consumer reaches a new equilibrium point, E_2 on IC_2. Similarly, if his income increases further, he moves from equilibrium E_2 to E_3 and then to E_4. Thus, with each successive upward shift in the budget line, the equilibrium position of the consumer moves upward showing increase in consumption of both the goods *X* and *Y*. The successive equilibrium-combinations of goods (*X* and *Y*) at four different levels of income are indicated by points E_1, E_2, E_3 and E_4 in Fig. 4.15. If these points of equilibrium are joined by a curve, we get the path of increase in consumption resulting from the increase in income. This curve is called the ***income consumption curve*** **(*ICC*).** The income-consumption curve may be defined as *the locus of points representing various equilibrium quantities of two commodities consumed by a consumer at different levels of income, all other things remaining constant*. The movement from point E_1 towards point E_4 indicates increase in the consumption of the normal goods *X* and *Y*. This is called *income effect*.

NOTES

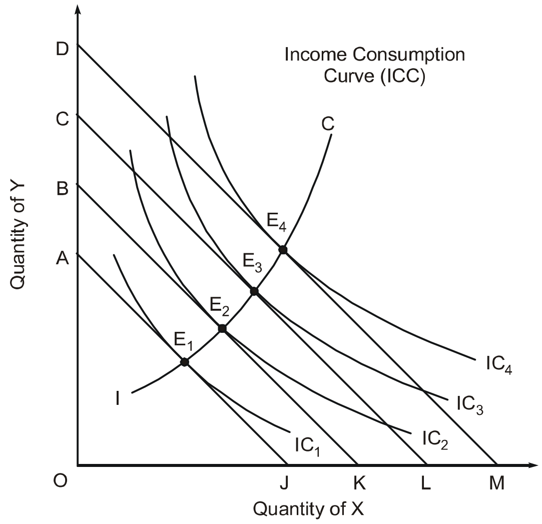


Fig. 4.15 *Income Consumption Curve for Normal Goods*

Income-Effect on Inferior Goods

As mentioned above, the income-effect on the consumption of different kinds of commodities is not uniform. It can be *positive* or *negative* or *even neutral*. Whether the income effect is positive or negative depends on the nature of a commodity. In case of normal goods, income-effect is positive and in case of inferior goods, it is negative. By definition, an *inferior good* is one whose consumption decreases when income increases. Fig. 4.16 and 4.17 present the case of *negative income effect*. In Fig. 4.16, *X* is an inferior good—its consumption decreases when consumer's income increases. The income-effect on consumption of *X* is, therefore, negative. Similarly, in Fig. 4.16, income-effect on *Y* is negative as consumption of *Y* decreases with increase in income.

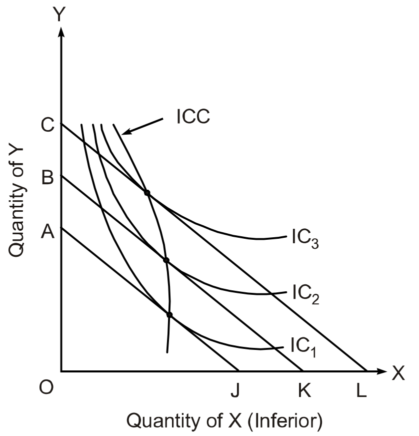


Fig. 4.16 *Income-Consumption Curve for Inferior Good X*

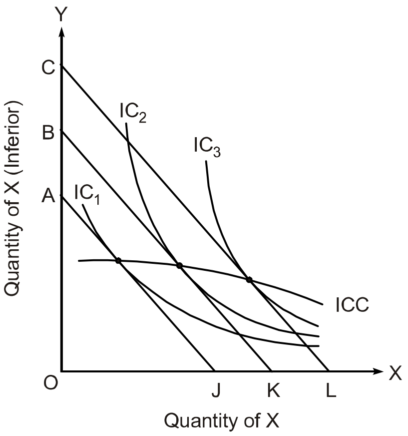


Fig. 4.17 *Income-Consumption Curve for Inferior Good Y*

NOTES

Finally, it is important to bear in mind that whether a commodity is a 'normal good' or an 'inferior good' depends on whether income-effect on its consumption is positive or negative. If income-effect is positive, the commodity is considered to be a 'normal good' and if it is negative, the commodity is said to be an 'inferior good'. Thus, the income-consumption-curve may take various shapes depending on whether a commodity is a 'normal good' or an 'inferior good'.

4.4.6 Effects of Change in Price on Consumer Demand

Let us now examine the effects of change in price on consumer demand. In two-commodity model, when price of one commodity changes, price of other commodity remaining constant, the slope of the budget line changes, which disturbs the consumer's equilibrium. A rational consumer adjusts his consumption basket with a view to maximizing his satisfaction under the new price conditions. The change in consumption basket is called *price-effect*.

Price-Effect. *The price-effect may be defined as the total change in the quantity consumed of a commodity due to a change in its price.*

Suppose that the consumer is initially in equilibrium at point E_1. Now let the price of X fall, *ceteris paribus*, so that the consumer's budget line shifts from its initial position LR to the position LS. As a result, the consumer reaches a higher indifference curve IC_2 and his new equilibrium point is E_2. Here, his consumption of X increases by UR. This is the *price-effect* on the consumption of commodity X.

As shown in Fig. 4.18, with a successive fall in the price of X, consumer's equilibrium shifts from E_2 to E_3 and from E_3 to E_4. By joining the points of equilibrium E_1, E_2, E_3 and E_4, we get a curve called *price-consumption-curve* (*PCC*). *Price-consumption-curve is a locus of points of equilibrium on indifference curves, resulting from the change in the price of a commodity*. The price-consumption-curve (*PCC*) shows the change in consumption basket due to a change in the price of commodity X. It can be seen from Fig. 4.18 that the quantity of X consumed goes on increasing whereas that of Y first decreases and then increases.

In must be noted here that the overall *price effect* of change in the price of one good includes both income and substitution effects. The initial decrease in the consumption of Y is the substitution effect and increase in its consumption at a later stage represents the income effect.

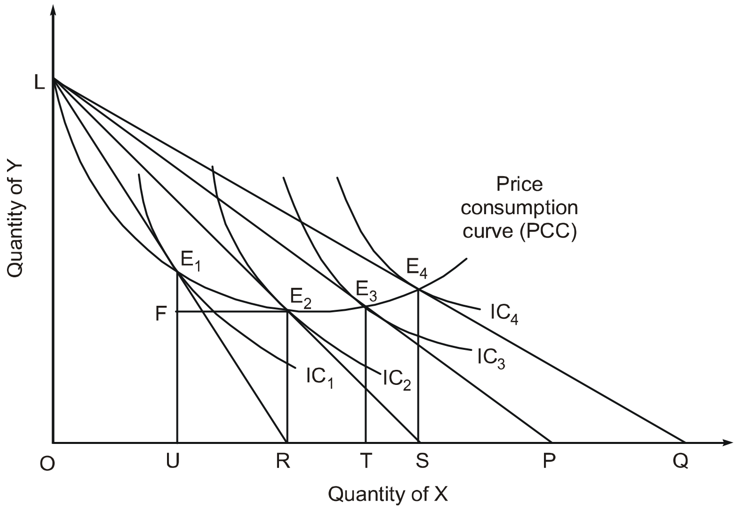


Fig. 4.18 *Price-Consumption Curve*

NOTES

Check Your Progress

1. State the law of demand.
2. Define market demand for a commodity.
3. What is arc elasticity?
4. What is cross-elasticity of demand?

4.5 ANSWERS TO 'CHECK YOUR PROGRESS' QUESTIONS

1. 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 decreased when its price increases.
2. Market demand for a commodity is the sum of all individual demands for the commodity at a given price, per unit of time.
3. The measurement of elasticity of demand between any two finite points on a demand curve is known as arc elasticity.
4. Cross-elasticity of demand is the measure of responsiveness of demand for a commodity to the changes in the price of its substitutes and complementary goods.

4.6 SUMMARY

NOTES

- Demand can be defined as the desire for a good for which one has sufficient purchasing power and willingness to pay for the good. In simple words, *demand is a desire for a good, backed by ability and willingness to pay.*
- 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.*
- A *demand schedule* is a tabular presentation of different prices of a commodity and its corresponding quantity demanded per unit of time.
- A *demand curve* is a graphical presentation of the demand schedule. A demand curve is btained by plotting a demand schedule.
- Diminishing marginal utility is also responsible for increase in demand for a commodity when its price falls.
- 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.
- A commodity is deemed to be a *complement* of another when it complements the use of the other. For example, petrol is a complement to motor vehicles; butter and jam are complements to bread; milk and sugar are complement to tea and coffee and so on.
- *The elasticity of demand is defined as the sensitiveness of demand to the change in its determinants.* Specifically, *elasticity of demand can be defined as the percentage change in quantity demanded of a product due to 1 per cent change in its determinant.*
- 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.
- 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*.

4.7 KEY WORDS

- **Individual demand:** It can be defined as the quantity of a commodity that an individual is willing to buy at a given price over a specified period of time say a month or week.
- **Demand schedule:** It is a tabular presentation of different prices of a commodity and its corresponding quantity demanded per unit of time.

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

4.8 SELF-ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. Write short notes on the following:
   (a) Demand schedule (b) Demand curve
2. What are the exceptions to the law of demand?
3. State the uses of income-elasticity in business decisions.
4. List the main properties of indifference curve.
5. Write a short note on the concept of consumer equilibrium.

Long-Answer Questions

1. What are the determinants of market demand?
2. Explain the concept of point elasticity with the help of a diagram.
3. Discuss the meaning and nature of indifference curve.

4.9 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 5 THEORY OF PRODUCTION

NOTES

Structure

5.0 INTRODUCTION

Be it profit maximization or any other objective of business firms, achieving optimum efficiency in production and minimising 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 or, what is the same thing, maximize output from a given quantity of inputs. In their effort to minimize the cost of production, the fundamental questions which managers are faced with are:

(*i*) How can the use of production inputs be optimized or cost minimized?

(*ii*) How does output behave when quantity of inputs is increased?

(*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 production scale is increased or 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 solution to the real life problems. But it does provide tools and techniques to analyse the production conditions and to find solution to the practical business problems.

In this unit, you will study about the concept of production, factors of production and laws of returns.

NOTES

5.1 OBJECTIVES

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

- Define production
- Explain the kinds of laws of production
- List the factors behind the laws of production

5.2 BASIC CONCEPTS OF PRODUCTION

5.2.1 Meaning of Production

In economics, 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 general, production means transforming inputs (labour, machines, raw materials, time, etc.) into an output with value added. This concept of production is however limited to only 'manufacturing'.

In economic sense, *production process* may take a variety of forms other than manufacturing. For example, transporting a commodity 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; a coal miner does virtually nothing more than transporting coal from coal mines to the market place. Similarly, a fisherman only transports fish to the market place. Their activities too are 'production'. Transporting men and materials from one place to another is 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 or farming 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 input and output are intangible; lawyers, doctors, social workers, consultants, hair-dressers, musicians, orchestra players are all engaged in producing intangible goods.

5.2.2 Input and Output

NOTES

An *input* is a good or service that is used 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 proceses." 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 have classified inputs as (*i*) labour, (*ii*) capital, (*iii*) land, (*iv*) raw materials, (*v*) time and (*vi*) technology. While capital and land are treated as 'stock' variable, all other variables are 'flow' variables, since they are measured per unit of time.

5.2.3 Fixed and Variable Inputs

For the purpose of theoretical analysis, inputs are classified under two categories:

(*i*) *fixed inputs* or *fixed factors*, and

(*ii*) *variable inputs* or *variable factors*.

Fixed and variable inputs are defined in economic sense and in technical sense. In **economic sense**, 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 in elastic, e.g., labour and raw material, etc. All the users of such factors can employ a larger quantity in the short-run. Technically, a variable input is one that changes with the change in output. In the long run, all inputs are treated to be variable.

5.2.4 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 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 fixed time period. While in some industries short run may be a matter of few weeks or few months, in some others (e.g., automobile, ship, plane and power industries), it may mean two, three or more years.

The *long run* refers to a period of time in which the supply of all the inputs is elastic, but the period is not long enough to permit a change in technology. That is, in the long run, all the inputs are variable. Therefore, in the long-run, 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 supposed to change. In the very long run, the production function also changes. The technological advances result in a larger output from a given quantity of inputs.

NOTES

5.3 LAWS OF PRODUCTION: MEANING AND KINDS

Before we proceed to discuss the laws of production, let us have a look at the meaning and the kinds of laws of production with reference to this frame.

The laws of production state the nature of relationship between output and input. The traditional theory of production studies the marginal input-output relationships under (*i*) short run, and (*ii*) long run conditions. In the short run, input-output relations are studied with one variable input, other inputs held constant. This changes the proportion of inputs. Therefore, the laws of production under these conditions are called 'The Laws of Variable Proportions' or the 'Laws of Returns to a Variable Input'. In the long-run, input-output relations are studied assuming all the input to be variable. The long-run input-output relations are studied under the 'Laws of Returns to Scale'.

5.3.1 Short-Run Laws of Production

1. Production with one Variable Input

Some factors of production have elastic supply even during the short period. Such factors are called **variable factors.** In the short-run, therefore, the firms can employ a large quantity of the variable factor. In other words, firms can employ in the short run, varying quantities of variable inputs against a given quantity of fixed factors. This kind of change in input combination leads to *variation in factor proportions.* The laws which bring out the relationship between varying factor proportions and output are therefore known also as the *Laws of Returns to a Variable proportions.* This law is more popularly known as the *Law of Diminishing Returns.* In this section, we explain the laws of returns to variable input.

The Law of Returns to a Variable Input : The Law of Diminishing Returns

The law of diminishing returns states that *when more and more units of a variable input are applied to a given quantity of fixed inputs, the total output may initially increase at an increasing rate and then at a constant rate but it will eventually increase at diminishing rates.* In other words, when a firm using two inputs—labour and capital—increases the number of labour, capital remaining constant, the marginal productivity of labour may initially increase, but it does decrease eventually. This is called the law of diminishing returns to the variable input.

NOTES

Assumptions. The law of diminishing returns is based on the following assumptions: (*i*) the state of technology is given, (*ii*) labour is homogeneous, and (*iii*) input prices—wages and interest—are given.

To illustrate the law of diminishing returns, let us assume (*i*) that 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 more of mine-workers to increase its coal production. Thus, the short run production function for the firm will take the following form.

$$Q_c = f(L)$$

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

$$Q_c = -L^3 + 15L^2 + 10L \qquad \ldots(5.1)$$

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 (say, thousand tonnes) that can be produced with different number of workers. For example, if $L = 5$, then by substitution,

$$\begin{aligned} Q_c &= -5^3 + 15 \times 5^2 + 10 \times 5 \\ &= -125 + 375 + 50 \\ &= 300 \text{ (thousand tonnes)} \end{aligned}$$

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 5.1 (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.

Tables 5.1 *Three Stages of Production*

No. of workers (N)	*Total product (TP_L) (000 tonnes)*	*Marginal Product* (MP_L)*	*Average Product (AP_L)*	*Stages of Production*
(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.

NOTES

Marginal Productivity of Labour (MP_L) can be obtained by differentiating the production function (5.1). Thus,

$$MP_L = \frac{\partial Q}{\partial L} = -3L^2 + 30L + 10 \qquad \ldots(5.2)$$

By substituting numerical value for labour (L) in Eq. (5.2), MP_L can be obtained at different levels of labour employment. However, this method can be used only where labour is perfectly divisible and δL 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, i.e., $\Delta L = 1$, 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 5.1.

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

$$AP_L = \frac{-L^3 + 15L^2 + 10L}{L} = -L^2 + 15L + 10 \qquad \ldots(5.3)$$

Now AP_L can be obtained by substituting the numerical value for L in Eq. (5.3). AP_L obtained by this method is given in col. 4 of Table 5.1.

The information contained in Table 5.1 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.) Beyond the 6th worker, TP_L increases (until the 10th worker) but the rate of increase in TP_L (i.e., marginal addition to TP_L) begins to fall and turns negative 11th worker onwards. This shows the operation of the law of diminishing returns, i.e., as units of labour increase, capital remaining constant, the total output increases initially at an increasing rate but finally at a decreasing rate.

2. Three Stages in Production

Table 5.1 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 worker. Given the production function (Eq. 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.

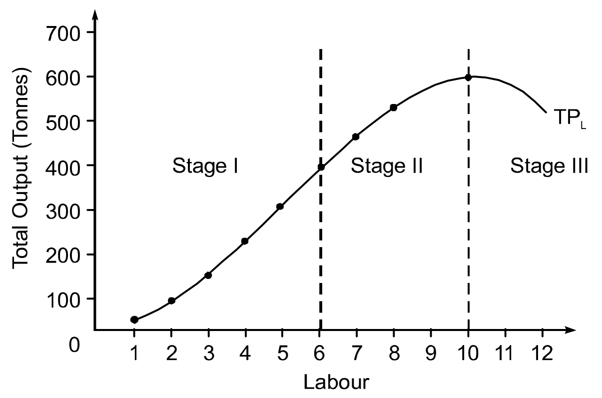
In Stage II, TP_L continues to increase but at diminishing rates, *i.e.*, MP_L begins to decline. This stage in production shows the law of diminishing returns to the variable factor. Total output reaches its maximum level at the employment of the 10th worker. Beyond this level of labour employment,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

NOTES

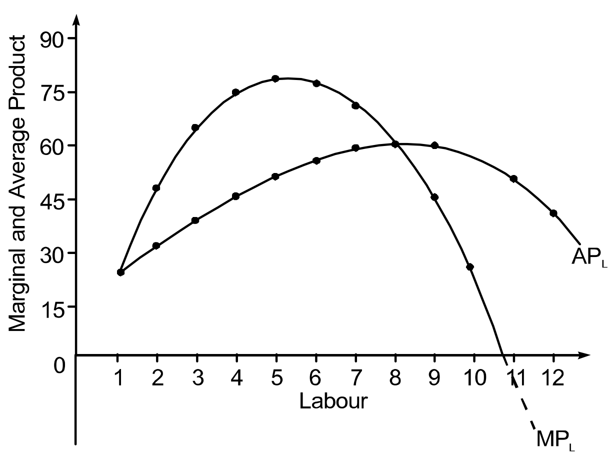
To conclude, given the employment of fixed factor (capital), when more and more workers are employed, the return from the additional worker may initially increase but will eventually decrease.

3. Factors Behind the Laws of Returns

As shown in Fig. 5.1, the marginal productivity of workers (MP_L) increases in Stage I and it decreases in Stage II. Stage I shows the Law of Increasing Returns and Stage II shows the Law of Diminishing Returns.



(a)



(b)

Fig. 5.1 *Total, Average and Marginal Products*

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). It results in under-utilisation of capital if labour is less than its optimum number. 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 underutilised. When more and more workers are added, utilization of capital increases and also the productivity of additional worker. Another reason for increase in labour productivity is that employment of additional workers leads to advantages of ***division of labour***, until optimum capital-labour combination is reached.

Once the optimum capital-labour ratio is reached, employment of additional workers amounts to substitution of capital with labour. But, technically, one factor can substitute another only upto a limited extent. Therefore, with increase in labour, capital per unit of labour decreases. This causes decrease in the productivity of the marginal labour. That is, employment of more and more labour against a given capital, causes decrease in MP_L.

Empirical Validity 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 the law is not as true as the law of gravitation. In some productive activities, it may operate quickly, in some its operation may be delayed; and in some others, it may not appear at all. This law has been found to operate in agricultural production more regularly than in industrial production. The reason is, in agriculture, natural factors play a predominant role whereas man-made factors play the major role in industrial production. Despite these variations and 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.

5.3.2 The Law of Diminishing Returns and Business Decisions

We have discussed above the law of diminishing returns in a theoretical framework. Let us now look at the applicability of this law to business decision-making.

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 provide answer to such questions as (*i*) how much labour to employ to maximise the output; and (*ii*) what number of workers to apply to a given fixed input so that per unit cost in minimized when output is maximumized. Fig 5.1 exhibits the three stages of production. In Stage III, has a very high labour-capital ratio. As a result, employment of additional workers proves not only unproductive but also causes a decline in the *TP*. Similarly, in Stage I, capital is presumably underutilized. So a firm operating in Stage I is required to increase labour, and a firm operating in Stage III is required to reduce labour, with a view to maximising its total production. From the firm's point of view, setting an output target in Stages I and III 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 questions 'how many workers to employ'.

Figure 5.1 shows also that the firm should employ a minimum of 6 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 6 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.

NOTES

5.3.3 Long-Term Laws of Production–I: Tools of analysis

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. This is a long-run phenomenon. 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 and 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 scales***. 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 have, therefore, first introduced and elaborated on this tool of analysis—the isoquant curve.

Production Isoquant: A Tool of Production Analysis

The Isoquant Curve. *An isoquant curve is 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*) while an indifference curve measures 'utility', an isoquant measures 'output'.

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*) the two inputs—L and K—can be substituted one for another but at a diminishing rate; and

(*iii*) the technology of production is given and labour and capital can be substituted only to a certain extent.

Given these assumptions, it is always possible to produce a given quantity of commodity X with various combinations of capital and labour. The input 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. 5.2. 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 5.2, 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., $IQ_1 = 100$.

NOTES

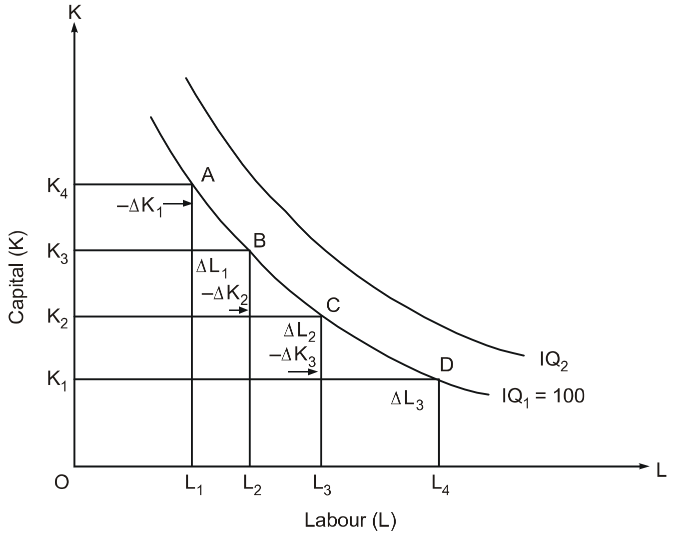


Fig. 5.2 *Isoquant Curves*

Table 5.2 *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

5.3.4 Long-Term Laws of Production–II: Laws of Returns to Scale

The laws of returns to scale state 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 of returns to scale.

NOTES

1. Increasing Returns to Scale

When a certain proportionate increase in both the inputs, *K* and *L*, leads to a more than proportionate increase in output, it exhibits *increasing returns to scale.* For example, if both the inputs, *K* and *L*, are successively doubled and the corresponding output is more than doubled, the returns to scale is said to be increasing. The increasing returns to scale is illustrated in Fig. 5.3. The movement from point *a* to *b* on the line *OB* means doubling the inputs. It can be seen in Fig. 5.3 that the combination of inputs *L* and *K*, 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.*, a 100 per cent increase in inputs results in 120 per cent increase in output. Simi-larly, the movement from point *b* to point *c* indicates a 50% increase in inputs as a result of which the output increases from 25 units to 50 units, i.e., by 200%. This kind of relationship between the inputs and output shows *increasing returns to scale*.

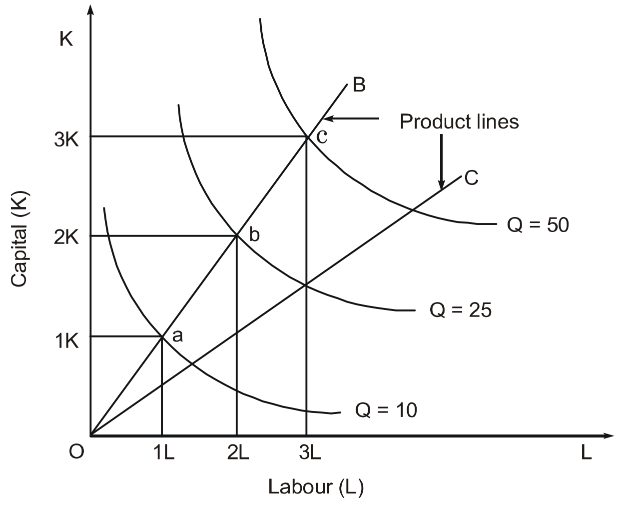


***Fig. 5.3** Increasing Returns to Scale*

Factors Leading to Increasing Returns to Scale: There are at least three plausible reasons for increasing returns to scale, called *economics of seale*.

(i) ***Indivisibility of Machinery and Managerial Manpower.*** Certain inputs, particularly mechanical equipments and managerial manpower, 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 to produce electricity and one-third of a composite

harvester and earth-movers cannot be used productively. 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 labour, manager and machinery, which becomes possible with increase in scale of production. The use of specialized labour suitable to job needs and composite machinery increases productivity 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 ($15' \times 10' = 150$ sq. ft.) are doubled, then the size of the room is more than doubled. It increases to $30' \times 20' = 600$ sq. ft. which is more than double the room size. Similarly, 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 proportional to the increase in inputs, it exhibits *constant returns to scale*. For example, if both the inputs, *K* and *L*, are doubled subsequently and output is also doubled, subseqently then the returns to scale are said to be constant. Constant returns to scale are illustrated in Fig. 5.4. The lines *OA* and *OB* are 'product lines' indicating two hypothetical techniques of production. 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., the output increases from 10 to 20, i.e., a 50% increase in output.

NOTES

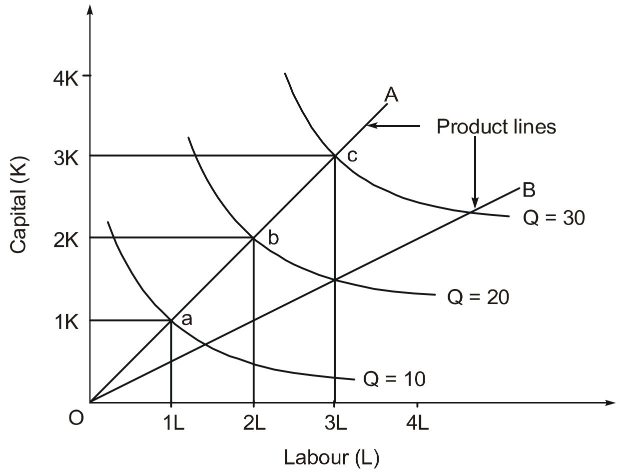


Fig. 5.4 *Constant Returns to Scale*

Similarly, movement from point *b* to *c* indicates a 50 per cent increase in labour as well as 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 a 50 per cent increase in output. This relationship between the proportionate change in inputs and proportional change in output may be summed up as follows:

$$1K + 1L \Rightarrow 10$$
$$2K + 2L \Rightarrow 20$$
$$3K + 3L \Rightarrow 30$$

This 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 labour-saving 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 also take place 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 proportional 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. 5.5. As the figure shows, when the inputs *K* and *L* are doubled, i.e., where capital-labour combination is increased from $1K + 1L$ to $2K + 2L$, the output increases from 10 to 18 units, which is less that the proportionate increase. The 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*.

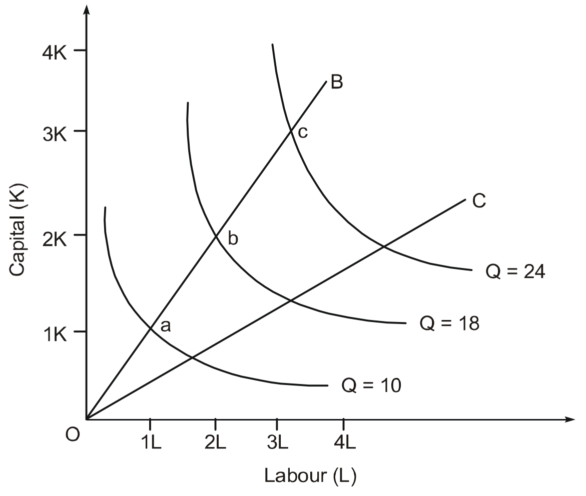


Fig. 5.5 *Decreasing Returns to Scale*

Causes of Diminishing Returns to Scale. The decreasing returns to scale are attributed to the diseconomies of scale. The most important factor causing diminishing returns to scale is 'the dimini-shing 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 coalmining 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.

5.3.5 Laws of Returns to Scale through Production Function

The laws of returns to scale may be explained more precisely through a production function. Let us assume a production function involving two variable inputs (*K* and *L*) and one commodity *X*. The production function may then be expressed as

$$Q_x = f(K, L) \qquad \ldots(5.4)$$

where Q_x denotes the quantity of commodity *X*. Let us also assume that the production function is homogeneous. A production function is said to be homogeneous when all the inputs need to be increased in the same proportion and the proportion can be factored out. And, if all the inputs are increased by a certain proportion (say, *k*) and output increases in the same proportion (*k*), then production is said to be homogeneous of degree 1. This kind of production function may be expressed as follows.

$$kQ_x = f(kK, kL) \qquad \ldots(5.5)$$

NOTES

or $$= k\ (K,\ L)$$

A homogeneous production function of degree 1, as given in Eq. (5.5), implies *constant returns* to *scale*. Eq. (5.5) shows that increase in inputs, K and L, by a multiple of k, increases output, Q_x, by the same multiple (k). This means constant returns to scale.

The constant returns to scale may not be applicable in all forms of production. Increasing both inputs K and L in the same proportion may result in increasing or diminimishing returns to scale. In simple words, it is quite likely that when all the inputs are increased by a certain proportion, output does not increase in the same proportion. For example, if the number of buses is doubled in a city, the number of passengers may not be doubled—it may increase by less than or more than double. Then the production function may be expressed as

$$hQ_x = f\ (kK,\ kL) \qquad \ldots(5.6)$$

where h denotes h-times increase in Q_x, as a result of k-times increase in inputs, K and L. The proportion h may be greater than k, equal to k, or less than k. Accordingly, it reveals the three laws of returns to scale:

(*i*) If $h = k$, production function reveals constant returns to scale.

(*ii*) If $h > k$, it reveals increasing returns to scale.

(*iii*) If $h < k$, it reveals decreasing returns to scale.

This aspect has been elaborated in the following section.

1. Degree of Production Function and Returns to Scale

In case of a homogeneous production function of degree 1 (Eq. 5.5), k has an exponent equal to 1, i.e., $k = k^1$. It means that if k has an exponent equal to 1, the production function is homogeneous of degree 1. But, all the production functions need not be homogeneous of degree 1. They may be homogeneous of a degree less than 1 or greater than 1. It means that the exponent of k may be less than 1 or greater than 1. Let us assume that exponent of k is r, where $r \neq 1$. A production function is said to be of degree r when all the inputs are multiplied by k and output increases by a multiple of k^r. That is, if then function (5.7), is homogeneous of degree r.

$$Q = f(kK,\ kL) = k^r(K,\ L) = k^r\ Q \qquad \ldots(5.7)$$

From the production function (5.17), we can again derive the laws of returns to scale.

(*i*) If $k > 1$, and $r < 1$, it reveals decreasing returns to scale;

(*ii*) If $k > 1$ and $r > 1$, it reveals increasing returns to scale; and

(*iii*) If $k > 1$ and $r = 1$, it means constant returns to scale.

For example, consider a multiplicative form of production function i.e.,

$$Q = K^{0.25}\ L^{0.50} \qquad \ldots(5.8)$$

If K and L are multiplied by k, and output increases by a multiple of h then

$$hQ = (kK)^{0.25} \ (kL)^{0.50}$$

By factoring out k, we get

$$hQ = k^{0.25+0.50} \ [K^{0.25} \ L^{0.50}]$$

$$= k^{0.75} \ [K^{0.25} \ L^{0.50}] \qquad \ldots(5.9)$$

In Eq. 5.9, $h = k^{0.75}$ and $r = 0.75$ [see Eq. (5.7) also]. This means that $r < 1$ and, thus, $h < k$. Production function (5.8), therefore, shows *decreasing returns to scale*.

Now consider another production function with an additional factor X, given as

$$Q = K^{0.75} \ L^{1.25} \ X^{0.50} \qquad \ldots(5.10)$$

If inputs K, L and X are multiplied by k, Q increases by a multiple of h then

$$hQ = (kK)^{0.75} \ (kL)^{1.25} \ (kX)^{0.50}$$

By factoring out k, we get

$$hQ = k^{(0.75+1.25+0.50)} \ [K^{0.75} \ L^{1.25} \ X^{0.50}]$$

$$= k^{2.5} \ [K^{0.75} \ L^{1.25} \ X^{0.50}]$$

Here $h = k^{2.5}$ and $r = 2.5 > 1$. So $h > k$. Therefore, function (5.10) gives *increasing returns to scale*. Similarly, if in a production function, $h = k$ or $r = 1$, the production function shows *constant returns to scale*.

2. Power Function: Cobb-Douglas Production Function

One of the widely used production functions is the *power function*. The most popular production function of this category is *'Cobb-Douglas Production Function*, expressed as

$$Q = AK^a \ L^b \qquad \ldots(5.11)$$

where A is a positive constant; a and b are positive fractions; and $b = 1 – a$.

The Cobb-Douglas production function is often used in its following form.

$$Q = AK^a \ L^{1-a} \qquad \ldots(5.12)$$

Properties of Cobb-Douglas Production Function: A power function of this kind has several important properties.

First, the multiplicative form of the power function (5.11) can be changed into its log-linear form as

$$\log Q = \log A + a \log K + b \log L \qquad \ldots(5.13)$$

In its logarithmic form, the function becomes simple to handle and can be empirically estimated using linear regression analysis.

NOTES

Second, power functions are homogeneous and the degree of homogeneity is given by the sum of the exponents a and b. If $a + b = 1$, then the production function is homogeneous of degree 1 and implies constant returns to scale.

Third, a and b represent the elasticity co-efficient of output for inputs K and L, respectively. The output elasticity co-efficient ($\in$) in respect of capital may be defined as proportional change in output as a result of a given change in K, keeping L constant. Thus,

$$\in_k = \frac{\partial Q/Q}{\partial K/K} = \frac{\partial Q}{\partial K} \cdot \frac{K}{Q} \qquad \text{... (5.14)}$$

By differentiating the production function $Q = AK^a L^b$ with respect to K and substituting the result in Eq. (5.14), we can find the elasticity co-efficient.

$$\frac{\partial Q}{\partial K} = a\ AK^{a-1}\ L^b$$

Substituting the values for Q and $\partial Q/\partial K$ in Eq. (5.14), we get

$$\in_k = a\ AK^{a-1}\ L^b \left(\frac{K}{AK^a L^b}\right) = a \qquad \text{...(5.15)}$$

Thus, output-elasticity coefficient for K is 'a'. The same procedure may be adopted to show that b is the elasticity co-efficient of output for L.

Fourth, constants a and b represent the relative distributive share of inputs K and L in total output Q. The share of K in Q is given by

$$\frac{\partial Q}{\partial K} \cdot K$$

Similarly, the share of L in Q may be obtained as

$$\frac{\partial Q}{\partial L} \cdot L$$

The relative share of K in Q can be obtained as

$$\frac{\partial Q}{\partial L}\ K \cdot \frac{1}{Q} = \frac{a\,AK^{a-1}L^b.K}{AK^a\ L^b} = a$$

Similarly, it can be shown that b represents the relative share of L in the total output.

Finally, Cobb-Douglas production function in its general form, $Q = K^a L^{1-a}$ implies that at zero cost, there will be zero production.

Some Input-Output Relationships: Some of the concepts used in production analysis can be easily derived from the Cobb-Douglas production function as shown below.

(i) Average Product (AP) of L and K:

$$AP_L = A\ (K/L)^{1-a}$$
$$AP_K = A\ (L/K)^1$$

(*ii*) Marginal Product of L and K

$$MP_L = a.A\ (K/L) - a = a\ (Q/L)$$
$$MP_K = (a - 1)\ A\ (L/K)a = (1 - a)\ Q/K$$

(*iii*) Marginal Rate of Technical Substitution

NOTES

$$MRTS_{L,\,k} = \frac{MP_L}{MP_K} = \left[\frac{a}{1-a} \cdot \frac{K}{L}\right]$$

Check Your Progress

1. Mention the two categories of inputs.
2. State the law of diminishing returns with reference to variable input.
3. Name the three kinds of returns to scale.

5.4 ANSWERS TO 'CHECK YOUR PROGRESS' QUESTIONS

1. Inputs are classified under two categories:
   (a) Fixed inputs or fixed factors
   (b) Variable inputs or variable factors
2. The law of diminishing returns states that when more and more units of a variable input are applied to a given quality of fixed inputs, the total output may initially increase at an increasing rate and then at a constant rate but it will eventually increase at diminishing rates.
3. The three kinds of returns to scale are the following:
   (i) Increasing returns to scale
   (ii) Constant returns to scale
   (iii) Diminishing returns to scale

5.5 SUMMARY

- In economics, 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 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.
- The traditional theory of production studies the marginal input-output relationships under (*i*) short run, and (*ii*) long run conditions.

NOTES

- The law of diminishing returns states that when more and more units of a variable input are applied to a given quantity of fixed inputs, the total output may initially increase at an increasing rate and then at a constant rate but it will eventually increase at diminishing rates.
- An isoquant curve is locus of points representing various combinations of two inputs—capital and labour—yielding the same output.
- The laws of returns to scale state 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.

5.6 KEY WORDS

- **Production:** In economics, the term production means a process by which resources are transformed into a different and more useful commodity or service.
- **Input:** An input is a good or service that is used into the process of production.
- **Short-run:** In economics, the short run refers to a period of time in which the supply of certain inputs is fixed or is inelastic.

5.7 SELF-ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. State the empirical validity of the law of diminishing returns.
2. How does the law of diminishing returns influence business decisions?
3. What are the differences between an isoquant curve and an indifferent curve?
4. Write a short note on the Cobb-Douglas production function.

Long-Answer Questions

1. 'The laws of production state the nature of relationship between output and input.' Discuss the statement.
2. Analyse the significance of law of returns to scale with reference to production function.
3. Discuss the factors leading to the increasing returns to scale.
4. 'The constant returns to scale are attributed to the limits of the economies of scale.' Explain the statement.

5.8 FURTHER READINGS

NOTES

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 THEORIES OF POPULATION

NOTES

Structure

6.0 INTRODUCTION

In the previous unit, you studied about factors of production and law of returns. In this unit, you will study about the Marxist theory of Population and the Demographic Transition Theory. The current world population of 7.6 billion is expected to reach 8.6 billion in 2030, 9.8 billion in 2050 and 11.2 billion in 2100, according to a new United Nations report. With roughly 83 million people being added to the world's population every year, the upward trend in population size is expected to continue, even assuming that fertility levels will continue to decline. *The World Population Prospects: The 2017 Revision*, published by the UN Department of Economic and Social Affairs, provides a comprehensive review of global demographic trends and prospects for the future. The new projections include some notable findings at the country level. China (with 1.4 billion inhabitants) and India (1.3 billion inhabitants) remain the two most populous countries, comprising 19% and 18% of the total global population. In roughly seven years, or around 2024, the population of India is expected to surpass that of China.

6.1 OBJECTIVES

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

- Discuss the main features of the Marxist theory of Population
- Explain the theory of Demographic Transition
- Analyse Marxist's argument against Malthus' Population theory

6.2 MARXIST THEORY OF POPULATION

Karl Marx was a renowned German philosopher who introduced the idea of modern communism. He explained the law of population with relation to the concept of surplus labour. The outline theory of population and the theory of population are one of his most important works where his theory has been discussed in great detail. Marx suggested that a capitalist society consists of the following classes:

- **Capitalists:** The people who are the owners of the means of production
- **Workers:** The people who do not possess any ownership but contribute in terms of labour

He held the opinion that in a capitalist society, the capitalist society controls the production in an economy. Their source of earnings is the profit which they incur from the production activities. The labour theory of value states that the workers create the value addition through their labour which means that the means of production cannot carry any activity, it is because of the contribution of the workers that a value is created. The workers are in return given the wages which is always less as compared to the contribution which they put in. The difference (surplus) is expropriated by the capitalists and added to fixed capital under their possession.

In other words, capital consists of the following parts:

- Constant capital: It is the type of capital which includes the means of production, raw materials and the various instruments of labour.
- Variable capital: It is the type of capital which is represented by the power of the labour.

It has also been concluded that the constant capital remains fixed in terms of quantitative value whereas the variable capital produces its own value and a surplus.

Accumulation of capital, though originally appearing as its quantitative extension only, results eventually in the change in its composition, under a constant increase of its constant constituent, and a constant decline in variable constituent. On several occasions when the total capital increases, its variable part may also increase, however, it will only increase in a constantly diminishing proportion. Technological development, rise in productivity of labour and centralization tend to decrease the ratio of variable capital further. Since in a capitalist society the demand for labour depends on the variable constituent only, therefore, it falls progressively, and the labouring population, therefore, produces, along with the accumulation of capital, the means by which it is rendered superfluous and surplus to an increasing extent. Accordingly, the advancement of modern industry leads to unemployment and underemployment.

For Marx, therefore, the correlation between accumulation of capital and rate of wages is nothing else than the correlation between the unpaid labour

NOTES

transformed into capital and the additional paid labour necessary for the setting in motion of this additional capital. It cannot be observed as a relation between two independent factors such as population and capital. If the quantity of unpaid labour supplied by the working-class, and accumulated by the capitalist class, increases so rapidly, that its conversion into capital requires an extraordinary addition of paid labour, which will lead to an increase in the wages and, all other circumstances remaining equal the unpaid labour diminishes in proportion. However, as soon as this diminution touches the point where the surplus labour goes beyond the limit of the normal quantity, as a consequence of which the smaller part of revenue is capitalized, accumulation lags, and the movement of rise in wages receives a check.

With progressive advancement of capital, labourers are set free more rapidly than the reduction in the variable part of capital as compared to the constant because it enables the capitalists to exploit the labour power. It also leads to progressive replacement of superior labour power by inferior labour power. Ultimately, the overwork of labour, reduction in the variable constituent of capital, and greater exploitation expand the ranks of industrial reserve army and force the workers to subjugate under the dictates of capital, independently of the natural increase of population. The development in this way increases both the demand and supply of labour by setting them free.

Marxist Argument Against Malthus' Population Theory

Karl Marx and Friedrich Engels were against the theory which was provided by Malthus as they believed it to be a theory which was an explanation for the status quo. By observing the scientific view point which they have raised against Malthus's theory, one can distinguish it according to the two level of analysis.

At first level, in the Marxist theory, reify is considered as a means of changing the concrete historical social relations and route into universal categories or eternal natural laws. This development of reification of social relations characterizes the nature of scholarly production under conditions of capitalist production. Through this process, he emphasised on the fact that a man's reflections of the forms of social life and, subsequently, also his scientific analysis of those forms, acquire a course which is in sharp contrast with the actual historical development. He believed that Malthus focussed on only certain aspects such as hunger, poverty, and unemployment, however, he ignored the social relations which existed in the society which further, created a sense of exploitation and competition.

Secondly, at a more precise level, Marx's answer to Malthus' principle of population is the principle of the reserve army of labour or relative surplus population in which he focuses on the general law of capital accumulation.

The accumulation and development of capital constitutes the energetic force of capitalism and it becomes possible as long as capitalists incur profit. Profits originate in the requisition, by the capitalist, of the surplus value created by the labor power he buys. Accumulation takes place when capitalists exchange a portion of their surplus value into capital; and this allows them to appropriate more surplus value which will further lead to processes such as accumulation and expansion and so on.

6.3 THEORY OF DEMOGRAPHIC TRANSITION

Demographic Transition (DT) refers to the transition from high birth and death rates to lower birth and death rates as a country develops from a pre-industrial to an industrialized economic system. The theory was proposed in 1929 by the American demographer Warren Thompson, who observed changes, or transitions, in birth and death rates in the industrialized societies over the previous 200 years. Most developed countries have completed the demographic transition and have low birth rates; while most of the developing countries are in the process of this transition. The major (relative) exceptions are some poor countries, mainly in sub-Saharan Africa and some Middle Eastern countries, which are poor or affected by government policy or civil strife, notably Pakistan, Palestinian territories, Yemen, and Afghanistan.

The demographic transition model, in isolation, can be taken to predict that birth rates will continue to go down as societies grow increasingly wealthy. However, recent data contradicts this, suggesting that beyond a certain level of development birth rates increase again. In addition, in the very long term, the demographic transition should be reversed via evolutionary pressure for higher fertility and higher mortality. The existence of some kind of demographic transition is widely accepted in the social sciences because of the well-established historical correlation linking dropping fertility to social and economic development.

Scholars debate whether industrialization and higher incomes lead to lower population, or whether lower populations lead to industrialization and higher incomes. Scholars also debate to what extent various proposed and sometimes inter-related factors such as higher per-capita income, higher female income, lower mortality, old-age security, and rise of demand for human capital are involved.

What is Demographic Transition?

Over the course of human history, there have been many people who have been interested in the characteristics of the human population and the future of population growth. After analysing how western populations have changed over time, one pattern was discovered that indicated a connection between population growth and the economic development of a country. It was observed that in countries with high standards of living, the population grew at a slow rate while in countries with low standards of living, the population grew more rapidly.

This discovery resulted in the creation of the *concept of demographic transition*, which is a series of stages that a country goes through when transitioning from a non-industrial to industrial society. The demographic transition concept involves four stages that are based on changes to population size and social behaviours.

Stages of the Demographic Transition Theory

NOTES

According to the demographic transition theory, human societies are categorized into one of four stages of industrial development. Let's have a look at them step by step:

- **Stage one, pre-industrial society:** It is the preliminary stage of economy where a country has an agrarian economy and low level of income and there are high birth rates and high death rates, but it is stable, so the overall population does not change very dramatically. From the commencement of human history until the first Industrial Revolution in the 1700s, this was the basic prototype for all human societies. Uses of family planning and contraception were non-existent; consequently, birth rates were essentially only limited by the skill of women to bear children. Children added to the economy of the household from an early age by bringing water, firewood, and messages, loving younger siblings, sweeping, washing dishes, getting ready food, and working in the field's. Only feeding the child was cost of raising a child; there were no education or entertainment expenses. Thus, the total cost of raising children hardly went beyond their contribution to the household. In addition, when they became adults, they participated in the family business, mainly farming, and were the primary form of social security for adults in their old age. Thus in this stage, birth rate is high out of economic requirements and death rate is also high due to insufficient food and adequate medical facility.
- **Stage two:** In this stage, the country is in the beginning of its industrial development. The agriculture is more productive due to the new change in technology and healthcare is more efficient, so there is additional food and less disease. This results in a sharp decrease in the death rates. Although birth rates stay the same, but the fact that there are fewer deaths means that population increases very quickly. This stage in Europe was initiated at the time of agricultural revolution of the 18th century and was initially quite slow. In the 20th century, the falls in the death rates started in developing countries also but tended to be substantially faster. The decline in the death rate in the beginning was due to two factors:
    - o First, upgradation in the food supply brought about by higher capitulates in agricultural practices and improved transportation averted death due to starvation and lack of water. Agricultural advances included crop rotation, selective breeding, and seed drill technology.
    - o Second, considerable improvements in public health decreased mortality, particularly in childhood as Europe surpassed through stage two before the progress of the mid-20th century; although there was

important medical advancement in the 19th century, such as the vaccination. Development in this stage in the areas of water supply, sewerage, food handling, and general personal hygiene were improved and started more of following of the scientific knowledge, of the causes of disease and there was improvement in the education and social status of mothers. Law and order situations were improved, introduction of immunization programmes, development of antibiotics, medical innovations and advancement, have led to considerable decrease in the incidence of disease and death. But at the same time, in this stage the birth rate persisted to remain very high in-spite of extensive drop in death rates leading to accelerated growth of population.

NOTES

- **Stage three:** As societies become more and more industrialized, they enter stage three of the demographic transition theory. By stage three, death rates are still low, but birth rates begin to decline as well. So, why are people in stage three having fewer kids? For one, fewer children are dying, so parents can stop having children earlier. Also, more industrial technology means that societal values tend to shift away from rural and agricultural lives to urban, industrial ones. In other words, people move to the city and buy food, rather than growing it themselves. Other factors include a higher number of educated, professional women, increased costs of childcare as a result of child labour laws and mandatory education, and increased social pressure to essentially spoil the children you have. Also, birth control is usually readily available in stage three, although this is a pretty recent change. So, that's a lot of social change, and as a result, the population growth of stage two tends to level out in stage three. In this stage with the gradual accomplishment of economic development, the economy of the country starts experiencing change in its structure from agrarian to an industrialized one. Throughout this stage, people become aware about the size of the family and also on limiting the size of the family. There is a mass departure of population from rural to urban areas in search of food and job.

  In order to avoid the menace of large family, the age of marriage starts postponing as well as to limited reproduction takes place. With the developing industrialization of the economy, the small family norm becomes very much popular among the people of higher sections of society and then it starts to penetrate among the lower sections of society. Thus at this stage, the country experiences reduction in the birth rate, low death rate and as a result, a fall in the rate of growth of population.

- **Stage four:** In this stage, both birth and death rates are low, which often results in a population that shrinks at first but then levels off at lower levels. Many stage four countries rely on immigration to keep their population up, which is generally not a problem since fully-industrialized societies tend to

NOTES

have more job opportunities. In fact, these opportunities are the reason that birth rates tend to be lowest in stage four countries. Almost everyone has equal education and work opportunities, so people choose not to have kids but instead work hard and live out long lives spending their money. The United States, Canada, Australia, South Korea, Iran, China, Brazil, and most of Europe all display these trends.

This occurs where birth and death rates are both low, leading to a total population which is high and stable. Death rates are low for a number of reasons, primarily lower rates of diseases and higher production of food. The birth rate is low because people have more opportunities to choose if they want children; this is made possible by improvements in contraception or women gaining more independence and work opportunities. Some theorists consider that there are only four stages and that the population of a country will remain at this level. The DTM is only a suggestion about the future population levels of a country, not a prediction.

Countries that are at this stage (Total Fertility Rate of less than 2.5 in 1997) include: United States, Canada, Argentina, Australia, New Zealand, most of Europe, Bahamas, Puerto Rico, Trinidad and Tobago, Brazil, Sri Lanka, South Korea, Singapore, Iran, China, Turkey, Thailand and Mauritius.

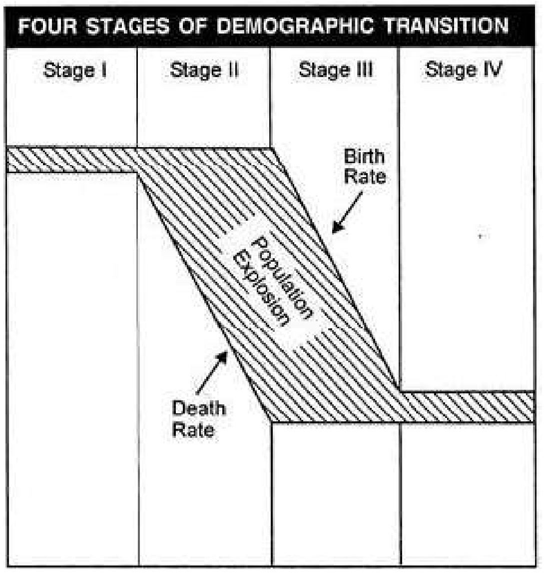


Fig. 6.1 *Four Stages of Demographic Transition*

Due to the accomplishment of economic development, standard of living of the people attains a high level during this fourth stage. During this stage, a noteworthy change in the social outlook of the people has taken place beneath the impact of urbanization, industrialization and high rate of literacy. Therefore, at this stage population becomes stationary at a low rate.

- **Stage 5 (Debated):** It is a stage where birth rates are below death rates, and the population ages and begins to decline. The original Demographic

Transition model has just four stages, but additional stages have been proposed. Both more-fertile and less-fertile futures have been claimed as a Stage Five.

NOTES

Some scholars delineate a separate fifth stage of below-replacement fertility levels. Others hypothesize a different stage five involving an increase in fertility. The United Nations Population Fund (2008) categorizes nations as high-fertility, intermediate-fertility, or low-fertility. The United Nations (UN) anticipates the population growth will triple between 2011 and 2100 in high-fertility countries, which are currently concentrated in sub-Saharan Africa. For countries with intermediate fertility rates (the United States, India, and Mexico all fall into this category), growth is expected to be about 26 per cent. And low-fertility countries like China, Australia, and most of Europe will actually see population declines of approximately 20 per cent.

Some countries have sub-replacement fertility that is fertility below 2.1-2.2 children per woman. Replacement fertility is usually 2.1-2.2. Many European and East Asian countries now have higher death rates than birth rates.

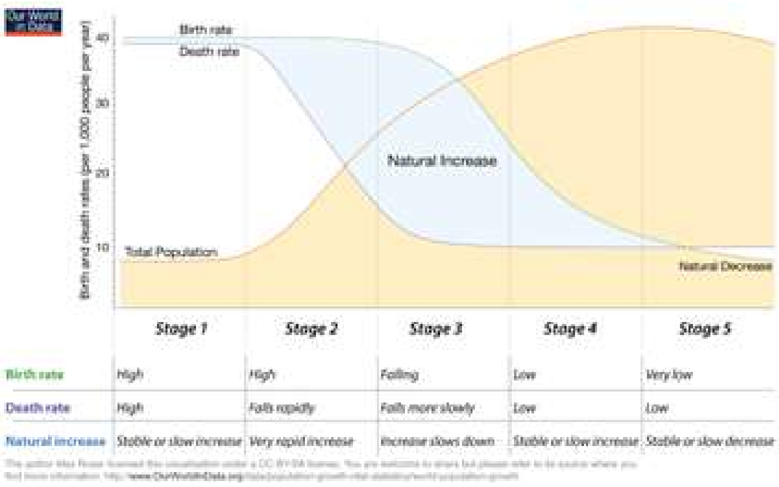


Fig. 6.2 Overview of Demographic Transition

Criticism of the Theory of Demographics Transition

The theory of demographic transition is widely accepted as a useful aid in describing the demographic history. Its contribution is, however, considered to be of limited value. Several questions are raised in this context: Can this theory provide theoretical explanation of the forces that caused demographic changes, especially as they related to fertility?

Does it have any predictive value? In other words, can it be used for predicting the sequence through which developing countries would pass? Such questions have inevitably led to a great deal of unfavourable criticism of the theory of demographic transition.

NOTES

It is to be noted that this theory is based on the actual experience of the changes in the vital rates in Western countries during the various stages of their industrial and economic development.

The critics of this theory, however, point out that the experiences of the various European countries were not uniform, in the sense that the sequences of the stages as described in the statement of the theory were not the same.

Recent studies reveal that, in Spain and in some countries of Eastern and Southern Europe, fertility decline occurred even when mortality was very high.

In some countries like the United States, the growth rate in the post-transition stage was probably higher than in Stage II Stage III of the demographic transition. Notestein's claim about the fertility declining initially in the urban areas is found to be untrue to some countries.

For instance, countries with predominantly retaliations, such as France, Sweden, Finland, Bulgaria, did experience decline in birth rates to the same extent as did some highly urbanised countries, such as England and Wales.

Such exceptions suggest that the theory of demographic transition is only a broad generalisation, which does not encompass the experiences of even all the Western countries.

This theory also cannot explain the phenomenon of 'the baby boom' in Western countries after the economic recovery and the Second World War.

Another criticism of this theory arises out of the fact that it does not provide a theoretical explanation of an important force, viz., fertility, which brought about the demographic transition.

Though it is a fact that fertility did decline in all the Western countries, the conditions under which it declined were diverse. Demographers have recently arrived at the conclusion that the decline in fertility in Europe is a very complex phenomenon which has not yet been fully understood.

David Glass, the British demographer, despairingly points out that even the English people do not have an adequate knowledge of their own demographic transition.

It has, moreover, been asserted that the theory of demographic transition cannot really be called a 'theory,' for it does not fulfil an important criterion of any theory, that is, to extract fundamental processes from a phenomenon and identify crucial variables.

This theory does not provide fundamental explanations for fertility decline, nor does it identify the crucial variables involved in the process of fertility decline. Therefore, it does not have any predictive value.

NOTES

In all fairness, it must be mentioned at this point that Notestein, who propounded this theory, was aware of its limitations. Nonetheless, he was of the opinion that the principle drawn from the European experience would be applicable to other parts of the world.

The most crucial question to be considered is: 'Can the theory demographic transition be applied to developing countries?' It is well known that developing countries have recently experienced a phenomenal reduction in death rates, as a result of which there has been a tremendous increase in the rates of population growth.

This rapidly increasing population is an obstacle in the path of the programmes of the developing countries, which are making concentrated efforts to eliminate poverty and to rise of living for the masses by launching large-scale plans for economic development, industrialisation and modernisation.

In such a situation, it may be asked will these countries have to wait for economic and social development till they bring down the birth rate and bring about a reduction in the growth rate.

According to the theory of demographic transition, the reduction in the birth rate is a by-product of industrialisation and modernisation. In the face of a very rapidly increasing population, however, it is not possible to wait for industrialisation and modernisation to bring about the required reduction in fertility.

Many developing countries have, therefore, adopted family planning programmes directly geared to influence fertility negatively.

In such a situation, it is difficult to maintain with any degree of confidence that the theory of demographic transition is also applicable to developing countries and that what happened in the west, in respect of population growth, would be duplicated in developing countries.

Thus, in the strictest sense of the term the theory of demographic transition cannot really be considered as a theory, though it does provide a satisfactory framework and means for wider empirical generalisations.

In conclusion, it can be said that the demographic transition theory is an excellent benchmark for describing the how populations in societies have changed remarkably.

Check Your Progress

1. Name the two main classes of a capitalist society.
2. What is variable capital?
3. In which year was the demographic transition theory proposed and by whom?

6.4 ANSWERS TO 'CHECK YOUR PROGRESS' QUESTIONS

NOTES

1. A capitalist society consists of capitalists and workers.
2. Variable capital is a type of capital which is represented by the power of the labour.
3. The Demographic Transition Theory was proposed in 1929 by the American demographer Warren Thompson, who observed changes, or transitions, in birth and death rates in the industrialized societies over the previous 200 years.

6.5 SUMMARY

- Karl Marx was a renowned German philosopher who introduced the idea of modern communism. He explained the law of population with relation to the concept of surplus labour.
- He held the opinion that in a capitalist society, the capitalist society controls the production in an economy. Their source of earnings is the profit which they incur from the production activities.
- Accumulation of capital, though originally appearing as its quantitative extension only, results eventually in the change in its composition, under a constant increase of its constant constituent, and a constant decline in variable constituent.
- Karl Marx and Friedrich Engels were against the theory which was provided by Malthus as they believed it to be a theory which was an explanation for the status quo.
- Demographic transition (DT) refers to the transition from high birth and death rates to lower birth and death rates as a country develops from a pre-industrial to an industrialized economic system.
- The demographic transition model, in isolation, can be taken to predict that birth rates will continue to go down as societies grow increasingly wealthy.
- According to the demographic transition theory, human societies are categorized into one of four stages of industrial development.
- The theory of demographic transition is widely accepted as a useful aid in describing the demographic history. Its contribution is, however, considered to be of limited value.
- In conclusion, it can be said that the demographic transition theory is an excellent benchmark for describing the how populations in societies have changed remarkably.

6.6 KEY WORDS

NOTES

- **Constant Capital:** It is the type of capital which includes the means of production, raw materials and the various instruments of labour.
- **Demographic Transition:** It refers to the transition from high birth and death rates to lower birth and death rates as a country develops from a pre-industrial to an industrialized economic system.

6.7 SELF-ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. Write a short note on the Marxist theory of Population.
2. What are the criticisms raised against the Demographic Transition Theory?
3. How does Marx interpret the correlation between accumulation of capital and rates of wages?

Long-Answer Questions

1. 'The Marxist theory is based on the idea of history of class struggle.' Explain the statement.
2. Analyse the Marxist argument against Malthus' theory of Population.
3. Discuss the stages of the Demographic Transition Theory.

6.8 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 7 MALTHUSIAN AND OPTIMUM THEORIES

NOTES

Structure

7.0 INTRODUCTION

In the previous unit, you studied about the Marxist theory of Population and the Demographic Transition Theory. In this unit, you will study about the famous theories of population namely, Malthus theory of Population and Optimum theory of Population. In the present scenario, where we are witnessing rapid growth of population in the world especially with India and China leading in this arena; the importance and significance of theories of population becomes all the more relevant.

7.1 OBJECTIVES

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

- Discuss Malthusian theory of Population
- Differentiate between the Malthusian theory and the Neo-Malthusian theory of population
- Explain the Optimum theory of Population

7.2 MALTHUSIAN THEORY

A renowned economist, Thomas Malthus was one of the first economist who in the year 1978, provided a methodical doctrine of the population. However, his theory was criticized too as it had a lot of incorrect data about the economic changes which occurred in Europe during the period of nineteenth and twentieth century. The theory propounded by him was known as 'Malthusian Theory' which explained the changes which occurred in a population of a given area and the causes for the changes. Malthus, thus became quite popular for his contribution

NOTES

and his famous essay, 'An essay on the principle of population', stated that the population of a particular area will change eventually as a result of famine or a disaster. Thus, it was concluded that the population will change continuously over a period of a given time.

He proposed that human population undergoes a change in each cycle while the production of the food grows at an arithmetic rate i.e. a uniform increment is added in a uniform interval of time. Therefore, population was capable of changing in a geometric way such as 1, 2, 4, 8, 16, and 32 and so on, whereas the food output was likely to increase in a series of twenty-five year intervals in the arithmetic progression 1, 2, 3, 4, 5, 6, 7, 8, 9 and so on. This change in population at a geometric manner and the change in food output in an arithmetic manner created a possibility of a future where human beings will not have sufficient resources in order to survive in this world. In order to avoid such circumstance, Malthus proposed that there should be some sort of control on the growth of the population.

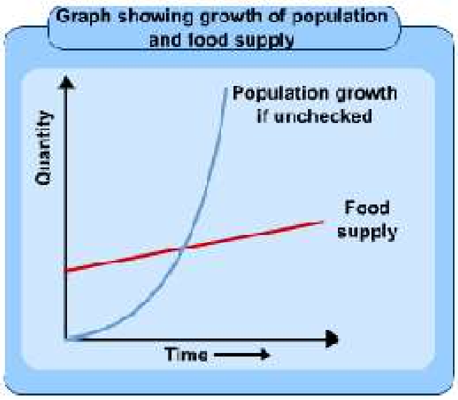


Fig. 7.1 *Relationship between Population, Growth and Food Supply*

Figure 7.1 shows the relationship between the population growth and food supply.

Thus, according to his ideas, population increases in a geometrical ratio whereas the food supply increases in an arithmetic ratio. For instance, a family continues to expand with each generation, however, the food production changes in an arithmetic manner, it will thus, increase at particular points only. Therefore, this situation according to him will create an imbalance which will eventually lead to starvation and poverty which would further be aggravated in case of natural disasters.

Malthus stated that there are mainly two kind of checks which can help in reducing the growth of the population. They are stated as follows:

- **Preventive checks**: These are the actions which are taken by people voluntarily that can help in reducing the population. He believed in the concept of moral restraint and the main idea of this concept was that people resist their urge to marry and reproduce until they are not capable enough to support their own family. It was basically considered as a way in which

NOTES

people gets married at a later stage. He further stated that there are several immoral ways too which can help in controlling the population such as adultery, prostitution and birth control. He thus, firmly believed in the idea of moral restraint and did not supported such immoral practices. Therefore, these measures aimed at controlling the birth rate and uses of various birth control methods.

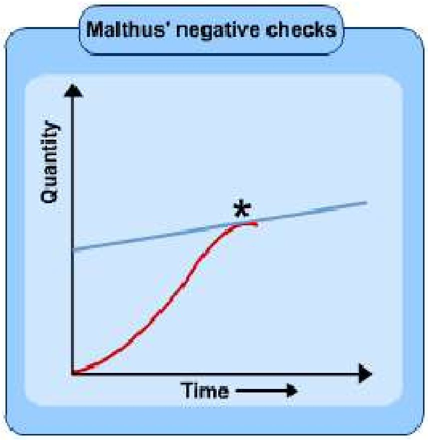


Fig. 7.2 *Preventive Measures for Population Growth*

Figure 7.2 shows the preventive measures for the population growth.

- **Positive checks:** These are the types of checks which may shorten the average lifespan of a person naturally such as disaster, famine, disease, and poor living and working conditions. These checks would eventually lead to Malthusian catastrophe**,** which will make the individuals return to basic survival.

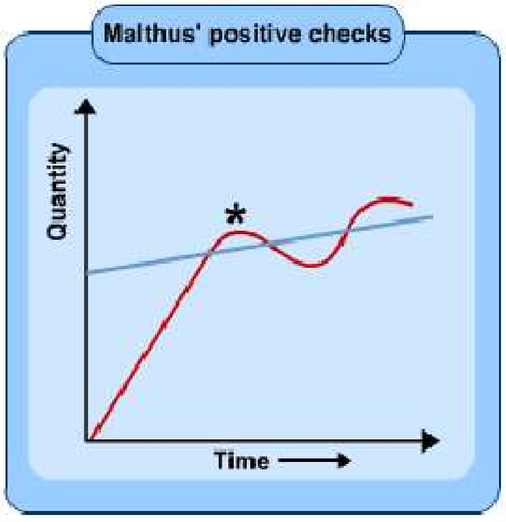


Fig. 7.3 *Positive Check on Population Growth*

Figure 7.3 shows the positive check of the population growth.

NOTES

Malthusian Catastrophe

The Malthusian controversy is also popularly known as 'Malthusian check' which is a prediction that the population by forcible circumstances would return to a subsistence-level situation, if it will be more than the agriculture production. He believed that famine is the worst resource of the nature and the increase in population and such worse circumstances, would lead to premature deaths.

In the year 1978, Malthus in his essay, concluded that the plants and animals can produce far more progeny than can survive. He thus, held the view that if reproduction will be left unchecked, then even the humans can overproduce. It is therefore, necessary that the human population should be kept in check, otherwise situation like famine will occur which would destroy the humans.

Though Malthus firmly did not believed that it is in the destiny of mankind to face the 'catastrophe' but to a certain extent, he held the opinion that it should be controlled in order to avoid harmful circumstances. Human beings should thus use preventive measures to control the rate of population.

He considered only two factors i.e. food supply and population growth while explaining his theory, however, his theory was also criticized as it ignored other factors, such as the advancements in technology or developments in the agriculture sector.

Principles of Malthus Theory

The following are the main principles of Malthus's theory:

- There are several genetic traits in human beings which help them to reproduce at a faster rate. As a result, population increases at a statistical rate if unimpeded two-folds itself after every twenty-five years.
- The supply of the food increases at a very slow pace because of the working of the law of diminishing returns which is based on the presumption that the supply of land is invariable.
- Since population increases in statistical progression and the food supply in the numerical sequence, population is likely to elude food supply. Thus, an imbalance is created which directs to over populace.
- It is necessary to keep a check on the population in order to avoid circumstances where there would be an imbalance between the population and the food supply.

NOTES

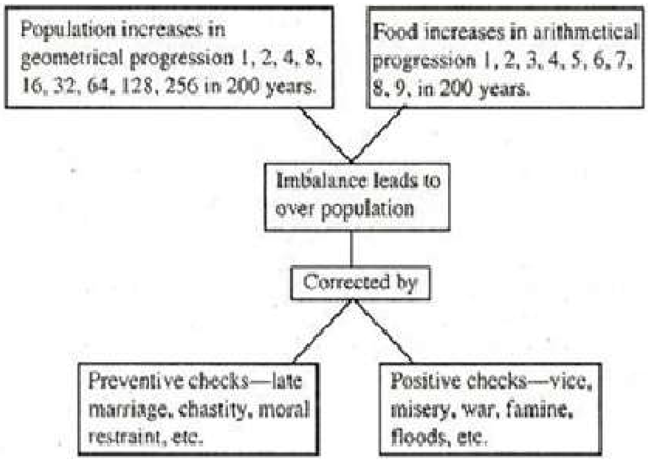


Fig. 7.4 *Mathusian Theory of Population*

Figure 7.4 shows the model of Malthusian Theory of Population.

Criticism of the Malthusian Theory

In the 19th century and early 20th century, the theory which was propounded by Malthus was severely criticized. Let us now discuss some of these criticisms.

- **The mathematical assumptions of the theory were incorrect:** During the span of twenty-five years, the theory has failed to prove the fact that population increases in a geometrical pattern whereas food supply increase in an arithmetic way. It has been observed that food supply has increased significantly in an arithmetical manner whereas population growth has failed to do so. However, this criticism is not rational as he included this mathematical formulation in his first edition of the essay, later deleted it in the second edition.
- **The theory was not universally applicable:** Malthus, observed the case of England only, when formulating the basis of his theory. He therefore, did not consider the cases of other countries where the availability of favorable agricultural conditions will increase the food supply, thus, reducing the chances of famine or poverty. Even the countries, which have a sufficient agricultural produce like Australia or Argentina exported cheap food to the countries where shortages were present. All these improvements have been made possible because of changes in the transportation facilities which has been totally ignored by Malthus. Thus, a country cannot have a fear of dying out of starvation as agricultural sector has advanced significantly.
- **Applied a Static Economic Law:** The static law which is the law of diminishing returns holds the same assumption like the Malthusian that the food supply increases in an arithmetical way. Malthus failed to predict the change in circumstances such as increase in the awareness related to the field of science or the developments in the agricultural field, which has thus, stated the law as a static. The food supply has increased in a much better

way and the contribution of green revolution has proved his theory to be absolutely insignificant.

- **Ignorance of the manpower aspect in the population growth:** In his theory, he failed to recognize the manpower aspect which contributes to one of the powerful weaknesses. He failed to remember the fact that a person being born into this world, comes with a pair of hands. This means that with the rise in population, the manpower also increases which further not only contributes in the agricultural field but also leads to improvements in industrial invention. Seligman has opined that the issue of population is not only about size but it also deals with the issues, such as proficient production and equitable distribution. Therefore, it is necessary that the population should increase over a period of a given time.
- **Population contributes to the overall wealth of the nation:** The theory propounded by Malthus is based the relationship between food supply and the population growth. It has been observed that the theory should have been formulated by linking the growth of the population with the total wealth of the country. This would have led to the optimum utilization of the theory as the argument which has been cited by a number of economists is that if a country fails to produce the sufficient food, it can export it from other countries in exchange of its own products or money. The Great Britain is a true example of such a case as it imports all of its agricultural produce from other countries such as Holland, Belgium, Denmark as it is a country which has enormous wealth and in lieu of which it meet is food production. Therefore, the foundation of his theory has been completely incorrect.
- **Population growth reduces the death rate:** The population increases because of the increase in the birth rate, is considered as a part of his theory, however the true fact is that population has increased because of the decrease in the death rate. He could not foresee the developments which have taken place in the field of medicine which has brought about a decline in certain cases of disease and at the same time has improved the living conditions. These advancements have increased the life period of a person especially in case of developing countries like India.
- **Preventive checks are not considered as a part of moral restraint:** In order to control the population growth, Malthus introduced the idea of moral restraint, celibacy, marriage at a later stage of life. However, he failed to realize that human beings could invent contraceptive and other methods too. Thus, he failed to realize the distinction between sexual desires and the desire to have his or her family.

However, despite so many drawbacks, the theory is still well acknowledged. The theory propounded by him is widely used even in the existing times. Its influence has been observed over two-third of this universe though it may not be applicable to its place of origin.

NOTES

NOTES

Neo-Malthusianism

The successors of Malthus are Neo-Malthusians who believed in the idea of family planning. They believed that Malthus was one of the foremost thinkers who opined that the over growth in population will lead to poverty and thus, the idea of birth control must be popularized.

The term 'Neo-Malthusian' theory is derived from the principles of Thomas Malthus which states that if there are limited resources in the environment, it will help in controlling the population levels and thereby, reducing the economic growth.

The concept of Neo-Malthusian however, differed from the Malthusian's concept on account of the following points:

- It aimed at popularizing various birth control methods and believed that the working class is responsible for an increase in the population.
- It encouraged the idea of racial discrimination and considered birth control as a method which promoted the idea of capitalism in the society.

They firmly believed that citizens belonging to all the strata's of the society can equally contribute towards the society. However, the main concern of the socialists was the unequal access to resources rather than the increase in the population. They believed that there are abundance of resources which should be used in such a way that it gets distributed equally. The main problem that results in unequal share of resources is the fact that the bourgeois and the propertied class are unwilling to give up their large share of resources which is under their control. Neo- Malthusianism also reinforced the ideology of private property, individualism and capitalism.

The Neo-Malthusian position was favored by the elites on the issue of overpopulation as the elites were threatened by the increase in the number of commoners. They feel scared that their resources can be over taken by the people belonging to the lower strata of the society. It can be observed that such an argument is true, as in societies today, many individuals seek to install security systems on their homes, cars and other fixed assets, these individuals are usually members of the upper class. They do this as a means to try to prevent members of the proletariat from obtaining their assets. Theft is also a common issue in most of the societies today and all of these issues are blamed on the proletariat as they are not able to provide for their large families. For instance, in third world countries today, it is usually children of the poor that drop out of school at an early age and then they that engage in either violent activities and or they beg. A prominent Neo-Malthusian, Paul Ehrlich in his book, *Population Bomb,* warned that a coming mass starvation is bound to occur because of overpopulation and pushed population control (abortion and birth control). He therefore, believed that the government should interfere in such an issue.

The Neo-Malthusians also advocated the idea of racial discrimination too, it can be observed that in the existing times, societies are not only divided on the basis of class differences but on the basis of race too. It is usually found that the blacks work under extreme circumstances whereas the white people work in extremely favorable conditions.

NOTES

Thus, it can be concluded that the Neo-Malthusian theory is derived from the Malthusian theory states that the resources of the world cannot support the population after a certain period of time. Even they believed that a kind of relationship exists between the supply of the food and the population growth, and they believed that abortion and birth control were the only means by which the population can be kept under control.

7.3 OPTIMUM THEORY

The optimum theory of population was propounded by Edwin Cannan in his book *Wealth* published in 1924 and popularised by Robbins, Dalton and Carr-Saunders. Unlike the Malthusian theory, the optimum theory does not establish relationship between population growth and food supply. On the contrary, it is concerned with the relation between the size of population and production of wealth. The Malthusian theory studies the population problem of a country by focussing on its economic conditions. Hence, the optimum theory is more realistic than the Malthusian theory of population.

The optimum population is the ideal population. This population combined with the other available resources or means of production of the country will yield the maximum returns or income per head. Robbins, Carr-Saunders and Dalton have differently defined the concept of optimum population. Robbins defines it as "the population which just makes the maximum returns possible is the optimum population or the best possible population." Carr-Saunders defines it as "that population which produces maximum economic welfare." According to Dalton, "Optimum population is that which gives the maximum income per head." After examining these views, we can conclude that Dalton's view is more scientific and realistic which we follow.

Statement

The optimum population is that ideal size of population which provides the maximum income per head. Any increase or decrease in the size of the population above or below the optimum level will reduce income per head. In keeping with the stock of natural resources, the techniques of production and the stock of capital in a country, there is a definite size of population corresponding to the highest per capita income.

If there is any deviation from this optimum-sized population, it will lead to a reduction in the per capita income. If the increase in population is followed by the increase in per capita income, it means that the country is under-populated and it can afford to increase its population till it reaches the optimum level. On the other hand, if the increase in population leads to decrease in per capita income, it means that the country is over-populated and needs a decline in population till the per capita income is maximised.

NOTES

But the optimum level is not a fixed point. It changes if there is any change in any of the factors assumed to be given. For instance, if the methods and techniques of production are improved, the output per head will rise and the optimum point will shift upward. The optimum point for the country that is today, may not be the same tomorrow if the stock of natural resources increases and the optimum point is higher than before Hence, the optimum is not a fixed but a movable point.

Check Your Progress

1. In which year did Thomas Malthus propose a methodical doctrine of population?
2. State the main principles of Malthus theory of Population.
3. Who proposed the Optimum theory of Population?

7.4 ANSWERS TO 'CHECK YOUR PROGRESS' QUESTIONS

1. In the year 1978, Thomas Malthus propose a methodical doctrine of population.
2. The main principles of Malthus theory of Population are the following:
    - There are several genetic traits in human beings which help them to reproduce at a faster rate. As a result, population increases at a statistical rate if unimpeded two-folds itself after every twenty-five years.
    - The supply of the food increases at a very slow pace because of the working of the law of diminishing returns which is based on the presumption that the supply of land is invariable.
    - Since population increases in statistical progression and the food supply in the numerical sequence, population is likely to elude food supply. Thus, an imbalance is created which directs to over populace.
    - It is necessary to keep a check on the population in order to avoid circumstances where there would be an imbalance between the population and the food supply.
3. The Optimum theory of Population was propounded by Edwin Cannan in his book *Wealth* published in 1924 and popularised by Robbbins, Dalton and Carr-Saunders.

NOTES

7.5 SUMMARY

- A renowned economist, Thomas Malthus was one of the first economist who in the year 1978, provided a methodical doctrine of the population.
- Malthus, thus became quite popular for his contribution and his famous essay, 'An essay on the principle of population', stated that the population of a particular area will change eventually as a result of famine or a disaster.
- He proposed that human population undergoes a change in each cycle while the production of the food grows at an arithmetic rate i.e. a uniform increment is added in a uniform interval of time.
- The Malthusian controversy is also popularly known as 'Malthusian check' which is a prediction that the population by forcible circumstances would return to a subsistence-level situation, if it will be more than the agriculture production.
- In the 19th century and early 20th century, the theory which was propounded by Malthus was severely criticized.
- The successors of Malthus are Neo-Malthusians who believed in the idea of family planning. They believed that Malthus was one of the foremost thinkers who opined that the over growth in population will lead to poverty and thus, the idea of birth control must be popularized.
- The Neo-Malthusian position was favored by the elites on the issue of overpopulation as the elites were threatened by the increase in the number of commoners.
- The optimum theory of population was propounded by Edwin Cannan in his book Wealth published in 1924 and popularised by Robbins, Dalton and Carr-Saunders.
- The optimum population is the ideal population which combined with the other available resources or means of production of the country will yield the maximum returns or income per head.
- The optimum population is the ideal population which combined with the other available resources or means of production of the country will yield the maximum returns or income per head.

7.6 KEY WORDS

- **Capitalism:** It is an economic and political system in which a country's trade and industry are controlled by private owners for profit, rather than by the state.
- **Proletariat:** This terms refers to working-class people regarded collectively (often used with reference to Marxism).

- **Per capita income:** It is also known as *income per* person, which *means* an *income* of the people in an economic unit such as a country or city.

NOTES

7.7 SELF-ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. Mention the checks proposed by Malthus which are helpful in reducing the growth of population.
2. Write a short note on the Optimum theory of Population.
3. What is the difference between the Malthus theory and the Optimum theory of Population?

Long-Answer Questions

1. Discuss the main proponents of Malthus theory of Population.
2. What are the criticisms raised against the Malthus theory of Population?
3. Differentiate between the Malthus and Neo-Malthus theory of Population.

7.8 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

OPTIMUM THEORIES LAW OF SUPPLY

UNIT 8 DIVISION OF LABOUR

Structure

8.0 INTRODUCTION

In the previous unit, you studied about Malthusian and Optimum theories of population.

The entrepreneur who is a business leader is always on the lookout for new ideas and puts them into effect in fostering economic growth and development. Entrepreneurship provides one of the most significant contributions to the economic development of a country. According to economist, Joseph Alois Schumpeter (1883-1950), 'entrepreneurs are not necessarily motivated by profit but regard it as a standard for measuring achievement or success.' In this unit, you will study about the significance of capital formation in the economy of a country as well as the essential role and functions of an entrepreneur.

8.1 OBJECTIVES

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

- Define capital formation
- List the functions of an entrepreneur
- State the important qualities of an entrepreneur

8.2 CAPITAL FORMATION

NOTES

Capital formation is addition to productive capacity of the economy. It is also known as investment in national accounting. Terms like capital formation, or additions to physical stock, or investment are synonymous in economic parlance. Capital formation forms the backbone of an economy. India has steadily improved its capital stock since independence. We try to explain the theory and trends of investments in India.

- **Gross Capital Formation Steadily Improved:** Gross capital formation in the 1950s was as low as 7.8% of GDP. This improved in the following decades with the Five-Year Plans constantly focusing on improving physical stock.
- **Measures of Capital Formation:** Gross capital formation, gross fixed capital formation and gross domestic capital formation are few other definitions to study capital formation. In India, the Central Statistical Organisation provides data on capital formation by institutions and sectors.
- **Savings:** Savings provide necessary means for investment in the economy. Savings rate has improved post-liberalisation to 36% of GDP, which points to rise in economic activity and national income in India. The country currently is among the high-saving economies of the world. However, India's saving rate is still far lower compared with China's, which is around 50% of GDP.
- **Contribution To Gross Capital Formation:** Currently, private sector leads in investments in the economy at 37% of total investments. Public investments are around 26% and household investments account for 32%. Public share in investments has fallen over time, which indicates the government's inability to stimulate investments in the country. The way forward is to boost investments through private-public partnerships that can help India solve its infrastructure problems.
- **Private Investments:** With liberalisation of the economy and a favourable business environment, private sector now leads in investments in the economy. Private investments are largely funded by household savings. Over the years, private savings have also improved, which have further aided private sector capital formation. The private sector, being the biggest contributor to gross capital formation, now has a crucial role to play in leading India's economic growth.
- **Public Investments:** Public investments have declined as government lacks enough resources to boost investment in a big way. Also, the government's propensity for dissavings, because of poor expenditure management, has left it with fewer resources to fund investments in the country. As government lacks enough muscle to raise investments further, the private sector and foreign fund inflows have become critical for increasing investment levels in the country.

- **Household Investments:** Household investments account for 32% of total investments. Household includes individuals, non-corporate business bodies, private and charitable institutions such as educational and religious organisations. Therefore, investments by these bodies in terms of physical capacity creations such as on land, buildings, factories, etc, are termed as household capital formation. Household investments are funded by rising household savings. Household savings accounts for nearly 70% of total savings and are the chief source of investments for both private and public investments.

Key Aspects

- **Measurement:** Investment activity in an economy refers to addition to physical capital stock. It is measured by gross capital formation. In India, the Central Statistical Organisation provides data on different components of gross capital formation, broadly defined as gross fixed capital formation (GFCF) plus change in stocks. Gross fixed capital formation provides a picture of gross value of goods added to fixed domestic capital stock during a year. GFCF includes plant, machinery, equipment and improvements to land. Change in stocks provides value of inventory and work in progress.
- **Transition Since Independence:** Capital formation in the 1950s was low, but it improved in the following decades. Capital formation improved more than four times to a high of 36% of GDP in 2010. With investments rising over the years, India's economic growth rate reached a high of 9% before the 2008 global crisis. Although all economic growth cannot be attributed to investments, importance of capital formation remains paramount in economic development.
- **Key To Long-Term Growth:** Importance has been laid on capital formation in economic development as it can lead to sustainable long-term growth. Although capital formation has increased in India since independence to touch 36% of GDP, it still remains below rates achieved in high-growth economies, such as China. Investment levels in China have risen to a high of 50% of GDP, which also underlines the high economic growth it has been able to sustain for the past 25 years.
- **Constraints To Investments:** Another aspect of capital formation that needs mention is constraints to finance it. Investments are typically funded from domestic savings (although foreign capital flows also contribute towards growth in investments). So, inadequate growth in savings rate can be a constraint for investments. Savings rate in China has increased to around 50% of GDP, which makes it possible for the Chinese economy to sustain high levels of investment. China also attracts huge foreign direct investment every year.
- **Only Investment Not Enough:** It is an oversimplification to say high investments can result in long-term high economic growth. There are many

NOTES

economies that initially showed a rapid rise in economic growth through improvement in capital formation, but ultimately slowed down. Examples of Soviet Union in the 1960s and the East Asian experience in the 1980s and 1990s provide evidence that high rates of investments aren't the only way forward. Investments have to be followed with improvement in productivity. Excess of capital without productivity results in slower growth and also lower returns on investments. Simply put, mobilisation of labour and capital are not sufficient for sustainable high long-term growth. Incremental capital output ratio (ICOR), a critical ratio that measures the amount of incremental capital needed to produce one incremental level of output, is a key measure of capital productivity. ICOR levels in India have remained unchanged at 4.5 in recent years. A lower ICOR is critical to achieve a high rate of growth with a given level of capital formation.

- **Perspectives For India:** Many studies on investment behaviour in India have pointed out that economic growth, rising incomes and economic liberalisation have led to a rise in private investments in India. Public sector investments have lost share in recent years, which is manifest in infrastructure deficit and other bottlenecks to economic development. Private sector investments hold the key for economic growth and should be encouraged with conducive business and policy environment.

8.3 FUNCTIONS OF AN ENTREPRENEUR

The basic characteristics of an entrepreneur are:

- **Desire for responsibility:** This is a fundamental characteristic of an entrepreneur who feels a deep sense of personal responsibility for the result of the projects that he starts. He prefers to have control over the resources, which he uses for achieving self-determined goals.
- **Preference for moderate risk:** This means that entrepreneurs do not take wild risks; instead, they take calculated risks. In other words, successful entrepreneurs are not as much risk takers as they are risk eliminators, removing obstacles that come in the way of the successful launch of their ventures.
- **Confidence in their ability to succeed:** The National Federation of Independent Businesses (NFIB) found that business owners rated the success of their companies as quite high—an average of 7.3 on a scale of 1 (*total failure*) to 10 (*extreme success*). This high level of optimism may explain why some of the most successful entrepreneurs have failed in business—often more than once—before finally succeeding. Hence, entrepreneurs are typically confident about succeeding.
- **High level of energy:** Entrepreneurs are more energetic than the average person, which sometimes proves to be vital in launching a start-up company.

NOTES

That energy may be a critical factor given the incredible effort required to launch a start-up company. Hard work and long working hours are common, and the pace can be gruelling. For example, during the nascent phase of Infosys, Narayana Murthy and a few members of his team worked tirelessly to build a company that would fulfill all the software needs and wants of customers.

- **Desire for immediate feedback:** Most entrepreneurs feel a sense of pride in being entrepreneurs. They take pleasure in facing the challenges that go with the running of businesses and constantly invite feedback as they are eager to find out how they are performing. They use this feedback to overcome their negative points by working on them.
- **Future orientation:** An entrepreneur has a definite sense of probing for opportunities. He looks beyond the present and concerns himself with considering what can be done in the future rather than brooding on what was done in the past. Not satisfied to sit back and revel in their success, real entrepreneurs stay focused on the future. An example would be 'home shopping' via television channels—a medium that the Indian audience is already familiar with. For some time now, Indian viewers have watched products by Asian Sky Shop being aired on different channels. However, what is new is the notion of a 24-hour channel that completely caters to home shopping. First on the block is HomeShop18, a Network 18 venture, which was launched in April 2008. Growing at a rate of 30 per cent per month, the company is targeting 'a run-rate breakeven' in the first year of its launch. With more channels wanting to grab a share in this market, one can expect more action in this area over the next year, a clear indication of the growing retail space in India. In an interview with *Brand Line* Sundeep Malhotra, CEO, HomeShop18 said, 'Retail trends are changing drastically. However, for e-commerce to evolve in the consumer retail space, consumers at large have to be aware of the benefits of the medium.'
- **Skill at organizing:** An entrepreneur knows how to bring together the right people for accomplishing a task. He effectively combines jobs and people and transforms his vision into reality.
- **Value of achievement over money:** One of the most widely held misconceptions about entrepreneurship is that it is the desire for money that drives entrepreneurs. In fact, entrepreneurs are chiefly motivated by the urge to achieve something; money merely serves as a symbol of achievement, a means of 'keeping score' of the accomplishments.
- **Flexibility:** Entrepreneurs are capable of adapting themselves to the varying demands of businesses and customers. Rigidity can lead to failure in today's swiftly altering global economy. For example, IBM's strategy of adopting email advertising in early 1990 evinced a good response and increased its microprocessors business.

NOTES

- **Tolerance for ambiguity:** Entrepreneur expert Amar Bhide says that entrepreneurs exhibit 'a willingness to jump into things when it's hard to even imagine what the possible set of outcomes will be'. Entrepreneurs have a tendency to be highly tolerant to forever shifting, ambiguous circumstances, which is often the environment they frequently operate in. It is essential for them to be able to deal with uncertainty as these business builders have to continuously make decisions using new and at times even conflicting information gleaned from a variety of unknown sources.
- **High degree of commitment:** Entrepreneurship is hard work, and launching an organization successfully requires total commitment from an entrepreneur. Business founders often immerse themselves completely in their companies. Most entrepreneurs have to overcome seemingly insurmountable barriers to launch an organization and to keep it growing. That requires commitment.

8.3.1 Functions of an Entrepreneur

An entrepreneur is expected to perform the following functions:

1. **Assumption of risk:** The entrepreneur assumes all possible risks of business which emerge with changes in the tastes of consumers, techniques of production and new inventions. Such risks are not insurable and the entrepreneur has to bear the loss, if any. Thus, risk-assumption and risk-bearing remain the most important functions of an entrepreneur, which he tries to reduce with his initiative, knowledge and skill and good decisions.
2. **Business decisions:** The entrepreneur has to decide the nature and type of business to undertake the nature and type of goods that must be produced or services that must be provided to customers. He enters the particular industry which offers him the best prospects and produces whatever commodities he thinks will pay him the most and employs those methods of production which seem to him to be the most profitable. He effects suitable changes in the size of the business, its location, techniques of production and does everything that is needed for the development of his business.
3. **Managerial functions:** There are different types of managerial functions that an entrepreneur has to perform and these are based on the size and activities of an enterprise. The managerial functions include formulation of production plans, raising of finance, dealing with suppliers for procurement of raw materials and other materials, providing production facilities, organizing sales, and conducting and so on. Administrative functions such as manpower planning, selection, recruitment, etc.

Generally, an entrepreneur performs many useful functions for the development of society and to satisfy the needs of fellow citizens. The entrepreneur can identify opportunities to start a business either as a manufacturer or as a distributor, for entrepreneurship exists in every field of economic endeavour.

Manufacturing activities require a relatively high capital investment and more entrepreneurial abilities than distribution activities. An entrepreneur has strong motivation and desire to achieve success by undertaking a venture and bearing risk for earning profit.

8.3.2 Qualities and Competencies of an Entrepreneur

Essential qualities of entrepreneurs are as follows:

1. **Success and achievement:** Entrepreneurs are determined to achieve high goals in business. This achievement motive gives them the strength to surmount obstacles, suppress anxieties, overcome misfortune and desire expedients to run a successful business.
2. **Risk-bearing attitude:** Rather than gambling or avoiding risks, entrepreneurs take moderate strategic and financial risks.
3. **Opportunity seeking:** An entrepreneur is always seeking new opportunities. He seizes opportunities and converts them into realistic achievable goals by adopting a constructive approach.
4. **Perseverance:** Entrepreneurs make extreme efforts and work hard till their goals are accomplished. They are undeterred by the uncertainties, risks and difficulties that come in the way of achievement of their goal. So obstacles do not stop an entrepreneur from reaching his goal.
5. **Facing uncertainty**: Achievement-oriented people tend to tackle unfamiliar but interesting situations. They go ahead with solutions for the problems even without guidelines. With their analytical ability they investigate a situation and act accordingly.
6. **Feedback**: Entrepreneurs want prompt feedback of their performance so that they can correct themselves without delay.
7. **Independence:** Entrepreneurs like to be their own master and want to be responsible for their own decisions. An entrepreneur is a job giver and not a job seeker.
8. **Flexibility:** Entrepreneurs make decisions based on the prevailing situations. Successful entrepreneurs do not hesitate in revising their decisions. An entrepreneur is a person with an open mind and is not rigid in his ideas.
9. **Planning:** Entrepreneurs frame realistic business plans and follow them rigorously to achieve their objectives in a stipulated time limit.
10. **Self-confidence:** Entrepreneurs direct their abilities towards the accomplishment of goals with the help of their strengths.
11. **Motivation:** Entrepreneurs influence and initiate people and make them think in their way and act accordingly.
12. **Composure:** Entrepreneurs take many decisions that may involve lot of physical and emotional stress. They keep their cool under a lot of tension while decision making.

8.3.3 Skills Required for Entrepreneurs

NOTES

An entrepreneur needs to have many types of skills to enables him or her to deal with complex situations and different types of social and economic scenarios. The skills required by entrepreneurs can be classified into the following categories:

- **Technical skills:** Technical skills like proficiency in communication, writing, monitoring environment, etc. are assets that every entrepreneur showed bank on.
- **Business management skills:** Business management skills are those which are required for starting, developing and managing an enterprise. Skills like planning, decision, making, goal setting, negotiating, marketing controlling, and managing growth, etc., are some of the business management skills that all entrepreneurs need to be successful.
- **Personal entrepreneurial skills:** Personal entrepreneurial skills are one of the essential skills required by an entrepreneur. These skills differentiate the manager from the entrepreneur. Inner control, risk-taking capabilities, change, orientation, being persistent, being a leader and a visionary are essential personal skills that a successful entrepreneur must have.

Check Your Progress

1. Define capital formation.
2. List the important functions of an entrepreneur.

8.4 ANSWERS TO 'CHECK YOUR PROGRESS'

1. Capital formation is a term used to describe the net capital accumulation during an accounting period for a particular country.
2. The important functions of an entrepreneur are the following:
    - Business decisions
    - Managerial functions
    - Assumption of risk

8.5 SUMMARY

- Capital formation is addition to productive capacity of the economy. It is also known as investment in national accounting.
- Savings provide necessary means for investment in the economy.
- With liberalisation of the economy and a favourable business environment, private sector now leads in investments in the economy.

- Incremental capital output ratio (ICOR), a critical ratio that measures the amount of incremental capital needed to produce one incremental level of output, is a key measure of capital productivity.
- Entrepreneurship is hard work, and launching an organization successfully requires total commitment from an entrepreneur.
- Generally, an entrepreneur performs many useful functions for the development of society and to satisfy the needs of fellow citizens.
- Entrepreneurs want prompt feedback of their performance so that they can correct themselves without delay.
- An entrepreneur needs to have many types of skills to enables him or her to deal with complex situations and different types of social and economic scenarios.
- Manufacturing activities require a relatively high capital investment and more entrepreneurial abilities than distribution activities. An entrepreneur has strong motivation and desire to achieve success by undertaking a venture and bearing risk for earning profit.

8.6 KEY WORDS

- **Physical Capital:** In economics, it is a factor of production.
- **Incremental Capital Output Ratio (ICOR):** It is a critical ratio which measures the amount of incremental capital needed to produce one incremental level of output.
- **Entrepreneurship:** It is both the study of how new businesses are created as well as the actual process of starting a new business.

8.7 SELF-ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. List the basic characteristics of an entrepreneur.
2. Mention the important skills required for entrepreneurs.
3. What are the measures of capital formation?

Long-Answer Questions

1. Discuss the significance of capital formation in the economy of a country.
2. Why is ‘risk-bearing attitude’ considered an essential quality of an entrepreneur? Explain with an example.
3. How does an entrepreneur contribute to the development of the society?

8.8 FURTHER READINGS

NOTES

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 9 INTERNAL AND EXTERNAL ECONOMIES

NOTES

Structure

9.0 INTRODUCTION

In the previous unit, you studied about capital formation and the qualities and functions of an entrepreneur. In this unit, you will be introduced to various cost concepts namely, business cost, actual cost, opportunity cost, explicit and implicit costs, fixed and variable costs and so forth. Moreover, you will also study about internal economies and external economies. Internal economies are those which arise from the expansion of the plant-size of the firm. External diseconomies are the disadvantages that originate outside the firm.

9.1 OBJECTIVES

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

- Explain cost of production
- Define average and marginal cost
- Differentiate between fixed costs and variable costs

9.2 ECONOMIES OF SCALE

NOTES

In this section, we will study economies of scale. Economies of scale are classified under:

1. Internal or real economies; and
2. External economies or pecuniary economies.

9.2.1 Internal or Real Economies

Internal economies are those which arise from the expansion of the plant-size of the firm. The economies are internal in the sense that economies are internalized to the expanding firms—not available to non-expanding firms. Internal economies can be classified under the following categories (a) Economies in production; (b) Economies in marketing; (c) Managerial economies; and (d) Economies in transport and storage.

A. Economies in Production. The economies in production arise from two sources: (i) technological advantages, and (ii) advantages of division of labour and specialisation.

(i) **Technological advantages.** Large scale production provides opportunity to avail the advantages of technological advances. The modern technology is highly specialised. 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 of planned production, the firm may not find it economical to have all the plants. And hence, it would not be in a position to have the full advantage of a composite technology. But, when scale of production expands and firms hire more of capital and labour, their total output increases more than proportionate increase in inputs, till the optimum size of the firm is reached. Beyond that level of output the technical advantages diminish and lead to diminishing returns to scale.

(ii) **Advantages of division of labour and specialisation.** When 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 employ labour with specialised qualification and skill and to place them in the process of production where they are best suited. This is known as division of labour. Division of labour leads to professional specialisation. It increases efficiency. Besides, specialised workers develop more efficient tools and techniques and gain speed of work. These advantages of division of labour improve productivity

NOTES

of labour—per unit of cost and time. But there is always a limit to which division of labour is possible. *This limit is given by the level of technology.* Beyond this limit the advantage of division of labour may not exist. Further expansion of the firm may therefore lead to decreasing returns to scale.

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

The economies in marketing 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-scale 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 firm's production. The firm also gains on large scale distribution through better utilisation of 'sales force, distribution of samples, etc.' This kind of economy however does not directly affect the production conditions.

C. Managerial Economies. Managerial economies arise from (a) specialisation of management, and (b) division of managerial functions. For a large size firm it becomes possible for the management to divide itself into specialised departments under specialised personnel, such as production manager, sales manager, personnel manager, *HRD* managers, labour officers, etc. This increases efficiency of management at all levels because of decentralisation of decision-making. Large scale firms have the opportunity to use advanced techniques of communication, telephones and telex machines, computers, and their own means of transport. These all lead to quick decision- making help in saving the valuable time of the management, and thereby improve the managerial efficiency.

D. Economies in Transport and Storage. Economies in transportation and storage costs arise from fuller utilisation 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 cost of transportation compared to the market rate, at least to the extent of profit margin of the transport companies. Besides, own transport facility prevents the delays in transporting goods. Some large scale firms have their own railway track from the nearest railway point to the factory, and thereby they reduce the cost of transporting goods in and out. For example, Bombay Port Trust has its own railway tracks, 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.

9.2.2 External or Pecuniary Economies

External or pecuniary economies arise to large size firms from the discounts available to them due to (i) large scale purchases of raw materials; (ii) large scale acquisition of external finance at lower rate of interest, particularly from the commercial banks; (iii) lower advertising rate charged by the advertisement media for large scale advertising; (iv) concessional rates charged by the transport companies on bulk transport of goods; and (v) lower wage rates if a large scale firm uses its monopsony power in certain kinds of specialised labour market. The economies of scale discussed above are based on experience of the large scale firms. These economies of scale, however, may not be necessarily available to all the large scale firms. On the contrary, one may find well-managed small scale firms more efficient, producing goods at a relatively lower cost.

9.3 COST OF PRODUCTION: AVERAGE AND MARGINAL COST

Let us now study the various concepts related to cost.

9.3.1 Actual Cost and Opportunity Cost

Actual costs are those which are actually incurred by the firm in payment for labour, material, plant, building, machinery, equipments, travelling and transport, etc. The total money expenses recorded in the books of accounts are, for all practical purposes, the actual costs. Actual cost concept comes under the accounting cost concept.

Opportunity cost is another *fundamental* cost concept used in business decisions. The concept of 'opportunity cost' is related to scarcity concept. The opportunity cost is the return expected from the second best use of the resources, which is foregone for availing the gains from the best use of the resources. For example, suppose, a businessman with his limited resources can buy either a printing machine or a lathe. From the printing machine he expects an annual income of ` 20,000, and from the lathe he expects ` 15,000. The rational businessman will obviously invest his money in the printing machine and forego the expected income from the lathe. The opportunity cost of his income from printing machine is the expected income from the lathe, i.e., ` 15,000. The opportunity cost arises because of the foregone opportunity. In assessing the opportunity costs, both *explicit* and *implicit* costs are taken into account.

Associated with the concept of opportunity cost is the concept of *economic rent* or *economic profit*. In our example, economic rent of the printing machine is the excess of its earning over the income from the lathe. Given the returns from printing machine and lathe, economic rent = ` 20,000 – ` 15,000 = ` 5,000. The business implication of this concept is that investing in 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, the choice of the best investment avenue will be a problem.

9.3.2 Business Cost and Full Cost

Business costs include all the expenses which are incurred in carrying out a business. The concept of business cost is similar to the actual or real cost. Business costs "include all the payments and contractual obligations made by the firm together with the book cost of depreciation on plant and equipment." Both these concepts are used in calculating the profits and losses in the business in filing returns for income-tax, and for other legal purposes.

The concept of **full cost** includes two other costs: *opportunity cost* and *normal profit*. Opportunity cost, as noted above, includes the expected earning from the next best use of the resources or the market rate of interest on the total money capital, and also the value or entrepreneurs own services which are not charged in the current business. Normal profit is a necessary minimum earning, in addition to opportunity cost, which a firm must get to remain in its present occupation.

9.3.3 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, salaries, utilities, interest, rent, purchase of materials, licence fee, insurance premium and depreciation charges are the examples of explicit costs. These costs involve cash payment and are clearly reflected by the normal accounting practices. In contrast with these 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. Implicit costs may be defined as the earning of owner's resources employed in their best alternative uses. For example, suppose an entrepreneur does not utilise his services in his own business and works as a manager in some other firm on a salary basis. If he joins his own business, he foregoes his salary as manager. This loss of salary which is opportunity cost of his services utilised in his own firm becomes an implicit cost of his own business. It is implicit because the income foregone by the entrepreneur is not charged as the explicit cost of his own business. The implicit cost includes implicit wages, implicit rent, implicit interest etc. Although implicit costs are not taken into account while calculating the loss or gains of the business, these costs do figure in business decisions.

9.3.4 Total, Average and Marginal Costs

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

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

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

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

9.3.5 Fixed and Variable Costs

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

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

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

9.3.6 Short-run and Long-run Costs

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

9.3.7 Private and Social Costs

We have so far discussed the cost concepts that are related to the functioning of the firm as a production unit, and are used in the cost-benefit analysis of the business decisions. There are however certain other costs which arise due to functioning of the firm but do not normally figure in the business decisions nor are such costs explicitly paid by the firms. Some such costs are paid by the society. Thus, the total cost generated by the firm's decision may be divided into two categories: (a) those paid out or provided for by the firms; (b) those not paid by the firms, including use of resources freely available *plus* the disutility created in the process of production. The costs of the category (a) are known as **private costs**, and costs of category (b) are known as **external** or **social costs**. For instance, Private firms, situated closer Yamuna river discharge their wastes into the Yamuna river causing water-pollution; mills and factories located in a city cause air-pollution by emitting smoke; plying cars, buses, trucks, etc. causes both air and noise pollution. Such pollutions cause tremendous health hazards which involve cost to the society as a whole. Such costs do not figure in the cost structure of the firms and hence are termed **external costs** from the firm's point of view, and *social cost* from society's point of view.

NOTES

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 internalised in the sense that "the firm must compensate the resource owner in order to acquire the right to use the resource." It is only the internalised cost that is incorporated in the firm's total cost of production.

Social cost, on the other hand, implies the cost which a society bears on account of production of a commodity. Social cost includes both *private cost* and the *external cost*. External cost includes (a) the cost of 'resources for which the firm is not compelled to pay a price,' e.g., 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' caused by air, water, and noise pollution, etc. The cost of category (*b*) is generally assumed to be equal to 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. But private and public expenditure serve only as an indicator of trends in 'public disutility', it does not give the exact measure of the public disutility.

9.4 COST FUNCTIONS AND COST CURVES

In this section, we introduce the different types of cost functions of empirical nature and illustrate the derivation of the cost curves.

As mentioned above, the shape of the cost curves depends on the nature of the cost functions. Cost functions are derived from actual cost data of the firms. Given the cost data, cost functions may take a variety of forms, yielding different kinds of cost curves. The cost curves produced by **linear**, **quadratic** and **cubic cost functions** have been illustrated subsequently in this unit.

9.9.1 Linear Cost Function

NOTES

A linear cost function takes the following form

$$TC = a + bQ \qquad ....(9.1)$$

where TC = total cost, Q = output, a = fixed cost and b is a constant.

Given the cost function (Eq. 9.l), AC and MC can be obtained as follows:

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

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

Note that since *'b'* is a constant, MC remains constant throughout in case of a linear cost function. The cost curves (TC, TVC and TFC) are graphed in Figs. 9.1 and 9.2 assuming an actual cost function given as

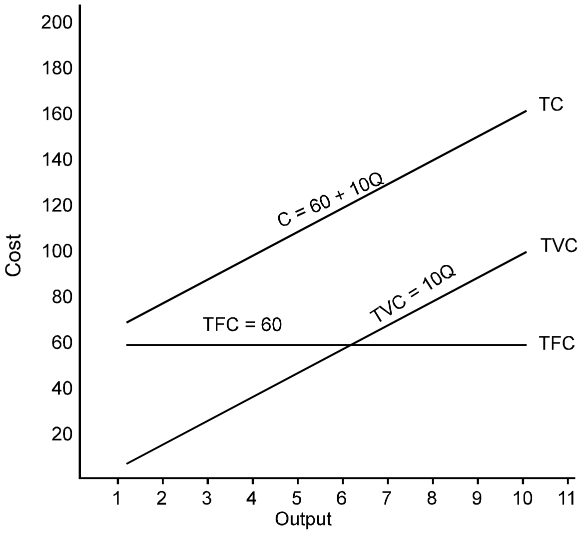


Fig. 9.1 *Linear Cost Function*

$$TC = 60 + 10Q$$

Given this function,

$$\text{AC} = 60/Q + 10$$

and $$MC = 10$$

Figure 9.1 shows the behaviour of total cost curves. The straight horizontal line shows TFC and line marked $TVC = 10Q$ shows the movement in TVC. The total cost function is shown by $TC = 60 + 10Q$.

NOTES

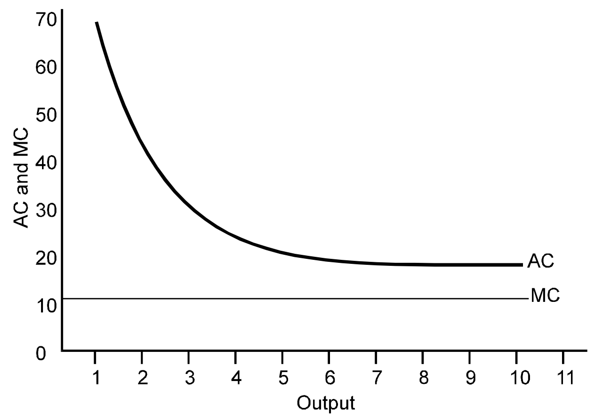


Fig. 9.2 *AC and MC Curves Derived from Linear Cost Function*

More important is to notice the behaviour of *AC* and *MC* curves in Fig. 9.2. Note that, in case of a linear cost function, $MC = AVC$ and it remains constant, while *AC* continues to decline with the increase in output. This is so simply because of the logic of the linear cost function.

9.4.2 Quadratic Cost Function

A quardratic cost function is of the following form:

$$TC = a + bQ + Q^2 \qquad \text{...(9.2)}$$

where a and b are constants and TC and Q are total cost and total output, respectively. Given the cost function (9.2), *AC* and *MC* can be obtained as follows:

$$AC = \frac{TC}{Q}$$

$$= \frac{a + bQ + Q^2}{Q} \qquad \text{...(9.3)}$$

$$= \frac{a}{Q} + b + Q$$

$$\text{and } MC = \frac{\partial TC}{\partial Q} = b + 2Q \qquad \text{...(9.4)}$$

Let the actual (or estimated) cost function be given as

$$TC = 50 + 5Q + Q^2 \qquad \text{...(9.5)}$$

Given the cost function (9.5),

$$AC = \frac{50}{Q} + Q + 5$$

$$\text{and } MC = \frac{\partial TC}{\partial Q} = 5 + 2Q$$

NOTES

The cost curves that emerge from the cost function (9.2) are graphed in Fig. 9.3(a) and (b). As shown in panel (a), while fixed cost remains constant at 50, *TVC* is increasing at an increasing rate. The rising *TVC* sets the trends in the total cost (*TC*). Panel (b) shows the behaviour of *AC*, *MC* and *AVC* in a **quadratic cost function**. Note that *MC* and *AVC* are rising at a constant rate whereas *AC* first declines and then tends to increase.

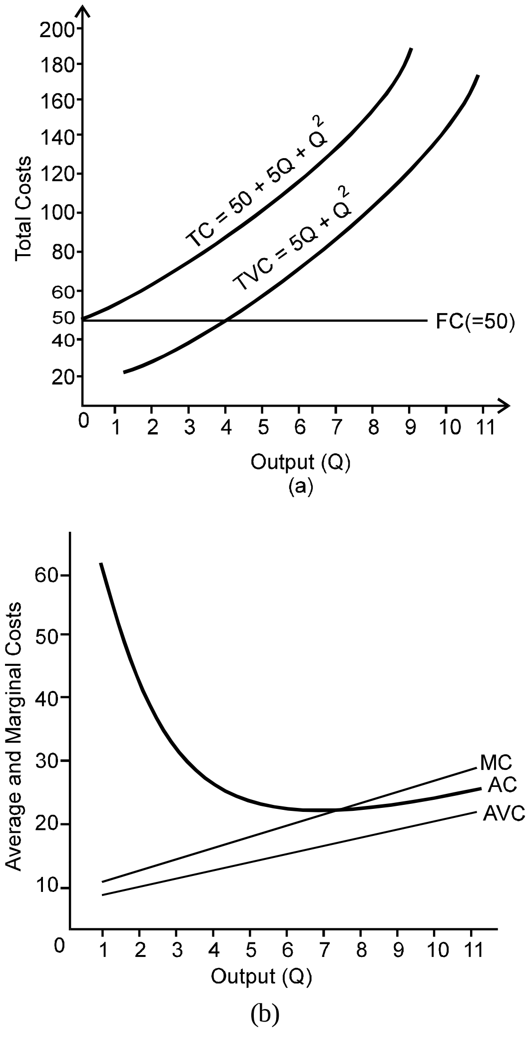


Fig. 9.3 *Cost Curves Derived from a Quadratic Cost Function*

9.4.3 Cubic Cost Function

A cubic cost function is of the following form:

$$TC = a + bQ - cQ^2 + Q^3 \qquad ...(9.6)$$

where a, b and c are the parametric constants. From the cost function (9.6), *AC* and *MC* can be derived as follows:

$$AC = \frac{TC}{Q}$$

$$= \frac{a + bQ - cQ^2 + Q^3}{Q}$$

$$= \frac{a}{Q} + b - cQ + Q^2$$

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

$$= b - 2cQ + 3Q^2$$

NOTES

Assuming an estimated cubic cost function of the form

$$TC = 100 + 55Q - 10Q^2 + Q^3$$

we can generate the cost data as given in Table 9.1.

Table 9.1 *Cost Data Obtained form Cubic Cost Function*

Q	*C*	*FC*	*TVC*	*AC*	*MC*	*AVC*
0	100	100	—	—	—	—
1	146	100	46	146	46	46
2	178	100	78	89	32	39
3	202	100	102	67	24	34
4	224	100	124	56	22	31
5	250	100	150	50	26	30
6	286	100	186	48	36	31
7	338	100	238	48	52	34
8	412	100	312	52	74	39
9	514	100	414	57	102	46
10	650	100	550	65	136	55

Check Your Progress

1. What are actual costs?
2. Define explicit costs.
3. Mention the various categories of internal economies.

9.5 ANSWERS TO 'CHECK YOUR PROGRESS' QUESTIONS

1. Actual costs ae those which are actually incurred by the firm in payment for labour, material, plant, building, machinery, travelling, transport and so forth.
2. Explicit costs are those which fall under actual or business costs entered in the books of accounts. The payment on account of wages, salaries, utilities, interest, rent, purchase of materials, licence fee, insurance premium and depreciation charges are the examples of explicit costs.

3. The various categories of internal economies are (a) Economies in production (b) Economies in marketing (c) Managerial economies (d) Economies in transport and storage.

NOTES

9.6 SUMMARY

- Actual costs are those which are actually incurred by the firm in payment for labour, material, plant, building, machinery, equipments, travelling and transport, etc. The total money expenses recorded in the books of accounts are, for all practical purposes, the actual costs.
- Business costs include all the expenses which are incurred in carrying out a business. The concept of business cost is similar to the actual or real cost.
- Total cost represents the value of the total resources used in the production of goods and services. It refers to the total outlays of money expenditure, both explicit and implicit on the resources used to produce a given output.
- Marginal cost is the addition to the total cost on account of producing one additional unit of the product.
- 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.
- Internal economies are those which arise from the expansion of the plant-size of the firm. The economies are internal in the sense that economies are internalized to the expanding firms—not available to non-expanding firms.
- External or pecuniary economies arise to large size firms from the discounts available to them due to (i) large scale purchases of raw materials; (ii) large scale acquisition of external finance at lower rate of interest, particularly from the commercial banks; (iii) lower advertising rate charged by the advertisement media for large scale advertising; (iv) concessional rates charged by the transport companies on bulk transport of goods; and (v) lower wage rates if a large scale firm uses its monopsony power in certain kinds of specialised labour market.

9.7 KEY WORDS

- **Business costs:** These include all the expenses which are incurred in carrying out a business.
- **Internal economies:** These are those which arise from the expansion of the plant-size of the firm. The economies are internal in the sense that economies are internalized to the expanding firms—not available to non-expanding firms.
- **Implicit costs:** These are defined as the earning of owner's resources employed in their best alternative uses.

9.8 SELF-ASSESSMENT QUESTIONS AND EXERCISES

NOTES

Short-Answer Questions

1. Define marginal cost.
2. Write short notes on the following:
   (a) Private cost (b) Social cost
3. Briefly mention external economies.

Long-Answer Questions

1. Differentiate between fixed costs and variable costs.
2. Graphically represent the cost curves produced by linear, quadratic and cubic cost functions.
3. Give examples to differentiate between actual costs and opportunity costs.

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.

BLOCK - IV
TIME MANAGEMENT AND THEORIES OF WAGES AND DISTRIBUTION

NOTES

UNIT 10 EXCHANGE (THEORY OF PRICING)

Structure

10.0 INTRODUCTION

Some empiricists have, however, produced the evidence, inadequate though, to prove that the firms follow a pricing rule other than the one suggested by the marginality rules. Besides, in a complex business world, business firms follow a variety of pricing rules and methods depending on the conditions faced by them. In this unit, you will study about the theory of pricing, market average revenue, marginal revenue and law of supply.

10.1 OBJECTIVES

NOTES

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

- Discuss the theory of pricing
- Differentiate between market average revenue and marginal revenue
- State the law of supply

10.2 THEORY OF PRICING: AN OVERVIEW

Let us now study the various concepts related to pricing.

10.2.1 Cost-plus Pricing

Cost-plus pricing is also known as **'mark-up pricing'**, **'average cost pricing'** and **'full cost pricing'**. The cost-plus pricing is the most common method of pricing used by the manufacturing firms. The general practice under this method is to add a 'fair'[1] percentage of profit margin to the average variable cost (*AVC*). The formula for setting the price is given as

$$P = AVC + AVC\,(m) \qquad \ldots(10.1)$$

where *AVC* = average variable cost, and *m* = mark-up percentage, and *AVC*(*m*) = gross profit margin (*GPM*).

The mark-up percentage (*m*) is fixed so as to cover average fixed cost (*AFC*) and a net profit margin (*NPM*). Thus,

$$AVC\,(m) = AFC + NPM \qquad \ldots(10.2)$$

The general procedure followed by the firms for arriving at *AVC* and price fixation may be summarized as follows.

The **first step** in price fixation is to estimate the average variable cost. For this, the firm has to ascertain the volume of its output for a given period of time, usually one accounting or fiscal year. To ascertain the output, the firm uses figures of its 'planned' or 'budgeted' output or takes into account its normal level of production. If the firm is in a position to compute its optimum level of output or the capacity output, the same is used as *standard output* in computing the average cost.

The **next step** is to compute the total variable cost (*TVC*) of the 'standard output.' The *TVC* includes direct cost, i.e., the cost of labour and raw material, and other variable costs e.g., electricity and transportation cost, etc. These costs added together give the *total variable cost*. The 'Average Variable Cost' (*AVC*) is then obtained by dividing the total variable cost (*TVC*) by the 'standard output' (*Q*), i.e.,

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

NOTES

After AVC is obtained, a 'mark-up' of some percentage of AVC is added to it as profit margin and the price is fixed. While determining the mark-up, firms always take into account 'what the market will bear' and the competition in the market.

Mark-up Pricing and Marginalist Rule Compared

The mark-up pricing method appears to be a 'rule of thumb' totally different from the marginalist rule of pricing. Fritz Machlup has, however, shown that mark-up pricing is not incompatible with the marginalist rule of pricing. Rather, it is very much compatible with marginalist rule of pricing. According to Machlup, when we look into the logic of mark-up pricing, it appears quite similar to the marginalist rule of pricing. Let us now compare the two rules of pricing.

Recall that, according to the marginalist rule, profit is maximum at the level of output where $MC = MR$ and that the mark-up pricing method is given by

$$P = AVC + AVC\ (m)$$

or

$$P = AVC\ (1 + m) \qquad \text{…(10.3)}$$

Let us now show that the mark-up pricing ultimately converges with the marginalist rule of pricing, at least under constant cost conditions.

Recall that profit is maximum at the level of output at which

$$MC = MR$$

and

$$MR = P\left(1 - \frac{1}{e}\right) \qquad \text{…(10.4)}$$

or

$$MR = P\left(\frac{e-1}{e}\right) \qquad \text{…(10.5)}$$

where e is price elasticity of demand.

By substituting Eq. (10.5) in Eq. (10.4), we may restate the necessary condition of profit maximization as

$$MC = MR = P\left(\frac{e-1}{e}\right) \qquad \text{…(10.6)}$$

If MC is constant, then $MC = AVC$. By substituting AVC for MC, Eq. (10.6) may be rewritten (ignoring MR) as,

$$AVC = P\left(\frac{e-1}{e}\right) \qquad \text{…(10.7)}$$

By rearranging the terms in Eq. (10.7), we get

$$P = AVC \div \left(\frac{e-1}{e}\right)$$

or

$$P = AVC\ \left(\frac{e}{e-1}\right) \qquad \text{…(10.8)}$$

Now, consider Eq. (10.6). If $MC > 0$, then $P\left(\frac{e-1}{e}\right)$ must be greater than 0. For $P\left(\frac{e-1}{e}\right)$ to be greater than 0, e must be greater than 1. This implies that profit can be maximized only when $e > 1$. The logic of this conclusion can be provided as follows.

Given the Eq. 10.5 and Eq. 10.6, if $e = 1$, $MR = 0$, and if $e < 1$, $MR < 0$, it means that if $MR < 0$ and $MC > 0$, or in other words, when $MR \neq MC$, then the rule of profit maximization breaks down. Thus, profit can be maximized only if $e > 1$, and $MC > 0$.

Now if $e > 1$, then the term $e/(e - 1)$ will always be greater than 1. Let e be greater than 1 by an amount, say m. Then

$$\frac{e}{e-1} = (1 + m) \qquad \ldots(10.9)$$

By substituting term $(1 + m)$ from Eq. (10.9) for $e/(e - 1)$ in Eq. (10.8), we get

$$P = AVC\ (1 + m) \qquad \ldots(10.10)$$

where m denotes the mark-up rate.

Note that Eq. (10.10) is exactly the same as Eq. (10.3). This means that the mark-up rule of pricing converges with the marginalist rule of pricing. In other words, it is proved that the mark-up pricing method leads to the marginalist rule of pricing. However, m in Eqs. (10.3) and in (10.9) need not be the same.

Limitations of Mark-up Pricing Rule

The cost-plus pricing has certain limitations, which should be borne in mind while using this method for price determination.

First, cost-plus pricing assumes that a firm's resources are optimally allocated and the standard cost of production is comparable with the average of the industry. In reality, however, it may not be so. The cost estimates based on these assumptions may be an overestimate or an underestimate. Under these conditions, pricing may not be commensurate with the objective of the firm.

Secondly, in cost-plus pricing, generally, historical cost rather than current cost data are used. This may lead to under-pricing under increasing cost conditions and to over-pricing under decreasing cost conditions, which may go against the firm's objective.

Thirdly, if variable cost fluctuates frequently and significantly, cost-plus pricing may not be an appropriate method of pricing on regular basis.

Finally, it is also alleged that cost-plus pricing ignores the demand side of the market and is solely based on supply conditions. This is, however, not true, because the firm determines the mark-up on the basis of 'what the market can bear' and it does take into account the elasticity aspect of the demand for the product, as shown above.

10.2.2 Multiple Product Pricing

The price theory or microeconomic models of price determination are based on the assumption that a firm produces a single, homogeneous product. In actual practice, however, production of a single homogeneous product by a firm is an exception rather than a rule. Almost all firms have more than one product in their line of production. Even the most specialized firms produce a commodity in multiple models, styles and sizes, each so much differentiated from the other that every model or size of the product may be considered a different product. For example, the various models of refrigerators, TV sets, cell phones, computers and car models etc. produced by the same company may be treated as different products for at least pricing purpose. The various models are so differentiated that consumers view them as different products and, in some cases, as close substitutes for each other. It is, therefore, not surprising that each model or product has different *AR* and *MR* curves and that one product of the firm competes against the other product. The pricing under these conditions is known *as multi-product pricing* or *product-line pricing.*

The major problem in pricing multiple products is that each product has a separate demand curve. But, since all the products are produced under one establishment by interchangeable production facilities, they have only one joint and one inseparable marginal cost curve. That is, while revenue curves, *AR* and *MR*, are separate for each product, cost curves, *AC* and *MC*, are inseparable. Therefore, the marginal rule of pricing cannot be applied straightaway to fix the price of each product separately. The problem, however, has been provided with a solution by E.W. Clemens. The solution is similar to the technique employed to illustrate third degree price discrimination under profit maximization assumption. As a discriminating monopoly tries to maximize its revenue in all its markets, so does a multi-product firm in respect of each of its products.

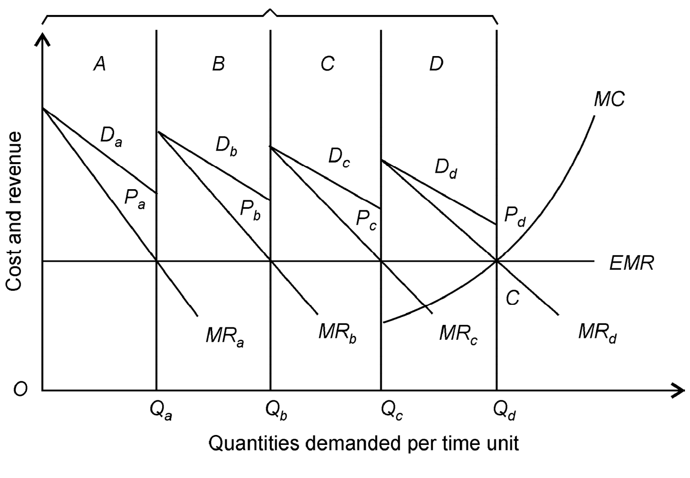


Fig. 10.1 *Multi-Product Pricing*

To illustrate the multiple product pricing, let us suppose that a firm has four different products—*A*, *B*, *C* and *D* in its line of production. The *AR* and *MR* curves for the four branded products are shown in four segments of Fig. 10.1. The marginal cost for all the products taken together is shown by the curve *MC*, *which is the factory marginal cost curve*. Let us suppose that when the *MRs* for the individual products are horizontally summed up, the aggregate *MR* (not given in the figure) passes through point *C* on the *MC* curve. If a line parallel to the *X*-axis, is drawn from point *C* to the *Y*-axis through the *MRs*, the intersecting points will show the points where *MC* and *MRs* are equal for each product, as shown by the line *EMR*, the *Equal Marginal Revenue* line. The points of intersection between *EMR* and *MRs* determine the output level and price for each product. The output of the four products are given as OQ_a of product *A*; Q_aQ_b of *B*; Q_bQ_c of *C*; and Q_cQ_d of *D*. The respective prices for the four products are: P_aQ_a for product *A*; P_bQ_b for *B*; P_cQ_c for *C*, and P_dQ_d for *D*. These price and output combinations maximize the profit from each product and hence the overall profit of the firm.

10.2.3 Pricing in Life-cycle of a Product

The life-cycle of a product is generally divided into five stages: (*i*) Introduction or initial stage, (*ii*) Growth, (*iii*) Maturity, (*iv*) Saturation and (*v*) Decline. Fig. 10.2 presents the five stages of a product's life-cycle through a curve showing the behaviour of the total sales over the life cycle. The ***introduction*** phase is the period taken to introduce the product to the market. The total sale during this period is limited to the quantity put on the market for trial with considerable advertisement. The sales during this period remain almost constant. ***Growth*** is the stage, after a successful trial, during which the product gains popularity among the consumers and sales increase at an increasing rate as a result of cumulative effect of advertisement over the initial stage. ***Maturity*** is the stage in which sales continue to increase but at a lower rate and the total sale eventually becomes constant. During the ***saturation period*** the total sale saturates—there is no considerable increase or decrease in the sales. After the saturation stage, comes the stage of ***decline*** in which total sales begin to decline for such reasons as (*i*) increase in the availability of substitutes and (*ii*) the loss of distinctiveness of the product.

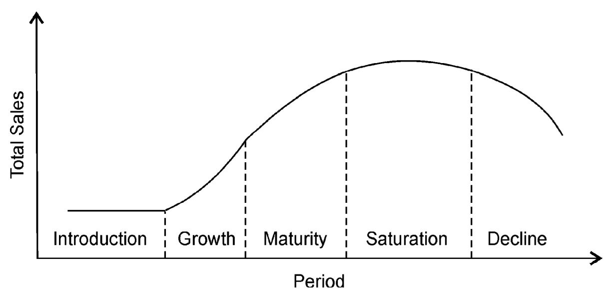


Fig. 10.2 *Life-Cycle of a Product*

The pricing strategy varies from stage to stage over the life-cycle of a product, depending on the market conditions. From the pricing strategy point of

NOTES

view, growth and maturity stages may be treated likewise. We have first discussed the pricing of a product in its initial stage as pricing of a new product and then the pricing method in the 'maturity' and 'decline' stage.

10.2.4 Pricing a New Product

A new product may simply be either another brand name added to the existing ones or an altogether new product. Pricing a new brand for which there are many substitutes available in the market is not as big a problem as pricing a new product for which close substitutes are not available. For, in case of the former, market provides adequate information regarding cost, demand, and availability of market, etc. Pricing in this case depends on the nature of the market. However, problems arise in pricing a new product without close substitutes because, for lack of information, there is some degree of uncertainty.

Thus, pricing policy in respect of a new product depends on whether or not close substitutes are available. Depending on whether or not close substitutes are available, generally two kinds of pricing strategies are suggested in pricing a new product, viz., (*i*) skimming price policy and (*ii*) penetration price policy.

(*i*) *Skimming price policy.* The *skimming price policy* is adopted where close substitutes of a new product are not available. This pricing strategy is intended to skim the cream off the market, i.e., consumer's surplus, by setting a high initial price, three or four times the ex-factory price, and a subsequent lowering of prices in a series of reduction, especially in case of consumer durables. The initial high price would generally be accompanied by heavy sales promoting expenditure. This policy succeeds for the following reasons.

First, in the initial stage of the introduction of the product, demand is relatively inelastic because of consumers' desire for distinctiveness by the consumption of a new product.

Second, cross-elasticity is usually very low for lack of a close substitute.

Third, step-by-step price-cuts help *skimming consumers'* surplus available at the lower segments of demand curve.

Fourth, high initial prices are helpful in recouping the development costs.

The *post-skimming strategy* includes the decisions regarding the time and size of price reduction. The appropriate occasion for price reduction is the time of *saturation* of the total sales or when strong competition is apprehended. As regards the rate of price reduction, when the product is on its way to losing its distinctiveness, the price-cut should be appropriately larger. But, if the product has retained its exclusiveness, a series of small and gradual price reductions would be more appropriate.

(*ii*) *Penetration price policy.* In contrast to skimming price policy, the penetration price policy involves a reverse strategy. This pricing policy is adopted generally in the case of new products for which substitutes are available. This policy requires fixing a lower initial price designed to penetrate the market as quickly as possible and is intended to maximize the profits in the long-run. Therefore, the firms pursuing the penetration price policy set a low price of the product in the initial stage. As the product catches the market, price is gradually raised up. The success of penetration price policy requires the existence of the following conditions.

First, the short-run demand for the product should have an *elasticity greater than unity.* It helps in capturing the market at lower prices.

Secondly, *economies of large-scale production* should be available to the firm with the increase in sales. Otherwise, increase in production would result in increase in costs which might reduce the competitiveness of the price.

Thirdly, the *potential market* for the product ought to be fairly large and have a good deal of future prospects.

Fourthly, the product should have *a high cross-elasticity* in relation to rival products for the initial lower price to be effective.

Finally, the product, by nature should be such that it can be easily accepted and adopted by the consumers.

The choice between the two strategic price policies depends on (*i*) the rate of market growth; (*ii*) the rate of erosion of distinctiveness; and (*iii*) the cost-structure of the producers. If the rate of market growth is slow for such reasons as lack of information, slow growth of purchasing power, consumers' hesitation, etc., penetration price policy would be unsuitable. The reason is a low price will not mean a large sale. If the pioneer product is likely to lose its distinctiveness at a faster rate, skimming price policy would be unsuitable. Penetration pricing policy should be followed when lead time, i.e., the period of distinctiveness, is fairly long. If cost-structure shows a decreasing trend over time, penetration price policy would be more suitable, since it enables the producer to reduce his cost and prevents potential competitors from entering the market in the short-run.

10.2.5 Pricing in Maturity Period

Maturity period is the second stage in the life-cycle of a product. It is a stage between the growth period and decline period of sales. Sometimes *maturity period* is bracketed with *saturation period*. Maturity period may also be defined as the period of decline in the *growth rate* of sales (not the total sales). It can be defined for all practical purposes as the period of zero growth rate. The concept of maturity period is useful to the extent it gives out signals for taking precaution with regard to pricing policy. However, the concept itself does not provide guidelines for the pricing policy. Joel Dean suggests that the "first step for the manufacturer whose speciality is about to slip into the commodity category is to reduce real … prices as soon as the system of deterioration appears." But he warns that "this does not mean that the manufacturer should declare open price war in the industry". He should rather move in the direction of "product improvement and market segmentation".

10.2.6 Pricing a Product in Decline

NOTES

The product in decline is one that enters the post-maturity stage. During this stage, the total sale of the product starts declining. The first step in pricing strategy at this stage is obviously to reduce the price with the objective of retaining sales at some minimum level. The product should be reformulated and remodelled to suit the consumers' preferences. It is a common practice in the book trade. When the sale of a hard-bound edition reaches saturation, paper-back edition is brought into the market. This facility is, however, limited to only a few commodities. As a final step in the strategy, the advertisement expenditure may be reduced drastically or withdrawn completely, and the residual market may be relied on. This, however, requires a strong will of the producer.

10.3 TRANSFER PRICING (MARKET AVERAGE REVENUE AND MARGINAL REVENUE)

Large size firms often divide their production into different product divisions or their subsidiaries. Growing firms add new divisions or departments to the existing ones. The firms then transfer some of their activities to other divisions. The goods and services produced by the new divisions are used by the parent organization. In other words, the parent division buys the product of its subsidiaries. Such firms face the problem of determining an appropriate price for the product transferred from one division or subsidiary to the parent body. This problem becomes much more difficult when each division has a separate profit function to maximize. Pricing of intra-firm 'transfer product' is referred to as 'transfer pricing'. One of the most systematic treatments of the transfer pricing technique has been provided by Hirshleifer. We will discuss here briefly his technique of transfer pricing.

To begin with, let us suppose that a refrigeration company established a decade ago used to produce and sell refrigerators fitted with compressors bought from a compressor manufacturing company. Now the refrigeration company decides to set up its own subsidiary to manufacture compressors. Let us also assume:

(*i*) both parent and subsidiary companies have their own profit functions to maximize and

(*ii*) the refrigeration company has the option of using all the compressors produced by its subsidiary and/or to sell the compressors in a competitive market and its demand is given by a straight horizontal line.

Given the model, transfer pricing is discussed under two conditions.

(*i*) The parent company uses the entire output of its subsidiary and there is no external market for the compressors and

(*ii*) There does exist a competitive market for the compressor and refrigeration company sells also in the open market.

Let us begin our analysis of transfer pricing model by assuming that there is no external market for the compressors. We will later drop this assumption and assume that there is an external market for the compressors and discuss the technique of transfer pricing under both the alternative conditions.

10.3.1 Transfer Pricing without External Market

When refrigeration company uses its entire compressor ouput, it has to set an appropriate price for the compressors so that the profit of its subsidiary too is maximum. To deal with the 'transfer pricing' problem, let us first look into the pricing and output determination of the final product, i.e., refrigerators. Since the refrigeration company sells its refrigerators in a competitive market, the demand for its product is given by a straight horizontal line as shown by the line $AR_r = MR_r$ in Fig. 10.3.

The marginal cost of intermediate good, i.e., compressor, is shown by MC_c curve and that of the refrigerator body by MC_b. The MC_c and MC_b added vertically give the combined marginal cost curve, the MC_t. At output OQ, for example, $TQ + MQ = PQ$. The MC_t intersects line $AR_r = MR_r$ at point P. An ordinate drawn from point P down to the horizontal axis determines the most profitable outputs of refrigerator bodies and compressors, each at OQ. Thus, the output of both—refrigerator bodies and compressors is simultaneously determined. Since at OQ level of output, the firm's $MC_t = MR_r$, the refrigerator company maximizes its profits from the final product, the refrigerators.

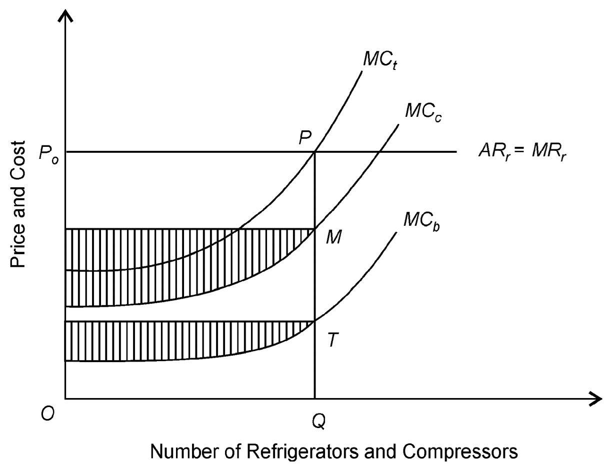


Fig. 10.3 *Price Determination of the Final Product (Refrigerators)*

Now, let us find the price of the compressors. The question that arises is: what should be the price of the compressors so that the compressor manufacturing division too maximizes its profit? The answer to this question can be obtained by applying the profit maximization rule, profit is maximum where $MC = MR$. This rule requires equalizing MC and MR in respect of compressors. The marginal cost curve for the compressors is given by MC_c in Fig. 10.3. The firm therefore has to obtain the marginal revenue for its compressors. The marginal revenue of the compressors (MR_c) can be obtained by subtracting the non-compressor marginal cost of the final good from the MR_r. Thus,

$$MR_c = MR_r - (MC_t - MC_c) \qquad \ldots(10.11)$$

NOTES

For example, in Fig. 10.3, at output OQ, $MR_r = PQ$, $MC_t = PQ$, and $MC_c = MQ$. By substituting these values in Eq. (10.11), we get

$$MR_c = PQ - (PQ - MQ)$$
$$PQ - PM = MQ$$

or, since in Fig.10.3, $PQ - MQ = PM$, and $PM = TQ$, therefore,

$$MR_c = PQ - TQ = PT \quad \text{and} \quad PT = MQ$$

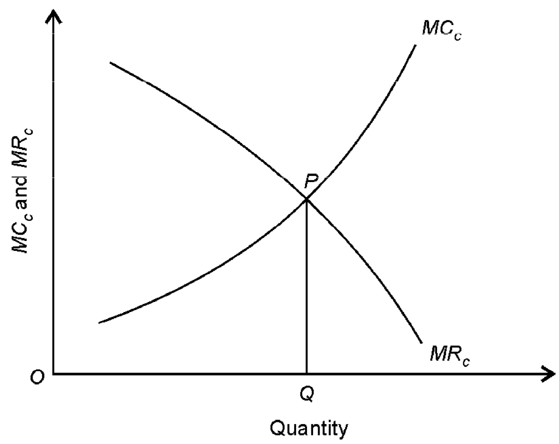


Fig. 10.4 *Determination of Transfer Price*

We may recall that, when price is constant, $AR_r = MR_r$, i.e., MR_r is constant as shown in Fig. 10.3. In contrast, however, MC_t is a rising function. Thus, $MR_r - MC_c$ will be a decreasing function. Notice the vertical distance between $AR_r = MR_r$ line, and MC_c curve is decreasing as shown in Fig. 10.3. When MR_c (which equals $MR_r - MC_c$) is obtained for different levels of output and graphed, it yields a curve like MR_c curve shown in Fig. 10.4. The MC_c curve (which is the same as MC_c curve in Fig 10.3) intersects the MR_c at point P. At point P, $MR_c = MC_c$ and output is OQ. Thus, the price of compressors is determined at PQ in Fig. 10.4. This price enables the compressor division to maximize its own profit.

10.3.2 Transfer Pricing with External Competitive Market

We have discussed above the transfer pricing under the assumption that there is no external market for the compressors. It implies that the refrigeration company was the sole purchaser of its own compressors and that the compressor division had no external market for its product. Let us now discuss the transfer pricing technique assuming that there is an external market for the compressors. The existence of the external market implies that the compressor division has the opportunity to sell its surplus production to other buyers and the refrigeration company can buy compressors from other sellers if the compressor division fails to meet its total demand. For simplicity sake, let us assume also that the external market is perfectly competitive. Determination of transfer price under these conditions is a little more complicated task.

The method of transfer pricing with external market is illustrated in Fig. 10.5. Since the compressor market is perfectly competitive, the demand for compressors is given by a straight horizontal line as shown by the line P_2D. In that case $AR = MR$. The marginal cost curve of the compressors is shown by MC_c. The MR_c curve shows the marginal net revenue from the compressors, (see Fig. 10.5). Note that in the absence of the external market, the transfer price of compressors would have been fixed at $OP_1 = P'Q_2$, i.e., the price where $MR_c = MC_c$. At this price the parent company would have bought compressors from its subsidiary only. But, since compressors are to be produced and sold under competitive conditions, the effective marginal cost of the compressor to the refrigeration company is the market price of the compressor, i.e., OP_2. Besides, the price OP_2 is also the potential MR for the compressor division. Therefore, in order to maximize its profit, the firm sets compressor's price at point P where $MR_c > MC_c$. Thus, the transfer price of compressor will be fixed at PQ_1 and the refrigeration company would buy OQ_1 compressors from the compressor division.

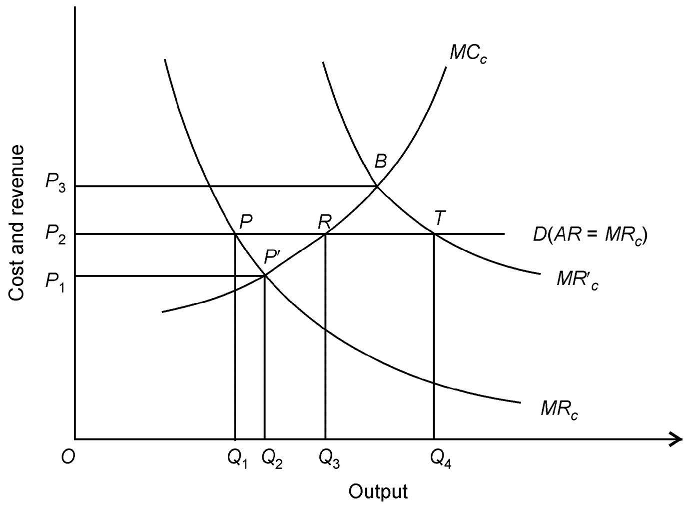


Fig. 10.5 *Determination of Transfer Price with External Market*

The total output of compressors is determined at a level where MC_c intersects the demand line, D ($AR = MR$). That is, profit maximizing level of output is determined by point R. At point R, the total output of compressors is OQ_3. Of this, OQ_1 is bought by the refrigeration company itself and the remaining output, Q_1Q_3 is sold in the external market, both at price OP_2. At this level of output and price, the compressor division maximizes its profit.

NOTES

Shift in MR_c and Transfer Price

Let us now consider how transfer price is determined when MR_c shifts upward to the right. The MR_c may shift upward because of an increase in demand causing an upward shift in $AR = MR$. Let the MR_c in Fig. 10.5 shift to MR'_c which intersects with MC_c at point B. In the absence of an external market, the refrigeration company would have set transfer price of compressors at OP_3—a price higher than the free market price OP_2. But, since there is an external market in which a lower price is given at OP_2, the transfer price cannot exceed the market price or else the refrigeration company would not be in a position to maximize its profit. Nor can the transfer price be less than the market price, otherwise the compressor division would not be able to maximize its profit. Thus, if there is an external market in which market price of an intermediary product is given, then the problem is to determine the quantity to be produced by the subsidiary and the quantity to be purchased from the external market. Fig. 10.5 shows that after the shift in MR_c curve to MR'_c, the demand for compressors by the refrigeration company increases to OQ_4 where $AR = MR = MR'_c$. But the subsidiary company cannot produce OQ_4 units of compressors, given its MC_c and the market price. It will, therefore, produce only OQ_3 number of compressors, which equalizes MC_c with MR at point R. Given the market price, OQ_3 is the most profitable output of compressors. Therefore, the difference between the total demand and the total internal supply from the subsidiary, i.e., $OQ_4 - OQ_3 = Q_3Q_4$, will be bought in the external market, at price $OP_2 = TQ_4$. Thus, the refrigeration company will buy OQ_3 compressors from its compressor division and buy Q_3Q_4 in the external market.

10.3.3 Transfer Pricing Under Imperfect External Market

When the refrigerator market is imperfect, the compressor division faces a demand curve with a negative slope in the external market, instead of a straight horizontal demand line. The downward sloping demand curve makes transfer pricing a much more complicated task. To illustrate the transfer pricing technique under imperfect market conditions in the external market, let us suppose (*i*) that the average and marginal revenue curves for the compressors are given by AR_x and MR_x respectively, in Fig. 10.6 and (*ii*) that the 'marginal net revenue' from the internal use of compressors and the marginal cost of producing compressors are represented by MR_c and MC_c respectively. With a view to maximizing the overall profit, the refrigeration company will determine the output of compressors where $MC_c = MR_c + MR_x$, i.e., where marginal cost of compressors equals the composite marginal revenue. The composite marginal revenue is obtained through horizontal summation of the MR_c and MR_x curves as shown by MR_t in Fig. 10.6.

NOTES

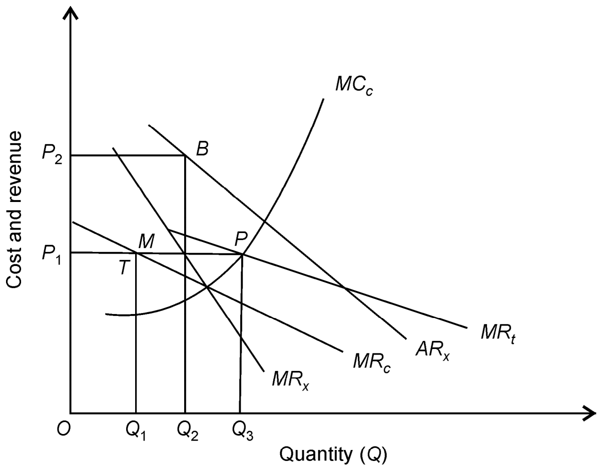


Fig. 10.6 *Transfer Pricing and Imperfect External Market*

As shown in Fig. 10.6, MC_c intersects MR_t at point P which determines the profit maximizing output of compressors at OQ_3. The compressor division can maximize its profit by dividing its output between the refrigeration company and the external market so as to equalize its MC and MR in both the markets—internal and external. If a line (PP_1) is drawn from point P parallel to the horizontal axis to the vertical axis, it intersects MR_x at point M and MR_c at point T. The points of intersection (T and M) determine the share of refrigeration company and the external market in the total output OQ_3. At point M, $MC_c = MR_x$ and at point T, $MC_c = MR_c$. Thus, the refrigeration company, the parent body, will buy OQ_1 for internal use and sell OQ_2 in the open market. Note that $OQ_1 + OQ_2 = OQ_3$. The profit maximizing price in the external market is OP_2 (= BQ_2) and the profit maximizing transfer price is set at OP_1. With these prices and output, both refrigeration company and compressor division maximize their respective profits.

10.4 PEAK LOAD PRICING

There are certain non-storable goods, e.g., electricity, telephones, transport and security services, etc., which are demanded in varying measures during the day as well as night. For example, consumption of electricity reaches its peak in day time. It is called 'peak-load' time. It reaches its bottom in the night. This is called 'off-peak' time. Electricity consumption peaks in daytime because all business establishments, offices and factories come into operation. Electricity consumption decreases during nights because most business establishments are closed and household consumption falls to its basic minimum. In Delhi, demand for electricity peaks during summers due to use of *ACs* and coolers, and it declines to its minimum level during winters. Similarly, consumption of telephone services is at its peak at day time and at its bottom at nights. Another example of 'peak' and 'off-peak' demand is of railway and air services. During festivals, summer holidays, 'Pooja' vacations, etc., the demand for railway and air travel services rises to its peak.

A technical feature of such products is that they cannot be stored. Therefore, their production has to be increased in order to meet the 'peak-load' demand and reduced to 'off-peak' level when demand decreases. Had they been storable, the excess production in 'off-peak' period could be stored and supplied during the 'peak-load' period. But this cannot be done. Besides, given the installed capacity, their production can be increased but at an increasing marginal cost (*MC*).

10.4.1 Problems in Pricing

Pricing of goods like electricity is problematic. The nature of the problem in a short-run setting is depicted in Fig. 10.7. The 'peak-load' and 'off-load' demand curves are shown by D_P and D_L curves, respectively. The short-run supply curve is given by the short-run marginal cost curve, *SMC*. The problem is 'how to price electricity'.

As Fig. 10.7 shows, if electricity price is fixed in accordance with peak-load demand, OP_3 will be the price and if it is fixed according to off-load demand, price will be OP_1. The problem is: what price should be fixed? If a 'peak-load' price (OP_3) is charged uniformly in all seasons, it will be unfair because consumers will be charged for what they do not consume. Besides, it may affect business activities adversely. If electricity production is a public monopoly, the government may not find it advisable to charge a uniform 'peak-load' price.

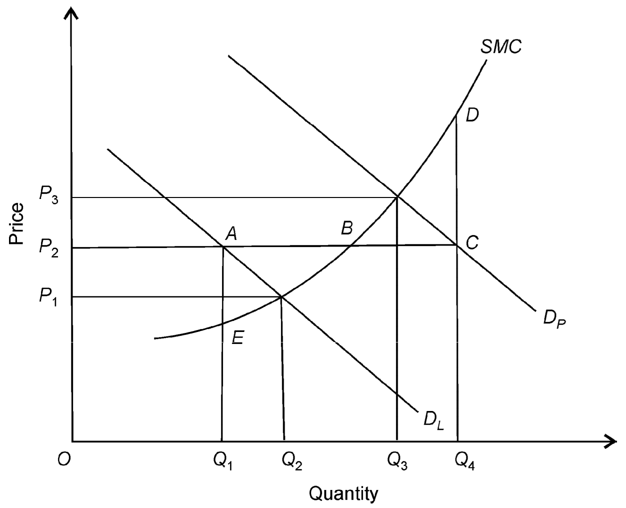


Fig. 10.7 *Peak-Load Pricing of Electricity*

On the other hand, if a uniform 'off load' price (OP_1) is charged, production will fall to OQ_2 and there will be acute shortage of electricity during peak hours. It leads to 'breakdowns' and 'load-shedding' during the peak-load periods, which disrupt production and make life miserable. This is a regular feature in Delhi, the capital city of India. This is because electricity rates in Delhi are said to be one of the lowest in the country.

Alternatively, if an average of the two prices, say P_2 is charged, it will have the demerits of both 'peak-load' and 'off-load' prices. There will be an excess production to the extent of *AB* during the 'off-load' period, which will go waste as it cannot be stored. If production is restricted to OQ_1, price P_2 will be unfair. And, during the 'peak-load' period, there will be a shortage to the extent of *BC*, which can be produced only at an extra marginal cost of *CD*.

NOTES

10.4.2 Double Pricing System

For the above reasons, generally, a double pricing system is adopted. A higher price, called 'peak-load price' (OP_3) is charged during the 'peak-load' period and a lower price (OP_1) is charged during the 'off-peak' period. During the 'peak-load' period, production is increased to OQ_3 at which D_P intersects *SMC*, and production is reduced to OQ_1 during the 'off-peak' period. However, the system has its own advantages and disadvantages.

Advantages: Peak-load pricing system has two advantages.

(*i*) It results in an efficient distribution of electricity consumption. Housewives run their dishwashers and washing machines during the 'off-peak' period.

(*ii*) It helps in preventing a loss to the electricity company and ensures regular supply of electricity in the long-run.

Disadvantages: This system has two disadvantages too.

(*i*) The businesses which are by nature day-business pay higher rates than those which can be shifted to 'off-peak' period.

(*ii*) Billing system is the greatest problem. Each consumer will have to install two meters—one for 'peak-load' and another for 'off-load' period with an automatic switch-over system. This can be done.

Alternatively, the problem can be resolved by adopting a progressive tariff rate for the use of electricity. But, in a country like India, all pervasive corruption will make it inefficient. Delhi Vidyut Board (DVB), even after privatization of electricity distribution, is reportedly able to collect only about 48 per cent of its cost of production. The rest goes to the unauthorized users of electricity.

10.5 LAW OF SUPPLY

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

NOTES

Check Your Progress

1. Name the common method of pricing mostly used by the manufacturing firms.
2. State the stages of the life-cycle of a product.
3. Mention the pricing strategies normally adopted in pricing a new product.

10.6 ANSWERS TO 'CHECK YOUR PROGRESS' QUESTIONS

1. The cost-plus pricing is the common method of pricing mostly used by the manufacturing firms.
2. The four stages of the life-cycle of a product are (i) Introduction or initial phase (ii) Growth (iii) Maturity (iv) Saturity (v) Decline.
3. The pricing strategies normally adopted in pricing a new product are (i) Skimming price policy and (ii) Penetration price policy.

10.7 SUMMARY

- Cost-plus pricing is also known as 'mark-up pricing', 'average cost pricing' and 'full cost pricing'. The cost-plus pricing is the most common method of pricing used by the manufacturing firms.
- The first step in price fixation is to estimate the average variable cost. For this, the firm has to ascertain the volume of its output for a given period of time, usually one accounting or fiscal year.
- The price theory or microeconomic models of price determination are based on the assumption that a firm produces a single, homogeneous product.
- The skimming price policy is adopted where close substitutes of a new product are not available. This pricing strategy is intended to skim the cream off the market, i.e., consumer's surplus, by setting a high initial price, three or four times the ex-factory price, and a subsequent lowering of prices in a series of reduction, especially in case of consumer durables.
- In contrast to skimming price policy, the penetration price policy involves a reverse strategy. This pricing policy is adopted generally in the case of new products for which substitutes are available.
- The product in decline is one that enters the post-maturity stage. During this stage, the total sale of the product starts declining.
- There are certain non-storable goods, e.g., electricity, telephones, transport and security services, etc., which are demanded in varying measures during the day as well as night. For example, consumption of electricity reaches its

peak in day time. It is called 'peak-load' time. It reaches its bottom in the night. This is called 'off-peak' time.

10.8 KEY WORDS

NOTES

- **Law of supply:** It states that the supply of a product increases with the increase in its price and decreases with decrease in its price, other things remaining constant.
- **Skimming price policy:** This pricing strategy is intended to skim the cream off the market, that is, consumer's surplus, by setting a high initial price, three or four times the ex-factory price, and a subsequent lowering of prices in a series of reduction, especially in case of consumer durables.
- **Transfer pricing:** It is the setting of the price for goods and services sold between controlled (or related) legal entities within an enterprise.

10.9 SELF-ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. What are the limitations of the mark-up pricing rule?
2. State the conditions necessary for the successful implementation of the penetration price policy.
3. Mention the advantages and disadvantages of the peak-load pricing system.

Long-Answer Questions

1. Discuss the theory of pricing.
2. Explain the pricing of a new product.
3. Explain the concepts of market average revenue and marginal revenue with the help of examples.

10.10 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.

NOTES

UNIT 11 PERFECT COMPETITION AND PRICE DETERMINATION

Structure

11.0 INTRODUCTION

Price, under conditions of perfect competition is determined by the interaction of demand and supply. Before Marshall, there was disagreement among economists as to whether the force of demand (i.e., marginal utility) or the force of supply (i.e., cost of production) played a vital role in determining price. Marshall assigned equal importance to both demand (and marginal utility) and supply (or cost of production) in the determination of value or price. In this unit, you will study about Marshall's time analysis, perfect competition and price determination.

11.1 UNIT OBJECTIVES

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

- Discuss Marshall's time analysis
- Define perfect competition
- List the features of perfect competition
- Explain price and output determination under perfect competition

11.2 MARSHALL'S TIME ANALYSIS

NOTES

Alfred Marshall was a dominant figure in British economics from about 1890 until his death in 1924. His specialty was *microeconomics* with the focus on the study of individual markets and industries, as opposed to the study of the whole economy. In his most important book, *Principles of Economics,* Marshall emphasized that the price and output of a good is determined by both *supply* and *demand*— the two curves are like scissor blades that intersect at equilibrium. Modern economists who are trying to understand why the price of a good changes are still referring to Marshall's approach.

Marshall is credited for having propounded the concept of price elasticity of demand, which quantifies buyers' sensitivity to price.

Marshall has also contributed to an understanding of the concept of consumer surplus. He noted that the price is typically the same for each unit of a commodity that a consumer buys. However, the value to the consumer of each additional unit declines. A consumer will buy units up to the point where the marginal value equals the price. Therefore, on all units previous to the last one, the consumer reaps a benefit by paying less than the value of the good to himself. The size of the benefit equals the difference between the consumer's value of all these units and the amount paid for the units. This difference is called the consumer surplus, for the surplus values or utility enjoyed by consumers. Marshall also introduced the concept of producer surplus. It is the amount that the producer is actually paid minus the amount that he would willingly accept. Marshall used these concepts to measure the changes in well-being from government policies such as taxation. Although economists have refined the measures since Marshall's time, his basic approach to what is now called welfare economics still stands.

To understand how markets adjust to changes in supply or demand over time, Marshall introduced the idea of three periods. First, is the market period— it is the amount of time for which the stock of a commodity is fixed. Second, the short period is the time in which the supply can be increased by adding labour and other inputs but not by adding capital (Marshall's term was "appliances"). Third, the long period is the amount of time taken for capital ("appliances") to be increased.

To make economics dynamic rather than static, Marshall used the tools of classical mechanics, including the concept of optimization. With these tools he, like neoclassical economists who have followed in his footsteps, took as givens technology, market institutions, and people's preferences. But Marshall was not satisfied with his approach. He once wrote that "the Mecca of the economist lies in economic biology rather than in economic dynamics." In other words, Marshall was arguing that the economy is an evolutionary process in which technology, market institutions, and people's preferences evolve along with people's behaviour.

11.3 PERFECT COMPETITION AND PRICE DETERMINATION

NOTES

The theory of firms came into existence during the 1930s with Joan Robinson's *The Economics of Imperfect Competition* and Edwin H. Chamberlin's *The Theory of Monopolistic Competition,* both written independently in 1933. Earlier, the theory related to price determination was in the form of the 'Theory of Value' attributed to Alfred Marshall and his *Principles of Economics.* The theory of value propounded by Marshall, on the assumptions of perfect competition and a static equilibrium system, was regarded to provide answer to all questions regarding price and output determination. The existence of perfect competition was however challenged by Piero Sraffa. He showed that perfect competition was not logically consistent with partial equilibrium analysis. This led to the abandonment of the assumption of the perfect competition. Robinson and Chamberlin developed independently the theory of imperfect competition and theory of monopolistic competition, respectively. Joan Robinson and Chamberlin have demonstrated that price and output are determined by individual decisions under the condition of imperfect competition. They had however retained the earlier assumption of profit maximisation. This assumption was later challenged, and many new theories of firms were suggested. None of the theories has however received a universal acceptance.

We begin our discussion on the theory of firm with the analysis of price and output determination in a perfectly competitive market.

11.3.1 The Features of Perfect Competition

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

1. **Large Number of Buyers and Sellers.** Under perfect competition, the number of sellers is assumed to be so large that the share of each seller in the total supply of a product is so small that no single firm can influence the market price by changing its supply. Therefore, firms are *price-takers* not *price-makers*. Similarly, the number of buyers is so large that the share of each buyer in the total demand is so small that no single buyer or a group of buyers can influence the market price by changing their individual or group demand for a product.
2. **Homogeneous Product.** The commodities supplied by all the firms of an industry are assumed to be *homogeneous* or approximately identical. Homogeneity of the product implies that buyers do not distinguish between

products supplied by the various firms of an industry. Product of each firm is regarded as a perfect substitute for the products of other firms. Hence, no firm can gain any competitive advantage over the other firms. This assumption limits the power of any firm to charge a price which is even slightly higher than the market price.

3. **Perfect Mobility of Factors of Production.** Another important characteristic of perfect competition is that the factors of production (especially, labour and capital) are freely mobile between the firms. Labour can freely change the firms as there is no barrier on labour mobility—legal, language, climate, skill, distance or otherwise. There is no trade union. Capital can also move freely from one firm to another. No firm has any kind of monopoly over any industrial input. This assumption guarantees that factors of production—labour, capital, and entrepren-eurship—can enter or quit a firm or the industry whenever it is found desirable.
4. **Free Entry and Free Exit.** There is no legal or market barrier on entry of new firms to the industry. Nor is there any restriction on exit of the firms from the industry. That is, a firm may enter the industry and quit it at its will. Thus, when normal profit of the industry increases, new firms enter the industry and if profits decrease and better opportunities are available, firms leave the industry.
5. **Perfect Knowledge about the Market Conditions.** There is perfect knowledge about the market conditions. All the buyers and sellers have full information regarding the prevailing and future prices and availability of the commodity. As Marshall put it, "... though everyone acts for himself, his knowledge of what others are doing is supposed to be generally sufficient to prevent him from taking a lower or paying a higher price than others are doing." Information regarding market conditions is available free of cost. There is no uncertainty.
6. **No Government Interference.** Government does not interfere in any way with the functioning of the market. There are no taxes or subsidies; no licencing system, no allocation of inputs by the government, or any kind of other direct control. That is, the government follows the *free enterprise* policy. Where there is intervention by the government, it is intended to correct the market imperfections.
7. **Absence of Collusion and Independent Decision-Making.** Perfect competition assumes that there is no collusion between the firms, i.e., they are not in league with one another in the form of *guild* or *cartel.* Nor are the buyers in collusion between themselves. There are no consumers' associations, etc. This condition implies that buyers and sellers take their decisions independently and they act independently.

NOTES

11.3.2 Perfect vs. Pure Competition

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

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

11.4 PRICE AND OUTPUT DETERMINATION UNDER PERFECT COMPETITION

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

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

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

11.4.1 Price Determination in Very Short-Run

The market period or very short run refers to a time period in which quantity supplied is absolutely fixed or, in other words, supply response to change in price is nil. In the market period, therefore, the total output of the product is fixed. Each firm has a given quantity of commodity to sell. The aggregate supply of all the firms makes the market supply. The supply

NOTES

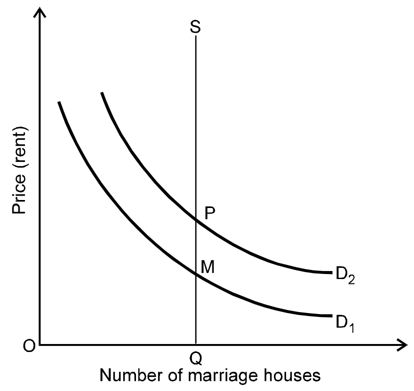


Fig. 11.1 *Determination of Market Price*

curve is *perfectly inelastic*, as shown by line *SQ* in Fig. 11.1. In this situation, price is determined entirely by the demand conditions. For instance, suppose that the number of marriage-houses (or tents) available per month in a city is given at *OQ* (Fig. 11.1), so that the supply curve takes the shape of a vertical straight line *SQ*. Let us also suppose that the monthly demand curve for marriage-houses is given by the demand curve, D_1. Demand and supply curves intersect each other at point *M*, determining the rental at *MQ*. Let us now suppose that during a particular month demand for marriage-houses suddenly increases because a relatively large number of parents decide to celebrate the marriage of their daughters and sons due to, say, non-availability of auspicious dates for some time to come. Consequently, the demand curve shifts upward to D_2. The demand curve D_2 intersects the supply curve at point *P*. The equilibrium rate of rental is thus determined at *PQ*. This becomes parametric price for all the buyers. Note that the rise in the rental from *MQ* to *PQ* is caused by the upward shift in the demand curve and that market supply curve remains perfectly inelastic in the market period. The other example of very short-run markets may be of perishable commodities like fish, milk, vegetable, etc. and of non-perishable commodities like shares and bonds.

11.4.2 Pricing in the Short-Run

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

Under competitive conditions the process of price determination and output adjustment in the short-run is given in Fig. 11.2(a) and 11.2(b). Figure 11.2(a) shows demand curve *DD* and supply curve *SS* intersect at point *P* determining the price at OP_1. This price is fixed for all the firms in the industry.

Given the price *PQ* (= OP_1), in Fig. 11.2(a), an individual firm can produce and sell any quantity at this price. But any quantity will however not yield maximum profit. The firms will have to adjust their output to the price OP_1. The process of output determination is presented through Fig. 11.2(b).

NOTES

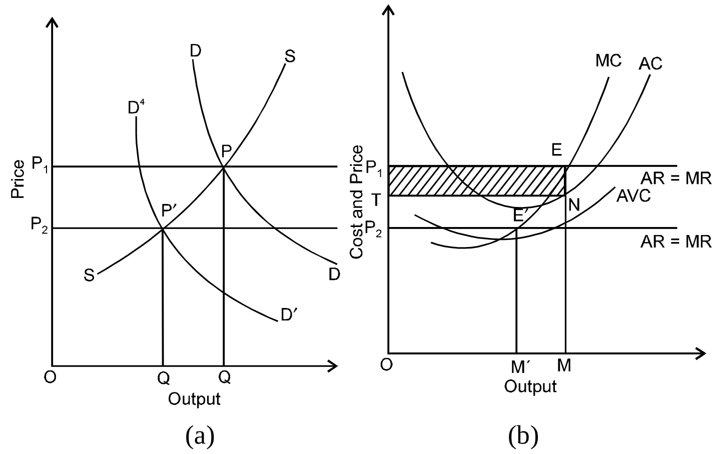


Fig. 11.2 *Pricing under Perfect Competition: Short-run*

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

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

In Fig. 16.7(b),

$$AR = EM;$$

$$AC = NM;$$

and $\quad Q = OM.$

By substituting the values from Fig. 11.2 (b), we get

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

Since $EM - EN = EN$,

$$\eta = EN.OM$$

This is the maximum profit that a firm can make, given the cost and revenue conditions as presented in Fig. 11.2 (b).

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

NOTES

11.4.3 Short-Run Equilibrium of the Industry

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

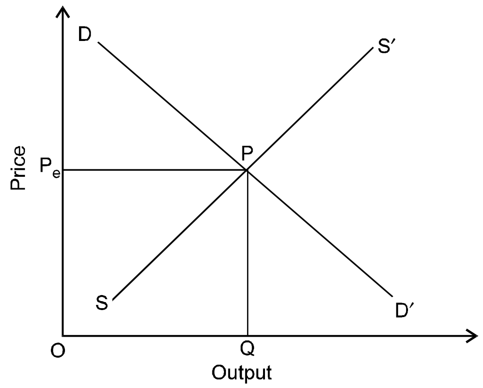


Fig. 11.3 *Equilibrium of the Industry*

An industry is in equilibrium in the short-run when market is cleared at a given price, i.e., when the total supply of the industry equals the total demand for its product. The price at which market is cleared is equilibrium price. The industry being in equilibrium, there is no tendency to expand or to contract the output. The equilibrium of industry is shown at point *P* in Fig. 11.3. The industry demand and supply curves intersect at point *P*, determining equilibrium price OP_e. The industry is supplying as much as consumers demand. In the short run equilibrium of the industry, individual firms may make pure profits, normal profits or losses, depending on their cost conditions.

11.5 PRICE AND OUTPUT DETERMINATION IN THE LONG RUN

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

NOTES

11.5.1 Price Determination in the Long Run

As in the short-run, market price is determined in the long-run by the market forces of demand and supply. Let us suppose that the market demand curve is given by DD' which is relevant for both short-run and long-run, and short-run supply curve is given by SS_1 in Fig. 11.4(a). The market demand curve DD' and market supply curve SS_1 intersect each other at point P_1 and the short-run market price is determined at OP_0. At this price, the firms find their short-run equilibrium at point E_1 and each of them produces[2] output OQ_1. The total market supply equals $OQ_1 \times$ No. of firms = ON_1 [in panel (a) of Fig. 11.4] and the industry is in short-run equilibrium.

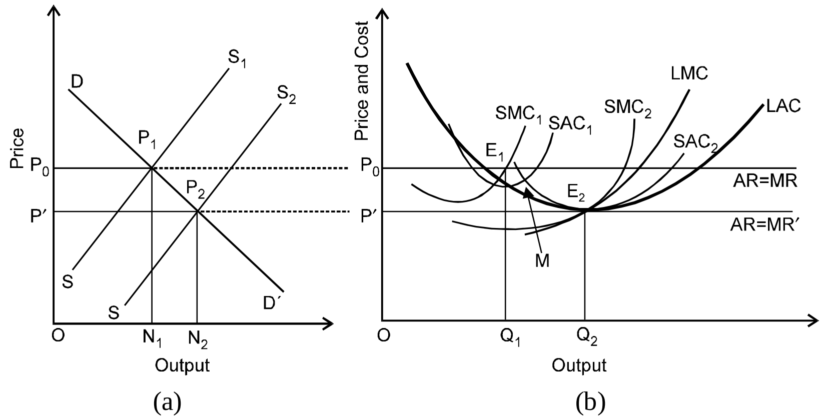


Fig. 11.4 *Long-run Equilibrium of the Firm*

Given the cost and revenue conditions in Fig. 11.4(b), the firms are making super normal profit of E_1M per unit. The existence of super normal profit in the short run leads to increase in the market supply on two accounts: *one*, new firms will enter the industry attracted by the super normal profits, and *two*, the existing firms would expand their plant-size because returns to scale would increase as shown by the *LAC*. As a result, the market supply would increase so that supply curve shifts rightward to SS_2 [Fig. 11.4(a)]. The shift in supply curve brings down the market price to *OP*¢ which is the long-run equilibrium price. Thus, equilibrium price is once again determined in the market.

11.5.2 Equilibrium of the Firm in the Long Run

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

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

That is, the firms of an industry reach their equilibrium position in the long-run where both a short-run and long-run equilibrium conditions coincide. In a

NOTES

perfectly competitive market, the cost and revenue conditions are given for the firms. What the firms can do, therefore, is to adjust their output to the given revenue and cost conditions in order to maximise their profit. Let us now illustrate the process of adjustment of output so as to reach the equilibrium in the long run.

Suppose that the firms are in equilibrium at point E_1 in Fig. 11.4 (a) where they make excess profits $AR - SAC_1 = EM$ per unit. This gives incentives to the firms to expand their scale of production, i.e., they add more plants to the existing ones. As a result, market supply increases. Besides, supply increases also because new firms enter the industry. Therefore, the market supply curve SS_1 tends to shift rightward causing a fall in price to OP'. On the other hand, due to increase in demand for inputs, cost tends to rise. But so long as economies of scale are greater than the diseconomies of scale, the *LAC* tends to decrease and it pays firms to expand their plant-size. When a stage is arrives where $P < LAC$, firms incur losses. The firms which are not able to make adjustment in the plant-size or scale of production leave the industry. This works in two directions. On the one hand, supply decreases and price increases, and on the other, demand for inputs decreases which causes a decrease in the input prices. This process of adjustment continues until *LAC* is tangent to $P = AR = MR$ for each firm in the industry. This position is shown at point E_2 in Fig. 11.4(b). Eventually, at point E_2, i.e., at the point of equilibrium,

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

Since $P = LAC$, the firms make only normal profits in the long-run. If firms deviate from point E_2, due to some short-run disturbances, the market forces will make them return to this point only.

11.5.3 Equilibrium of the Industry

An industry is in equilibrium when its market demand equals its market supply. When an industry is in equilibrium, all its firms are supposed to be in equilibrium [as shown in Fig. 11.4(b)]. When an industry is in equilibrium, all its firms earn only normal profits, because under the conditions of perfect competition all the firms are assumed to achieve the same level of efficiency in the long run. Since industry yields only normal profits, there is no incentive for new firms to enter the industry. These conditions are fulfilled at price OP' in Fig. 11.4(a) and (b).

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

Since $P = LAC$, all the firms are earning only normal profits. At industry's equilibrium output ON_2, market demand equals market supply [Fig. 11.4(a)]. At price *OP*¢, therefore, market is cleared. The output ON_2 has a fair chance to remain stable in the long run. For, there is no incentive for new firms to enter the industry and for existing ones to leave the industry. The industry is therefore in equilibrium.

NOTES

Check Your Progress

1. Who challenged the theory of perfect competition?
2. Who developed the theory of imperfect competition and the theory of monopolistic competition?
3. State the difference between perfect competition and pure competition.

11.6 ANSWERS TO 'CHECK YOUR PROGRESS'

1. The theory of perfect competition was challenged by Italian economist, Piero Sraffa.
2. Robinson and Chamberlin developed independently the theory of imperfect competition and theory of monopolistic competition, respectively.
3. Pure competition has no element of monopoly enabling a producer to change more whereas perfect competition has the same element of monopoly.

11.7 SUMMARY

- The theory of firms came into existence during the 1930s with Joan Robinson's *The Economics of Imperfect Competition* and Edwin H. Chamberlin's *The Theory of Monopolistic Competition*, both written independently in 1933.
- Robinson and Chamberlin developed independently the theory of imperfect competition and theory of monopolistic competition, respectively.
- A perfectly competitive market is characterised by complete absence of rivalry among the individual firms. In fact, under perfect competition as conceived by the economists, competition among the individual firms is so widely dispersed that it amounts to no competition.
- While 'perfect competition' has all the features mentioned above, 'pure competition' does not assume *perfect mobility* of factors and *perfect knowledge*. That is, *perfect competition* less *perfect mobility and knowledge* is pure competition. 'Pure competition' is 'pure' in the sense that it has absolutely no element of monopoly.
- Under perfect competition, market price in a perfectly competitive market is determined by the market forces, *viz*., demand and supply.
- The role of market forces and the mode of price determination depends on the time taken by supply position to adjust itself to the changing demand conditions. Price determination is analysed under three different time periods: (i) Market period or very short-run; (ii) short-run; and (iii) long-run.

NOTES

- Unlike in the short-run, the supply curve in the long run is supposed to be more elastic. Long-run brings in two additional factors in operation which make the supply curve more elastic.
- As in the short-run, market price is determined in the long-run by the market forces of demand and supply.

11.8 KEY WORDS

- **Welfare Economics:** It lays emphasis on the optimal allocation of resources and goods and how the allocation of these resources affects social welfare.
- **Homogeneity:** It is the quality or state of being all the same or all of the same kind.
- **Monopoly:** In economics, it denotes a market structure characterized by a single seller, selling a unique product in the market.

11.9 SELF-ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. Write a short note on Marshall's time analysis.
2. List the features of perfect competition.
3. Why is the supply curve in the long run supposed to be more elastic as compared to the short-run?

Long-Answer Questions

1. 'A perfect competitive market is characterized by complete absence of rivalry among the individual firms.' Discuss the statement.
2. Explain price determination in very short-run time period.
3. Give one example to illustrate the state of equilibrium in an industry.

11.10 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 12 MONOPOLY, MONOPOLISTIC COMPETITION AND OLIGOPOLY

Structure

12.0 INTRODUCTION

In Unit 11 you studied the price and output determination under the condition of perfect competition. In this unit, we explain price and output determination under another extreme form of market structure, i.e., monopoly. In contrast to perfect competition, monopoly is an extremely opposite form of a product market. In case of perfect competition, the number of sellers is so large that no one of them has any power whatsoever to influence the market price. A monopoly firm, on the other hand, has the sole power to influence the market price. While under perfect competition, no seller can afford to discriminate between the buyers of different categories, the monopolists practice price discrimination as a matter of policy. In this unit, you will study about discriminating monopoly, monopolistic competition, selling cost and oligopoly.

NOTES

12.1 UNIT OBJECTIVES

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

- Define monopoly
- Discuss demand and revenue curves under monopoly
- State the nature of monopolistic competition
- Prepare an analysis of selling cost and firm's equilibrium
- List the characteristics of oligopoly

12.2 MONOPOLY: DEFINITION AND SOURCES

12.2.1 Monopoly Defined

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

Moreover, the precise definition of monopoly has been a matter of opinion and purpose. For instance, in the opinion of Joel Dean, a monopoly market is one in which 'a product of lasting distinctiveness is sold.' The monopolised product has distinct physical properties recognised by its buyers and the distinctiveness lasts over many years. Such a definition is of practical importance if one recognises the fact that most of the commodities have their substitutes varying in degree and it is entirely for the consumers or users to distinguish between them and to accept or reject a commodity as the substitute. Another concept of pure monopoly has been advanced by D.H. Chamberlin who envisages the control of all goods and services by the monopolist. But such a monopoly has hardly ever existed, hence his definition is unrealistic. In the opinion of some others, any firm facing a sloping demand curve is a monopolist. This definition however includes all kinds of firms except those under perfect competition. We will, however, adopt for our purpose here a general definition of a pure monopoly: a pure monopoly means an absolute power to produce and sell a commodity which has no close substitute.

12.2.2 Sources and Kinds of Monopolies

The emergence and survival of a monopoly is attributed to the factors which prevent the entry of other firm into the industry. The barriers to entry are therefore the

NOTES

sources of monopoly power. The major sources of barriers to entry to a monopolised market are described here briefly.

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

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

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

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

12.2.3 Demand and Revenue Curves Under Monopoly

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

1. Demand Curve under Monopoly

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

2. Relation between *AR* and *MR*

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

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

$$TR = P \cdot Q$$

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

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

$$= P\left(1 + \frac{Q}{P}\,\frac{\partial P}{\partial Q}\right) \qquad ...(12.1)$$

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

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

By substituting $-\frac{1}{e}$ for $\frac{Q}{P} \cdot \frac{\partial P}{\partial Q}$ in Eq. 12.1, we get

$$MR = P\left(1 - \frac{1}{e}\right) \qquad ...(12.2)$$

NOTES

or $$MR = P - \frac{P}{e} \quad ...(12.3)$$

Since $P = AR$

$$MR = AR - \frac{AR}{e} \quad ...(12.4)$$

This relationship between *MR* and *AR* can be derived geometrically. Consider the *AR* and *MR* curves in Fig. 12.1.

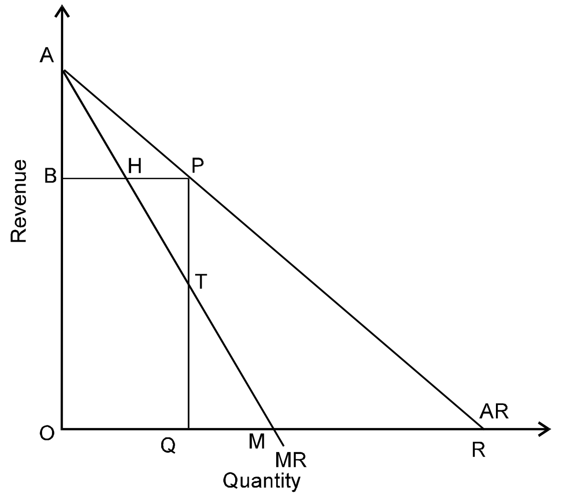


***Fig. 12.1** Relationship between AR and MR*

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

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

where *e* = elasticity of demand curve.

Since $OB = PQ$,

$$\therefore \quad e = \frac{PQ}{AB} \quad ...(12.5)$$

It can be proved that *AB* = *PT*. By substituting *PT* for *AB* in Eq. (12.5), we get

$$e = \frac{PQ}{PT} \quad ...(12.6)$$

Since *PT* = *PQ* – *TQ*, Eq. (12.6) may be written as

$$e = \frac{PQ}{PQ - TQ} \quad ...(12.7)$$

It can be seen from (Fig. 12.1) that at price *OB*, *PQ* = *AR* and *TQ* = *MR*. Therefore, Eq. (12.7) can be expressed as

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

and $MR = AR - \dfrac{AR}{e}$ ...(12.8)

NOTES

Note that Eq. (12.4) is the same as Eq. (12.8).

Given the Eq. 12.7, *AR* can be easily obtained.

Since $\quad MR = AR - \dfrac{AR}{e}$

or $\quad MR = AR\left(1 - \dfrac{1}{e}\right)$ ...(12.9)

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

or $\quad AR = MR\left(\dfrac{1}{e-1}\right)$ ...(12.10)

The general relationships between *AR* and *MR* are given by Eq. (12.9) and Eq. (12.10). A general pattern of relations between *AR* and *MR* can be easily obtained from Eq. (12.9) as follows. Given the negative slope of the demand curve,

when

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

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

12.2.4 Cost and Supply Curves under Monopoly

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

NOTES

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

12.2.5 Profit Maximisation under Monopoly

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

1. Monopoly Equilibrium in the Short Run

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

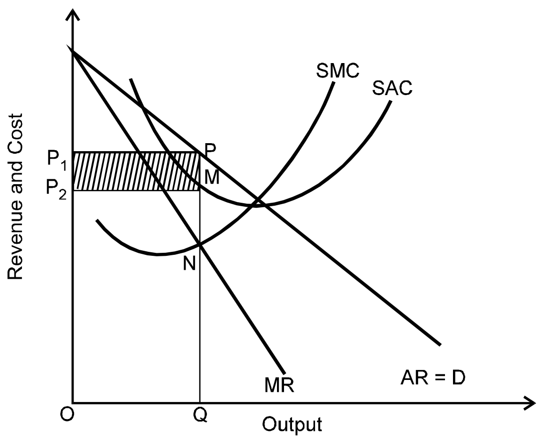


Fig. 12.2 *Price Determination under Monopoly: Short-run*

The price and output determination under monopoly, and also the firm's equilibrium, are demonstrated in Fig. 12.2. The *AR* = *D* and *MR* curves show the revenue conditions, while *SMC* and *SAC* present the short-run cost conditions faced by the monopoly firm. Given the revenue and cost curves, the decision rule for selecting profit maximising output and price is the same as for a firm in the competitive industry, i.e., firm's *MR* = *MC* and slope of *MC* > the slope of *MR*. Therefore, the monopoly firm chooses a price-output combination for which *MR*

NOTES

= *SMC*. The *MR* and *SMC* curves intersect each other at point *N*. Thus, the profit maximising output for the firm is *OQ*, since at this output firm's *MR* = *SMC*. Given the demand curve *AR* = *D*, the output *OQ* can be sold per time unit at only one price, i.e., *PQ*(= OP_1). Thus, the determination of equilibrium output simultaneously determines the price for the monopoly firm. Once price and output are determined, the total profits are also simultaneously determined.

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

2. Two Common Misconceptions

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

One of the misconceptions is that a monopoly firm necessarily makes super normal profits. There is however no guarantee that monopoly firm will always make profits in the short run. In fact, whether a monopoly makes profits or losses in the short run depends on its revenue and cost conditions. It is quite likely that its *SAC* lies above its *AR* as shown in Fig. 12.3. The monopoly firm then makes losses to the extent of $PM \times OQ = P_2MPP_1$. The firm may yet continue to produce and sell in the hope of making profits in the long-run. The monopoly firm, like a competitive firm, will however stick to the maximisation rules (i.e., *MR* = *MC*) in order to minimise its losses.

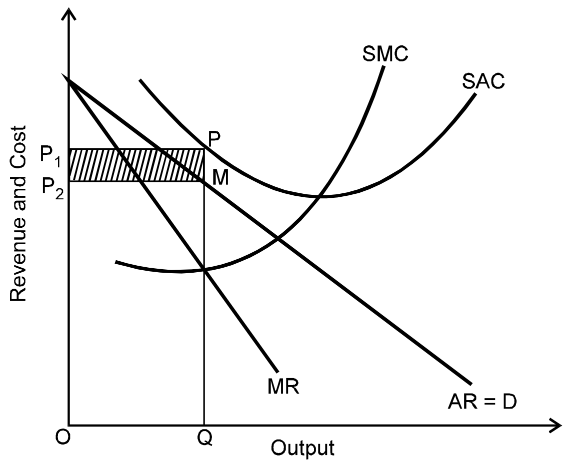


Fig. 12.3 *Monopoly Equilibrium in the Short-run: Losses*

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

3. Monopoly Equilibrium in the Long Run

NOTES

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

Protected by barriers to entry, a monopoly firm gets an opportunity to expand the size of its plant with a view to maximising its long-run profits. The expansion of the plant-size may however be subject to such conditions as (a) size of the market; (b) expected economic profits; and (c) risk of inviting legal restrictions. Assuming none of these conditions limits the expansion of monopoly firm, **the general case** of monopoly equilibrium in the long-run is illustrated in Fig. 12.4. The *AR* and *MR* curves show the market demand and marginal revenue conditions faced by the monopoly firm. The *LAC* and *LMC* curves show the log-run cost conditions. The profit maximising monopoly firm equalises its *LMC* and *MR* at output OQ_2. The price at which the total output OQ_2 can be sold is P_2Q_2. Thus, in the long run equilibrium, price is P_2Q_2 and equilibrium output is OQ_2. This output-price combination maximises the monopolist's long-run profits. The total monopoly profit is shown by the area LP_2SM.

It may be noted at the end that if there are barriers to entry, the monopoly firm would not reach the optimal scale of production in the long-run, nor will make full use of its existing capacity. This case can be verified from Fig. 12.4. The optimum size of the plant is given by point *B*, i.e., at the minimum *LAC*. But the monopoly firm settles at less than optimal output because optimum size of the plant will not yield the maximum profit.

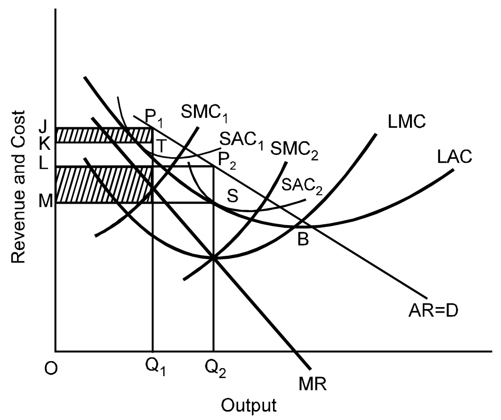


Fig. 12.4 *Monopoly Equilibrium in the Long-run*

Also, if the size of the market and the cost conditions permit, a profit maximising monopoly firm may even exceed the optimum size of the plant and overutilise its long-run capacity. Figure 12.5 depicts the more-than-optimal size of the plant and its overutilisation. The optimum size of the plant is given at point *B*, the point of intersection between *LAC* and *LMC*, whereas the monopoly firm chooses output at *M* where his profit is maximum. Alternatively, the monopoly firm may find its equilibrium just at the optimum size of the plant. This is possible only when the market-size is just large enough to permit optimisation and full utilisation of the plant size. This possibility has been illustrated in Fig. 12.6.

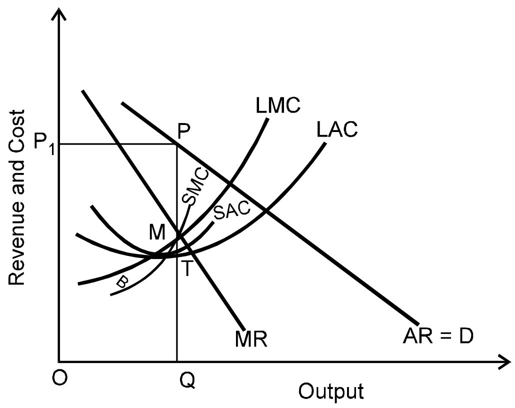


Fig. 12.5 *Monopoly Equilibrium: Overutilization of Point Size*

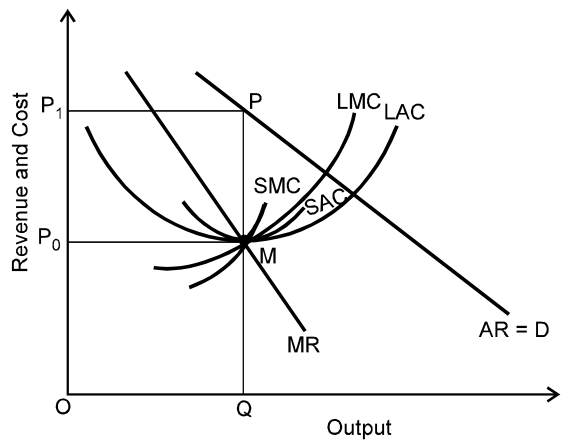


Fig. 12.6 *Monopoly Equilibrium at Optimal Size of the Plant*

12.2.6 Why Absence of Supply Curve under Monopoly

As already mentioned, there is no unique or precise supply curve under monopoly. Let us now examine this fact by using the concept of equilibrium output. We know that supply curve presents a unique relationship between price and quantity demanded. This unique relationship between market price and quantity supplied does not exist under monopoly. The reason is, a profit maximising monopoly firm does not determine its output where $P = MC$ or where $AR = MC$. Rather, it determines its equilibrium output where $MR = MC$. Therefore, a unique relationship between price ($AR = P$) and quantity supplied cannot be traced. It is therefore quite possible to trace (i) that given the *MC*, the same output is supplied at different prices, and (ii) that at a given price, different quantities are supplied if the two downward sloping demand curves have different elasticities. The two cases are illustrated in Figs. 12.7 and 12.8, respectively.

NOTES

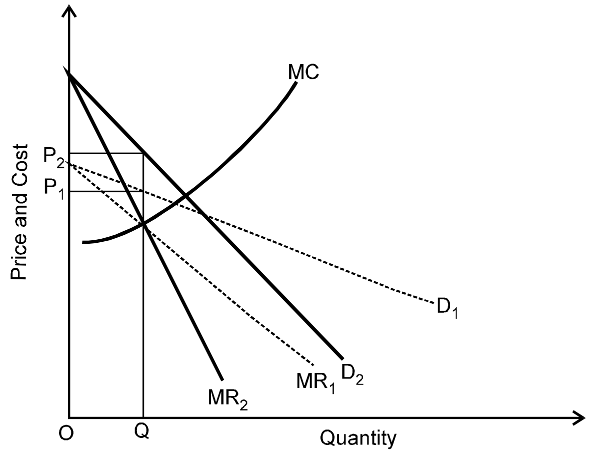


Fig. 12.7 *The Same Quantity Supplied at Two Different Prices*

As Figure 12.7 shows, given the *MC*, the same quantity *OQ* can be supplied at two different prices—OP_1 when demand curve is D_1 and OP_2 when demand curve is D_2. Obviously, there is no unique relationship between price and quantity supplied.

Figure 12.8 presents the case of two different quantities supplied at the same price, *OP*. Given the *MC*, quantity OQ_1 is supplied when demand curve is D_1 and quantity OQ_2 is supplied when demand curve is D_2 at the same price *OP*. In this case too, there is no unique relationship between price and quantity supplied. It is thus clear that there is no unique supply curve under monopoly.

12.2.7 Monopoly Vs. Perfect Competition: Comparison of Long-run Price and Output

We will confine ourselves to only long-run price and output under monopoly and perfect competition.

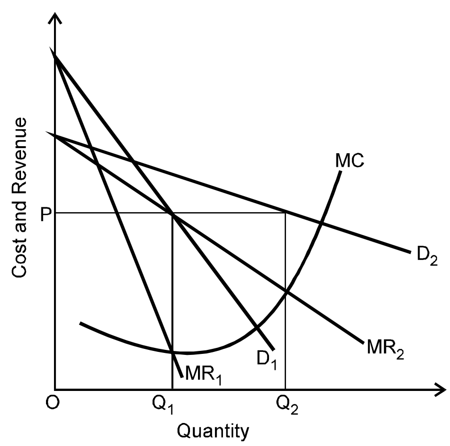


Fig. 12.8 *Different Quantities Supplied at the Same Price*

NOTES

Figure 12.9 presents a comparative analysis of equilibrium price and output under perfect competition and monopoly in the long-run. Let us assume that *LMC* and *LAC* are identical for both a competitive industry and a monopoly.

The equilibrium condition for a competitive industry in the long-run requires that all its firms are in equilibrium. That is, all the firms have their $AR = MR = LAC = LMC$. This condition is satisfied at point P' in Fig. 12.9. Thus, in a competitive industry, equilibrium price will be OP_1 and equilibrium output will be OQ_2. Now, if this industry were to be monopolised, the revenue conditions (*AR* and *MR*) and profit maximisation rule will be different. The monopoly firm will maximise its profits at the level of output where $MR = MC$. The equilibrium condition for the monopoly firm in fulfilled at point *B*. Therefore, the equilibrium output under monopoly will be OQ_1 and the equilibrium price will be OP''.

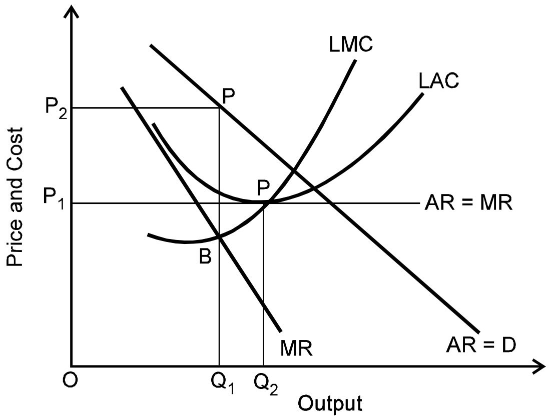


Fig. 12.9 *Comparison of Price and Output: Monopoly Vs Competitive Industry*

12.2.8 Two Major Conclusions

Two important conclusions can be drawn from the comparison of equilibrium price and output of monopoly and competitive industry—(i) monopoly results in an inoptimal output, and (ii) monopoly causes loss of social welfare. These conclusions are illustrated in Fig 12.10, assuming a constant cost industry.

Inoptimal of Output. As Fig. 12.10 shows, if both monopoly and competitive industries are faced with identical cost conditions, the equilibrium output under competitive conditions will be higher than under monopoly and price in the competitive industry will be lower than in monopoly. In other words, output under monopoly is lower and price higher compared to competitive industry.

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

NOTES

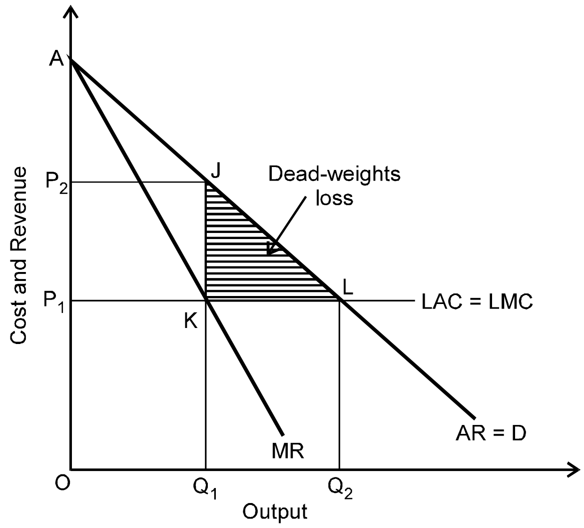


Fig. 12.10 *Price and Output under Monopoly and Perfect Competition*

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

12.2.9 Loss of Social Welfare

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

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

$$AP_1L - AP_2L = P_1JLP_1$$

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

12.3 THE NATURE OF MONOPOLISTIC COMPETITION

NOTES

The model of monopolistic competition developed by Edward H. Chamberlin presents a more realistic picture of the actual market structure and the nature of competition. In this unit, we discuss briefly the nature of the market structure and monopolistic competition among the firms.

Monopolistic Competition is a market structure in which a *large number* of sellers sell *differentiated* products which are close, but not perfect, substitutes for one another. Monopolistic competition combines the characteristics of perfect competition and monopoly.

The assumptions of the monopolistic competition are the same as those of pure competition, with an exception of homogeneity of products. While pure competition model assumes that products are *homogeneous* in every possible dimension, monopolistic competition model assumes that products are *differentiated*. The product of each firm is so differentiated from those of other firms that consumers are able to distinguish the product of a firm from those of others. For example, consumers know for sure the difference between different brands of mobile phones—Nokia, Sony, Samsung, Reliance, etc. Since each firm produces a product distinguishable from that of other firms, each firm holds a monopoly power over its own products.

Although products are differentiated, they remain a close substitute for one another. This creates condition for competition among the firms which are monopolists in their own rights. This kind of competition is the genesis of monopolistic competition.

12.3.1 Foundations of the Monopolistic Competition model

(i) Assumptions

Chamberlin's model of monopolistic competition is based on the following assumptions:

1. There is a large number of buyers and sellers in the market.
2. Each seller sells a product differentiated from that of others.
3. The differentiated products are close, not perfect, substitute for one another.
4. There is free entry and free exit of firms.
5. The firms seek to maximise their profits in both short and long runs.
6. Technology and factor prices are given and the firms are aware of revenue and cost curves.

NOTES

(ii) Product Differentiation and the Demand Curve

Chamberlin has defined *product differentiation* in the following words: "A general class of product is differentiated if any significant basis exists for distinguishing the goods (or services) of one seller from those of others. Such a basis may be real or fancied, so long as it is of any importance whatever to buyers, and leads to a preference for one variety of the product over another.... Differentiation may be based upon certain characteristics of the product itself, such as exclusive patented features, trade marks, trade names, peculiarities of the package or container, if any, or singularity in quality, design, colour or style. It may also exist with respect to the conditions surrounding its sales. In retail trade, these conditions include such factors as the convenience of the seller's location, the general tone or character of his establishment, his way of doing business, his reputation for fair dealing, courtesy, efficiency, and all the personal links which attach his customers either to himself or to those employed by him." So far as these and other tangible and intangible factors create consumers' preference for one product over the others, the products are virtually differentiated.

Thus, product differentiation is primarily intended to make consumers distinguish the product of one producer from that of the other producers in the industry. When the consumers are able to distinguish one product from the others, they may develop a preference or brand loyalty for one product over the others. Once preference for a product is created, it alters the course of demand curve for the product. In ultimate analysis, product differentiation leads to a change in demand curve for the product from a horizontal demand line (as under pure competition) to a downward sloping demand curve. The downward sloping demand curve enables the sellers to exercise some discretion in determining the price of his product.

(iii) Cost Curves and Selling Cost

In his model of monopolistic competition, Chamberlin has assumed the traditional U-shaped cost curves—*AC*, *AVC* and *MC*. In addition, he has introduced a new cost, i.e., *selling cost.* "Selling costs are defined as costs incurred in order to alter the position or the shape of the demand curve for a product." Selling costs include all the expenses that are intended to promote the sales, including cost of advertisement, salesmen's salaries, expenses of sales department, margins granted to dealers—wholesalers and retailers—and on window displays and demonstration of new goods. Selling costs affect demand curve in two ways.

First, selling costs make the demand curve for the product shift upward by informing consumers about the availability of the product and by increasing consumer's preference for the product.

Second, selling cost makes the demand curve less elastic by strengthening the consumers' preference for the product.

NOTES

Chamberlin assumes average selling cost to be U-shaped, that is, selling cost per unit of sales initially decreases but eventually increases. Thus, the average selling-cost curve has a shape similar to the *AC* curve.

(iv) The Concept of Industry and Product Groups

Under monopolistic competition, products are so differentiated that each product is distinguishable from others, and each firm is, in a sense, an industry in itself, exactly as a monopoly firm is an industry in itself. The heterogeneity of the products, therefore, causes a problem in analytical treatment of the industry. It may be recalled that, in case of homogeneous products, demand curve for an industry can be obtained by adding individual demand curve. But in case of heterogeneous products, the demand for individual products cannot be added to obtain market demand and supply curves.

For this reason, Chamberlin attempted to redefine the industry for his analytical purpose. He defined the monopolistically competitive industry as a 'group' of firms producing a 'closely related' commodity, referred to as *product group*. The product of the 'group' must be *close, technological and economic substitutes.* The two products are *technological substitutes* for each other if they technically satisfy the same want, e.g., personal computers, soaps, toothpastes, automobiles, TV sets, etc. The two products are considered as *economic substitutes* for each other if they satisfy the same want and have *more or less* the same price. For example, all brands of TV sets are economic substitutes for one another. But flat TV sets are not economic substitutes for ordinary ones since their prices are widely different, though they remain technological substitutes. Operationally, the product group may be defined as the group of firms whose products have between themselves high *price* and *cross elasticities*. This definition, although theoretical plausible, involves the problems of measuring cross-elasticities and of determining the degree of cross-elasticities that can make a commodity admissible to the group. Determining the product group would therefore involve subjective judgement.

12.3.2 Price and Output Determination under Monopolistic Competition

Chamberlin's theory of price and output determination under monopolistic competition is basically the same as that under monopoly with a difference, of course. While under monopoly demand and cost curves are both assumed to be given, under monopolistic competition, firms are assumed to indulge in competition to change the slope of the demand curve or to make it shift rightward, given the cost curves. They seek to make these changes in the demand curve by any or all of the following measures:

(i) change in the price of the product;

(ii) change in the nature of the product; and

(iii) change in the advertisement outlays.

NOTES

As to price change, since a monopolistically competitive firm faces an elastic demand curve similar to one faced by a monopoly firm, it has the option to raise the price and sell less or to lower the price and sell more. But, it fixes a price that maximises its profits. As to change in product, the sales can be promoted by change in the quality of product through technical changes, introduction of a new design, use of better material, use of new package or containers, prompt and courteous services, credit facilities, etc. Also, the firm may influence its volume of sales by increasing advertisement expenditure so that more consumers are attracted to the product. Increase in advertisement expenditure also increases the selling price. The firm is therefore required to so adjust its price and output that its profits is maximum.

While adjustment between price and output for profit maximisation is a short-run phenomenon, changes in the quality of the product and advertisement expenses are long-run phenomena. We will therefore explain Chamberlin's theory of price and output determination under monopolistic competition under both short-run and long-run conditions.

1. Firm's Short-run Equilibrium

While monopolistic competition is characteristically closer to perfect competition, it is closer to monopoly in regard to pricing and output determination. Like a monopolist, a monopolistic competitor faces a downward sloping demand curve having a smaller slope. This demand curve is the product of (i) strong preference of a section of consumers for a particular product; and (ii) the quasi-monopoly of the seller over the supply. The strong preference or loyalty of the consumers gives the seller an opportunity to raise the price and yet retain some customers. And, since each product is a close substitute for another, they attract the consumers of other products by lowering down their prices.

As mentioned above, short-term analysis of pricing and output determination under monopolistic competition is similar to price and output determination under monopoly. The short-term equilibrium analysis is primarily the adjustment of price and output to the given cost and revenue conditions. The short-run price and output adjustment is illustrated in Fig. 12.11. The $AR = D$ and MR curves show the revenue conditions and SAC and SMC curves show the cost conditions faced by the firm in the short-run.

As shown in Fig. 12.11, the necessary condition of profit maximisation, i.e., MR must be equal to MC, is fulfilled at output OQ. This output can be sold at price PQ, so the price is also determined. At this output and price, the firm earns a maximum economic profit, shown by the rectangle P_1PMP_2.

The economic profit per unit (PM) exists in the short-run because new firms cannot enter the industry. But the rate of profit would not be the same for all the firms under monopolistic competition because of difference in the elasticity of demand. For the same reason, product price will be different for the different products, though price differentials will be only marginal.

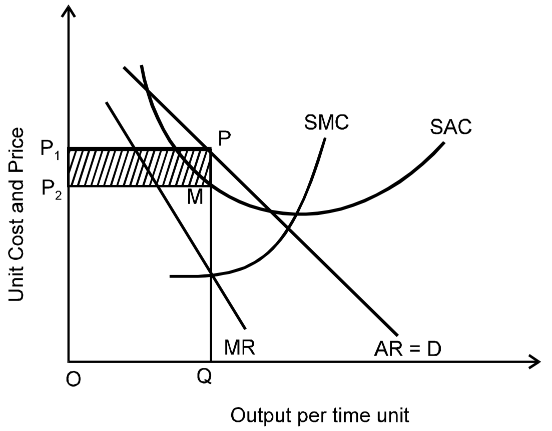


Fig. 12.11 *Short-run Equilibrium under Monopolistic Competition*

In the short run, the firms may attempt to maximise their profits by changing the nature of the product and by increasing advertisement expenditure. But, since there are many close substitutes, neither of the strategies would be of much avail in the short run. If the firms do adopt these strategies, they would do so only to maximise their profits.

2. Long-run Equilibrium under Monopolistic Competition

The conditions faced by a firm of the 'product-group' under monopolistic competitions are different from those in the short run. In the long run, a firm can expand its plant-size and if there are no barriers to entry, new firms will enter the product group, especially when existing firms are making pure profits. Besides, the firms get the opportunity to change the nature and position of the demand curve for their product by (i) changing the quality of the product; and (ii) incurring a large amount of advertisement expenditure. Since conditions change in the long run, the firms can maneuver their price and output in order to maximise their profits in the long run too.

Basic Assumptions

Chamberlin has made the following explicit and implicit assumptions in order to develop his theory of monopolistic competition under long-run conditions.

First, the basic assumption is that there is a large number of firms producing differentiated products which are close substitutes for one another.

Second, the number of firms in the product group is so large that each firm expects its maneuvering of prices and output to go unnoticed by the rival firms.

Third, one of the heroic assumptions of Chamberlin is that both demand and cost curves for all the products are uniform throughout the group. That is, all firms have identical revenue and cost curves.

Fourth, his second heroic assumption is that consumer's preferences are evenly distributed among the different varieties, and that differences between products are not such as to give rise to differences in cost.

Under these assumptions, Chamberlin develops three models of equilibrium:

(i) to analyse equilibrium with free entry of new firms to the industry with non-price competition;

(ii) to analyse equilibrium under price competition, assuming no entry or exit; and

(iii) to present a combined analysis of the first and the second models.

We will explain briefly the three models in the subsequent subsections.

(i) Long-run Equilibrium with Free Entry of New Firms

The long-run equilibrium of firms under the assumption of free entry of new firms is illustrated in Fig. 12.12. The *LAC* and *LMC* are the cost curves faced by the firms. The initial *AR* and *MR* curves (i.e., prior to the entry of new firms) are given by $AR = D_2$ and MR_2. Given the cost and revenue curves, the firms will be in short-run equilibrium at point *E* where price is OP_2 and output OQ_2. The firms are making super-normal profits to the extent of *EB* per unit of output.

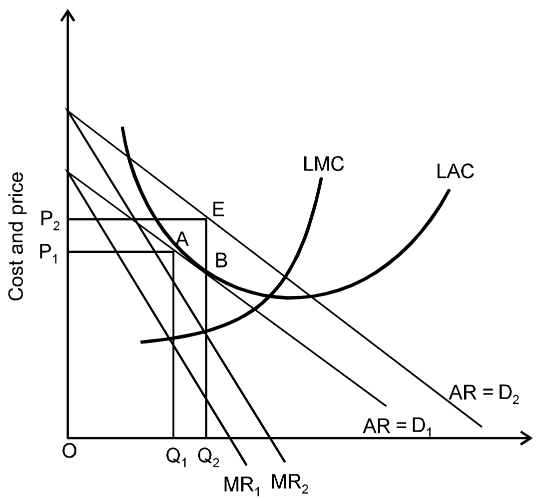


Fig. 12.12 *Introduction of Second Demand Curve*

The existence of super-normal profits attracts new firms to the product group. With the entry of new firms, the sale of each firm in the group decreases. Consequently, the demand curve (or *AR* curve) for the firm shifts leftward and so does the *MR* curve. This forces the firms to adjust their price and output to the new *AR* and *MR* conditions, cost conditions remaining the same. If still there exists super-normal profit more new firms join the product group, and existing firms will be forced to readjust their price and output to another equilibrium position. This process will continue until the demand curve becomes tangent to the *LAC*, and all the firms earn only normal profits. The final equilibrium position of the firms in the long-run under monopolistic competition is shown at point *A*, where price is OP_1 and output OQ_1. At this price and output, all firms make only normal profit. Therefore, there is no incentive for the new firms to enter the industry. The equilibrium will therefore be stable at point *A*.

(ii) Long-run Equilibrium with Price Competition

NOTES

In his model of long-run equilibrium with price competition Chamberlin assumes that the number of existing firms in the product group is optimal. That is, the number of firms is compatible with long-run equilibrium of the industry. There is no entry or exit of the firms. In this case, the equilibrium analysis has been accomplished in two stages. In the first stage, a second demand curve is introduced to the model to incorporate the effects of competitive change in prices. In the second stage, the long-run equilibrium under the condition of optimal number of firms has been analysed.

To commence the analysis of stage one, let us assume that demand curve for the product of a firm is given by DD_1, as shown in Fig. 12.13, and the firm is in equilibrium at point *E*, with price *OP* and output *OQ*. Now suppose, that the firm contemplates a price reduction and assumes that the rival firms will not react to its price cut. By reducing its price, the firm expects to expand its sales on account of two factors: *first*, the demand for its product is elastic, and *second*, the consumers of other products will switch over to the product of this firm if other firms do not simultaneously reduce their prices. Thus, the firm can expect a substantial increase in its total sales.

For example, if the firm *reduces* its price by *EM*, the demand for its product increases by *MB*. Of this, *MN* is due to the elasticity of its original demand DD_1 and *NB* is due to the fact that some customers of other firms switch over to the product of this firm. On the other hand, if the firm *raises* its price, and the rival firms do not raise their prices it looses its market partly because its demand curve is elastic and partly because its customers switch over to other products which become automatically cheaper. Thus, another demand curve DD_2 emerges. This is called as the *second demand curve*. But, if all the firms change their prices simultaneously this advantage to an individual firm is lost and all the firms return to the original demand curve DD_1.

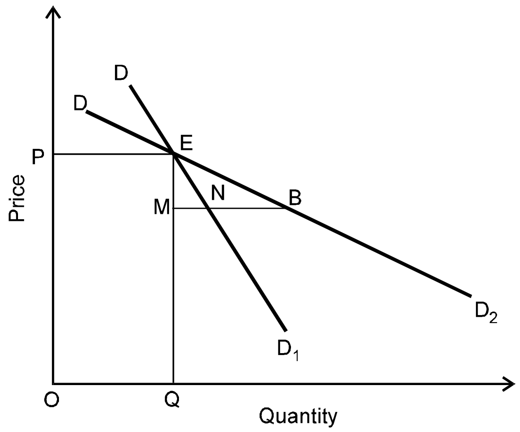


Fig. 12.13 *Introduction of Second Demand Curve*

NOTES

Having introduced the second demand curve, we can proceed to present Chamberlin's long-run equilibrium with price competition, i.e., *second stage* of his analysis. The long-run equilibrium with price competition is presented in Fig. 12.14. The curves DD' and dd_1 are the two demand curves and *LAC* is the long-run average cost curve of a 'typical' firm of the group. Let the initial short-run equilibrium of the firms of the group be at point *P* with price OP_2 and output OQ_1. At this price and output, the firm makes the total abnormal profit represented by the area P_2PMC.

Although in equilibrium, each firm regards dd_1 to be its demand curve and believes that it can increase its profits by reducing the price and expanding the output. The basis of this belief is the elasticity of their assumed demand curve dd_1. In an attempt to increase profits, each firm reduces its price expecting to move along the demand curve dd_1. However, instead of moving along dd_1, the firms move along the market demand curve DD', because all of them reduce their prices simultaneously. However, according to Chamberlin, the firms do not learn from their past experience and each firm sticks to its own belief that the demand curve (dd_1) for its product is more elastic than the market demand curve (DD'). Therefore, the firms go on reducing their prices independently and their assumed demand curve (dd_1) shifts downward. This process continues until the downward shift in dd_1 makes it tangent to the *LAC* curve, as shown by dd_2. A further reduction in price will make firms incur loss. Therefore, reduction in price below OP_1 is not desirable. Thus, the long-run equilibrium of firms takes place at *E*, where each firm produces OQ_2 and fixes its price at OP_1.

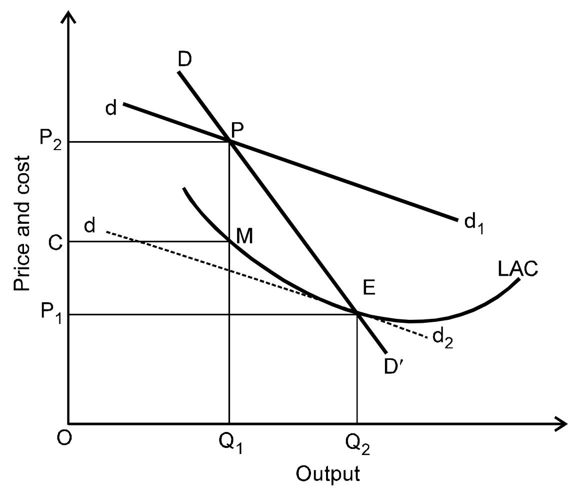


Fig. 12.14 *Long-run Equilibrium Price Competition*

(iii) Long-run Equilibrium with Free Entry and Price Competition

We have explained above the equilibrium of the firm with free entry and with price competition separately. In this section, we bring together the two equilibrium analyses, and explain Chamberlin's third and the final model of firms' equilibrium under monopolistic competition. According to Chamberlin, the ultimate equilibrium

NOTES

under monopolistic competition can be achieved through an integrated analysis of effects of free entry and price adjustments. The integrated analysis of equilibrium is presented in Fig. 12.15.

In Fig. 12.15, DD_1 is assumed to represent the initial demand curve and *LAC* to represent the long-run average cost curve. Let us suppose that the firms are initially in equilibrium at point *B*, and they make abnormal profits to the extent of vertical distance between DD_1 and the *LAC*. Since entry to the 'product group' is free, new firms are attracted by the industry. When new firms with slightly differentiated products enter the 'product group', the market share of each existing firm is reduced. Hence, their demand curve DD_1, shifts leftward. Given the *LAC*, the leftward shift in the demand curve will continue until it becomes tangent to *LAC*, as shown by DD_3 in Fig. 12.15, because till this point of tangency is reached, firms make abnormal profits and new firms continue to enter the 'product group'.

Thus, it might seem that the long-run equilibrium is attained at point *A* with output QQ_1 and price OP_3. This is however not the case. This is only half of the story, i.e., the influence of free entry. Let us now consider the competitive maneuvering of price and its role in determining to the long-run equilibrium.

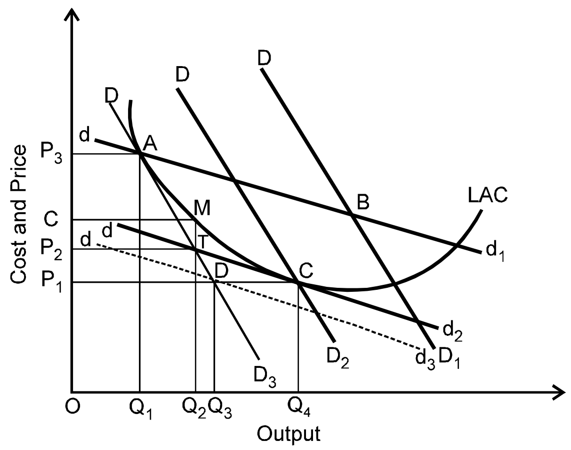


Fig. 12.15 *Long-run Equilibrium with Free Entry and Price Competition*

Once the firms reach point *A*, each firm thinks that its demand curve is dd_1, not DD_3. Each firm believes that it can increase its profit by reducing the price and thereby increasing the sales. Therefore, in their attempt to increase their profits, they reduce prices of their product simultaneously because each firm has the same incentive to do so. As a result, their subjective demand curve (dd_1) slides downward to dd_2 and they incur losses. For example, if price is reduced to OP_2, the total loss equals the rectangle $CMTP_2$. It might seem that the firms could eliminate their loss by reducing the price to OP_1. But when all the firms reduce their price to OP_1— and they will do so under the assumption—their subjective demand curve dd_2 slides further down to dd_3, the dotted line that lies below the *LAC*. As a result, the firms make increasing losses. A temporary equilibrium will be attained at point *D* with output OQ_3, where all firms incur heavy losses. Consequently, the firms which

NOTES

are unable to sustain losses will eventually leave the industry. The remaining firms find their share in the market increasing. Therefore, DD_3 and dd_3 move to the right until DD_3 shifts to DD_2 and dd_3 shifts upward to the position of dd_2. Note that dd_2 intersects DD_2 at point C the point where dd_2 is tangent to LAC. Thus, the long-run equilibrium is attained at point C where it is stable because all firms earn only normal profits and, therefore, there is no entry or exit of the firms.

12.4 ANALYSIS OF SELLING COST AND FIRM'S EQUILIBRIUM

As noted above, under monopolistic competition, products are differentiated; market is imperfect; consumers are not fully aware of existence of a particular variety of the product. Therefore, producers always have an opportunity to advertise their product, attract more customers to their product, and create brand loyalties in the minds of consumers and thereby increase their share in the market. In fact, the basic objective of advertising is to change the position and shape of the demand curve for the product of the advertising firm. Whether a firm succeeds in achieving these objectives depends also on the effectiveness of competitive advertising by the rival firms. But, one thing is obvious: advertising involves additional costs which pushes the *AC* curve upward.

Apart from advertisement expenses, monopolistically competitive firms, incur other costs on competitive promotion of their sales, e.g., salary to sales personnel, allowance to dealers, discounts to customers, expenses on displays, gifts to customers and free samples, additional costs on attractive packaging of goods, etc. All such expenses plus advertisement expenditure make a firm's *selling cost*.

Incurring selling cost increases sales, but at a varying degree. In the initial stage, the increase in sales may be greater than the proportionate increase in the selling cost. But eventually, it decreases. Consequently, the unit selling cost or the average selling cost (*ASC*) initially decreases (because of economies of scale) but ultimately increases. The *ASC* curve is therefore U-shaped, similar to the conventional short-run *AC* curve. It implies that total sales are subject to diminishing returns to increasing selling costs.

Selling Costs and Group Equilibrium

To analyse group equilibrium of firms with selling cost, let us recall that the main objective of the firm is to maximise its total profits. When they incur selling costs, they do so with the same objective in mind. All earlier assumptions regarding cost and revenue curves remain the same. The analysis of group equilibrium is presented in Fig. 12.16. Suppose *APC* represents the average production cost and price is given at OP_3. None of the firms incurs any selling cost.

NOTES

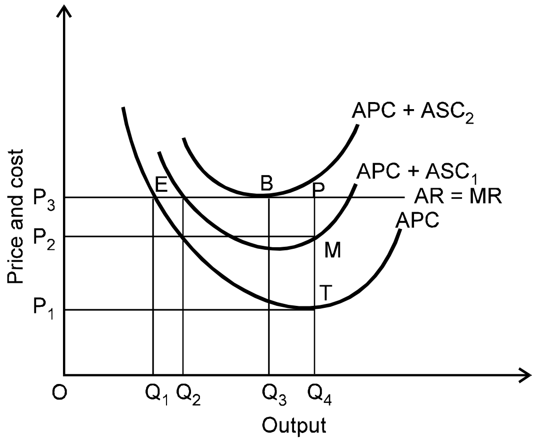


Fig. 12.16 *Selling Costs and Group Equilibrium*

Also, let all the firms be in equilibrium at point *E* where they make only normal profits. Now suppose that one of the firm's incurs selling cost so that its *APC* added with average selling costs (*ASC*) rises to the position shown by the curve $APC + ASC_1$ and its total sales increases to OQ_4. At output OQ_4 the firm makes supernormal profits of P_3PMP_2. This profit is however possible only so long as other firms do not advertise their own products. If other firms do advertise their products and incur the same amount of selling cost, the initial advantage to firm advertising first will disappear and its output will reduce to OQ_2. In fact, all the firms produce only OQ_2 units. But their short sightedness impels them to increase their selling cost because they expect to reduce their *APC* by expanding their output. With increased selling cost, their $APC + ASC$ curve shifts further upward. This process continues until $APC + ASC_2$ becomes tangent to the $AR = MR$ line at point *B*. Beyond point *B*, advertising is of no avail to any firm, even if other firms do not advertise. The equilibrium will be stable at point *B* where each firm produces OQ_3. Note that the equilibrium output OQ_3 is greater than the initial output of OQ_1. In equilibrium, however, firms make only normal profits.

12.4.1 Critical Appraisal of Chamberlin's Theory of Monopolistic Competition

Chamberlin's theory of monopolistic competition has been criticised on both theoretical and empirical grounds. Let us first look into its theoretical or methodological weaknesses.

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

Second, Chamberlin's model implicitly assumes that monopolistically competitive firms do not learn from their past experience. They continue to commit

NOTES

the mistake of reducing their prices even if successive price reductions lead to increase in their losses. Such an assumption can hardly be accepted.

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

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

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

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

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

12.5 OLIGOPOLY: MEANING AND CHARACTERISTICS

Oligopoly is a form of market structure in which a *few* sellers sell differentiated or homogeneous products. 'How *few* are the sellers' is not easy to define numerically in the oligopolistic market structure. The economists are not specified about a definite number of sellers for the market to be oligopolistic in its form. It may be two, three, four, five or more. In fact, the number of sellers depends on the size of the market. Given the size of the market, if number of sellers is such that each seller has command over a sizeable proportion of the total market supply[3] then there exists oligopoly in the market.

The products traded by the oligopolists may be differentiated or homogeneous. Accordingly, the market may be characterised by **heterogeneous**

oligopoly or **homogeneous (or pure) oligopoly**. In automobile industry, Maruti Zen, Hyudai's Santro, Daewoo's Matis, Fiat's Palio and Tata's Indica, etc., are the outstanding examples of differentiated oligopoly. Similarly, cooking gas of Indane and of Burshane are the examples of homogeneous oligopoly. Differentiated oligopolies include automobiles, cigarettes, refrigerators, TV industries. Pure oligopoly includes such industries as cooking gas, cement, baby food, vegetable oils, cable wires, dry batteries, etc. Other examples of oligopolistic industries are aluminium, paints, tractors, steel, tyres and tubes, etc.

12.5.1 Characteristics of Oligopoly

The basic characteristics of oligopolistic market structure are following:

1. **Intensive Competition.** The characteristic fewness of their number brings oligopolist in intensive competition with one another. Let us compare oligopoly with other markets structures. Under perfect competition, competition is non-existent because the number of sellers is so large that no seller is strong enough to make any impact on market conditions. Under monopoly, there is a single seller and, therefore there is absolutely no competition. Under monopolistic competition, number of sellers is so large that degree of competition is considerably reduced. But, under oligopoly, the number of sellers is so small that any move by one seller immediately affects the rival sellers. As a result, each firm keeps a close watch on the activities of the rival firms and prepares itself with a number of aggressive and defensive marketing strategies. To an oligopolist, business is a 'life' of constant struggle as market conditions necessitate making moves and counter-moves. This kind of competition is not found in other kinds of market. *Oligopoly is the highest form of competition.*
2. **Interdependence of Business Decisions.** The nature and degree of competition among the oligopolists makes them interdependent in respect of decision-making. The reason for inter- dependence between the oligopolists is that a major policy change made by one of the firms affects the rival firms seriously and immediately, and forces them to make counter-moves to protect their interest. Therefore, each oligopolist, while making a change in his price, advertisement, product characteristics, etc. takes it for granted that his actions will cause reaction by the rival firms. Thus, interdependence is the source of action and reaction, moves and counter-moves by the competing firms. An illuminating example of strategic maneuvering by the oligopoly firm has been given by Robert A. Meyer. To quote the example, one of the US automobile companies announces in September an increase of $ 180 in the list price of its new car model. Following it, a few days later, a second company announces an increase of only $80 and a third announces increase of $91. The first company makes a counter-move: it suddenly reduces the increase in list price to $71 from $180 announced earlier. One can now expect that other firms will follow

NOTES

the first in price-cutting. Obviously, there is a good deal of uncertainty in the behaviour of firms.

3. **Barrier to Entry.** An oligopolistic market structure is also characterised, in the long run, by strong barriers to entry of new firms to the industry. If entry is free, new firms attracted by the super-normal profits, if it exists, enter the industry and the market eventually becomes competitive. Usually barriers to entry do exist in an oligopolistic market. Some common barriers to entry are economies of scale, absolute cost advantage to old firms, price-cutting, control over important inputs, patent rights and licencing, preventive price and existence of excess capacity. Such factors prevent the entry of new firms and preserve the oligopoly.

12.5.2 Oligopoly Models: An Overview

The uncertainty in respect of behaviour pattern of a oligopoly firms arising out of their unpredictable action and reaction makes systematic analysis of oligopoly extremely difficult. Under the circumstances, a wide variety of behaviour pattern has been observed: they may come in collusion with each other or 'may try to fight each other to the death'. The agreement may last or may breakdown soon. Indeterminateness of price and output therefore becomes the basic feature of oligopolistic markets. In accordance with the variety of behaviours, economists have developed a variety of analytical models based on different behavioural assumptions. Among notable models are Cournot's Duopoly model (1838), Bertrand's model (1883), Edgeworth's model (1897), Stackelberg's leadership model (1930), Hotelling's model (1930s), Chamberlin's model (1933), Sweezy's kinked-demand curve model (1939), Neumann and Morgenstern's game theory model (1944), and Baumol's sales maximisation model. None of these models, however provide a universally acceptable analysis of oligopoly, though these models do provide insight into the behavioural pattern of oligopolists. Moreover, these models are studied for their pedagogic importance.

The analytical models of oligopoly, suggested by the economists, may be classified under two broad categories:

(i) duopoly models, and

(ii) general oligopology models.

Check Your Progress

1. What is the main objective of a monopoly firm?
2. List any two characteristics of an oligopolistic market structure.

12.6 ANSWERS TO 'CHECK YOUR PROGRESS' QUESTIONS

NOTES

1. The main objective of a monopoly firm is profit maximization.
2. Two characteristics of an oligopolistic market structure are the following:
    - Intensive competition
    - Interdependence of business decisions

12.7 SUMMARY

- The term *pure monopoly* signifies an absolute power to produce and sell a product which has no close substitute.
- The emergence and survival of a monopoly is attributed to the factors which prevent the entry of other firm into the industry. The barriers to entry are therefore the sources of monopoly power.
- The nature of revenue curves under monopoly depends on the nature of demand curve a monopoly firm faces.
- The objective of a monopoly firm, like all other firms, is assumed to be profit maximisation. Profit maximisation is however not necessarily the sole objective of the firm.
- *Monopolistic Competition* is a market structure in which a *large number* of sellers sell *differentiated* products which are close, but not perfect, substitutes for one another.
- Under monopolistic competition, products are so differentiated that each product is distinguishable from others, and each firm is, in a sense, an industry in itself, exactly as a monopoly firm is an industry in itself.
- Oligopoly is a form of market structure in which a *few* sellers sell differentiated or homogeneous products.

12.8 KEY WORDS

- **Monopoly Market:** A monopoly market is one in which there is only one seller of a product having no close substitute.
- **Oligopoly:** It is a form of market structure in which a few sellers sell differentiated or homogeneous products.
- **Patent Rights:** These rights are granted by the government to a firm to produce a commodity of specified quality and character or to use a specified technique of production.

12.9 SELF-ASSESSMENT QUESTIONS AND EXERCISES

NOTES

Short-Answer Questions

1. List the major barriers to entry to a monopolised market.
2. Write a short note on the nature of monopolistic competition.
3. Name the prominent oligopoly models.

Long-Answer Questions

1. Explain the relationship between average revenue and marginal with the help of a diagram.
2. 'Monopoly causes loss of social welfare and distortion in resource allocation.' Discuss the statement.
3. Discuss the foundations of the monopolistic competition model.
4. Prepare an analysis of selling cost and firm's equilibrium.
5. Critically analyse Chamberlin's theory of Monopolistic Competition.

12.10 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 13 DISTRIBUTION: THEORY OF FACTOR PRICING

NOTES

Structure

13.0 INTRODUCTION

The theory of factor pricing is not fundamentally different from the product pricing. Both factor and commodity prices are determined essentially by the demand and supply forces, though the factors which determine demand for and supply of commodities and factors are different. Demand curves for both commodities and factors are *derived* demand curves. While demand for a commodity is derived from its *marginal utility* schedule, demand for a factor is derived from its *marginal productivity* schedule. There are however differences on the supply side. While supply of a product depends mainly on its marginal cost, the supply of factors depends on a number of factors which vary from factor to factor. In this unit, we will discuss the demand and supply sides of the factor market.

As mentioned above, factor prices are determined by the interaction of demand and supply schedules. The demand and supply schedules of factors of production have certain peculiarities in contrast to those of commodities. It is therefore essential to derive the demand and supply schedules of factors of production. Derivation of demand curve for a factor of production is based on its marginal productivity. We will therefore discuss first the *marginal productivity theory* and derive the demand curve for a factor. The theory of marginal productivity applies to all the factors of production. However, for the sale of simplicity, we will consider labour as the variable factor. The demand and supply curves derived with reference to labour will be used to explain price determination of other factors also wherever relevant.

13.1 OBJECTIVES

NOTES

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

- Define derived demand
- Discuss the marginal productivity theory of distribution
- Explain derivation of individual labour supply curve

13.2 MARGINAL PRODUCTIVITY THEORY OF DISTRIBUTION

Demand for a factor is derived demand. It is derived on the basis of the marginal productivity of a factor. Firms demand factors of production—land, labour and capital—because they are productive. Factors are demanded not merely because they are productive but also because the resulting product has a market value. Thus, demand for a factor of production depends on the existence of demand for the goods and services that a factor of production can create. For example, land is demanded because there is demand for what land can produce—foodgrains, vegetables, fruits, etc. The services of doctors, engineers, managers are in demand because there is a demand for their services. Services of coalminers are demanded because there exists demand for coal. It is in this sense that the demand for a factor of production is a derived demand.

The demand curve for factors is derived under two different conditions:

(*i*) when a single variable factor is used in the process of production; and

(*ii*) when more than one variable factor is used in the production process.

Let us first derive the demand curve for a variable factor assuming that only one factor is involved in production process. We will next derive the demand curve for a variable factor when two factors—labour and capital—are used in the production process.

For deriving the demand curve for a factor, we assume that *labour* is the only variable factor. One may however assume any other factor. The process of deriving the demand curve for labour, presented in this chapter, can be applied to any other factor also. Furthermore, we outline, in this chapter, the theory of wage determination under perfectly competitive conditions. The same theory can be applied to explain the determination of price of all other factors.

13.2.1 Demand For a Single Variable Factor: Labour

The demand for a variable factor depends on the value of its marginal productivity. Therefore, we shall first derive the *value of marginal productivity* (*VMP*) curve of labour. The *VMP* curve is derived on the basis of its *marginal productivity* which is subject to the *law of diminishing returns*. The law of diminishing returns states that as more and more units of a factor are employed, all other factors remaining constant, the productivity of the marginal unit goes on diminishing. The marginal productivity curve of labour (MP_L) is derived on the basis of this law. The MP_L curve is shown in Fig. 13.1. The curve MP_L shows diminishing marginal returns to the variable factor—labour. The MP_L at each level of labour employment multiplied by a *constant* price of a commodity X, i.e., P_x, gives the *value of marginal physical product curve*, as shown by the curve $VMP_L = MP_L.P_x$. It is this curve which is the basis of demand curve for labour. Now the question is: how does the VMP_L curve help in deriving the demand curve for labour? Derivation of labour demand curve is illustrated and explained in the following section.

13.2.2 Derivation of Labour Demand Curve

Before we proceed to derive a firm's demand curve for labour, let us make the following assumptions for the sake of simplicity in the analysis.

(*a*) Firm's objective is to maximise profit.

(*b*) The firm uses a single variable factor, labour, whose market is perfectly competitive and hence the price of labour, wage rate (w), is constant and given for all the firms. This implies (*i*) that supply of labour for an individual firm is perfectly elastic; and (*ii*) that firm's $MC = w$ which is constant.

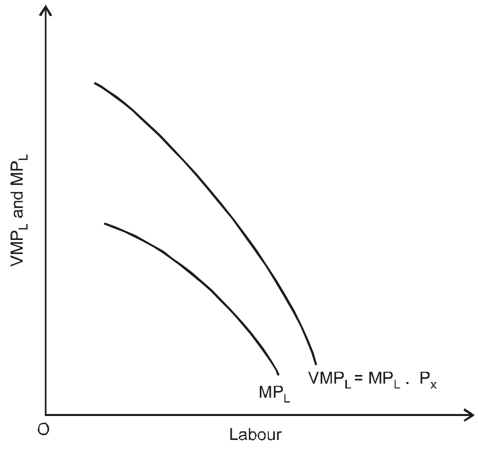


Fig. 13.1 *MP_L and VMP_L*

(*c*) The firm produces a single commodity, say X, whose price is constant at P_x.

Given the assumptions and the VMP_L curve, as shown in Fig. 13.1, we can now derive the firm's demand curve for labour. We have noted earlier that a profit maximising firm produces a quantity of output at which its $MR = MC$. This profit-maximisation rule can be interpreted as *a profit-maximising firm increases its output upto the point at which the marginal cost of labour employed equals the value of its product.* In other words, a profit-maximising firm employs a factor till the marginal cost of the variable factor (labour) equals the value of its marginal product (i.e., VMP_L).

The short-run equilibrium of the profit-maximising firm is illustrated in Fig. 13.2. The VMP_L curve shows the value of marginal product of labour. The S_L

lines represent the labour supply curves for an individual firm [assumption (b)], at different constant wage rates. The VMP_L curve and SL_3 line intersect at point E_3, where $VMP_L = W_3$, that is, where the value of marginal product equals marginal cost of labour. The profit-maximising firm will therefore employ only OL_1 units of labour. By employing OL_1 units of labour, the firm maximises its profit. Thus, given the VMP_L and SL_3, the profit maximising firm will demand only OL_1 units of labour.

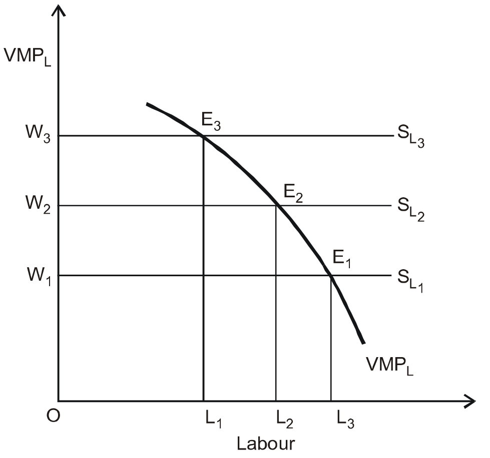


Fig. 13.2 *MP_L and VMP_L Curves*

If wage rate falls to OW_2 firm's equilibrium point shifts from point E_3 to E_2 increasing the demand for labour from OL_1 to OL_2. Similarly, when wage rate falls to OW_1, firm's equilibrium shifts downward to E_1 causing an increase in the demand for labour to OL_3.

To summarise, when wage rate is OW_3, demand for labour OL_1; when wage rate falls to OW_2, demand for labour increases to OL_2; and when wage rate falls further to OW_1, labour demand increases to OL_3. Obviously, as wage rate falls, demand for labour increases. This relationship between the wage rate and labour demand gives a usual downward sloping demand curve for labour, which is, by definition, the same as VMP_L curve. It may now be concluded that *individual demand* curve for a single variable factor (e.g., labour) is given by its value of marginal product curve (VMP_L) or its marginal revenue product curve (MRP_L).

13.2.3 Demand for Labour with More than One Variable Factor: Two Factor Case

When more than one variable factor is used by a firm in the process of production, the *VMP* curve of a variable factor is not the firm's demand curve for the factor. The reason is that *VMP* curve is drawn on the assumption that there is only one factor, other factors held constant. When more than one factor is variable, a change in the price of one variable factor leads to a change in the demand for other factors through its three kinds of different effects, viz., *substitution effect, output effect* and *profit effect*. Let us first discuss the substitution and output effects.

NOTES

Measuring Substitution and Output Effects: Substitution effect arises due to change in relative factor prices. If price of one factor changes, relative prices of factors change: while one actor becomes relatively cheaper, others become relatively costlier. The profit maximising firms then substitute cheaper factor for the costlier ones. Therefore, the firm's demand curve for a factor will be more elastic than the *VMP* curve. How high is the elasticity of substitution depends on the ease with which one factor can be substituted for another. The ease with which one factor can be substituted for another depends on the technical condition of production.

The influence of *substitution effect* on the demand for a variable factor is illustrated in Fig. 13.3 by using *isoquant* and *isocost* curves. For the sake of convenience, let us assume (*i*) that there are only two variable factors, labour (*L*) and capital (*K*) used in the production process of a commodity; (*ii*) that the initial wage rate is given as w_1, and price of capital as r_1; and (*iii*) that initial isocost is given by the line *AB* and initial isoquant by the curve Q_1.

Goving by the rule of outut maximization, the firm finds its equilibrium where isocost, *AB*, is tangent to isoquant Q_1. Thus, as Fig 13.3 shows, the firm is in equilibrium at point *P* where it uses OL_1 of labour and OK_2 of capital to produce Q_1 units of a commodity. Now let the wage rate fall from w_1 to w_2 so that the new isocost is *AD*. Consequently, the firm moves on to a new equilibrium point *R* at isoquant Q_2. The movement from point *P* to *R* indicates increase in the use of labour by L_1L_3. This increase in labour employment is the result of *substitution effect* and *output effect* of change in wage rate. Thus, the movement from point *P* to *R* can be decomposed into *substitution* and *output* effects. The two effects can be split by drawing an imaginary isocost ($A'B'$) parallel to the isocost *AD* and tangent to the original isoquant Q_1. The significance of the isocost $A'B'$ is that it removes the *output effect* of the fall in wage rate and gives the measure of the *substitution effect.* It tells how much of labour is substituted for how much of capital. As can be seen from Fig. 13.3, the movement from point *P* to *N* on the isoquant Q_1 shows substitution of L_1L_2 units of labour for K_1K_2 units of capital or *MN* units of labour for *PM* units of capital. It is clear that L_1L_2 is the substitution effect of decrease in price of labour.

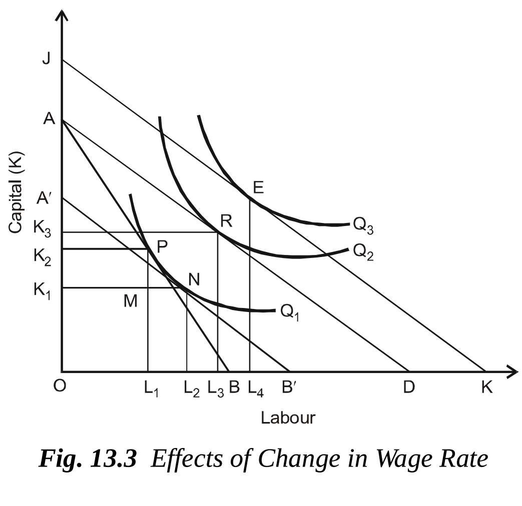


Fig. 13.3 *Effects of Change in Wage Rate*

NOTES

The *output effect* can now be obtained as wage effect *less* substitution effect. Thus, output effect = $L_1L_3 - L_1L_2 = L_2L_3$. It shows that the movement from point *N* to *R* is the *output effect*. Since the firm will ultimately settle at equilibrium point *R*, it will use more of both labour and capital. The increase in capital employment with the increase in labour increases marginal product of labour. Consequently, the VMP_L shifts to the right due to *output effect*.

Measuring Profit Effect: Let us now return to the *profit effect*. The movement from point *P* to *R* in Fig. 13.3, accounts for only *substitution* and *output effects*. These effects do not account for the *profit effect* of fall in wage rate. The profit effect of a fall in wage rate is illustrated in Fig. 13.4. The profit effect arises due to the downward shift in the *MC* curve due to fall in wage rate as shown in Fig. 13.4. For example, given the commodity price at *OP* and the marginal cost curve *MC* intersecting at point *E* determine the profit maximising output at *OQ*. Let us suppose that output *Q* is the same output level indicated by Q_2 in Fig. 13.3. When wage rate falls, the *MC* curve shifts downward to the position of *MC*′. In a perfectly competitive market, the equilibrium shifts from *E* to *E*′ and the profit maximising output increases from *OQ* to *OQ*′. The profit maximising firm will therefore have to increase its output by *QQ*′. The expansion of output requires an additional expenditure on labour and capital. The increase in expenditure will make the isoquant *AD* shift upward to *JK* (Fig. 13.3), and the firm will be finally in equilibrium at point *E* at isoquant Q_3. At equilibrium *E*, the total demand for labour is OL_4 of which L_2L_3 is the additional demand for labour caused by the output effect, and L_3L_4 is the *profit maximisation effect*. The output and profit maximisation effects, both being positive, lead to additional employment of capital. Thus, the employment of both labour and capital increases simultaneously as a result of output and profit maximisation effects. Increase in capital together with labour leads to increase in the marginal product of labour (MP_L). Consequently, in the final analysis, the decrease in wage rate causes a rightward shift in the VMP_L curve.

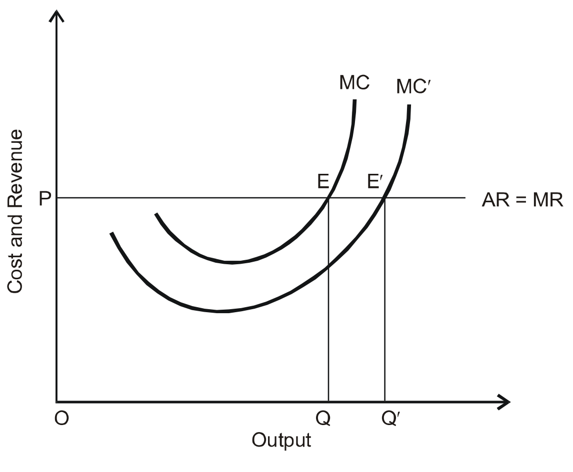


Fig. 13.4 *Profit Maximisation Effect*

13.2.4 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. 13.5. 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).

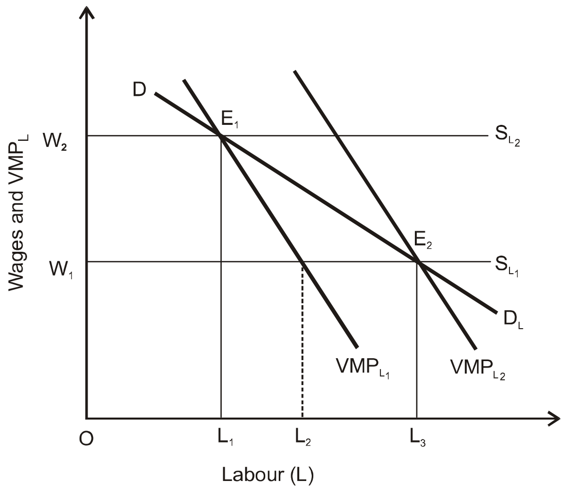


Fig. 13.5 *Labour Demand Curve when both L and K are Variable*

13.2.5 Market Demand Curve For a Variable Factor: Labour

As first approximation, the market demand for a variable factor is the horizontal summation of the individual demand curves for a factor. However, this process leaves out the *market* or *external* effects of change in the factor prices on factor demand. The market or external effects are the effects of change in the price of a factor on the price of the commodity which it produces and its repercussions on the demand for the factor. In a purely competitive market, when the price of a variable factor (say, labour) decreases, all the firms employ more of labour and hence the supply of the commodity they produce (say X) increases. As a result, the supply curve of the commodity X, shifts rightward which leads to a fall in the price of the commodity, P_x. The fall in P_x causes a decrease in the VMP_L. For example, if P_{x_1} and P_{x_2} are the two original and new prices, respectively, VMP_{L_1} =

$P_{x_1} . MPP$, and $VMP_{L_2} = P_{L_2} . MPP$. Since $P_{x_1} > P_{x_2}$, $VMP_{L_1} > VMP_{L_2}$. Thus, the fall in P_x causes VMP_L to shift downward. Therefore, market demand curve for a variable factor cannot be obtained directly by summing the individual demand curves for it.

The derivation of market demand curve for a variable factor is illustrated in Fig. 13.6. Suppose that curve dd_1 in Fig. 13.6 (a) is a typical individual demand curve for labour, and the initial wage rate is given at OW_2. The profit maximising firm is in equilibrium at point *P* and employs Ol_1 units of labour. Supposing all firms employ the same (Ol_1) units of labour, the market demand for labour can be obtained by multiplying Ol_1 with the number of firms which, let us suppose, equals OL_1 in Fig. 13.6 (b). Thus, we get point *J* as one point of market demand curve for the labour. Now let the wage rate fall to OW_1. Other things remaining the same, the firms would move down to point *T* on demand curve dd_1 and their employment of labour increases to OL_3. But other things will not remain the same. The increase in labour employment by all firms—each using l_1l_3 additional labour—the total supply of the commodity increases. Consequently, commodity supply curve would shift rightward causing a fall in commodity price, P_x. Following the fall in P_x, VMP_L shifts leftward to dd_2. So the new demand curve is dd_2 and the new equilibrium point is *M* where demand for labour is Ol_2 [Fig. 13.6 (a)]. Thus, market demand for labour equals Ol_2 multiplied by the number of firms, i.e., the total market demand for labour equals $Ol_2 \times P_{x2} = OL_2$. Thus, we get a new point, *K* in Fig. 13.6 (*b*). Point *K* represents the market demand for labour at wage rate OW_1. By joining points *J* and *K*, we obtain the market demand curve for labour.

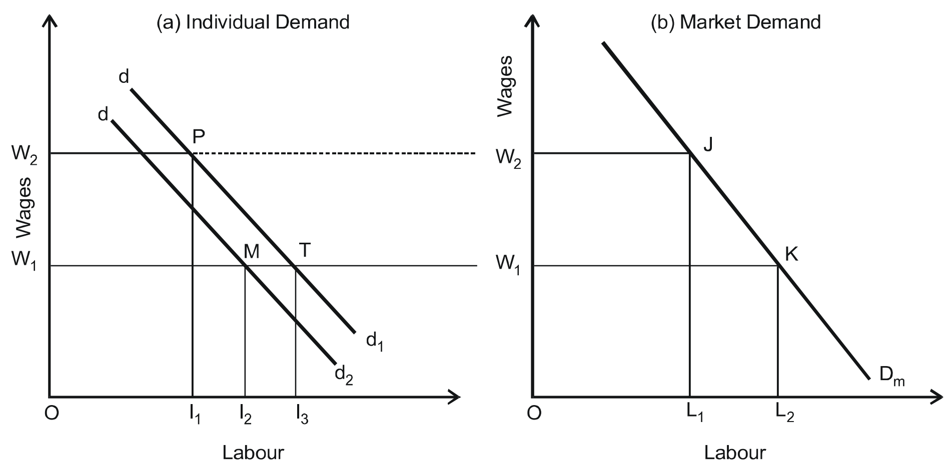


Fig. 13.6 *Derivation of Market Demand Curve for a Variable Factor: Labour*

13.3 THE FACTOR SUPPLY: AN OVERVIEW

We have explained above how demand curve for a factor is derived. In this section, we explain the supply aspects of factor inputs.

NOTES

The factors of production are traditionally classified as (*i*) land (including all natural resources); (*ii*) labour including all human (productive) efforts; and (*iii*) capital (including all man-made means of production). Let us look into the *m*eaning and supply aspects of each category of these factors of production.

(*a*) ***Land:*** The classical economists "treated land as a 'free gift of nature', a special factor of production distinct from man-made means of production and reproducible human labour." Land as a means of production has been used in three different senses: (*a*) the area of land within the territories of a country; (*b*) the area of fertile land available for cultivation; and (*c*) natural resources found under and above the ground.

(*b*) ***Capital:*** Capital is a man-made factor of production. It is reproducible. The stock of capital in a country consists of plant, machinery, building, etc. The stock of capital diminishes at the rate of its wear and tear in the process of production. Maintaining capital stock intact requires replenishment of the capital stock. The stock of capital can be increased by increasing investment in capital production. The production of capital goods, like that of consumer goods is subject to diminishing returns. That is, more of capital goods can be produced only at an increasing marginal cost. Here it suffices to note that capital supply curve is positively sloped like commodity supply curves.

(*c*) ***Labour:*** Labour may be defined as a physical and mental effort to produce goods, services, ideas and techniques.The manpower of a country consists of the persons in the age-group of 15–65 years. The *labour force* of a country consists of the number of its people willing to work at the prevailing wage rate. The total number of hours they are willing to work, given the reward per time unit, is called the *supply of labour.* The total supply of labour depends on (*a*) the size of population; (*b*) the proportion of population willing to work; and (*c*) number of hours which they are willing to work at a given wage rate. The supply of labour has some peculiarities in contrast to supply of other factors. It, therefore, requires a detailed treatment. The derivation of the labour supply curve is illustrated in the following section.

Derivation of Individual Labour Supply Curve

Let us assume that all labour is homogeneous; labour units are identical; and, unlike other factors, labour finds a trade-off between hours of work or leisure and income.

NOTES

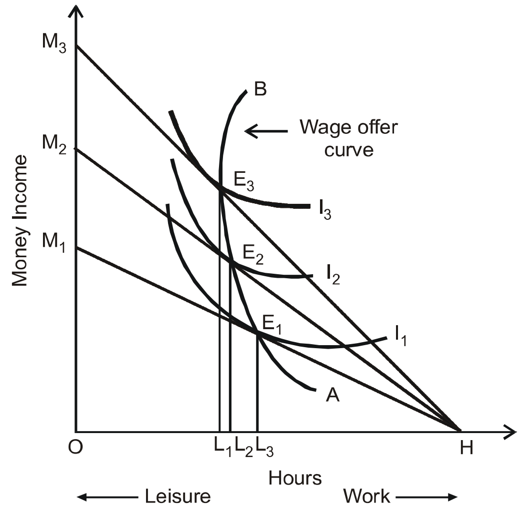


Fig. 13.7 *Labour Supply Curve*

Figure 13.7 illustrates the derivation of labour supply curve. Money income is measured on the vertical axis and the hours available to a worker per time-unit (per day or per week) are measured on the horizontal axis. One can also read the hours of work on the horizontal axis. The hours of work equals the total hours minus the hours of leisure. Let us now suppose that the total number of hours available to an individual is *OH* which he can use either for leisure or for work. If he works for *OH* hours and enjoys no leisure, then he makes OM_1 income. The wage rate (w_1) may then be obtained as

$$w_1 = \frac{OM_1}{OH} = \text{Slope of } M_1H \text{ line}$$

The indifference curve I_1 represents the leisure-income preference function of the individual. Let the individual be in equilibrium at point E_1 by working for L_3H hours and enjoying leisure for OL_3 hours. Note that $L_3H + OL_3 = OH$, the hours available to the individual. When the wage rate rises to $w_2 = OM_2/OH$, the individual moves to a new equilibrium point E_2. At this equilibrium point, he works for L_2H hours and has leisure of OL_2 hours and earns income E_2L_2. Note that with increase in wage rate the supply of labour by the individual increases. Similarly, when wage rate increases further to $w_3 = OM_3/OH$, the individual moves to equilibrium point E_3 where he works for L_1H hours, reduces his leisure from OL_2 to OL_1. By joining the equilibrium points E_1, E_2 and E_3, we get the *wage-offer curve*. The curve *AB* is essentially the labour supply curve, though its shape is apparently unusual. The information contained in curve *AB* may be regraphed as in Fig. 13.8 (a), to obtain the normal labour supply curve. The curve *SS′* is the normal positively sloped supply curve of labour. If wage-rate continues to increase above w_2, the labour supply curve *SS′* will be a backward bending one, as shown by curve *SL* in Fig. 13.8 (b). This is so because higher wages create disincentive for longer hours of work and incentive for increasing hours of leisure.

NOTES

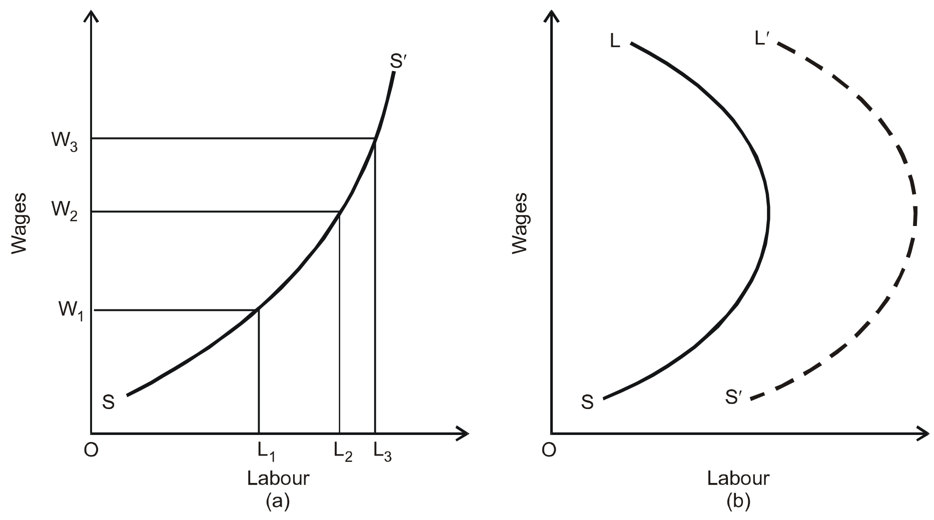


Fig. 13.8 *Labour Supply Curve*

13.4 DERIVATION OF MARKET LABOUR SUPPLY CURVE

The market labour supply curve is the sum of individual labour supply curves. Let us suppose that all the labour is homogeneous and each labour has the same labour supply curve as shown in Fig. 13.8 (b). In that case the total or market labour supply curve will also be of the same nature as shown in Fig. 13.8 (b). The only difference it makes is that summing up the individual labour supply curves makes the market labour supply curve shift rightward as shown by the dotted labour supply curve, $S'L'$.

Economists however disagree on the shape of the market supply curve for labour. Different shapes have been suggested for short- and long-run labour supply curves, depending on the occupational mobility, the type of labour and level of economic growth. In the short-run, if a firm uses specific type of labour, "nothing can be said about the slope or shape of the labour supply curve. It may be negative, or it may have segments of positive and negative slope." In the long-run, however, the market supply of specialised labour is likely to have a positive slope because (*a*) occupational mobility of labour increases; (*b*) with the increase in population and expansion of education and training facility, supply of all types of labour *increases*; and (*c*) with the increase in information about job market, young people are able to plan their education and career in accordance with job market conditions.

In case of non-specialised labour, the shape of labour supply curve is more clear. It is generally positively sloping for two reasons.

First, industries which plan to employ more labour, can draw it from other industries by offering higher wages. Hence the labour supply to such industries increases only at higher wage rates.

NOTES

Second, output decreases in the industries which lose their labour. The fall in the output of such industries leads to a rise in the price of their products causing an upward pressure on the demand for and price of labour. Thus, such industries too obtain larger supply of labour at increasing wage rates. For these reasons, the industries attempting to employ larger units of labour must face a labour supply curve with a positive slope. There is a likelihood that the market labour supply curve in an affluent country is a backward-bending one. For, when individuals become richer and richer, they begin to prefer smaller number of working hours, longer holidays, and fewer hours of work per day, and so on, as they need longer time to enjoy their high earnings. This likelihood would however be limited to the rich nations in which the scope for further expansion is limited. It is therefore generally believed that labour supply curve in the long-run must have a positive slope.

Check Your Progress

1. How is the law of diminishing returns related with the value of marginal productivity?
2. Mention the traditional classification of factors of production.

13.5 ANSWERS TO 'CHECK YOUR PROGRESS' QUESTIONS

1. The value of marginal productivity curve is derived on the basis of its marginal productivity which is subject to the law of diminish returns. The law of diminishing returns states that as more and more units of factor are employed, all other factors remaining constant, the productivity of the marginal unit goes on diminishing.
2. The factors of production are classified (i) land (ii) labour (iii) capital.

13.6 SUMMARY

- The demand curve for factors is derived under two different conditions:
  (*i*) when a single variable factor is used in the process of production; and
  (*ii*) when more than one variable factor is used in the production process.
- The factors of production are traditionally classified as (*i*) land (including all natural resources); (*ii*) labour including all human (productive) efforts; and (*iii*) capital (including all man-made means of production).
- Labour may be defined as a physical and mental effort to produce goods, services, ideas and techniques.The manpower of a country consists of the persons in the age-group of 15–65 years.

13.7 KEY TERMS

- **Isoquant:** It is a curve that shows all the combinations of inputs that yield the same level of output.
- **Labour:** In economics, labour is a primary factor of production.
- **Isocost:** In economics, an isocost line shows all combinations of inputs which cost the same total amount.

13.8 SELF-ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. Write a short note on the marginal productivity theory of production.
2. Prepare an overview of the supply aspects of factor inputs.
3. Define derived demand.

Long-Answer Questions

1. 'The demand for a variable factor depends on the value of its marginal productivity.' Discuss.
2. Explain the derivation of labour demand curve.
3. Illustrate the derivation of labour supply curve with the help of a diagram.

13.9 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 14 THEORIES OF RENT AND QUASI-RENT, WAGES, INTEREST AND PROFIT

Structure

14.0 INTRODUCTION

In the previous unit, you studied about the theory of pricing and marginal productivity theory of distribution. In this unit, you will study about the theories of rent and quasi-rent, theories of wages, theories of interest and theories of profit. 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.

14.1 UNIT OBJECTIVES

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

- Explain wage determination under perfect competition
- Define wage differentials
- State monopolistic exploitation of labour
- Discuss the Ricardian theory of Rent
- Define the term quasi-rent
- Discuss the various theories of interest and theories of profit

14.2 THEORY OF WAGE DETERMINATION

14.2.1 Wage Determination under Perfect Competition

Under the condition of perfect competition in both labour and product markets, wages are determined by the demand for and supply of labour.

Wage determination under the condition of perfectly competitive market structure is illustrated in Fig. 14.1. 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.

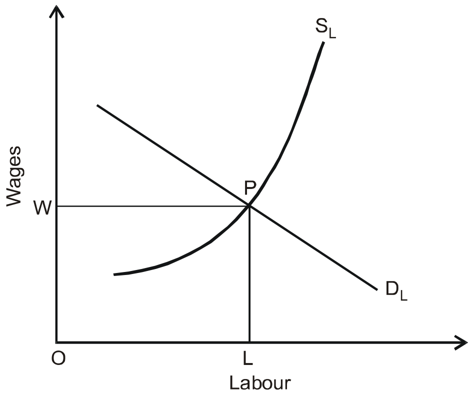


Fig. 14.1 *Determination of Wages in a Perfectly Competitive Market*

14.2.2 Wage Differentials

The theory of wage determination assumes that the units of a factor (labour) are homogeneous. If all the units of labour are identical and if non-monetary advantages given to labour 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.

There are two main reasons for wage differentials: (*i*) market disequilibrium causing dynamic and static wage differentials, and (*ii*) wage differentials arising due to non-homogeneity of labour.

14.2.3 Dynamic and Static Wage Differentials

Wage differentials due to market disequilibrium may be classified under two groups:

(*i*) Dynamic Wage Differentials, and

(*ii*) Static Wage Differentials.

(*i*) *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 of self-adjusting nature, 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*.

(*ii*) *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:

- qualitative differences in labour, i.e., non-homogeneity of labour;
- difference in the nature of occupations in which labour is employed;
- differences in the prices of product produced by labour; and
- 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***.

14.2.4 Determination of 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. It is important to remember that 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.

14.2.5 Monopolistic Exploitation of Labour

Monopoly in the product market and perfect competition in the labour market creates conditions for the *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.

The exploitation of labour (a variable factor) by an individual monopoly firm is illustrated in Fig. 14.2 and by the monopolistic firms as a group in Fig. 14.3. As shown in Fig. 14.2, 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 (= 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 $= EL_1$. 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

NOTES

competitive firm would have employed OL_2 units of labour. Not only that the lower level of employment by a monopolist also results in loss of output.

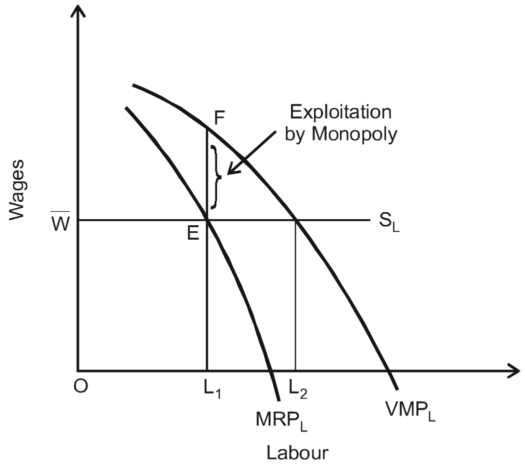


Fig. 14.2 *Exploitation of Labour by a Monopoly Firm*

Figure 14.3 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, especially under imperfect competition. In case of 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 lower than wage rate OW_C under a perfectly competitive product market. Thus, according to Joan Robinson, $OM_C – OW_M = W_CW_M$ is the extent of *monopolistic exploitation*.

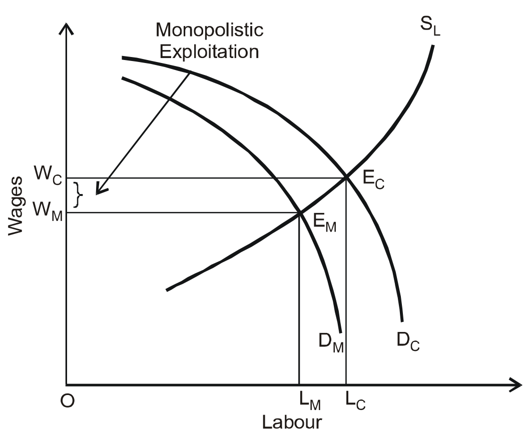


Fig. 14.3 *Exploitation of Labour under Monopolistic Competition*

NOTES

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.

14.3 THEORY OF RENT AND QUASI-RENT

Let us go through the various theories of rent and quasi-rent.

14.3.1 Ricardian Theory of Rent

Ricardo defined rent as "that portion of the produce of earth which is paid to the landlord for the use of the *original* and indestructible powers of soil". Ricardo considered payment of rent as an indication of niggardliness of nature. 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." 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.

14.3.2 Quasi-Rent: The Short-Term Rent on Fixed Factors

The *quasi-rent*, a concept used by Marshall, refers to the short-term earnings of factors which are in fixed supply in the short-run. To explain the concept of *quasi-rent*, let us make a distinction between the short-run and the long-run. In the long-run, all inputs are variable in large quantities as their supply is elastic. In the short-run, however, the supply of certain inputs is fixed. For example, the supply of plant and machinery in the short-run is inelastic.

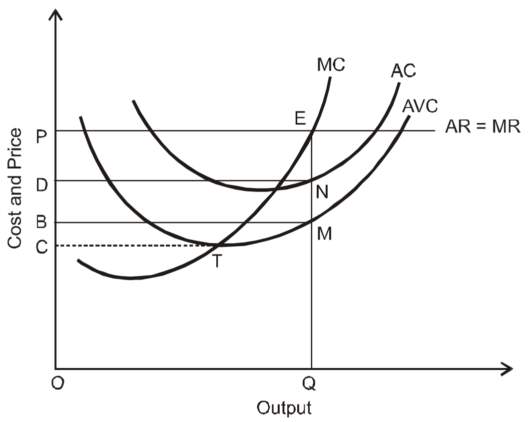


Fig. 14.4 *Determination of Quasi-Rent*

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. 14.4. Suppose, given the *AVC, AC* and *MC* curves, price is *OP,* and the firm is in equilibrium at point *E.*

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

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.,

$TR = OC \times CT$ 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 repre-sented by the area *PBME.* Of this, the area *DPEN* represents the difference between the *TR* and *TC* ($= OQ \times OD$). 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 condi-tions. Therefore, the area *BDNM* is the *opportunity cost* of fixed factors. Thus:

Quasi-rent = *TFC*+ Economic Profit

14.4 THEORIES OF INTEREST

Let us now go through the various theories of interest.

14.4.1 The Classical Theory of Interest

The classical theory of interest refers, according to Keynes, to the theories (or views) of Marshall, Cassel, Tausig, Walras, etc. In fact, none of these economists whom Keynes calls modern classical school, has given a precise or an explicit account of the interest theory. It was Keynes, in fact, who crystalized the scattered classical thoughts on interest and formulated the classical theory of interest, which is presented below.

Keynes provides an excellent summary of classical theory of interest. According to the classical theory, the rate of interest is determined by the demand for investment and willingness to save. According to Keynes, investment represents the demand for investible resources; savings represents the supply of investible funds; and the rate of interest is the 'price' of investible resources. The rate of interest is determined where the demand for investible funds is equal to the supply of savings. "Just as the price of a commodity is necessarily fixed at that point where the demand for it is equal to the supply, so the rate of interest necessarily comes to rest under the play of market forces at the point where the amount of investment at that rate of interest is equal to the amount of saving at that rate." In simple words, the market rate of interest is determined where demand for investible funds equals the supply of funds, i.e., supply of savings. The classical theory of interest is presented in Fig. 14.5. The demand for investible funds is shown by I, curve and the supply of savings is given by the curve *S*. Note that the investment demand curve, *I*, is negatively related to interest, *r*, whereas supply of savings schedule, *S*, is positively related to interest. The investment demand curve I_1, intersects the supply of savings schedule, *S*, at point *P*. The equilibrium rate of interest is thus determined at *PQ* = *Or* by the point of intersection of the two schedules. At this rate of interest the demand for investible funds, *OQ*, equals the supply of funds. The rate *Or* is therefore the equilibrium rate of interest. The equilibrium rate of interest will be disturbed only when there is a change in the demand for investible funds and in the supply of savings. But, investment and savings will automatically adjust themselves to a new equilibrium rate of interest.

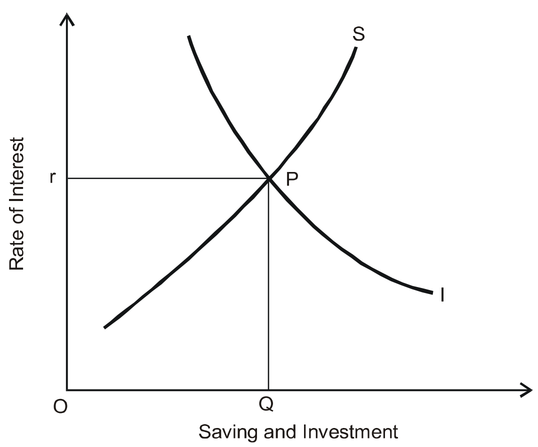


Fig. 14.5 *The Classical Theory of Rate of Interest*

Keynes's Criticism of Classical Theory: Keynes criticised classical theory of interest on the following grounds.

First, the classical theory implicitly assumes *income to be given* and saving to be a unique function of interest. Keynes argues that the classical assumption of

NOTES

income to be given itself implies that there exists an important relationship between savings and income, i.e., savings are not the function of interest alone but also of income. But the classical school neglects this important relationship between income and savings, which leads to a formal error in the analysis.

Second, the above error leads to the conclusion that, given the saving schedule, if investment demand curve shifts leftward, the interest rate will decrease and *vice versa*, as shown in Fig. 14.6. Suppose that rate of interest was initially determined at Or_1. If investment demand schedule shifts leftward to I_2, due to, say, a fall in the marginal efficiency of capital, interest rate will be determined at a lower rate, Or_3. Similarly, if both investment demand and savings schedules shift leftward, interest will be determined at Or_2. As Keynes puts it, "...the classical theory of the rate of interest seems to suppose that, if the demand curve for capital shifts or if the curve relating the rate of interest to the amount saved out of a given income shifts or if both these curves shift, the new rate of interest will be given by the point of intersection of the new position of the two curve." (Keynes, p. 179). "But this," according to Keynes, "is nonsense theory." The error in classical theory lies in its assumption that investment demand schedule (I) can shift without causing, at the same time, a shift in the saving schedule (S). In fact, when investment schedule shifts, it means a change in investment. Change in investment causes a change in income because income, $(Y) = f(I)$. When income changes, savings change too because $S = f(Y)$ too. It is, therefore, inconsistent to assume that investment demand and saving schedules can shift independent of each other. A shift in investment demand schedule does cause a shift in saving schedule. If investment demand and saving schedules keep shifting from one position to another, the whole classical scheme of interest determination breaks down. Interest rate cannot find its equilibrium. Thus, according to Keynes, *the classical theory of interest is indeterminate*.

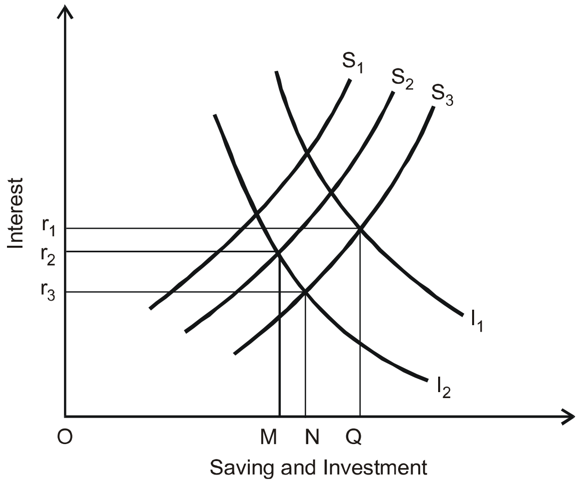


Fig. 14.6 *Change in Interest Rate and Levels of Income*

14.4.2 The Loanable Fund Theory of Interest

A variant of classical theory is the *loanable fund theory of interest* also called as neo-classical theory of interest. The neo-classical economists who have contributed to the growth of this theory include Wicksell, Ohlin, Robertson, Pigou and Viner. According to the loanable fund theory, rate of interest is determined by the intersection of demand schedule for loanable funds with the supply schedule of loanable funds. The demand for loanable funds consists of:

(*i*) investors' demand for funds (I_D);

(*ii*) consumers' demand for funds (C_D); and

(*iii*) demand for funds for hoarding (H_D).

All these three kinds of demand for funds are inversely related to the interest as shown by H_D, C_D and I_D schedules in Fig. 14.7. The horizontal summation of the three kinds of demand gives the aggregate demand for loanable funds (D_L) as shown in Fig. 14.7.

As regards the supply of loanable funds, in Robertsonian sense, it consists of:

(*i*) voluntary savings (V_s), i.e., savings out of disposable income;

(*ii*) bank credits (B_c), i.e., loans available from banks; and

(*iii*) activated idle cash balance or dishoardings (D_h).

All the three components of the supply of loanable funds are deemed to be positively related to interest as shown by D_h, B_c and V_s schedules in Fig. 14.7. A horizontal summation of the schedules V_s, B_c and D_h, gives the aggregate supply schedule of loanable funds (S_L).

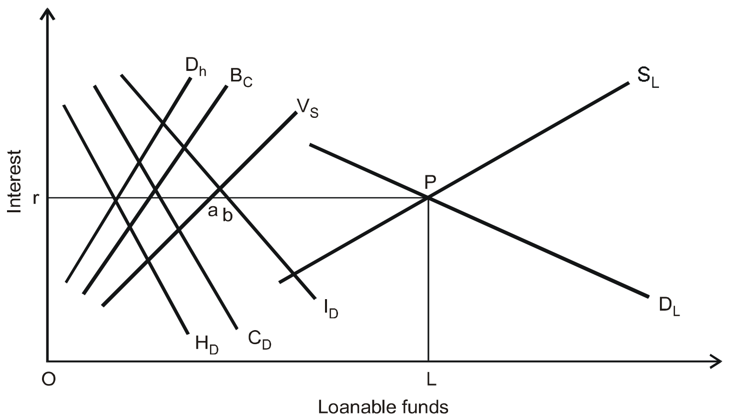


***Fig. 14.7** Loanable Fund Theory of Interest*

The loanable fund theory of interest is presented in Fig. 14.7. The demand schedule for loanable funds (D_L) intersects with the supply schedule of loanable funds (S_L) at point *P*, determining interest rate at *Or*.

The neo-classical theory is superior to the classical theory of interest as it considers the demand for funds also other than the demand for investment, and also the other sources of supply of funds than voluntary savings. Besides, the loanable fund theory, as presented in Fig. 14.7, reveals an important information that planned savings may not be equal to the planned investment at equilibrium rate of interest. The planned savings are represented by the schedule V_s, and planned investment by the schedule I_D. As Fig. 14.7 shows, at equilibrium rate of interest, *Or*, planned savings, *ra*, are less than planned investment, *rb*.

Criticism: Keynes's criticism of classical theory applies to the loanable-fund theory also. It means that loanable-fund theory is also indeterminate. In the loanable-fund theory, supply of investible funds includes voluntary savings, bank money, and the dishoarding of idle balances. Of the three components of the total supply of investible funds, 'savings' accounts for the largest proportion in the total. Saving is function of the 'disposable incomes'. It implies that the total supply of investible funds also depends on the disposable income. Therefore, interest cannot be known unless income is known. And, income cannot be known unless interest is known, since interest influences the investment. As in classical theory, the error in the neo-classical theory lies in the implicit assumption that demand for investible fund and supply of investible funds can vary independent of each other. It is, however, suggested that Keynes's criticism of classical theory does not apply to the neo-classical theory. For, unlike classical theory, the neo-classical theory considers savings to be the function of the preceding year's income which is known and cannot be influenced by the current investment.

14.4.3 Keynesian Theory of Interest

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.,

   $M_t = f(Y)$,   where Y 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 precautions 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

NOTES

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. 14.8. As can be seen in Fig. 14.8, 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.

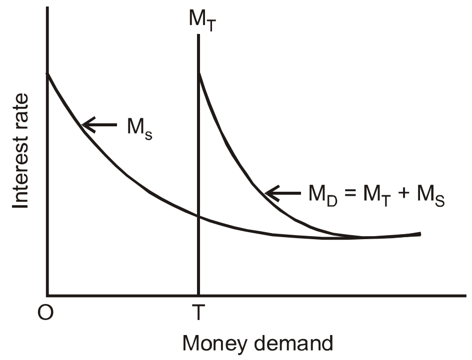


Fig. 14.8 *Demand for Money: Keynesian Approach*

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. 14.9. 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. 14.9, 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.

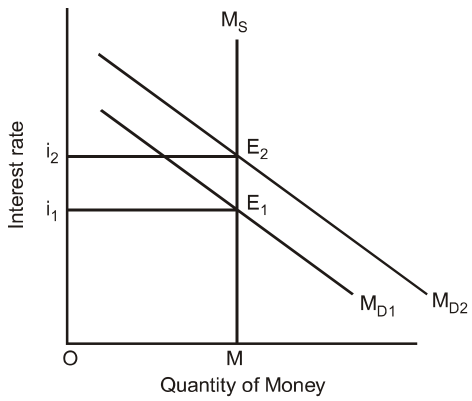


Fig. 14.9 *Determination of Interest Rate: Keynesian Approach*

Criticism of Keynesian Theory of Interest: Recall Keynes's criticism of classical 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 cannot 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.

14.5 THEORIES OF PROFIT

In this section, we will discuss some important theories of profit. Profit theories reveal, in fact, only the source of profit, not the determination of profit rate.

14.5.1 Walker's Theory: Profit as Rent of Ability

One of the most widely known theories advanced to explain the nature of profit was formulated by F.A. Walker. According to him, profit is rent of the exceptional abilities that an entrepreneur may possess over the least entrepreneur. Just as rent on land is the difference between the yields of the least fertile and super lands, pure profit is the difference between the receipts of the least efficient entrepreneur and that of those with greater efficiency or managerial ability.

Assumptions. In formulating his profit theory, Walker visualised a state of perfect competition in which all firms (or entrepreneurs) are presumed to possess equal managerial ability or entrepreneurship. There being no barrier to prevent the entry of new firms to the industry, the number of firms would increase until the remuneration of each was just enough to keep them in the industry. Each firm would then receive only the wages of management which, in Walker's view, formed no part of (pure) profit. He regarded wages of management as ordinary wages. Thus, under perfectly competitive conditions, there would be no pure profits and all firms would be no-profit firms.

However, when one departs from the realm of perfect competition, one finds, in almost every economic activity, some firms making only a bare living while other firms in the same industry are making pure profits. Walker regarded profits of profit-making firms arising out of what a more efficient firm is able to produce over and above what the least efficient firm is able to produce with the same amount of capital and labour. Walker attributed this surplus wholly to the greater efficiency of a firm, which distinguishes it from the least efficient ones.

Thus, to Walker, *profit is a reward for exceptional business ability* over and above the ordinary ability required for management of the organisation which could be taken as wage or salary. Just as rent is a reward for a higher productivity of land, so is the profit reward for superior managerial ability of an entrepreneur.

A natural corollary of this view is that profit did not enter the cost of production as is the case with rent. Therefore, according to Walker, profit does not enter the price determination. The logic that Walker gives for his argument runs as follows. Market price is determined by the cost of production of that portion of supply which is produced by the least efficient firms. Prices so determined make allowance for only wages of management not the surplus that accrues to the firms with greater efficiency.

14.5.2 Clark's Theory: Profit as Reward for Dynamic Entrepreneurship

The dynamic theory of profit is associated with the name of J.B. Clark, which he propounded in 1900. According to Clark, profits accrue in a dynamic world, not in a static world. Let us have a glance at the static and dynamic world and how profit arises in a dynamic world.

Static World. As visualised by Clark, a static world is one in which there exists absolute freedom of competition; but population and capital are stationary; there are no inventions; production process does not change; and the goods continue to remain homogeneous. Besides, in a static state there is perfect mobility of factors of production but there is no motion because marginal products of labour and capital are equal in all groups of industries. Also, in a static state, there is no uncertainty and hence, no risk. Whatever risks might arise due to natural calamities are covered by insurance.

No Profit in Static World. To show how profits were eliminated in a static state, Clark draws a distinction between the work of an entrepreneur and that of a manager of business. He believed that the task of a manager could be described as labour which can be paid for by wage. In a static state, profit would not arise because competition would not permit any business manager to earn more than his managerial wages which would be equal to the value of marginal product of management. Therefore, there would be no surplus available which could be called as profit.

Dynamic World. In contrast to a static word, a dynamic world is one in which the factors that remain constant in a static world undergo the process of change. Clark indicated certain generic changes that mark the transition of a society from a static to a dynamic state. Briefly speaking, generic changes include:

(*a*) increase in population;

(*b*) increase in capital;

(*c*) improvement in production techniques;

(*d*) changes in the forms of business organisation; and

(*e*) multiplication of consumer's wants.

Profit as Reward for Dynamic Enterprise. In Clark's view, the major functions of an entrepreneur in a dynamic society are related to these changes, i.e., to take the advantage of generic changes, to promote business, to expand sales, and to reduce cost of production. The typical changes that emerge out of these kinds of special effort of some entrepreneurs are inventions and improvements in the methods of production. Such changes lead to increase in production given the costs or reduction in costs given the output, which results in emergence of profits to the initial inventors.

Profits in Dynamic World are not there for ever. With the passage of time, profits resulting from the inventions and improvements in production methods dis-appear. What happens, in fact, is that competition forces other entrepreneurs to imitate or innovate the new technology. This leads to rise in demand for labour and capital. Consequently, wages and interest rise and cost of production increases. On the other hand, with larger employment of labour and capital, production increases leading to fall in product prices. The ultimate result is that profits disappear. In Clark's own words, "profit is an elusive sum which entrepreneurs grasp but cannot hold. It slips through their fingers and bestows itself on all members of the society."

Profits Disappear to Reappear. This however should not mean that, in a dynamic society, profits arise only once and disappear forever. In fact, under dynamic conditions, the generic changes continue to take place: it is a continuous process. The process of dynamic change gives entrepreneurs opportunities time and again to adjust their business to the changing conditions, make inventions and improve production methods, with a view to making pure profit. In fact, emergence and disappearance of profits is a continuous process.

On the question of risk involved in making inventions and improving production methods, Clark was of the view that profit does not arise due to risk. If risk is there, it affects the entrepreneurs because risk related income accrues to them. Profit, on the other hand, is the result of entrepreneurial functions under dynamic condi-tions. Therefore, profit does not result from risk-bearings.

To sum up, according to J.B. Clark, profit is a reward for coordinating managerial functions of entrepreneurs under dynamic conditions. It is a reward for dynamism. It is not a reward for risk bearing. Pure profit, according to him, is a residue that remains after interest and wages are paid. That is, the difference between the gross receipts and payments for wages and interest represents profit.

Criticism of Clark's Theory: Clark's theory, though impressive, has failed to win unqualified acceptance and has been criticised on the following grounds.

First, to some economists the division of firm's earning between the wage of management and profits is not acceptable. It has been contended, for instance, that even the routine conduct of a business calls for a prudent judgement and administrative ability just as these qualities are called for in the exploitation of a new invention or in any other manifestation of economic change. Clark's definition was therefore a matter of phraseology and no clear line could be drawn to show the functions which give wages of management and those which could be taken as profit.

Secondly, even if it is accepted that profits are accounted for by the coordinating functions of entrepreneur, it poses special difficulties in explaining the profits in the practical world. For instance, profits of companies are mainly paid to the shareholders. But these shareholders exercise no coordinating functions. One may say, for the sake of argument, that shareholders receive only a fair interest on their investment and that the profit is what remains after paying this 'interest'. Still, this sum after deducting the 'interest' paid to shareholders would continue to be their property, because they are the owners of retained earnings. Thus, Clark's theory fails to explain the profits in practice.

Thirdly, the basic tenet of Clark's theory is that profits result from the change in business conditions and are reward for dynamism and Clark's entrepreneur is the pioneer of this change. But in practice, one finds that profit exists under different conditions. There are many profitable business concerns engaged in forms of activity in which dynamic stage is long since past and in which no change takes place. In many lines of activity business has settled down to almost routine conditions and yet profits continue to be made despite competition.

Fourthly, it has been argued by F.H. Knight that all changes would not give rise to profits. Certain changes are predictable and others are not. So far as predictable changes are concerned they pose no managerial problems or uncertainty. Therefore, such changes cannot give rise to profit. Only the unpredictable changes would require the use of managerial talent as they, give rise to uncertainty. Clark's theory thus misses an important element of uncertainty and risk and their relation to profit.

14.5.3 Hawley's Risk Theory: Profit as Reward for Risk-Bearing

The risk theory of profit was propounded by F.B. Hawley in 1893. Hawley regarded risk-taking as the inevitable accompaniment of dynamic production and those who take risk have a sound claim to a separate reward, known as profit. Thus, according to Hawley, profit is simply the price paid by the society for assuming business risks. In his opinion, businessmen would not assume risk without expecting an adequate compensation in excess of acturial value. That is, the entrepreneur would always look for a return in excess of the expected losses. The reason why Hawley maintains that profit is over and above the acturial risk is that the assumption of risk is irksome; it gives rise to trouble, anxiety and disutilities of various kinds, which gives a claim to reward for all these pains in excess of acturial value of risk. Profit, according to Hawley, consists of two parts: ***first***, represents compensation for acturial or average loss incidental to the various classes of risks necessarily assumed by the entrepreneur; and ***second*** the remaining part represents, an inducement to suffer the problems of being exposed to the risk.

Hawley recognises that the coordination which Clark spoke of was important, but he believes that profit is attendant upon profit only when coordination happens to be an incident of ownership; and that profit arises from ownership only so long as ownership involves risk. Thus, risk has to be assumed to qualify for profit. If an entrepreneur shifts his risks by insuring against them, he would cease to be an entrepreneur and would not receive any profit. It is only from the uninsured risks that profits arise, and until the uncertainty ends with the sale of entrepreneur's products, the amount of the reward cannot be determined. Profit, therefore, is a residue. Hawley's theory is also called as a **residual theory of profit**.

Hawley was aware that his theory did not offer a complete explanation of all the gains arising from business activities. In monopoly undertakings, for example, many a time profit could not be attributed to the risks which were undertaken: profits in monopoly firms arise from the very fact of not undertaking the risks. Thus, monopoly gains fall outside his theory. To meet this flaw he placed monopoly gains in a distinct, separate category of business gains which might arise to other factors also. According to his view, monopoly gains could occur also to labour, landlords, capital suppliers. But since their respective incomes—wages, rent and interest—do not arise from the operation of productive forces, these are merely economic gains.

Criticism of Hawley's Theory: Perhaps no other theory of profit has attracted so much attention and generated so much discussion as the Risk Theory of Profit. It ranks today as one of the most widely accepted theories of profits. Nevertheless, Hawley's risk theory of profit has been criticised on the following grounds.

First, in his reaction to the risk theory of profit, Clark remarked that the profit visualised by Hawley was nothing but an interest on capital. Risk, in Clark's view, was risk of loss of capital. Therefore, the reward for assuming risk (of loosing of capital) was interest: it is not profit.

Secondly, it has also been argued that Hawley stressed on only the risk in terms of loss of capital: he did not give due consideration to the fact that risks arise also in the use of factors of production other than capital.

Thirdly, Hawley's theory of profit concentrates only on risk-bearing element, and ignores other entrepreneurial functions, viz., organisation and coordination, which also lead to emergence of profit.

Fourthly, it is also argued that Hawley failed to make a distinction between predictable and unpredictable risks. While predictable (or foreseeble) risks are insurable, unpredictable (or unforeseeable) risk are not. Since predictable risks can be insured, such risks do not give rise to profit because the risk is shifted on to the insurer. As Knight put it, it is in fact the uninsurable risk, which is uncertain and gives rise to profit. Thus, in his view, profit is a reward for uncertainty bearing rather than a reward for risk-bearing.

Fifthly, Carver observed that profits are reward for avoiding risk and not for bearing risk, because only those entrepreneurs who are able to avoid risk make profits.

Finally, if profits were the reward for risk bearing, then the greater the risk undertaken, the greater the profits. But, there is no empirical support to this inference which can be drawn from Hawley's theory.

14.5.4 Knight's Theory: Profit as a Return to Uncertainty Bearing

Frank H. Knight treated profit as a residual return to uncertainty bearing—not to risk bearing. Obviously, Knight made a distinction between **risk** and **uncertainty**. He divided risks into calculable and non-calculable risks. Calculable risks are those whose probability of occurrence can be statistically calculated on the basis of available data, e.g., risks due to fire, theft, accidents, etc. Such risks are insurable. There remains, however, an area of risks in which probability of risk occurrences cannot be calculated. For instance, there may be a certain element of cost which may not be accurately calculated. For example, the strategies of the competitors may not be accurately guessed. The risk element of such incalculable events are not insurable. The area of incalculable risks is thus marked by 'uncertainty'. It is in this area of uncertainty that decision becomes a peculiar responsibility of an entrepreneur. If his decisions are proved right by the subsequent events, the entrepreneur makes profit, and *vice versa*. Obviously, profit arises from the decisions taken and implemented under the conditions of uncertainty, as visualised by Knight. The profits may arise as a result of (*a*) decisions concerning the state of market; (*b*) decisions which result in increasing the degree of monopoly; (*c*) decisions with respect to holding stocks that give rise to windfall gains when prices increase; and (*d*) decisions taken to introduce new techniques or innovations that give rise to profit.

Criticism of Knight's Theory: Several objections have been raised against Knight's theory of profit too.

First, it has been contended that Knight's uncertainty theory lacks scientific precision. Uncertainty is a difficult concept to handle. Tausig, for instance, has shown that though certain risks are in the area of uncertainly, many are not. For example, suppose that a person is betting in a horse race. If he has the knowledge of age, training, rearing, etc., of different horses and their jockeys, he would be operating in the region of *risk.* And, if he does not have the knowledge about the horses and jockeys participating in the race, he would be regarded as operating in the area of *uncertainty.* But, if he has some knowledge about the horses and/or jockeys, it will be difficult to decide whether the person is operat-ing in the area of risk or in the area of uncertainty.

Secondly, by considering profit as a reward exclusively for uncertainty bearing, Knight has implicitly accorded it (uncertainty bearing) the status of a factor of production, whereas it is simply an element of real cost as distinguished from money cost. Therefore, uncertainty bearing cannot be accepted as a factor of production, and hence the sole cause of profit.

Thirdly, Knight's attempt to explain profits only by 'uncertainty' makes his theory unconvincing if one examines it in the light of real experience of the business world. If his theory is accepted, it would mean that the greater the degree of uncertainty, the greater the profits, and *vice versa*. But there are enterprises, e.g., agriculture, which are known for their high uncertainty and low returns.

14.5.5 Schumpeter's Theory: Profit as Reward for Innovations

The Innovation Theory of Profit was developed by Joseph A. Schumpeter. Throughout his life as an economist, he was preoccupied with the study of economic evaluation and development in capitalist system. He was of the opinion that issues like interest, profit, trade cycles and many others were only incidental to a distinct process of economic development: and certain principles which could explain this process would also explain these economic variables. His theory of profit is thus embedded in his theory of economic development.

Stationary Equilibrium: The Starting Point. To explain the phenomenon of economic development (and therefore of profit) Schumpeter starts from the state of stationary equilibrium which is characterised by full equilibrium in all the spheres. He assumes a closed, commercially organised, capitalist economy in which private property, division of labour and free competition prevail, along with constant population level. Everybody sells all his produce and insofar as he himself consumes, he is his own customer. The productive services may also be included in the same category of marketable things which are sold. But anyone who wants to purchase these goods or productive services must also have his own products or services to offer. Thus all goods and services are exchanged for one another. "Hence it follows that somewhere in the economic system a demand is, so to say, ready awaiting every supply, and nowhere in the system are there commodities without complements..." It, therefore, follows that sellers of all the commodities appear as buyers to acquire the goods. This maintains their consumption and also productive capacity in the next period at the existing level. As a result, there emerges, "an unchanging economic process which flows on at constant rates in time and merely reproduces itself."

Profit as the Reward for Innovations. Under these conditions of stationary equilibrium, total receipts from the business are exactly equal to the total outlay: there is no profit. Profit can be made by introducing innovations in manufacturing and methods of supplying the goods. Innovations include:

(*i*) introduction of a new product or a new quality of existing product;

(*ii*) introduction of a new method of production;

(*iii*) creating or finding a new market;

(*iv*) finding new sources of raw material; and

(*v*) organising the industry in a different manner.

When an entrepreneur introduces an innovation, there will be a surplus over cost under the following conditions.

1. When a new supply comes forth as a result of innovation, the price of commodity should not fall to such an extent that it eliminates all the gains from the larger product.
2. The cost per unit of output with new technique should be less than that of older method.
3. The increase in demand for the productive services due to innovation should not lead to such a rise in the cost of the productive services that it pushes per unit cost of the commodity beyond the expected price.

If these conditions do exist, the surplus realised will *ipso facto* become a net profit.

Profits Disappear Due to Imitation. The profits resulting from innovations exist only temporarily. This is so because when an entrepreneur introduces an innovation, others are likely to imitate it for its profitability. First a few and then many follow the lead, and produce the commodity in the same manner. This causes a keen competition for the productive services to be employed with the new techniques. The supply of productive factors remaining the same, their remuneration tends to increase. As a result, cost of production increases. On the other hand, with other firms adopting the innovations, supply of goods and services increases resulting in fall in their prices. Thus, on the one hand, cost per unit goes up and, on the other, revenue per unit of output decreases. Ultimately, a time comes when the difference between cost and receipt disappears. So the profit disappears. In the process, however, the economy reaches higher level of stationary equilibrium.

It is however quite likely that profits exist in spite of the process of profits being wiped out. Such profits are in the nature of *quasi-rent* arising due to some special characteristics of productive services. Furthermore, where profits arise due to factors like patents, trusts, cartels, etc., such profits would be in the nature of monopoly revenue rather than entrepreneurial profits.

It may be inferred from the above that *profit is the child as well as victim* of *economic development.* Economic development consists of increase in national output. When innovations occur the national output increases because the same output can be produced at lower costs, or what is the same thing, with the same amount of resources greater output can be produced. But producing at lower cost or producing more output with the same total cost results in profits. Thus, economic development gives birth to profits. But, when other producers also adopt the technique introduced by the innovator, the total national output increases, i.e., economic development catches pace. The widespread use of innovation, however, results in wiping out of profits, as was explained earlier. Hence, economic development itself is responsible for the disappearance of profits.

Criticism of Innovation Theory: The major criticism against Schumpeter's innovation theory of profit is that he ignored the risk and uncertainty, the two major sources of profit as shown in the traditional theories of profits. Although in his book *Capitalism, Socialism and Democracy,* he admits that innovations are made by the risk-taking entrepreneurs, he ignores uncertainty altogether. Besides, it has also been argued that innovation is not the only function of the entrepreneurs. As delineated in the dynamic theory of profit, entrepreneur's functions include organisational and coordinational activities also in response to the changing conditions and needs of the society.

Does Profit Enter the Cost of Production?

From the above description of profit theories, one is tempted to infer that profits do not enter the cost of production. In fact, whether profits enter the cost of production or not depend on the concept of profit under reference. 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.

We have already described the meaning of pure profit, in the beginning of this chapter. We describe here briefly the meaning of 'normal 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.

Check Your Progress

1. Name the theory of interest propounded by Keynes.
2. State the principle on which Ricardo has based his theory of rent.
3. Mention the Keynesian theory of interest.

14.6 ANSWERS TO 'CHECK YOUR PROGRESS' QUESTIONS

1. Keynes propounded his own Liquidity Preference Theory of Interest.
2. Ricardo has based his theory of rent on the principle of demand and supply.
3. 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.

14.7 SUMMARY

- The theory of wage determination assumes that the units of a factor (labour) are homogeneous. If all the units of labour are identical and if non-monetary advantages given to labour are the same in all uses, then the wage rate for each labour tends to be the same in a perfectly competitive market.
- Wage differentials due to market disequilibrium may be classified under two groups:
    - o Dynamic Wage Differentials, and
    - o Static Wage Differentials.

- Monopoly in the product market and perfect competition in the labour market creates conditions for the monopolistic exploitation of labour.
- 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.
- The classical theory of interest refers, according to Keynes, to the theories (or views) of Marshall, Cassel, Tausig, Walras, etc. In fact, none of these economists whom Keynes calls modern classical school, has given a precise or an explicit account of the interest theory.
- A variant of classical theory is the loanable fund theory of interest also called as neo-classical theory of interest. The neo-classical economists who have contributed to the growth of this theory include Wicksell, Ohlin, Robertson, Pigou and Viner.
- 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.
- One of the most widely known theories advanced to explain the nature of profit was formulated by F.A. Walker. According to him, profit is rent of the exceptional abilities that an entrepreneur may possess over the least entrepreneur.
- The dynamic theory of profit is associated with the name of J.B. Clark, which he propounded in 1900. According to Clark, profits accrue in a dynamic world, not in a static world.
- The risk theory of profit was propounded by F.B. Hawley in 1893. Hawley regarded risk-taking as the inevitable accompaniment of dynamic production and those who take risk have a sound claim to a separate reward, known as profit.
- The risk theory of profit was propounded by F.B. Hawley in 1893. Hawley regarded risk-taking as the inevitable accompaniment of dynamic production and those who take risk have a sound claim to a separate reward, known as profit.
- The Innovation Theory of Profit was developed by Joseph A. Schumpeter. Throughout his life as an economist, he was preoccupied with the study of economic evaluation and development in capitalist system.

14.8 KEY WORDS

- **Quasi-rent:** A concept used by Marshall, refers to the short-term earnings of factors which are in fixed supply in the short-run.

- **Wage:** In economics, the price paid to labour for its contribution to the process of production is called wages.
- **Liquidity:** It refers to the degree to which an asset or security can be quickly bought or sold in the market without affecting the asset's price.

14.9 SELF-ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. Write a short note on monopolistic exploitation of labour.
2. What are the criticisms raised against the Ricardian theory of Rent?
3. Briefly mention the Classical theory of Interest.

Long-Answer Questions

1. Illustrate the concept of wage determination under the condition of perfectly competitive market structure with the help of a diagram.
2. Discuss the main proponents of Ricardian theory of Rent.
3. Explain the various theories of profit.

14.10 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.