

**TEERTHANKER MAHAVEER UNIVERSITY
MORADABAD, INDIA**

**CENTRE FOR ONLINE & DISTANCE
LEARNING**



Accredited with NAAC **A** Grade

12-B Status from UGC

Programme: Master of Arts

Course: Advanced Microeconomics

Course Code: MAECC201

Semester-II

BLOCK - I

ECONOMICS OF SCALE, COST AND REVENUE

UNIT 1 ECONOMIES OF SCALE

NOTES

Structure

- 1.0 Introduction
- 1.1 Objectives
- 1.2 Economies of Scale: Meaning and Use
 - 1.2.1 Diseconomies of Scale
- 1.3 Answers to Check Your Progress Questions
- 1.4 Summary
- 1.5 Key Words
- 1.6 Self Assessment Questions and Exercises
- 1.7 Further Readings

1.0 INTRODUCTION

Economies of scale refers to the cost advantages which arises when average cost starts decreasing as the output increases. Economies of scale is a concept which applies to a variety of organisational and business situations and at various levels such as business or manufacturing unit, plant or an entire enterprise. The shape of LAC or the Long-run Average Cost function is determined by the economies and diseconomies of scale. Economies of scale give rise to lower per-unit cost for several reasons including specialisation, division of labour, integrated technology, larger advertising boost production volume. In this unit, we will discuss the concept of economies of scale in detail.

1.1 OBJECTIVES

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

- Discuss the concept of economies of scale
- Explain the diseconomies of scale

1.2 ECONOMIES OF SCALE: MEANING AND USE

Economies of scale refers to the cost advantage experienced by a firm when it increases its level of output. The advantage arises due to the inverse relationship between per-unit fixed cost and the quantity produced by the firm.

NOTES

While optimization of output in the long run is an important concern of business firms, *cost minimization* is an equally important decision area. Cost of production depends not only on *internal factor* – the productivity of inputs – but also on many *external factors* – the factors that arise out of the firm. In this section we give a detailed analysis of internal and external economies and diseconomies of scale and how they determine the trend in cost of production. To begin with, let us have a look at the trend of long-run average cost curve (*LAC*). It is to be noted that *LAC* decreases with the expansion of production scale up to OQ_2 and then it begins to rise. Decrease in *LAC* is caused by the economies of scale and increase in *LAC* is caused by diseconomies of scale. Economies of scale result in cost saving and diseconomies lead to rise in cost. *Economies and diseconomies of scale* determine also the returns to scale. Increasing returns to scale operate till economies of scale are greater than the diseconomies of scale, and returns to scale decrease when diseconomies are greater than the economies of scale. When economies and diseconomies are in balance, returns to scale are constant. In this section, we briefly discuss the various kinds of economies and diseconomies of scale and their effect on cost of production.

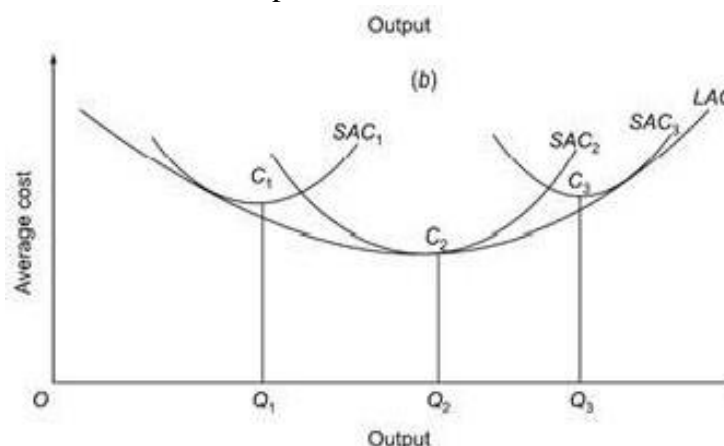


Fig. 1.1 Long-run Total and Average Cost Curves

Economies of Scale Classification

The economies of scale are classified as (a) Internal or Real Economies, and (b) External or Pecuniary Economies.

A. Internal Economies: Internal economies, also called ‘real economies’, are those that arise within the firm with addition of new production plants. This means that internal economies are available exclusively to the expanding firm. Internal economies may be classified under the following categories.

- (i) Economies in production;
- (ii) Economies in marketing;

- (iii) Managerial economies, and
- (iv) Economies in transport and storage.

- (i) **Economies in Production:** Economies in production arise from two sources: (a) technological advantages, and (b) advantages of division of labour based on specialization and skill of labour.

Technological advantages: Large-scale production provides an opportunity to the expanding firms to avail the advantages of technological advances. Modern technology is highly specialized. The advanced technology makes it possible to conceive the whole process of production of a commodity in one composite unit of production. For example, production of cloth in a textile mill may comprise such plants as (i) spinning; (ii) weaving; (iii) printing and pressing; and (iv) packing, etc. Likewise, a composite dairy scheme may consist of plants like (i) chilling; (ii) milk processing; and (iii) bottling. Under small-scale production, the firm may not find it economical to have all the plants under one roof. It would, therefore, not be in a position to take the full advantage of a composite technology. But, when scale of production expands and firms hire more capital and labour, their total output increases more than proportionately till the optimum size of the firm is reached. It results in lower cost of production.

Advantages of division of labour and specialization: When a firm's scale of production expands, more and more workers of varying skills and qualifications are employed. With the employment of larger number of workers, it becomes increasingly possible to divide the labour according to their qualifications, knowledge, experience, expertise and skills and to assign them the function to which they are best suited. This is known as division of labour. *Division of labour* leads to a greater specialization of manpower. It increases productivity of labour and, thereby, reduces cost of production. Besides, specialized workers develop more efficient tools and techniques and gain speed of work. These advantages of division of labour improve productivity of labour per unit of labour cost and time. Increase in labour productivity decreases to per unit cost of production.

- (ii) **Economies in Purchase of Inputs:** Economies in input purchases arise from the large-scale purchase of raw materials and other material inputs and large-scale selling of the firm's own products. As to economies in the purchase of inputs, the large-size firms normally make bulk purchases of their inputs. The large scale purchase entitles the firm for certain discounts in input prices and other concessions that are not available on small purchases. As such, the growing firms gain economies on the cost of their material inputs. The internal economies arise also in marketing the firm's own product as (a) economies in advertisement cost; (b) economies in

NOTES

NOTES

large-scale distribution through wholesalers, etc.; and (c) other large-sale economies. With the expansion of the firm, the total production increases. But the expenditure on advertising the product does not increase proportionately. Similarly, selling through the wholesaleddealers reduces the cost of distribution of the firm's production. The firm also gains on large scale distribution through better utilization of _sales force, distribution of sample, etc.'

- (iii) **Managerial Economies:** Managerial economies arise from (a) specialization in managerialactivities, i.e., theuse of specializedmanagerial personnel, and (b) systemization of managerial functions. For a large-size firm, it becomes possible to divide its management into specialized departmentsunderspecializedpersonnel, such as productionmanager, sales manager, HR manager, financialmanager, etc. Themangement of different departmentsbyspecializedmanagersincreasestheefficiencyofmanagement at all the levels of management because of the decentralization of decision-making. It increases production, given the cost. Large-scale firms have the opportunity to use advanced techniques of communication, telephones and telex machines, computers, and their own means of transport. All these lead to quick decision-making, help in saving valuable time of the management and, thereby, improve the managerial efficiency. For these reasons, managerialcostincreaseslessthanproportionatelywiththeincrease in production scale upto a certain level, of course.
- (iv) **Economies in Transport and Storage:** Economies in transportation and storage costs arise from fuller utilization of transport and storage facilities. Transportation costs are incurred both on production and sales sides. Similarly, storage costs are incurred on both raw materials and finished products. The large-size firms may acquire their own means of transport and they can, thereby, reduce the unit cost of transportation, at least to the extent of profit margin of the transport companies. Besides, own transport facility prevents delays in transporting goods. Some large-scale firms have their own railwaytracks from the nearest railwaypoint to the factory, and thus they reduce the cost of transporting goods in and out. For example, Bombay Port Trust has its own railway tracks, oil companies have their own fleet of tankers. Similarly, large-scale firms can create their own godowns in various centres of product distribution and can save on cost of storage.

B. External or Pecuniary Economies of Scale

Externaleconomiesarethosethatariseoutsidethefirmandaccrue to theexpanding firms. External economies appear in the form of moneysaving on inputs, called *pecuniary economies*. Pecuniaryeconomies accrue to the large-size firms in the

form of discounts and concessions on (i) large scale purchase of raw material, (ii) large scale acquisition of external finance, particularly from the commercial banks; (iii) massive advertisement campaigns; (iv) large scale hiring of means of transport and warehouses, etc. These benefits are available to all the firms of an industry but large scale firms benefit more than small firms. Besides, expansion of an industry encourages the growth of ancillary industries that supply inputs. In the initial stages, such industries also enjoy the increasing returns to scale. In a competitive market, therefore, input prices go down. The benefit of decreasing input prices accrues to the expanding firms in addition to discounts and concessions. For example, growth of the automobile industry helps the development of tyre industry and other motor parts manufacturing units. The economies of scale reaped by such industries flow also to automobile industry. If Maruti Udyog Limited starts producing tyres for its own cars and ancillaries, cost of Maruti cars may go up. Consider another example: growth of computer industry encourages growth of firms that manufacture and supply computer chips and other software. Competition between such firms and law of increasing returns reduces the cost of inputs. Reduction in input costs is an important aspect of external economies.

NOTES

1.2.1 Diseconomies of Scale

The economies of scale have their own limits, i.e., scale economies exist only up to a certain level of production scale. The expansion of scale of production beyond that limit creates condition for *diseconomies of scale*. Diseconomies of scale are disadvantages that arise due to the expansion of production scale beyond its optimum level and lead to rise in the cost of production. Like economies, diseconomies may be internal and external. Let us describe the nature of **internal** and **external diseconomies** in some detail.

1. **Internal Diseconomies:** Internal diseconomies are those that are exclusive and internal to a firm as they arise within the firm. Like everything else, economies of scale have a limit too. This limit is reached when the advantages of division of labour and managerial staff have been fully exploited; excess capacity of plant, warehouses, transport and communication systems, etc., is fully used; and economy in advertisement cost tapers off. Although some economies may still exist, diseconomies begin to outweigh the economies and the costs begin to rise.

Managerial Inefficiency: Diseconomies begin to appear first at the management level. Managerial inefficiencies arise, among other things, from the expansion of scale itself. With fast expansion of the production scale, personal contacts and communications between (i) owners and managers, (ii) managers and labour, and (iii) between the managers of different departments or sections get rapidly reduced. The lack of fast or quick communication causes delays in decision-making affecting production

NOTES

adversely. *Secondly*, close control and supervision is replaced by remote control management. With the increase in managerial personnel, decision-making becomes complex and delays in decision-making become inevitable. *Thirdly*, implementation of whatever decisions are taken is delayed due to coordination problem in large scale organisations. *Finally*, with the expansion of the scale of production, management is professionalized beyond a point. As a result, the owner's objective function of profit maximization is gradually replaced by managers' utility function, like job security and high salary, standard or reasonable profit target, satisfying functions. All these lead to laxity in management and, hence to a rise in the cost of production.

Labour Inefficiency: Increasing number of labour leads to a loss of control over labour management. This affects labour productivity adversely. Besides, increase in the number of workers encourages labour union activities that cause loss of output per unit of time and hence, rise in the cost of production.

- 2. External Diseconomies:** External diseconomies are the disadvantages that arise outside the firm, especially in the input markets, due to natural constraints, especially in agriculture and extractive industries. With the expansion of the firm, particularly when all the firms of the industry are expanding, the discounts and concessions that are available on bulk purchases of inputs and concessional finance come to an end. More than that, increasing demand for inputs puts pressure on the input markets and input prices begin to rise causing a rise in the cost of production. These are *pecuniary diseconomies*. On the production side, the law of diminishing returns to scale come into force due to excessive use of fixed factors, more so in agriculture and extractive industries. For example, excessive use of cultivable land turns it into barren land; pumping out water on a large scale for irrigation causes the water table to go down resulting in rise in cost of irrigation; extraction of minerals on a large scale exhausts the mineral deposits on upper levels and mining further deep causes rise in cost of production; extensive fishing reduces the availability of fish and the catch, even when fishing boats and nets are increased. These kinds of diseconomies make the *LAC* move upward.

Check Your Progress

1. List the classification of economies of scale.
2. What are internal economies?

1.3 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. The economies of scale are classified as (a) Internal or Real Economies, and (b) External or Pecuniary Economies.
2. Internal economies, also called 'real economies', are those that arise within the firm with addition of new production plants. This means that internal economies are available exclusively to the expanding firm.

NOTES

1.4 SUMMARY

- Economies of scale refers to the cost advantage experienced by a firm when it increases its level of output. The advantage arises due to the inverse relationship between per-unit fixed cost and the quantity produced by the firm.
- Economies of Scale are classified as (a) Internal or Real Economics, and (b) External or Pecuniary Economics.
- Cost of production depends not only on *internal factor* – the productivity of inputs – but also on many *external factors* – the factors that arise out of the firm.
- Diseconomies of scale are disadvantages that arise due to expansion of production scale and lead to rise in the cost of production.

1.5 KEY WORDS

- **Economies of scale:** It refers to the cost advantage experienced by a firm when it increases its level of output.
- **Diseconomies of scale:** It refers to the disadvantages that arise due to expansion of production scale and lead to rise in the cost of production.

1.6 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short Answer Questions

1. Write a brief note on internal economies.
2. What are external economies?

Long Answer Questions

1. Compare and contrast internal economies and external economies.
2. Explain the concept of diseconomies in detail.

NOTES

1.7 FURTHER READINGS

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

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

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

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

UNIT 2 DIFFERENT CONCEPTS OF COST

*Different Concepts
of Cost*

NOTES

Structure

- 2.0 Introduction
- 2.1 Objectives
- 2.2 Types of Cost
 - 2.2.1 Accounting Cost Concepts
 - 2.2.2 Analytical Cost Concepts
- 2.3 Traditional Theory of Cost
 - 2.3.1 Short-Run Cost-Output Relations
 - 2.3.2 Long-Run Cost-Output Relations: Envelope Curve
- 2.4 Modern Theory of Cost
- 2.5 Answers to Check Your Progress Questions
- 2.6 Summary
- 2.7 Key Words
- 2.8 Self Assessment Questions and Exercises
- 2.9 Further Readings

2.0 INTRODUCTION

For any organization, maximization of profit is one of the primary objectives of operating. This increase of profit is dependent on two very important factors: cost and revenue. In layman terms, cost refers to the value paid for the goods or services acquired, but in economics, the cost concept assumes a greater importance owing to its definition which now as an element includes the value for the money, resources spent, the utilities of time consumed, the risk taken as well as the opportunities forgone for the production of goods and services. In this unit, you will study the different concepts of cost. This unit will also include a discussion in the traditional and modern theory of cost and output relations.

2.1 OBJECTIVES

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

- Explain the different types of costs
- Describe the traditional theory of cost
- Examine the modern theory of cost

NOTES

2.2 TYPES OF COST

The cost concepts which are relevant to business analysis and decision-making can be grouped, on the basis of their nature and purpose, under two overlapping categories: (i) concepts used for accounting purposes, and (ii) analytical cost concepts used in economic analysis of business activities. We will discuss here some important concepts of the two categories. It is important to note here that this classification of cost concepts is only a matter of analytical convenience.

2.2.1 Accounting Cost Concepts

Accounting cost concepts include the following:

- 1. Opportunity Cost and Actual Cost:** Resources available to any person, firm or society are scarce, but have alternative uses with different returns. In order to maximize resource owners put their scarce resources to their most productive use and thus, they forego the income expected from all other uses of the resources. The income foregone is called *opportunity cost*. While measuring the opportunity cost, the return from the second best use only is taken into account. The *opportunity cost* may be defined as the expected returns from the second best use of the resources were foregone due to the scarcity of resources. The opportunity cost is also called *alternative cost*. If resources available to a person, a firm or a society were unlimited there would be no opportunity cost.

For example, suppose that a firm has a sum of ₹ 100,000 for which it has only two alternative uses. It can buy either a printing machine or alternatively a lathe machine both having productive life of 10 years. From the printing machine, the firm expects an annual income of ₹ 20,000 and from the lathe, ₹ 15,000. A profit maximizing firm would invest its money in the printing machine and forego the expected income from the lathe. The opportunity cost of the income from printing machine is the expected income from the lathe, i.e., ₹ 15,000.

Associated with the concept of opportunity cost is the concept of **economic rent** or **economic profit**. In our example of expected earnings firm printing machine and economic rent of the printing machine is the excess of its earning over the income expected from the lathe. That is, economic rent equals ₹ 20,000 – ₹ 15,000 = ₹ 5,000. The implication of this concept for a business man is that investing in the printing machine is preferable so long as its economic rent is greater than zero. Also, if firms know the economic rent of the various alternative uses of their resources, it will be helpful in the choice of the best investment avenue.

In contrast to the concept of *opportunity cost*, **actual costs** are those which are actually incurred by the firm in payment for labour, material, plant, building, machinery, equipment, travelling and transport, advertisement, etc.

The total money expenses, recorded in the books of accounts are for all practical purposes, the **actual costs**. In our example, the cost of printing machine, i.e., ₹ 100,000 is the actual cost. Actual cost comes under the accounting cost concept.

- 2. Business Costs and Full Costs:** *Business costs* include all the expenses which are incurred to carry out a business. The concept of business costs is similar to the actual or real costs. Business costs include all the payments and contractual obligations made by the firm together with the book cost of depreciation on plant and equipment. These cost concepts are used for calculating business profits and losses and for filing returns for income-tax and also for other legal purposes.

The concept of **full cost**, includes business costs, opportunity cost and normal profit. The opportunity cost includes the expected earning from the second best use of the resources, or the market rate of interest on the total financial capital and also the value of an entrepreneur's own services which are not charged for in the current business. Normal profit is a necessary minimum earning in addition to the opportunity cost, which a firm must receive to remain in its present occupation.

- 3. Explicit and Implicit or Imputed Costs:** *Explicit costs* are those which fall under actual or business costs entered in the books of accounts. The payments on account of wages and salaries, materials, license fee, insurance premium, depreciation charges are the examples of explicit costs. These costs involve cash payment and are recorded in normal accounting practices.

In contrast to explicit costs, there are certain other costs which do not take the form of cash outlays, nor do they appear in the accounting system. Such costs are known as **implicit** or **imputed costs**. Opportunity cost is an important example of implicit cost. For example, suppose an entrepreneur does not utilize his services in his own business and works as a manager in some other firm on a salary basis. If he sets up his own business, he foregoes his salary as manager. This loss of salary is the opportunity cost of income from his own business. This is an implicit cost of his own business. Thus, implicit wages, rent, and implicit interest are the wages, rents and interest which an owner's labour, building and capital, respectively, can earn from their second best use.

Implicit costs are not taken into account while calculating the loss or gains of the business, but they do appear as an important consideration in whether or not to retain a factor in its present use. The explicit and implicit costs together make the **economic cost**.

- 4. Out-of-Pocket and Book Costs:** The items of expenditure which involve cash payments including, both recurring and non-recurring expenses, are known as *out-of-pocket costs*. All the explicit costs (e.g., wages, rent, interest, cost of materials and maintenance, transport expenditure, etc.) fall

NOTES

NOTES

in this category. On the contrary, there are certain actual business costs which do not involve cash payments, but a provision is therefore made in the books of account and they are taken into account while finalising the profit and loss accounts. For example, payments made by a firm to itself, depreciation allowances and unpaid interest on the owner's own fund are the example of *book costs*.

2.2.2 Analytical Cost Concepts

Analytical cost concepts include the following:

- 1. Fixed and Variable Costs:** *Fixed costs* are those which remain fixed in volume over a certain level of output. Fixed cost does not vary with variation in the output between zero and a certain level of output. In other words, costs that do not vary for a certain level of output are known as *fixed costs*. The fixed costs include (i) costs of managerial and administrative staff, (ii) depreciation of machinery, building and other fixed assets, (iii) maintenance of land, etc. The concept of fixed cost is associated with the short-run. In the long run, no cost is fixed.

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

- 2. Total, Average and Marginal Costs:** *Total Cost (TC)* is the total expenditure incurred on the production of goods and service. It refers to the total outlays of money expenditure, both explicit and implicit, on the inputs used to produce a given level of output. It includes both fixed and variable costs. That is,

$$TC = TFC + TVC$$

Average Cost (AC) is of statistical nature—it is not an actual cost. It is obtained by dividing the total cost (*TC*) by the total output (*Q*), i.e.,

$$AC = \frac{TC}{Q} \quad \dots(2.1)$$

Marginal Cost (MC) is the addition to the total cost on account of producing one additional unit of the product. Or, marginal cost is the cost of the marginal unit produced. Marginal cost is calculated as $TC_n - TC_{n-1}$ where *n* is the number of units produced. Alternatively, given the cost function, *MC* can be defined as

$$MC = \frac{\partial TC}{\partial Q} \quad \dots(2.2)$$

These cost concepts are discussed in further detail in the following section. The total, average and marginal cost concepts are used in the economic analysis of firm's production activities.

- 3. Short-Run and Long-Run Costs:** Short-run and long-run cost concepts are related to variable and fixed costs, respectively, and often figure in economic analysis interchangeably.

Short-run costs are the costs which have short-run implication in production process and vary with the variation in output, the size of the firm remaining the same. *Short-run costs* are the same as *variable costs*. *Long-run costs*, on the other hand, are the costs which are incurred on the fixed assets like plant, building, machinery, etc. It is important to note that the running cost and depreciation of the capital assets are included in the short-run or variable costs.

Long-run costs are by implication the same as *fixed costs*. In the long-run, however, even the fixed costs become variable costs as the size of the firm or scale of production increases. Broadly speaking, the short-run costs are those associated with variables in the utilization of fixed plant or other facilities whereas long-run costs are associated with the changes in the size and kind of plant.

- 4. Incremental Costs and Sunk Costs:** Conceptually, *incremental costs* are closely related to the concept of marginal cost, but with a relatively wider connotation. While marginal cost refers to the cost of the marginal unit of output, incremental cost refers to the total additional cost associated with the decisions to expand the output or to add a new variety of product, etc. The concept of incremental cost is based on the fact that in the real world, it is not practicable (for lack of perfect divisibility of inputs) to employ factors for each unit of output separately. Besides, in the long run, when firms expand their production, they hire more of men, materials, machinery and equipments. The expenditures of this nature are incremental costs and not the marginal cost (as defined earlier). Incremental costs arise also owing to the change in product lines, addition or introduction of a new product, replacement of worn out plant and machinery, replacement of old technique of production with a new one, etc.

The *sunk costs* are those which are incurred once for all. Such costs cannot be altered, increased or decreased, by varying the rate of output. For example, once it is decided to make incremental investment expenditure and the funds are allocated and spent, all the preceding costs are considered to be the sunk costs since they accord to the prior commitment and cannot be revised or reversed or recovered when there is a change in market conditions or change in business decisions.

- 5. Historical and Replacement Costs:** *Historical cost* refers to the cost of an asset acquired in the past whereas *replacement cost* refers to the expenditure made for replacing an old asset. These concepts owe their significance to the unstable nature of price behaviour. Stable prices over time, other things given, keep historical and replacement costs on par with

NOTES

NOTES

each other. Instability in asset prices makes the two costs differ from each other.

Historical *cost* of assets is used for accounting purposes, in the assessment of the net worth of the firm. The replacement cost figures in business decisions regarding the renovation of the plant.

6. Private and Social Costs: We have so far discussed the cost concepts that are related to the working of the firm and that are used in the cost-benefit analysis of business decisions. There are, however, certain other costs which arise due to the functioning of the firm but do not normally figure in the business decisions nor are such costs explicitly borne by the firms. The costs on this category are borne by the society. Thus, the total cost generated by a firm's working may be divided into two categories: (i) costs paid out or provided for by the firms, and (ii) costs not paid or borne by the firms including the use of resources freely available plus the disutility created in the process of production. The costs of the former category are known as *private costs* and costs of the latter category are known as *external* or *social costs*. To mention a few examples of social cost: Mathura Oil Refinery discharging its wastage in the Yamuna river causes water pollution. Mills and factories located in a city cause air pollution, and so on. Let us now look at these concepts of cost in some detail.

Private costs are those which are actually incurred or provided for by an individual or a firm on the purchase of goods and services from the market. For a firm, all the actual costs both explicit and implicit are private costs. Private costs are internalized costs that are incorporated in the firm's total cost of production.

Social costs, on the other hand, refer to the total cost borne by the society due to production of a commodity. Social cost includes both private cost and the external cost. Social cost includes (a) the cost of resources for which the firm is not required to pay a price, i.e., atmosphere, rivers, lakes, and also for the use of public utility services³ like roadways, drainage system, etc., and (b) the cost in the form of 'disutility' created through air, water and noise pollution, etc. The costs of category (c) are generally assumed to equal the total private and public expenditure incurred to safeguard the individual and public interest against the various kinds of health hazards created by the production system. The private and public expenditure, however, serve only as an indicator of 'public disutility'—they do not give the exact measure of the public disutility or the social costs.

Cost Function

Cost function is a symbolic statement of the technological relationship between cost and output. In its general form, it is expressed by an equation. Cost function can be expressed also in the form of a schedule and a graph. In fact, tabular, graphical, and algebraic equation forms of cost function can be converted in the

form of each other. Going by its general form, total cost (TC) function is expressed as follows.

$$TC = f(Q)$$

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

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

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

Table 2.1 *Kinds of Cost Functions and Change in TC*

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

Check Your Progress

1. Actual cost is a sub-type of which concept of cost?
2. Name the cost which is made together with explicit and implicit costs.
3. Where are historical cost and replacement cost used in a firm?

2.3 TRADITIONAL THEORY OF COST

In this section, we will deal with the traditional theory of cost.

2.3.1 Short-Run Cost-Output Relations

The theory of cost deals with the behaviour of cost in relation to a change in output. In other words, the cost theory deals with cost-output relations. The basic principle of the cost behaviour is that the *total cost increases with increase in output*. This simple statement of an observed fact is of little theoretical and practical

NOTES

NOTES

importance. What is of importance from a theoretical and managerial point of view is not the absolute increase in the total cost but the direction of change in the average cost (AC) and the marginal cost (MC). The direction of change in AC and MC —whether AC and MC decrease or increase or remain constant—depends on the nature of the cost function. The specific form of the cost function depends on whether the time framework chosen for cost analysis is short-run or long-run. It is important to recall here that some costs remain constant in the short-run while all costs are variable in the long-run. Thus, depending on whether cost analysis pertains to short-run or to long run, there are two kinds of cost functions:

(i) short-run cost functions, and (ii) long-run cost functions,

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

Cost Concepts used in Cost Analysis

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

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

$$TC = TFC + TVC \quad \dots(2.3)$$

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

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

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

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

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

and $AC = AFC + AVC \quad \dots(2.4)$

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

$$MC = \frac{\Delta TC}{\Delta Q} \quad \dots(2.5)$$

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

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

Short-Run Cost Functions and Cost Curves

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

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

$$TC = a + bQ \quad \dots(2.6)$$

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

Given the cost function (Eq. 2.6), AC and MC can be obtained as follows.

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

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

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

Assuming an actual cost function given as

$$TC = 60 + 10Q \quad \dots(2.7)$$

the cost curves (TC , TVC and TFC) are graphed in Fig. 2.7.

Given the cost function (Eq. 2.7),

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

$$\text{and } MC = 10$$

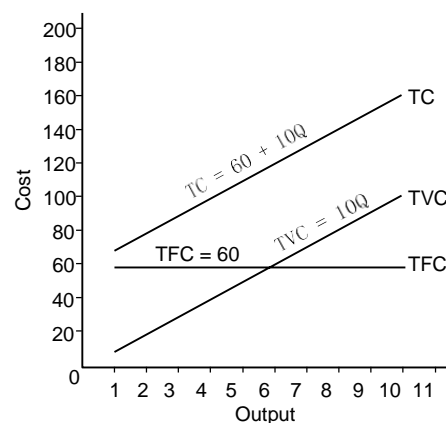


Fig. 2.1 Linear Cost Functions

NOTES

NOTES

Fig. 2.1 shows the behaviour of TC , TVC and TFC . The straight horizontal line shows TFC and the line marked $TVC = 10Q$ shows the movement in TVC . The total cost function is shown by $TC = 60 + 10Q$.

More important is to notice the behaviour of AC and MC curves in Fig. 2.2. Note that in case of a linear cost function MC remains constant, while AC continues to decline with the increase in output. This is so simply because of the logic of the linear cost function.

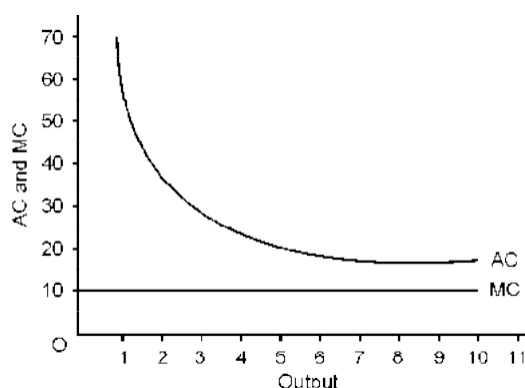


Fig. 2.2 AC and MC Curves Derived from Linear Cost Function

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

$$TC = a + bQ + Q^2 \quad \dots(2.8)$$

where a and b are constants.

Given the cost function (Eq. 2.8), AC and MC can be obtained as follows.

$$AC = \frac{TC}{Q} = \frac{a + bQ + Q^2}{Q} = \frac{a}{Q} + b + Q \quad \dots(2.9)$$

$$MC = \frac{\partial TC}{\partial Q} = b + 2Q \quad \dots(2.10)$$

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

$$TC = 50 + 5Q + Q^2 \quad \dots(2.11)$$

Given the cost function (Eq. 2.11),

$$AC = \frac{50}{Q} + Q + 5 \quad \text{and} \quad MC = \frac{\partial TC}{\partial Q} = 5 + 2Q$$

The cost curves that emerge from the cost function (2.11) are graphed in Fig. 2.3 (a) and (b). As shown in panel (a), while fixed cost remains constant at 50, TVC is increasing at an increasing rate. The rising TVC sets the trend in the total cost (TC). Panel (b) shows the behaviour of AC , MC and AVC in a quadratic cost function. Note that MC and AVC are rising at a constant rate whereas AC first declines and then increases.

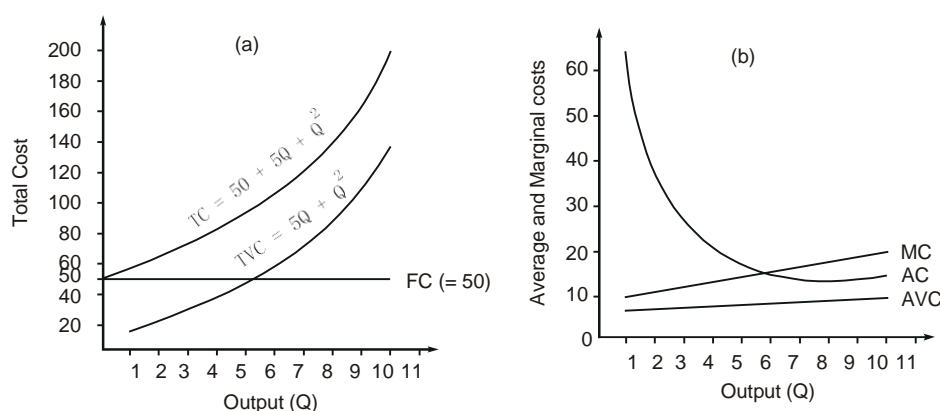


Fig. 2.3 Cost Curves Derived from a Quadratic Cost Function

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

$$TC = a + bQ - cQ^2 + Q^3 \quad \dots(2.12)$$

where a , b and c are the parametric constants.

From the cost function (2.12), AC and MC can be derived as follows.

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

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

Let us suppose that the cost function is empirically estimated as

$$TC = 10 + 6Q - 0.9Q^2 + 0.05Q^3 \quad \dots(2.13)$$

Given the cost function (2.35), the TVC function can be derived as

$$TVC = 6Q - 0.9Q^2 + 0.05Q^3 \quad \dots(2.14)$$

The TC and TVC , based on Eqs. (2.13) and (2.14), respectively, have been calculated for $Q = 1$ to 16 and presented in Table 2.2. The TFC , TVC and TC have been graphically presented in Fig. 2.4. As the figure shows, TFC remains fixed for the whole range of output, and hence, takes the form of a horizontal line— TFC . The TVC curve shows that the total variable cost first increases at a decreasing rate and then at an increasing rate with the increase in the output. The rate of increase can be obtained from the slope of the TVC curve. The pattern of change in the TVC stems directly from the law of increasing and diminishing returns to the variable inputs. As output increases, larger quantities of variable inputs are required to produce the same quantity of output due to diminishing returns. This causes a subsequent increase in the variable cost for producing the same output.

NOTES

NOTES

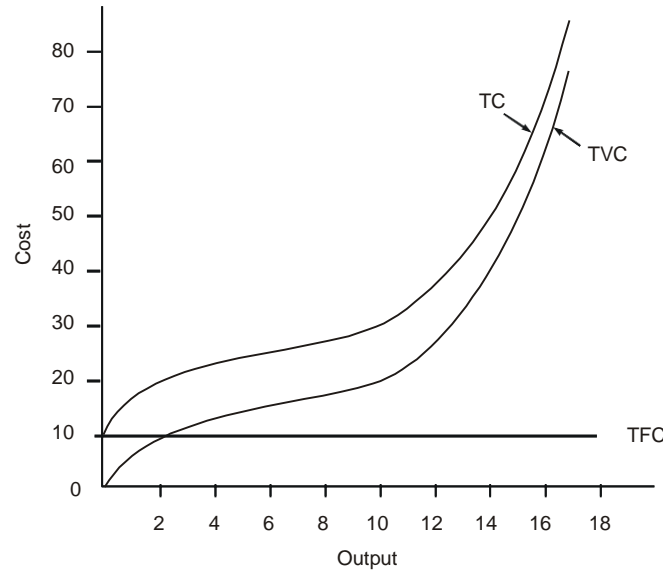


Fig. 2.4 TC, TFC and TVC Curves

Table 2.2 Cost-Output Relations

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

From Eqs. (2.13) and (2.14), we may derive the behavioural equations for *AFC*, *AVC* and *AC*. Let us first consider *AFC*.

Average Fixed Cost (AFC): As already mentioned, the costs that remain fixed for a certain level of output make the total fixed cost in the short-run. The fixed cost is represented by the constant term \underline{a} in Eq. (2.12) and $a = 10$ as given in Eq. (2.13). We know that

$$AFC = \frac{TFC}{Q} \dots\dots\dots (2.15)$$

Substituting 10 for TFC in Eq. 2.37, we get

$$AFC = \frac{10}{Q} \quad \dots(2.16)$$

Equation (2.16) expresses the behaviour of AFC in relation to change in Q . The behaviour of AFC for Q from 1 to 16 is given in Table 2.2 (col. 5) and presented graphically by the AFC curve in Fig. 2.5. The AFC curve is a rectangular hyperbola.

Average Variable Cost (AVC): As defined above, $AVC = TVC/Q$. Given the TVC function (Eq. 2.14), we may express AVC as follows.

$$\begin{aligned} AVC &= \frac{6Q - 0.9Q^2 + 0.05Q^3}{Q} \\ &= 6 - 0.9Q + 0.05Q^2 \end{aligned} \quad \dots(2.17)$$

Having derived the AVC function in Eq. (2.17), we may easily obtain the behaviour of AVC in response to change in Q . The behaviour of AVC for $Q = 1$ to 16 is given in Table 2.2 (col. 6), and graphically presented in Fig. 2.2 by the AVC curve.

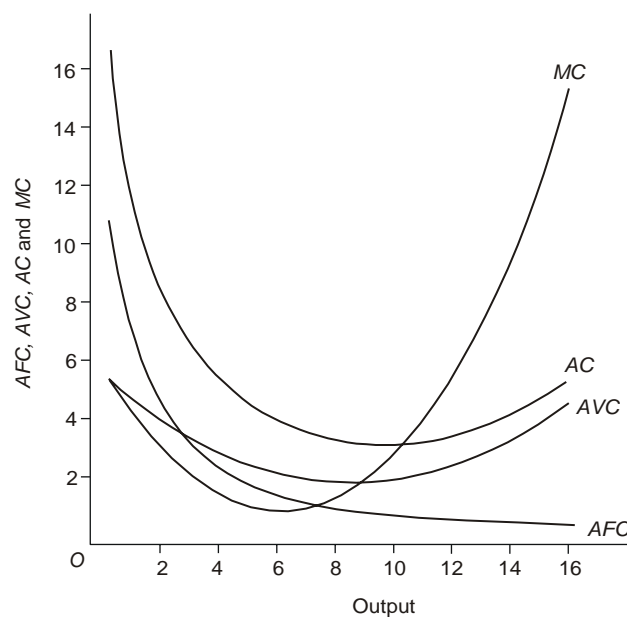


Fig. 2.5 Short-run AFC , AVC , AC and MC Curves

Critical Value of AVC: From Eq. (2.11), we may compute the critical value of Q in respect of AVC . *The critical value of Q (in respect of AVC) is one that minimizes AVC .* The AVC will be minimum when its rate of decrease equals zero. This can be accomplished by differentiating Eq. (2.17) and setting it equal to zero.

NOTES

NOTES

Thus, critical value of Q can be obtained as

$$\text{Critical value of } Q = \frac{\partial AVC}{\partial Q} = -0.9 + 0.10 Q = 0$$

$$0.10 Q = 0.9$$

$$Q = 9$$

In our example, the critical value of $Q = 9$. This can be verified from Table 2.2. The AVC is minimum (1.95) at output 9.

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

Substituting Eq. (2.13) for TC in the above equation, we get

$$\begin{aligned} AC &= \frac{10 + 6Q - 0.9Q^2 + 0.05Q^3}{Q} \\ &= \frac{10}{Q} + 6 - 0.9Q + 0.05Q^2 \end{aligned} \quad \dots(2.18)$$

The Eq. (2.18) gives the behaviour of AC in response to change in Q . The behaviour of AC for $Q = 1$ to 16 is given in Col. 7 of Table 2.2 and graphically presented in Fig. 2.5 by the AC curve. Note that AC curve is U -shaped.

Minimization of AC: One objective of business firms is to minimize AC of their product or, which is the same as, to optimize the output. The level of output that minimizes AC can be obtained by differentiating Eq. (2.18) and setting it equal to zero. Thus, the optimum value of Q can be obtained as follows.

$$\frac{\partial AC}{\partial Q} = \frac{10}{Q^2} - 0.9 + 0.1Q = 0$$

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

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

$$\text{or } Q^3 - 9Q^2 - 100 = 0 \quad \dots(2.19)$$

By solving⁴ equation (2.19) we get $Q = 10$.

Thus, the critical value of output in respect of AC is 10. That is, AC reaches its minimum at $Q = 10$. This can be verified from Table 2.2.

Marginal Cost (MC): The concept of marginal cost (MC) is useful particularly in economic analysis. MC is technically the first derivative of the TC function. Given the TC function in Eq. (2.13), the MC function can be obtained as

$$MC = \frac{\partial TC}{\partial Q} = 6 - 1.8Q + 0.15Q^2 \quad \dots(2.20)$$

Eq. (2.20) represents the behaviour of MC . The behaviour of MC for $Q = 1$ to 16 computed as $MC = TC_n - TC_{n-1}$ is given in Table 2.2 (col. 8) and graphically presented by the MC curve in Fig. 2.5. The critical value of Q with respect to MC is 6 or 7. This can be seen from Table 2.2.

Cost Curves and the Law of Diminishing Returns

Now, we will discuss the law of variable proportions the cost curves are related. Figures 2.4 and 2.5 represent the cost curves conforming to the short-term law of production, i.e., the law of diminishing returns. Let us recall the law: it states that when more and more units of a variable input are applied, other inputs held constant, the returns from the marginal units of the variable input may initially increase but it decreases eventually. The same law can also be interpreted in terms of decreasing and increasing costs. The law can then be stated as, if more and more units of a variable input are applied to a given amount of a fixed input, the marginal cost initially decreases, but eventually increases. Both interpretations of the law yield the same information—one in terms of **marginal productivity** of the variable input, and the other in terms of the **marginal cost**. The former is expressed through a production function and the latter through a cost function.

Figure 2.5 presents the short-run laws of return in terms of cost of production. As the figure shows, in the initial stage of production, both AFC and AVC are declining because of some internal economies. Since $AC = AFC + AVC$, AC is also declining. This shows the operation of the law of increasing returns to the variable input. However, beyond a certain level of output (i.e., 9 units in our example), while AFC continues to fall, AVC starts increasing because of a faster increase in the TVC . Consequently, the rate of fall in AC decreases. The AC reaches its minimum when output increases to 10 units. Beyond this level of output, AC starts increasing which shows that the law of diminishing returns comes into operation. The MC curve represents the change in both the TVC and TC curves due to change in output. A downward trend in the MC shows increasing marginal productivity of the variable input due mainly to internal economy resulting from increase in production. Similarly, an upward trend in the MC shows increase in TVC , on the one hand, and decreasing marginal productivity of the variable input, on the other.

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

- (a) Over the range of output both AFC and AVC fall, AC also falls because $AC = AFC + AVC$.
- (b) When AFC falls but AVC increases, change in AC depends on the rate of change in AFC and AVC .
 - (i) if decrease in $AFC >$ increase in AVC , then AC falls,
 - (ii) if decrease in $AFC =$ increase in AVC , AC remains constant, and
 - (iii) if decrease in $AFC <$ increase in AVC , then AC increase.
- (c) The relationship between AC and MC is of a varied nature. It may be described as follows:
 - (i) When MC falls, AC follows, over a certain range of initial output. When MC is falling, the rate of fall in MC is greater than that of AC ,

NOTES

NOTES

because in the case of MC the decreasing marginal cost is attributed to a single marginal unit while, in case of AC , the decreasing marginal cost is distributed over the entire output. Therefore, AC decreases at a lower rate than MC .

- (ii) Similarly, when MC increases, AC also increases but at a lower rate for the reason given in (i). There is, however, a range of output over which the relationship does not exist. Compare the behaviour of MC and AC over the range of output from 6 to 10 units (Fig. 2.5). Over this range of output, MC begins to increase while AC continues to decrease. The reason for this can be seen in Table 2.2: when MC starts increasing, it increases at a relatively low rate which is sufficient only to reduce the rate of decrease in AC —not sufficient to push the AC up. That is why AC continues to fall over some range of output even if MC increases.
- (iii) The MC curve intersects the AC at its minimum point. This is simply a mathematical relationship between MC and AC curves when both of them are obtained from the same TC function. In simple words, when AC is at its minimum, it is neither increasing nor decreasing; it is constant. When AC is constant, $AC = MC$. That is the point of intersection.

Output Optimization in the Short-Run

Optimization of output in the short-run has been illustrated graphically in Fig. 2.5. Optimization technique is repeated here for the sake of completeness.

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

$$TC = 200 + 5Q + 2Q^2 \quad \dots(2.21)$$

We have noted above that an optimum level of output is one that equalizes AC and MC . In other words, at optimum level of output, $AC = MC$. Given the cost function in Eq. (2.21),

$$AC = \frac{200 + 5Q + 2Q^2}{Q} = \frac{200}{Q} + 5 + 2Q \quad \dots (2.22)$$

and $MC = \frac{\partial TC}{\partial Q} = 5 + 4Q \quad \dots (2.23)$

By equating AC and MC equations, i.e., Eqs. (2.22) and (2.23), respectively, and solving them for Q , we get the optimum level of output. Thus,

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

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

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

Thus, given the cost function (2.21), the optimum output is 10.

2.3.2 Long-Run Cost-Output Relations: Envelope Curve

By definition, long-run is a period in which all the inputs—specifically, labour and capital—become variable. The variability of inputs is based on the assumption that in the long-run supply of all the inputs, including those held constant in the short-run, becomes elastic. The firms are, therefore, in a position to expand the scale of their production by hiring a larger quantity of both labour and capital. The long-run-cost-output relations, therefore, imply the relationship between the changing scale of the firm and the total output, whereas in the short-run this relationship is essentially one between the total output and the variable cost (labour).

Derivation of Long-run Cost Curves

To understand the long-run-cost-output relations and to derive long-run cost curves it will be helpful to imagine that a long-run is composed of a series of short-run production decisions. As a corollary of this, long-run cost curve is composed of a series of short-run cost curves. Based on this principle, we may now derive the long-run cost curves and study their relationship with output.

Long-run Total Cost Curve (LTC): In order to draw the long-run total cost curve, let us begin with a short-run situation. Suppose that a firm having only one plant has its short-run total cost curve as given by STC_1 , in panel (a) of Fig. 2.6. Let us now suppose that the firm decides to add two more plants to its size over time, one after the other. As a result, two more short-run total cost curves are added to STC_1 , in the manner shown by STC_2 and STC_3 in Fig. 2.6 (a). The LTC can now be drawn through the minimum points of STC_1 , STC_2 and STC_3 as shown by the LTC curve corresponding to each STC .

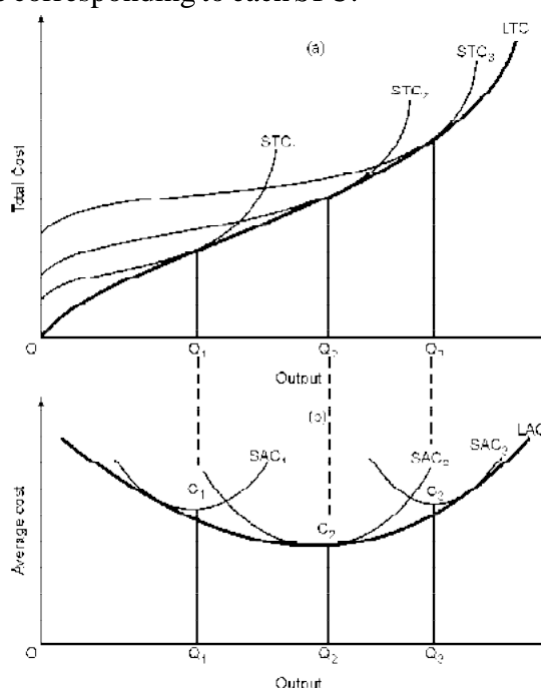


Fig. 2.6 Long-run Total and Average Cost Curves

NOTES

NOTES

Long-run Average Cost Curve (LAC): The long-run average cost curve (*LAC*) is derived by combining the short-run average cost curves (*SAC_s*). Note that there is one *SAC* associated with each *STC*. Given the *STC₁*, *STC₂*, *STC₃* curves in panel (a) of Fig. 2.6 there are three corresponding *SAC* curves as given by *SAC₁*, *SAC₂*, and *SAC₃* curves in panel (b) of Fig. 2.6. Thus, the firm has a series of *SAC* curves, each having a bottom point showing the minimum *SAC*. For instance, C_1Q_1 is minimum *AC* when the firm has only one plant. The *AC* decreases to C_2Q_2 when the second plant is added and then rises to C_3Q_3 after the addition of the third plant. The *LAC* curve can be drawn through the bottom points of the *SAC₁*, *SAC₂* and *SAC₃* as shown in Fig. 2.6(b). The *LAC* curve is also known as the **Envelope Curve** or **Planning Curve** as it serves as a guide to the entrepreneur in his plans to expand production.

Alternatively, the *SAC* curves can be derived from the data given in the *STC* schedule, from *STC* function or straightaway from the *STC* curve. Similarly, *LAC* can be derived from *LTC*-schedule, *LTC* function or from *LTC*-curve.

The relationship between *LTC* and output, and between *LAC* and output can now be easily derived. It is obvious from the *LTC* that the long-run cost-output relationship is similar to the short-run cost-output relation. With the subsequent increases in the output, *LTC* first increases at a decreasing rate, and then at an increasing rate. As a result, *LAC* initially decreases until the optimum utilization of the second plant and then it begins to increase. These cost-output relations follow the 'laws of returns to scale'. When the scale of the firm expands, unit cost of production initially decreases, but ultimately increases as shown in Fig. 2.6(b). The decrease in unit cost is attributed to the internal and external economies of scale and the eventual increase in cost is linked to the internal and external diseconomies. The economies and diseconomies of scale are discussed in the following section.

Long-run Marginal Cost Curve (LMC): The long-run marginal cost curve (*LMC*) is derived from the short-run marginal cost curves (*SMC_s*). The derivation of *LMC* is illustrated in Fig. 2.7 in which *SACs* and *LAC* are the same as in Fig. 2.6 (b). To derive the *LMC*, consider the points of tangency between *SACs* and the *LAC*, i.e., points *A*, *B* and *C*. In the long-run production planning, these points determine the output levels at the different levels of production. For example, if we draw perpendiculars from points *A*, *B* and *C* to the X-axis, the corresponding output levels will be OQ_1 , OQ_2 and OQ_3 . The perpendicular AQ_1 intersects the *SMC₁* at point *M*. It means that at output OQ_1 , *LMC* is MQ_1 . If output increases to OQ_2 , *LMC* rises to BQ_2 . Similarly, CQ_3 measures the *LMC* at output OQ_3 . A curve drawn through points *M*, *B* and *N*, as shown by the *LMC*, represents the behaviour of the marginal cost in the long-run. This curve is known as the long-run marginal cost curve, *LMC*. It shows the trends in the marginal cost in response to the changes in the scale of production.

Some important inferences may be drawn from Fig. 2.7. The *LMC* must be equal to *SMC* for the output at which the corresponding *SAC* is tangent to the

LAC . At the point of tangency, $LAC = SAC$. Another important point to notice is that LMC intersects LAC when the latter is at its minimum, i.e., point B . There is one and only one short-run plant size whose minimum SAC coincides with the minimum LAC . This point is B where

$$SAC_2 = SMC_2 = LAC = LMC$$

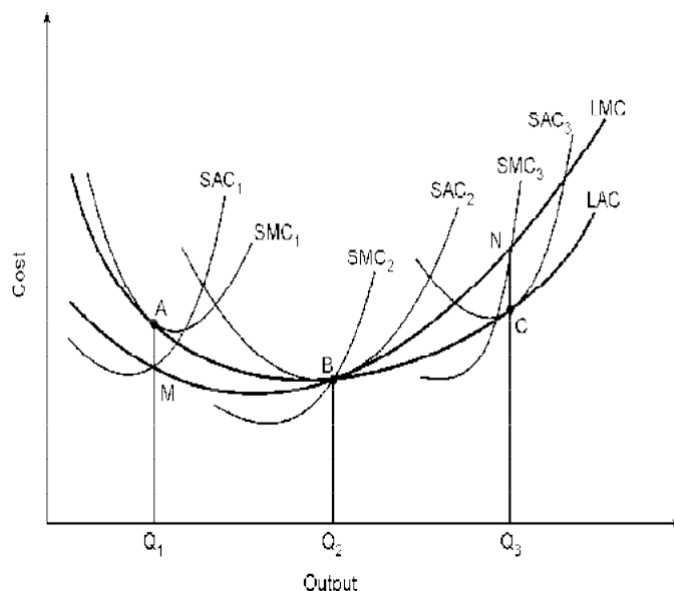


Fig. 2.7 Derivation of LMC

Conceptually, the optimum size of a firm is one which ensures the most efficient utilization of resources. Practically, the optimum size of the firm is one which minimizes the LAC . Given the state of technology over time, there is technically a unique size of the firm and level of output associated with the least-cost concept. In Fig. 2.7, the optimum size consists of two plants which produce OQ_2 units of a product at minimum long-run average cost (LAC) of BQ_2 . The downtrend in the LAC indicates that until output reaches the level of OQ_2 , the firm is of less than optimal size. Similarly, expansion of the firm beyond production capacity OQ_2 , causes a rise in SMC and, therefore, in LAC . It follows that given the technology, a firm aiming to minimize its average cost over time must choose a plant which gives minimum LAC where $SAC = SMC = LAC = LMC$. This size of plant assures the most efficient utilization of the resource. Any change in output level—increase or decrease—will make the firm enter the area of inoptimality.

2.4 MODERN THEORY OF COST

Some economists, especially the American economist George Stigler, have questioned theoretically as well as empirically the U-shaped cost curves of 'the traditional theory of cost' and have attempted to establish that the shape of the cost curves, at least in the long run, is L-shaped. However, this point of view does

NOTES

NOTES

not appear to have received a general recognition by the economists⁸ or as much attention as the traditional theory of cost, at least in the context of pricing theory. One possible reason is that the traditional theory of cost has a greater application to the theory of price determination and has a greater predicting power than the 'modern theory'. However, this section provides a brief description of the 'modern approach' to the theory of cost. Incidentally, like traditional theory of cost, modern theory too analyses cost-output relationships in the short-run and long-run framework.

Modern Approach to Short-run Cost Behaviour: Saucer-Shaped Cost Curve

Like traditional theory of cost, modern theory recognizes that in the short-run,

$$TC = TFC + TVC$$

and

$$AC = AFC + AVC$$

In traditional as well as in modern theory of cost, *TFC* includes the following elements of costs:

- (i) the salaries of administrative staff and related expenses;
- (ii) the salaries of direct production labour paid on fixed-term basis;
- (iii) standard depreciation allowance; and
- (iv) maintenance cost of land and building.

This point onwards, the modern theory deviates from the traditional theory. *Traditional theory* assumes optimum capacity of a plant to be technically given (where *SAC* is minimum) and a cost-minimizing firm has no choice, but to utilize the plant to its optimum capacity. On the other hand, *modern theory of cost* emphasizes that firms, in their production planning, choose a plant with flexible capacity, i.e., a plant with **built-in reserve capacity**. According to the modern theory, firms want to have some *reserve capacity*, as a matter of *planning* for the following reasons:

- (i) to meet the 'seasonal' and eventual increase in demand;
- (ii) to avoid loss of production due to break-down and repair works;
- (iii) to have provision for meeting anticipated growth in demand;
- (iv) to take the advantage of technology providing built-in reserve capacity;
- (v) to build excess capacity in land and building for expansion, if required; and
- (vi) to make full utilization of excess 'organizational and administrative' capacity.

Under these conditions, a firm does not necessarily choose a plant that gives the lowest cost of production. Instead, it chooses a plant (a set of machinery) that gives 'maximum flexibility' in production with minor adjustment in technique. For example, let us suppose that the firm has the option of setting up a plant which has an *absolute limit* to produce a commodity at the minimum cost. This absolute limit is shown by the quantity OQ_2 in Fig. 2.8. Note that if the firm

NOTES

chooses this plant, it can produce a maximum quantity of OQ_2 at the minimum AFC (average fixed cost) as shown in the figure by the boundary line BQ_2 . Since there is no excess capacity, the firm cannot produce any quantity beyond OQ_2 even if demand increases and hence the firm will not be able to take the advantage of rising demand for its product. Therefore, the firm chooses a flexible plant capable of producing more than OQ_2 with minor adjustment or alternation in the production technique. For example, let us suppose that the firm chooses a flexible plant with absolute limit of output OQ_1 as shown by the boundary line AQ_1 . Now let the firm anticipate a rise in demand for its product and add a small-unit machinery to its flexible plant at the output level OQ_1 . As Fig 2.8 shows, with the addition of a small-unit machinery, firm's AFC increases from point c to point a on the boundary line AQ_1 . But what is important from the firm's point of view is that the firm can increase its production beyond OQ_2 to meet the anticipated increase in demand. Though its AFC increases initially, it declines as production increases, as shown by the curve ab and goes below the limit set by the inflexible plant and the firm is a gainer.

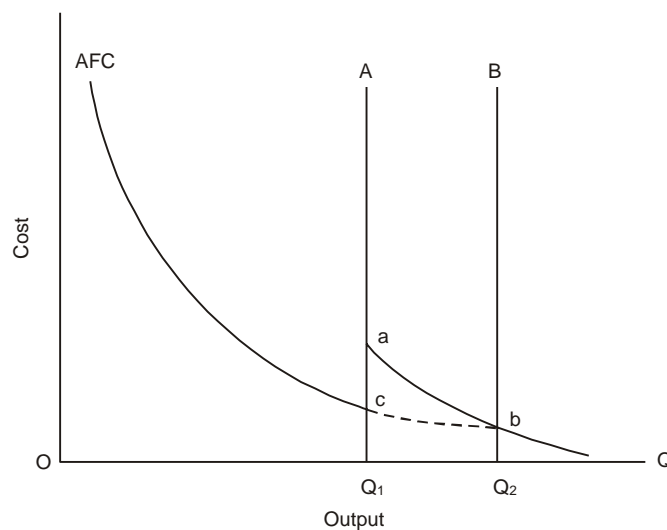


Fig. 2.8 The Built-in Reserve Capacity and AFC Curve

What Happens to the Average Variable Cost? The *average variable cost*, as in traditional theory, includes *average cost* of (a) direct labour, (b) raw materials, and (c) running cost of machinery. There is however a difference between the *short-run average variable cost (SAVC)* curves of the traditional and modern cost theories. While in traditional theory, the *SAVC* curve is U-shaped, in modern theory, it is saucer-shaped or bowl-shaped. Part (a) of Fig. 2.9 shows *SAVC* curve of the traditional theory, and part (b) shows *SAVC* curve of the modern theory.

NOTES

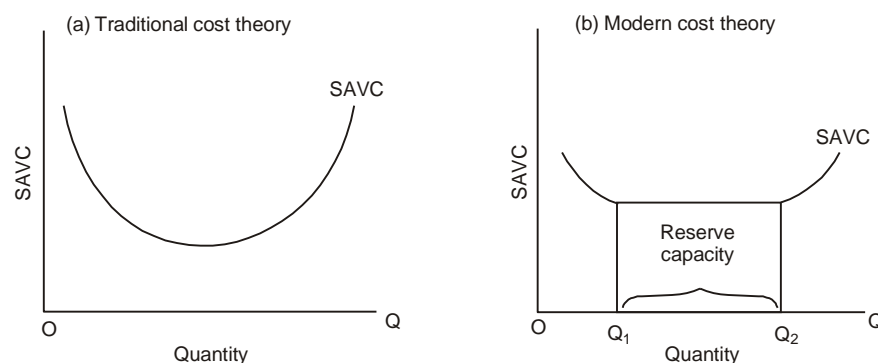


Fig. 2.9 The Traditional and Modern SAVC Curves

As part (b) of Fig. 2.9 shows, according to the modern theory cost, the *SAVC* remains constant over a long stretch of output between OQ_1 and OQ_2 . The constancy of *SAVC* in the modern theory is attributed to the *built-in reserve capacity* of the flexible plant. The utilization of the built-in reserve capacity keeps the *SAVC* constant. This is an 'innovative' aspect of the modern theory of cost. In the traditional theory, there is no such built-in reserve capacity and therefore *SAVC* begins to rise once the technically efficient level of output is reached.

SAVC and SMC curves: A more important aspect of the modern theory of cost is the nature of and relationship between the *SAVC* and the *SMC* curves. The derivation of *SAVC* and the *short-run marginal cost (SMC)* curves is illustrated in Fig. 2.10. The *SAVC* curve is the same as in Fig. 2.9. The *SMC* curve follows the pattern of the traditional theory. The *SMC* decreases with increase in output upto a certain level.

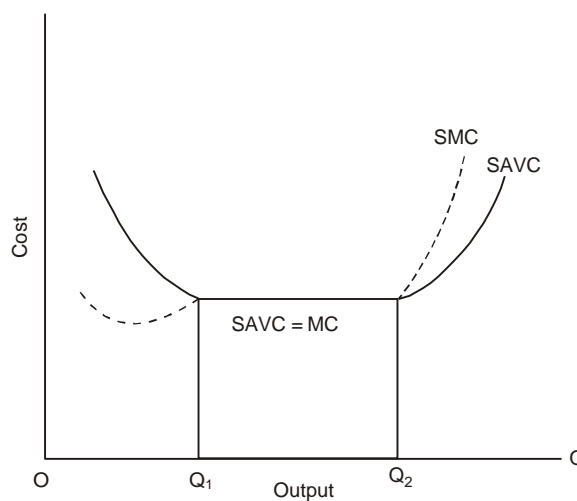


Fig. 2.10 Modern SAVC and SMC Curves

This behaviour of *SMC* curve is shown in Fig. 2.10 till the output OQ_1 . However, in the range of output, between OQ_1 and OQ_2 , the *SAVC* is constant. It is therefore equal to *SMC*. We know that when *SMC* begins to rise, it rises faster than *SAVC*. This behaviour of *SMC* is shown begin at output OQ_2 and

continue beyond. Beyond output OQ_2 , the SMC begins to rise and it rises faster than the $SAVC$ as is the case in the traditional-theory.

Short-run Average Cost (SAC) Curves: As in traditional theory, in modern theory of cost, $SAC = AFC + SAVC$. The AFC includes *normal profit*. Derivation of the SAC curve in the modern theory is illustrated in Fig. 2.11. The $SAVC$ curve (and also the SMC curve) is similar to one given in Fig. 2.10. For the derivation of the SAC curve, the AFC curve is added to Fig. 2.11. The SAC curve is the vertical summation of the $SAVC$ and AFC curves.

As Fig. 2.11 shows, AFC falls continuously whereas $SAVC$ decreases till output OQ_1 and remains constant between output OQ_1 and OQ_2 . Therefore, a vertical summation of AFC and $SAVC$ curves gives the SAC curve which declines continuously till output OQ_2 . Thus, in modern theory of cost, SAC decreases until the *built-in reserve capacity* is fully exhausted. The *reserve capacity* is exhausted at output OQ_2 .

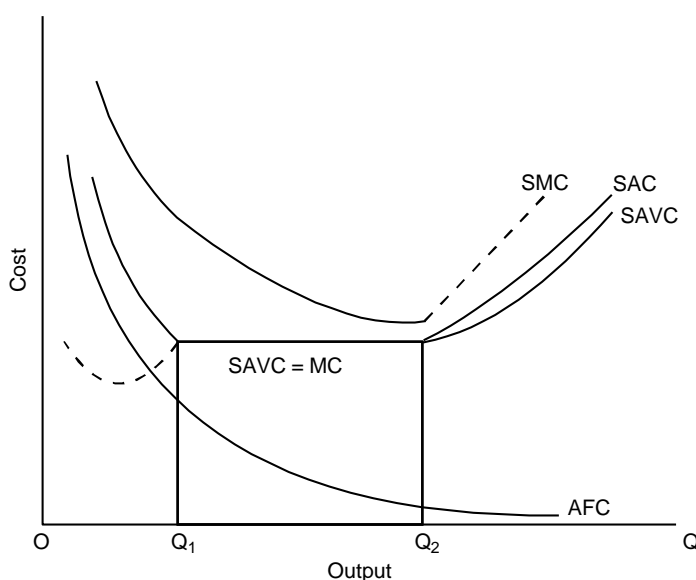


Fig. 2.11 Derivation of the Modern SAC Curve

Beyond output OQ_2 , therefore, SAC begins to increase and goes on increasing following the increase in $SAVC$ while decreasing AFC loses its significance.

Modern Approach to Long-run Cost Behaviour: The L-shaped Scale Curve

In respect of long-run cost behaviour, the modern theory of cost distinguishes between *production costs* and *managerial costs*. Both these costs are variable in the longrun. The behaviour of these costs determines the shape of the long-run average cost curve (LAC). According to the modern theory, the long-run LAC is broadly L-shaped. Let us now look at the behaviour of the *production* and *managerial costs* in the long run and how they determine the shape of the LAC curve.

NOTES

NOTES

Production Cost Behaviour: Production cost decreases steeply in the beginning with the increase in the scale of production but the rate of decrease slows down as the scale increases beyond a certain level of production. The decrease in the production costs is caused by the technical economies which taper off when the scale of production reaches its *technical optimum scale*. Nevertheless, some economies of scale are always available to the expanding firms due to (i) decentralization and improvement in skills; and (ii) decreasing cost of repairs per unit of output. In addition, in case of multi-product firms producing some of their raw materials and equipments have economies in material cost compared to purchases made from outside.

Managerial Cost Behaviour: The modern theory of cost assumes that, in modern management technology, there is a fixed managerial or administrative set up with a certain scale of production. When the scale of production increases, management set up has to be accordingly expanded. It implies that there is a link between the scale of production and the cost of management. According to the modern theory, the managerial cost first decreases but begins to increase as the scale of production is expanded beyond a certain level.

What Makes LAC L-shaped? *The net effect of decreasing production cost and increasing managerial cost determines the shape of the long-run average cost (LAC). Recall that production cost continues to decrease, though slowly beyond a certain scale of production and managerial cost too decreases initially but rises later. In the initial stage of production, therefore, the LAC decreases very steeply. Beyond a certain scale of production, however, while production cost continues to decline, management cost begins to rise. According to the modern theory of cost, the rise in managerial cost is more than offset by the decrease in the production costs. Therefore, the LAC continues to fall but very slowly. In case the decrease in production cost is just sufficient to offset the rise in the managerial cost, the LAC becomes constant. This makes LAC an L-shaped curve.*

Derivation of the LAC Curve: The derivation of the LAC curves is illustrated in Figs. 2.12 and 2.13. Fig. 2.12 shows the decreasing LAC curve. Let us suppose that, given the technology, the optimum scale of production consists of four plants and SAC curves from SAC_1 to SAC_4 in Fig. 2.13 represent the addition of four plants to the production scale in each period of time. Clower and Due have found that firms use 'normally' only $2/3$ to $3/4$ of the plant size. This is called 'load factor'. The *load factor* is the 'ratio of average actual rate of use to the capacity or best rate of use, and this load factor will generally be smaller than one'. The points A, B, C and D on the SACs mark the 'load factor' in case of each plant, respectively—it may be any value between $2/3$ and $3/4$ of the plant size. By drawing a curve through the 'load-factor' points, we get the LAC curve. If there is a larger

number of plants, we will get much larger number of load factor points and draw a smooth LAC curve as shown in Fig. 2.12.

*Different Concepts
of Cost*

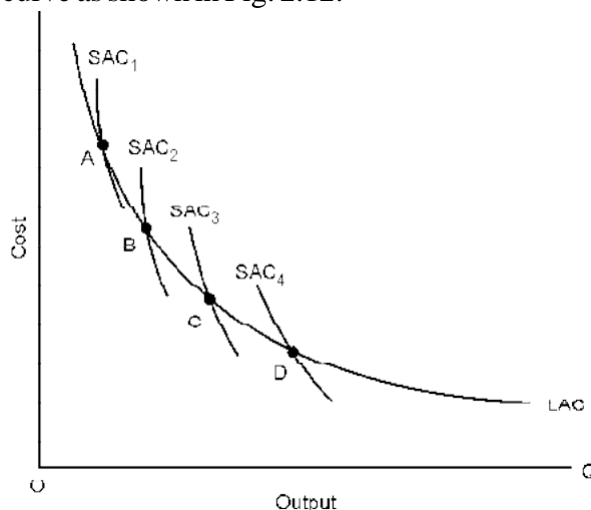


Fig. 2.12 Derivation of LAC Curve in Modern Theory

To compare the LAC of the modern and traditional theories of cost, two points need to be noted: (i) the LAC curve of the modern theory does not show the tendency to turn up even at a very large scale of production whereas the traditional LAC curve does turn up; (ii) unlike traditional LAC forming an *envelope curve*, modern LAC intersects the SAC s.

If scale of production has a *minimum optimal scale of plant*, as shown by output level OQ in part (b) of Fig. 2.13, all economies of scale are achieved at output OQ and the LAC becomes constant even if scale of production is expanded. In this case, the LMC lies below the LAC till the *minimum optimal scale of plant* is reached, as shown in part (a) of Fig. 2.13. When the firm operates in the range of no-scale-economies, i.e., beyond output OQ in part (b) of Fig. 2.13, the LAC becomes constant and the LAC curve takes the shape of a horizontal line. Both the parts (declining and constant) of the LAC curve put together make it roughly L-shaped.

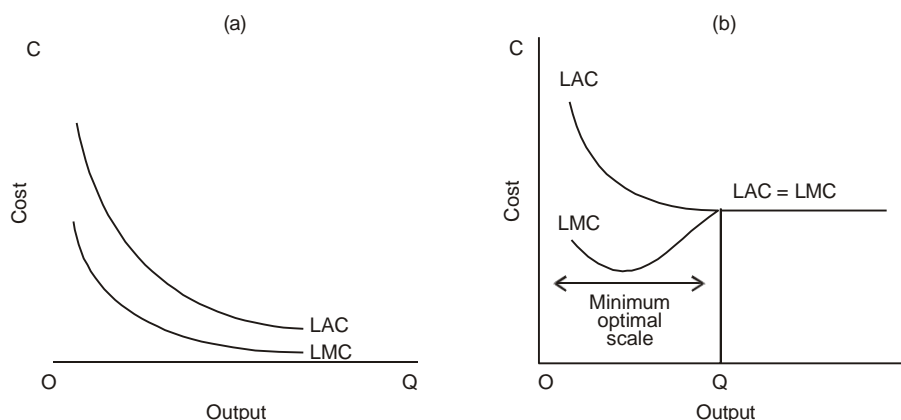


Fig. 2.13 Derivation of the L-shaped LAC Curve

NOTES

NOTES

From practical point of view, the modern *LAC* curve is regarded to be more realistic. But from analytical and prediction point of view, the traditional cost curves still hold the ground firmly. In fact, the so called ‘modern theory of cost’ is a modification of the traditional theory on the basis of empirical data in some manufacturing industries of some countries.

Learning Curve

In many industries, some firm specialize in one particular good or service and continue to produce it over time. As a result, they gain additional knowledge, expertise and efficiency and are able to make technological advances in the field of production. The economists have found that these factors result in continuous decline in their average cost with increase in output. A curve drawn to show the continuous decline in the average cost is called *learning curve*. The *learning curve* is, in fact, the long-run average cost curve showing a continuous decline in average cost with increase in output over time, as shown in Fig. 2.14 (a) and (b). The shape of the learning curve depends on the nature of the product. As shown by Mansfield *et. al.* Fig. 2.14 (a) shows the learning curve for the production of portable turbine and Fig. 2.14 (b) shows the learning curve for production of optical equipment. The continuous decrease in long-run average cost, as shown in Fig. 2.14, results from the continuous technological improvement, on the one hand, and knowledge, expertise and efficiency gained by the firms by producing the same commodity over time, on the other.

It is **important** to note here that average cost is estimated by dividing **cumulative total cost** by **cumulative total output** increasing at constant rate. This implies that when firm continues to produce the same commodity over time, its cumulative cost increases less than proportionate to a constant increase in output.

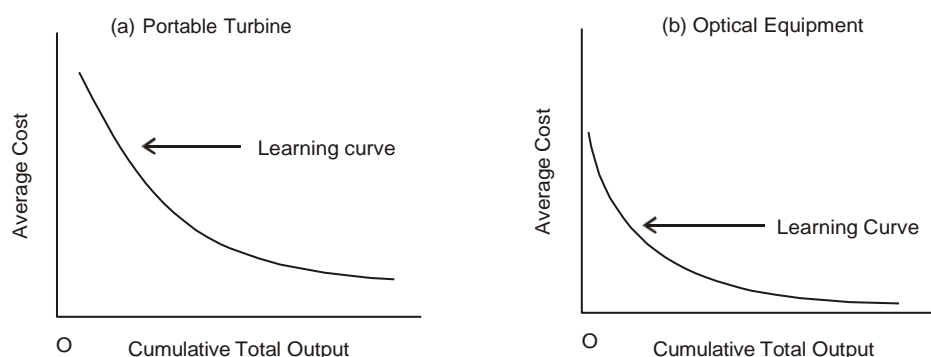


Fig. 2.14 The Learning Curve

The learning curve is generally expressed by the following form of AC function.

$$AC = BQ^a$$

where AC average cost; Q is output; B is the cost of the first unit of output; and power a is a *negative* constant.

In its double logarithmic form, the AC function for learning curve can be expressed as

$$\text{Log } AC = \log B + a \log Q$$

In its logarithmic form of the learning cost function, a gives the slope of the learning curve. The numerical values of parameters B and a are estimated by regression techniques by using historical data of cumulative output and cost.

Many firms in both manufacturing and service sectors have adopted pricing strategy based on learning curve, mainly by the firms manufacturing airplanes, ships, semi-conductor chips, domestic appliances, refined petroleum products and power plants. The learning curve is used to assess and forecast the future requirements for inputs including manpower, raw materials, and machinery.

Check Your Progress

4. What is the critical value of AVC ?
5. What is technically the first derivative of TC function?
6. Where does MC curve intersect the AC curve?
7. What keeps the $SAVC$ constant in the modern theory of cost?
8. Name the costs which are distinguished in respect of long-run cost behaviour in modern theory of cost.

2.5 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. Actual cost is a sub-type of accounting cost concept.
2. The explicit and implicit costs together make economic cost.
3. Historical cost of assets is used for accounting purposes, in the assessment of the net worth of the firm. The replacement cost figures in business decisions regarding the renovation of the plant.
4. The critical value of Q (in respect of AVC) is one that minimizes AVC . The AVC will be minimum when its rate of decrease equals zero.
5. MC is technically the first derivative of the TC function.

NOTES

NOTES

- 6. The MC curve intersects the AC at its minimum point.
- 7. The constancy of the SAVC in the modern theory is attributed to the built-in reserve capacity of the flexible plant.
- 8. In respect of long-run cost behaviour, the modern theory of cost distinguishes between production costs and managerial costs.

2.6 SUMMARY

- The cost concepts which are relevant to business analysis and decision-making can be grouped, on the basis of their nature and purpose, under two overlapping categories: (i) concepts used for accounting purposes, and (ii) analytical cost concepts used in economic analysis of business activities.
- Opportunity cost and actual cost, business cost and full cost, explicit and implicit cost, out of pocket and book cost constitute the accounting concepts whereas analytical cost concepts include fixed and variable cost, total, average and variable cost, short-run and long-run costs, incremental and sunk cost and historical and replacement and private and social costs.
- Cost function is a symbolic statement of the technological relationship between cost and output. In its general form, it is expressed by an equation. Cost function can be expressed also in the form of a schedule and a graph.
- The theory of cost deals with the behaviour of cost in relation to a change in output. In other words, the cost theory deals with cost-output relations. The basic principle of the cost behaviour is that the *total cost increases with increase in output*.
- The direction of change in AC and MC—whether AC and MC decrease or increase or remain constant—depends on the nature of the cost function. The specific form of the cost function depends on whether the time framework chosen for cost analysis is short-run or long-run. Thus, depending on whether cost analysis pertains to short-run or to long run, there are two kinds of cost functions: (i) short-run cost functions, and (ii) long-run cost functions.
- The long-run cost-output relations, therefore, imply the relationship between the changing scale of the firm and the total output, whereas in the short-run this relationship is essentially one between the total output and the variable cost (labour).

- Some economists, especially George Stigler⁷, have questioned theoretically as well as empirically the U-shaped cost curves of 'the traditional theory of cost' and have attempted to establish that the shape of the cost curves, at least in the long run, is L-shaped. However, this point of view does not appear to have received a general recognition by the economists⁸ or as much attention as the traditional theory of cost, at least in the context of pricing theory.

NOTES

2.7 KEY WORDS

- **Cost Function:** It is a symbolic statement of the technological relationship between cost and output.
- **Long-run:** It refers to a period in which all the inputs—specifically, labour, and capital—become variable.
- **Learning Curve:** It refers to the long-run average cost curve showing a continuous decline in average cost with increase in output over time.

2.8 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short Answer Questions

1. Differentiate between business and full costs.
2. Write a short note on incremental and sunk costs.
3. Briefly state some of the important relationships between costs in analysing short-run cost behaviour.
4. List the elements of costs included in the TFC in traditional as well as modern theory of cost.
5. Why do modern firms want to have some reserve capacity as a matter of planning?

Long Answer Questions

1. Explain the accounting cost concepts.
2. Discuss the analytical cost concepts.
3. Describe the cost curves made by linear, quadratic and cubic cost functions.
4. Examine long-run cost-output relations.

NOTES

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 CONCEPT OF REVENUE

Concept of Revenue

Structure

- 3.0 Introduction
- 3.1 Objectives
- 3.2 Types of Revenue
- 3.3 Price Elasticity and Revenue
- 3.4 Answers to Check Your Progress Questions
- 3.5 Summary
- 3.6 Key Words
- 3.7 Self Assessment Questions and Exercises
- 3.8 Further Readings

NOTES

3.0 INTRODUCTION

As mentioned in the previous unit, revenue is an important concept for firms in order for them to meet their objective of profit maximization. Revenue, as the name suggests, refers to the profit or income earned by the businesses for selling their goods and services at different prices. The concept of revenue, like cost, is related to the study of the total, average and marginal revenue and their relationship amongst themselves. This information allows the firm to plan the prices at which they should be selling their products or services. In this unit, you will be introduced to the concept of revenue and out of its varied applications, you will also learn how the concept of price elasticity and its relation to revenue is explored by the firm to meet their missions.

3.1 OBJECTIVES

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

- Discuss the various types of revenue
- Explain the relationship between price elasticity and revenue

3.2 TYPES OF REVENUE

The amount which a firm receives from the sale of the output is known as revenue. According to Prof. Dooley, 'the revenue of a firm is its sales receipts or income.'

Revenue of a firm can be divided into three types:

1. Total Revenue

2. Average Revenue

3. Marginal Revenue

Let us discuss each of these types.

NOTES

1. Total Revenue

Total Revenue is the amount receipt from the sale of the output. The total revenue depends on the total sale of the firm and the price of per unit output.

$$TR = Q \times P$$

Where TR = Total Revenue

Q = Quantity or unit sales

P = Per unit price of the commodity

According to Prof. Dooley, –the revenue of a firm is its sales receipts or income. ||

Average Revenue

We can obtain the average revenue by dividing the total revenue by total unit sold.

Where AR = Average Revenue

TR = Total Revenue

Q = Total unit sold

$$AR = TR / Q$$

It should be noted that AR or average revenue is equal to price, in the coming topics we shall discuss why the average revenue is equal to price.

2. Marginal Revenue

Marginal Revenue is the amount of money received from the sale of additional unit or in other words, additional revenue received by the sale of an additional unit.

$$MR = TR_n - TR_{n-1} \text{ Or } MR = \Delta TR / \Delta Q$$

Where MR = Marginal Revenue

ΔTR = Change in total Revenue

ΔQ = Change in the amount of sale

TR_n = Total Revenue of n units. TR_{n-1} = Total Revenue of n-1 units.

We can better understand the concept of TR, AR and MR with the help of a table that is given below,

Table 3.1 Different Concepts of Revenue

Concept of Revenue

Quantity Sold	Price per Unit	(Total Revenue) TR	Average Revenue AR	Marginal Revenue MR
1	6	6	6	6
2	5	10	5	4
3	4	12	4	2
4	3	12	3	0
5	2	10	2	-2
6	1	6	1	-4

NOTES

According to the table, we can get the total revenue by the multiplication of quantity sold and price per unit. Again by dividing the total revenue by quantity sold we can obtain the Average revenue. The marginal revenue as we have discussed earlier it the revenue receive by the sale of additional unit. We can better understand the properties of TR, AR and MR curves with the help of a diagram.

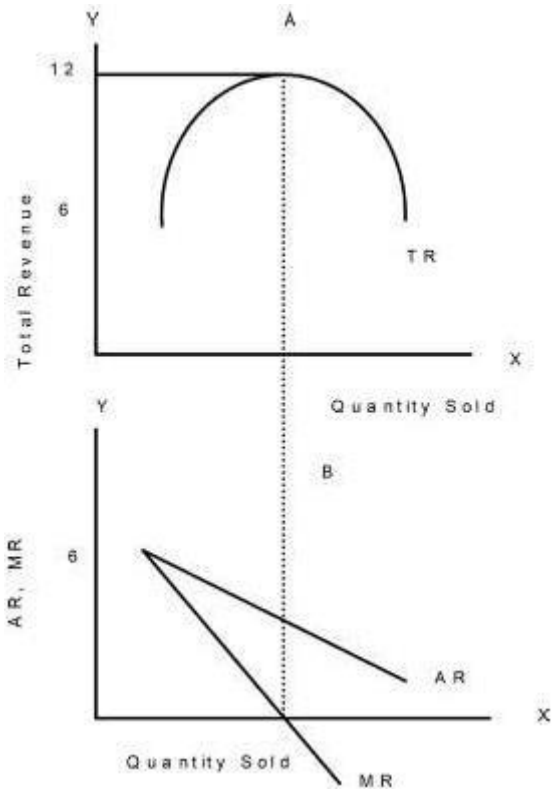


Fig. 3.1 Properties of TR, AR and MR

NOTES

1. When total revenue is at its maximum, Marginal revenue is equals to zero.
2. When total revenue is decreasing that time Marginal Revenue is negative.
3. Total revenue first increases and after a certain point it decreases.

Why Average Revenue Equals to Marginal Revenue Under Perfect Competition

When a firm can not change the price to affect the sale, in other words a firm under perfect competition can not decrease the price to increase its sale, in this case average revenue is always equals to marginal revenue.

Now according to the above table, the price is given to the firm is Rs. 5, and the quantity sold by the firm is 1 to 5. The total revenue of the firm is increasing from 5 to 25. With the help of total revenue we can calculate the amount of average revenue and marginal revenue, when we calculate them we found that the average revenue is equals to marginal revenue. It is because a firm can not change the price. When there is no change in price the average revenue is always equals to marginal revenue.

In imperfect competition firm is a price maker, no a price taker. A firm can decrease the price to increase the amount of sale or firm can increase the price to earn maximum profit. When there is a change in the price (in imperfect competition) the average revenue can not be equals to marginal revenue.

Quantity sold	Price	Total Revenue	Average Revenue	Marginal Revenue
1	5	5	5	5
2	5	10	5	5
3	5	15	5	5
4	5	20	5	5
5	5	25	5	5

3.3 PRICE ELASTICITY AND REVENUE

While the computation of revenue is helpful in many areas of economics, in this section, you will study one of the crucial aspects related to price elasticity.

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 due to

one per cent change in the price of the commodity. A formal definition of price elasticity of demand (e_p) is given as

$$e_p = \frac{\text{Percentage change in quantity demanded}}{\text{Percentage change in price}} \quad \dots(3.1)$$

A general formula for calculating coefficient of price elasticity, as given in Eq. (3.1), is given as follows:

$$\begin{aligned} e_p &= \frac{\frac{\Delta Q}{Q}}{\frac{\Delta P}{P}} = \frac{\Delta Q}{Q} \times \frac{P}{\Delta P} \\ &= \frac{\Delta Q}{\Delta P} \times \frac{P}{Q} \quad \dots(3.2) \end{aligned}$$

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

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

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

Price Elasticity and Total Revenue

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

$$Q = 100 - 5P$$

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

$$P = 20 - 0.2Q \quad \dots(3.3)$$

Given the price function, TR can be obtained as

$$\begin{aligned} TR &= P \cdot Q = (20 - 0.2Q)Q \\ &= 20Q - 0.2Q^2 \quad \dots(3.4) \end{aligned}$$

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

$$MR = \frac{\partial TR}{\partial Q} = 20 - 0.4Q \quad \dots(3.5)$$

NOTES

NOTES

The demand function and MR -function (3.5) are presented graphically in panel (a) and TR -function (3.4) in panel (b) of Fig. 3.2. As the figure shows, at point P on the demand curve, $e = 1$ where output, $Q = 50$. Below point P , $e < 1$ and above point P , $e > 1$. It can be seen in panel (a) of Fig. 3.2 that TR increases so long as $e > 1$; TR reaches its maximum level where $e = 1$; and it decreases when $e < 1$.

The relationship between price-elasticity and TR is summed up in Table 1.1. As the table shows, when demand is *perfectly inelastic* (i.e., $e_p = 0$ as is the case of a vertical demand line) there is no decrease in quantity demanded when price is raised and *vice versa*. Therefore, a rise in price increases the total revenue and *vice versa*.

As shown in panel (a), over the range of demand curve $e > 1$, quantity demanded increases by more than the proportionate decrease in price and hence the total revenue increase when price falls. The total revenue increases till price decreases till $e = 1$.

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

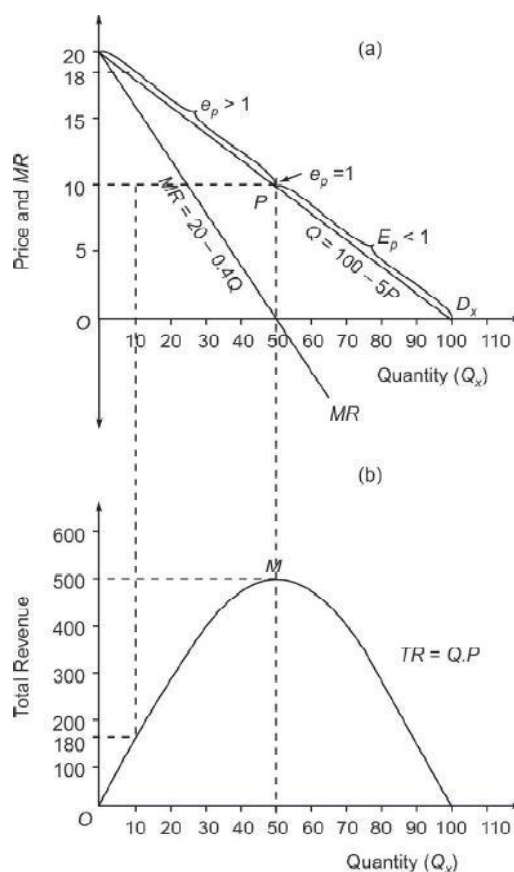


Fig. 3.2 Relationship between Demand Function, TR and MR

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

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

NOTES

Table 3.2 Elasticity, Price-change and Change in TR

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

*Subject to the size of the market.

Price Elasticity and Marginal Revenue

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

Since $TR = P \cdot Q$,

$$\begin{aligned}
 MR &= \frac{\partial(P \cdot Q)}{\partial Q} = P \left(\frac{\partial Q}{\partial Q} \right) + Q \left(\frac{\partial P}{\partial Q} \right) = P + Q \frac{\partial P}{\partial Q} \\
 &= P \left(1 + \frac{Q}{P} \cdot \frac{\partial P}{\partial Q} \right) \qquad \dots(3.6)
 \end{aligned}$$

Note that $\frac{Q}{P} \cdot \frac{\partial P}{\partial Q}$ in Eq. (3.6) gives the reciprocal of elasticity. That is,

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

Remember that e_p carries a ‘minus’ sign.

By substituting $\frac{1}{e}$ for $\frac{Q}{P} \cdot \frac{\partial P}{\partial Q}$ in Eq. (3.6), we get

NOTES

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

$$\text{or } MR = P [1 + (1/e_p)] \quad \dots(3.7)$$

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

Price Elasticity, AR and MR Given the Eq. (3.7), the formula for price elasticity (e_p) can be expressed in terms of AR and MR . We know that $P = AR$. So Eq. (3.7) can be written as

$$MR =$$

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

By rearranging the terms, we get

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

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

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

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

Application of Price Elasticity

Having explained the concept and measurement of elasticity, we discuss now the application of price elasticity of demand with respect to (a) manoeuvring price to maximize sales revenue and (b) determination of optimal price for profit maximization.

- 1. Manoeuvring of Price:** Price manoeuvring means changing price of the product to achieve business objective. The concept of elasticity of demand plays a crucial role in business-decisions regarding manoeuvring of prices for the benefit of the firm. For instance, when cost of production is increasing,

the firm would want to pass the rising cost on to the consumer by raising the price. Firms may decide to change the price even without any change in the cost of production. But, whether raising price following the rise in cost or otherwise proves beneficial or not depends on at least two factors:

- (a) The price-elasticity of demand for the product, i.e., how high or low is the proportionate change in its demand in response to a certain percentage change in its price; and
- (b) Price-elasticity of demand for its substitute, because when the price of a product increases, the demand for its substitutes increases automatically even if their prices remain unchanged.

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

2. Determination of Optimum Price: Another and a rather very important application of price elasticity is that it can be used, at least theoretically, to determine the optimum price with the objective of profit maximization for a firm facing downward sloping demand curve. How optimum price, i.e., profit maximizing price, can be determined by using the price elasticity is explained below. The necessary condition for profit maximization is given as $MR = MC$. That is, profit is maximized at the level of price and output at which

$$MR = MC \quad \dots(3.8)$$

As has been shown in Eq. (3.7),

$$MR = P [1 + (1/e_p)] \quad \dots(3.9)$$

By substituting Eq. (3.9) for MR in Eq. (3.8), the profit maximizing condition can be expressed as

$$P [1 + (1/e_p)] = MC \quad \dots(3.10)$$

The optimal price can be worked out from Eq. (3.10) as follows.

$$P = MC / [1 + (1/e_p)] \quad \dots(3.11)$$

Given the Eq. (3.11), if point elasticity of demand curve and marginal cost (MC) of a firm are known, the optimal price can be easily determined. For example, suppose point elasticity of demand at point on demand curve is estimated as $e_p = -2$ and firm's $MC = ₹50$. In that case, optimal price can

be worked out as follows.

$$\begin{aligned} P &= 50 / [1 + (1/-2)] \\ &= ₹100 \end{aligned}$$

NOTES

NOTES

Check Your Progress

1. What does the total revenue depend upon?
2. How can the average revenue be obtained?
3. Where can price elasticity be measured?
4. What happens to the total revenue in case the demand for a product is unit elastic?

3.4 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. The total revenue depends upon the total sale of the firm and the price of per unit output.
2. The average revenue can be obtained by dividing the total revenue by total units sold.
3. Price elasticity can be measured between any two points on a demand curve called (arc elasticity) or at a point (called point elasticity).
4. If demand for a product is unit elastic, quantity demanded increases (or decreases) in the proportion of decrease (or increase) in the price. It implies that a small change in price leaves the total revenue unchanged. Therefore, the total revenue remains unaffected.

3.5 SUMMARY

- The amount which a firm receives from the sale of the output is known as revenue. According to Prof. Dooley, –the revenue of a firm is its sales receipts or income.¶
- Revenue of a firm can be divided into three types:
 1. Total Revenue
 2. Average Revenue
 3. Marginal Revenue
- Total Revenue is the amount receipt from the sale of the output. We can obtain the average revenue by dividing the Total revenue by total unit sold. Marginal Revenue is the amount of money received from the sale of additional unit.
- The properties of the TR, AR and MR include: (i) When total revenue is at its maximum, the marginal revenue equals to 0, (ii) When total revenue is

decreasing that time the marginal revenue is negative, (iii) total revenue first increases and after a certain point it decreases.

Concept of Revenue

- When a firm can not change the price to affect the sale, in other words a firm under perfect competition can not decrease the price to increase its sale, in this case average revenue is always equals to marginal revenue.
- In imperfect competition firm is a price maker, not a price taker. A firm can decrease the price to increase the amount of sale or firm can increase the price to earn maximum profit. When there is a change in the price (in imperfect competition) the average revenue can not be equal to marginal revenue.
- Price elasticity of demand is generally defined as the responsiveness or sensitiveness of demand for a commodity to the changes in its price.
- The relationship between TR and MR and the price elasticity depends on the level of elasticity.
- Price elasticity is majorly used for two important reasons by the firms: (i) manoeuvring of price and (ii) determination of optimum price.

NOTES

3.6 KEY WORDS

- **Revenue:** It refers to the amount which a firm receives from the sale of output.
- **Marginal Revenue:** It is the amount of money received from the sale of additional unit.
- **Price Elasticity:** It is defined as the responsiveness or sensitiveness of demand for a commodity to the changes in its price.

3.7 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short Answer Questions

1. What are the three types of revenue?
2. State the properties of TR, MR and AR.
3. What are the applications of price elasticity?

Long Answer Questions

1. Explain the concept of TR, AR and MR with the help of diagrams.
2. What happens to average revenue under perfect and imperfect competition?
3. Examine the relationship between price elasticity, TR and MR.

NOTES

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

MARKET AND VARIOUS MARKET STRUCTURE

*Market Concept and
Classification*

NOTES

UNIT 4 MARKET CONCEPT AND CLASSIFICATION

Structure

- 4.0 Introduction
- 4.1 Objectives
- 4.2 Concept of Market and Market System
- 4.3 Different Market Structures
- 4.4 Answers to Check Your Progress Questions
- 4.5 Summary
- 4.6 Key Words
- 4.7 Self Assessment Questions and Exercises
- 4.8 Further Readings

4.0 INTRODUCTION

In layman terms, a market refers to a location where the buyers and sellers interact for the purchase and sale of goods and services. In economics, however, the term market is used in a different sense. Here the physical location for interaction is not important. Rather, this is assumed to determine the manner in which the transaction of goods and services is to take place. The market structure is affected and differentiated from each other through many characteristics including the number of buyers and sellers, the type of products transacted, barriers related to entry and exit, etc. In this unit, you will learn about the concept and classification of market and market structures.

4.1 OBJECTIVES

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

- Discuss the concept of market and market system
- Describe the market demand, supply and equilibrium
- Explain the different market structures

4.2 CONCEPT OF MARKET AND MARKET SYSTEM

NOTES

In general usage, the term ‘market’ means a place or locality where goods and services are bought and sold, e.g., Connaught Place market, Chandani Chowk market, Karol Bagh Market, etc. The term ‘market’ is also used with reference to a product, e.g., wheat market, vegetable market (Sabji mandi), gold market, stock market, etc.

In economics, the word ‘market’ is used in a rather abstract sense. The market means a system by which sellers and buyers of a commodity interact to settle its price and the quantity to be bought and sold. According to Samuelson and Nordhaus, —A market is a mechanism by which buyers and sellers interact to determine the price and quantity of a good or service. Market for a commodity consists of the buyers and sellers who interact to settle its price and quantity to be transacted. The sellers and buyers may be individuals, firms, factories, dealers and agents.

Some important aspects of the market concept are the following ones:

- (i) A market need not be situated in a particular place or locality. The geographical area of a market depends on how far and wide are the buyers and sellers scattered. It may be as small as a fish market in a corner of a city or as large as the entire world, e.g., the global markets for arms, cars, electronic goods, aeroplanes, computers, oil, medicines, etc.
- (ii) Buyers and sellers need not come into personal contact with each other. The transactions can be carried out through postal services or telecommunication system—telephone, fax, agents, or e-mail etc. People do buy many goods and services without ever meeting the supplier.
- (iii) The word ‘market’ may refer to a commodity or service (e.g., fruit market, car market, share market, money market, paper market, labour market, etc.) or to a geographical area, Bombay market, Indian market, Asian market, or the world market.
- (iv) The economists distinguish between markets also on the basis of (a) nature of goods and services, e.g., factor market, input market and output market; (b) number of firms and degree of competition, e.g., perfectly competitive market (very large number of firms), monopolistic market (many firms with differentiated products); oligopolistic market, and so on.

The Market System: An Overview

Market Concept and
Classification

Market system refers to the process by which buyers and sellers of a product interact to settle the price of the product and carry out the sale-purchase transactions. The market system works on the basis of a basic **market principle**. The market principle is based on the **fundamental laws of demand and supply**. Buyers create the demand and sellers create the supply of the product. Demand and supply work as two opposite **market forces**. The working of the market system is governed by the fundamental **laws of demand and supply**. The **laws of demand and supply** play a crucial role in determining the price of a product and the size of the market—total demand and total supply. A clear understanding of how markets work is essential for business decisions on production and future planning. As noted above, market system works on the basis of two **market forces—demand and supply**. The demand and supply forces represent two sides of the markets, viz., (i) **Demand side**, and (ii) **Supply side**. The demand and supply sides work on the laws of demand and supply, respectively.

In this section, you will study a brief review of market system, market mechanism and market structure, i.e., the different kinds of market settings.

Demand Side of the Market

The demand side of the market for a product refers to all its consumers and the price that they are willing to pay for buying a certain quantity of the product during a period of time. The quantity that consumers buy at a given price determines the market size. It is the size of the market that determines the business prospects of a firm and an industry. The demand side of the market is governed by the **Law of Demand**. The law of demand governs the market in the sense that when prices go up, demand goes down and size of the market is reduced, all other things remaining the same. Similarly, when prices decrease, demand increases causing a rise in sales and market size tends to increase. Let us now look at the law of demand in some detail and see how it governs the demand side of the market.

Law of Demand

The law of demand states the nature of relationship between the quantity demanded of a product and the price of the product. Although quantity demanded of a commodity depends also on many other factors, e.g., consumer's income, price of the substitutes and complementary goods, consumer's taste and preferences, advertisement, etc., the current price is the most important and the only determinant of demand in the short run. But the law of demand states the relationship between the demand for a product and its price only.

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, the originator of the law, has stated the law of

NOTES

NOTES

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*. The law of demand can be illustrated through a *demand schedule* and a *demand curve* as shown below.

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 annual market demand schedule for shirts is given in Table 4.1. This table presents price of shirts (P_c) and the corresponding number of shirts demanded (Q_c) in Delhi per week.

Table 4.1 illustrates the law of demand. As data given in the table shows, the demand for shirts (Q_c) increases as its price (P_c) decreases. For instance, at price ₹ 800 per shirt, only 8 thousand shirts are demanded in Delhi per week. When price decreases to ₹ 400, the demand for shirts increases to 30 thousand and when price falls further to ₹ 100, demand rises to 80 thousand. This relationship between price and quantity demanded gives the law of demand.

Table 4.1 Monthly Demand Schedule for Shirts

P_c (Price in ₹)	Q_c (Delhi in '000)
800	8
600	15
400	30
300	40
200	55
100	80

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 presented graphically as in Fig. 4.1, the resulting curve DD' is the demand curve. 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 shirts and their quantity demanded.

NOTES

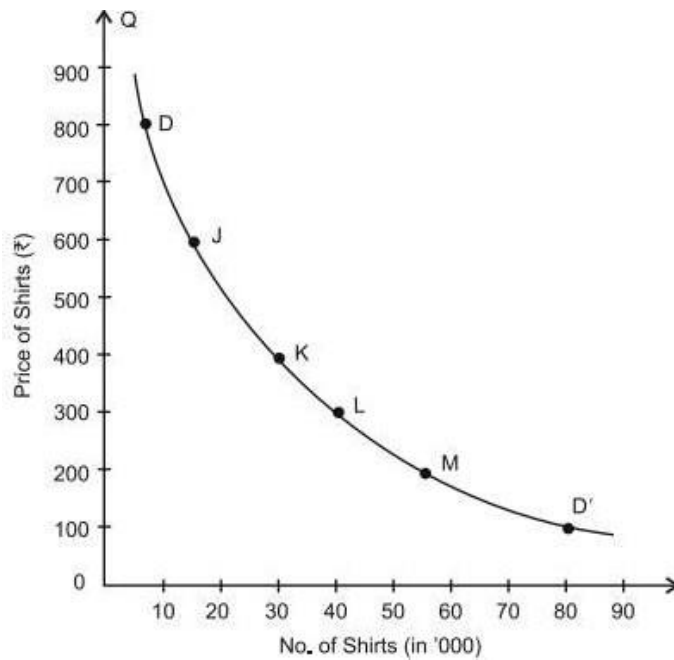


Fig. 4.1 The Demand Curve for Shirts

Supply Side of the Market

In a market economy, buyers' decision on 'what to buy' and 'how much to buy' constitute the demand side of the market, and sellers' decision to 'what to sell' and 'how much to sell' make the supply side of the market. In this section, we discuss the supply side of the market beginning with the law of supply.

Market Supply

Supply means the quantity of a commodity that its producers or sellers offer for sale at a given price, per unit of time. Market supply, like market demand, is the sum of supplies of a commodity made by all individual firms or their supply agencies. The market supply of a product is governed by the law of supply.

The 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

The Supply Schedule

A supply schedule is a tabular presentation of the law of supply. A supply schedule is a table showing different prices of a commodity and the corresponding quantity that suppliers are willing to offer for sale. Table 4.2 presents a hypothetical supply schedule of shirts, i.e., number of shirts supplied per week at different prices.

Table 4.2 Supply Schedule of Shirts

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

The Supply Curve

A supply curve is a graphical presentation of the supply schedule. The supply curve SS' given in Fig. 4.2 has been obtained by plotting the data in Table 4.2. The points S, P, Q, R, T and S' show the price-quantity combinations on the supply curve SS' . The supply curve, SS' , depicts *the law of supply*. The upward slope of the supply curve indicates the rise in the supply of shirts with the rise in its price and fall in the supply with fall in prices. For example, at price ₹ 200, only 35 thousand shirts are supplied per week. When price rises to ₹ 400, supply increases to 60 thousand shirts.

As shown in Fig. 4.2, a supply curve has a positive slope. The positive slope or upward movement of the supply curve is caused by the rise in cost of production and seller's effort to make a larger profit. The rise in cost of production results from the law of diminishing returns. In fact, supply curve is derived from the marginal cost curve.

NOTES

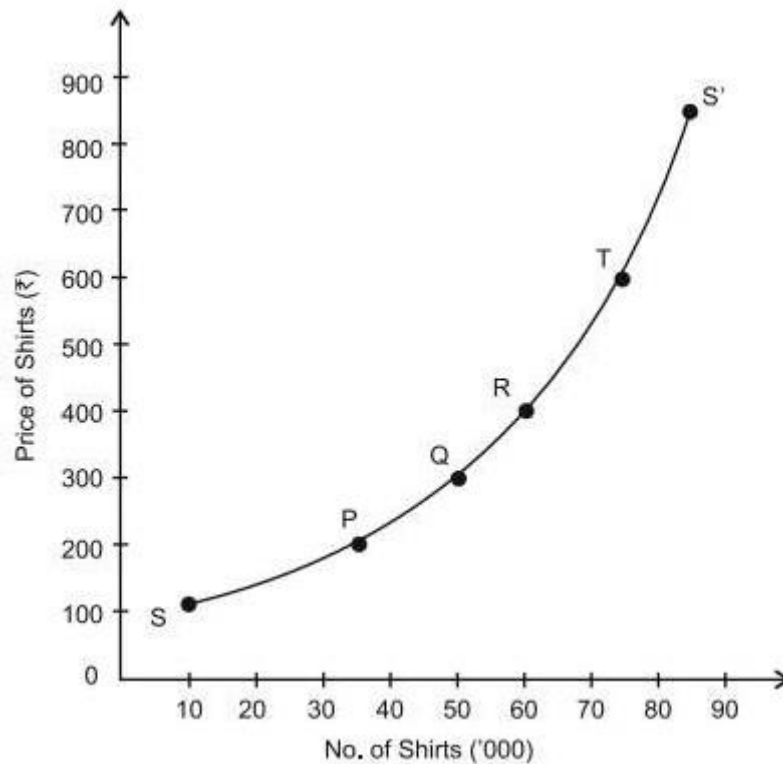


Fig. 4.2 Supply Curve of shirt

Shift in the Supply Curve

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

- (i) **Change in Input Prices.** When input prices decrease, the use of inputs increase. As a result, product supply increases and the supply curve SS shifts to the right to SS' , as shown in Fig. 4.3. Similarly, when input prices increase, product supply curve shifts leftward from SS to SS' .

NOTES

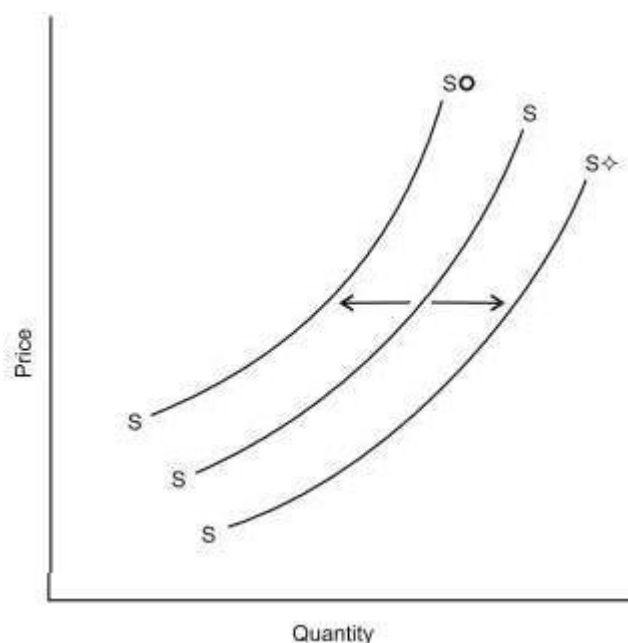


Fig. 4.3 Shift in the Supply Curve

- (ii) **Technological Progress.** Technological changes that reduce cost of production or increase efficiency in production cause increase in product supply. For instance, introduction of high yielding variety of paddy and new techniques of cultivation increased per acre yield of rice in India in the 1970s. Such changes make the supply curve shift to the right.
- (iii) **Price of Product Substitutes.** Given its technology and production capacity, a firm can produce more than one good which require a similar technology. For example, a refrigerator company can also produce ACs; Tatas, famous for truck production can also produce cars; Maruti Udyog can produce trucks, and so on. Fall in the price of one of the product substitutes may lead to the rise in the supply of other due to capacity utilization for profit maximization. This may cause shift in the supply curve.
- (iv) **Nature and Size of the Industry.** The supply of a commodity depends also on whether an industry is monopolized or competitive. Under monopoly, supply is fixed. When a monopolized industry is made competitive, the total supply increases. Besides, if size of an industry increases due to new firms joining the industry, the total supply increases and industry supply curve shifts rightward.
- (v) **Government Policy.** When government imposes restrictions on production, e.g., import quota on inputs, rationing of or quota imposed on input supply, etc., production tends to fall. Such restrictions make supply curve shift leftward.

(vi) **Non-economic Factors.** Factors like labour strikes and lock-outs, war, drought, flood, communal riots, epidemics, etc. also adversely affect the supply of commodities and make the supply curve shift leftward.

Market Concept and
Classification

Market Equilibrium: Equilibrium of Demand and Supply

NOTES

Determination of Price in a Free Market. In this section, we explain how demand and supply interact to strike a balance, how market attains *equilibrium*, and how *equilibrium price* is determined in a *free market*. A free market is one in which market forces of demand and supply are free to take their own course and there is no outside control on price, demand and supply.

The Concept of Market Equilibrium

In physical sense, the term equilibrium means the ‘state of rest’. In general sense, it means that forces working in opposite directions are in balance. In the context of market analysis, *equilibrium refers to a state of market in which quantity demanded of a commodity equals the quantity supplied of the commodity*. The equality of demand and supply produces an *equilibrium price*. The equilibrium price is the price at which quantity demanded of a commodity equals its quantity supplied. That is, at equilibrium price, demand and supply are in equilibrium. Equilibrium price is also called *market-clearing price*. Market is cleared in the sense that there is no unsold stock and no unsupplied demand.

Determination of Market Price

Equilibrium price of a commodity in a free market is determined by the market forces of demand and supply. In order to analyze how equilibrium price is determined, we need to integrate the demand and supply curves. For this purpose, let us use our earlier example of shirts. Let us suppose that the weekly market demand and supply schedules for shirts in Delhi are given as shown in Table 4.3.

Table 4.3 Weekly Demand and Supply Schedules for Shirts

Price per Shirt (₹)	Demand (`000 shirts)	Supply (`000 shirts)	Market Position	Effect on Price
100	80	10	Shortage	Rise
200	55	28	Shortage	Rise
300	40	40	Equilibrium	Stable
400	28	50	Surplus	Fall
500	20	55	Surplus	Fall
600	15	60	Surplus	Fall

As the table shows, there is only one price of shirts (₹ 300) at which quantity demanded per week equals the quantity supplied at 40 thousand shirts. It means that the shirt market in Delhi is in equilibrium at price ₹ 300. At all other prices, the shirt market is in *disequilibrium*—the state of imbalance between supply and demand. When market is in the state of disequilibrium, either demand exceeds

NOTES

supply or supply exceeds demand. As the table shows, at all prices below ₹ 300, demand exceeds supply showing *shortage* of shirts in the market. Likewise, at all prices above ₹ 300, supply exceeds demand showing *excess supply*.

Under the conditions of a free market, disequilibrium itself creates the condition for equilibrium. When there is excess supply, it forces downward adjustments in the price and quantity supplied. When there is excess demand, it forces upward adjustments in the price and quantity demanded. The process of downward and upward adjustments in price and quantity continues till the price reaches ₹ 300 and quantities supplied and demanded are in balance at 40 thousand shirts. This process is automatic.

Check Your Progress

1. What is the market principle based on?
2. In what manner are the supply of a commodity and its price related?
3. State another name for equilibrium price.

4.3 DIFFERENT MARKET STRUCTURES

In the previous section we have discussed and illustrated determination of price assuming simple market conditions and simple demand and supply functions. In reality, however, market structure is extremely complex and so is the system of price determination. This has led to formation of different theories of price determination under different kinds of market conditions. In this section, we discuss briefly the *market structure* – the market morphology.

Market structure refers to the number of firms in an industry and the degree of competition among the firms. The categorization of market structure is based on the following factors.

- Number of firms—the sellers
- Degree and nature of competition
- Level of product differentiation
- Possibility of entry and exit of the firms

The number of sellers of a product in a market determines the *nature and degree of competition* in the market. The nature and degree of competition make the *structure of the market*. Depending on the number of sellers and the degree of competition, the market structure is broadly classified as given in Table 4.4.

Table 4.4 Types of Market Structures

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

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

Market Structure and Pricing Decisions

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

Under **perfect competition**, a large number of firms compete against each other for selling their product. Therefore, the degree of competition under perfect competition is close to one, i.e., the market is highly competitive. Consequently, firm's discretion in determining the price of its product is close to none. In fact, in perfectly competitive market, price is determined by the market forces of demand and supply and a firm has to accept the price determined by the market forces. If a firm uses its discretion to fix the price of its product above or below its market level, it loses its revenue and profit in either case. For, if it fixes the price of its product above the ruling price, it will not be able to sell its product, and if it cuts the price down below its market level, it will not be able to cover its average cost.

NOTES

NOTES

In a perfectly competitive market, therefore, firms have little or no choice in respect to price determination.

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

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

Check Your Progress

4. What is the degree of freedom?
5. What is the degree of competition in oligopoly?

4.4 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. The market principle is based on the fundamental laws of demand and supply.
2. The supply of a commodity and its price are positively related.
3. Equilibrium price is also called market clearing price. Market is cleared in the sense that there is no unsold stock and no unsupplied demand.
4. With reference to market structures, the degree of freedom implies the extent to which a firm is free or independent of the rival firms in taking its own pricing decisions.
5. Under oligopoly, the degree of competition is quite low, lower than that under monopolistic competition. The firms, therefore, have a good deal of control over the price of their products and can exercise their discretion in pricing decisions, especially where product differentiation is prominent.

4.5 SUMMARY

*Market Concept and
Classification*

- In general usage, the term ‘market’ means a place or locality where goods and services are bought and sold.
- In economics, the word ‘market’ is used in a rather abstract sense. The market means a system by which sellers and buyers of a commodity interact to settle its price and the quantity to be bought and sold.
- Market system refers to the process by which buyers and sellers of a product interact to settle the price of the product and carry out the sale-purchase transactions. The market system works on the basis of a basic market principle. The market principle is based on the fundamental laws of demand and supply.
- The demand and supply forces represent two sides of the markets, viz., (i) Demand side, and (ii) Supply side. The demand and supply sides work on the laws of demand and supply, respectively.
- The demand side of the market for a product refers to all its consumers and the price that they are willing to pay for buying a certain quantity of the product during a period of time. The quantity that consumers buy at a given price determines the market size. It is the size of the market that determines the business prospects of a firm and an industry.
- In a market economy, buyers’ decision on ‘what to buy’ and ‘how much to buy’ constitute the demand side of the market, and sellers’ decision to ‘what to sell’ and ‘how much to sell’ make the supply side of the market.
- In physical sense, the term equilibrium means the ‘state of rest’. In general sense, it means that forces working in opposite directions are in balance. In the context of market analysis, equilibrium refers to a state of market in which quantity demanded of a commodity equals the quantity supplied of the commodity. The equality of demand and supply produces an equilibrium price.
- Equilibrium price of a commodity in a free market is determined by the market forces of demand and supply. In order to analyze how equilibrium price is determined, we need to integrate the demand and supply curves.
- Market structure refers to the number of firms in an industry and the degree of competition among the firms. The categorization of market structure is based on the following factors:
 - o Number of firms—the sellers
 - o Degree and nature of competition
 - o Level of product differentiation
 - o Possibility of entry and exit of the firms

NOTES

NOTES

4.6 KEY WORDS

- **Market:** It means a system by which sellers and buyers of a commodity interact to settle its price and the quantity to be bought and sold.
- **Market System:** It refers to the process by which buyers and sellers of a product interact to settle the price of the product and carry out the sale-purchase transactions.
- **Equilibrium:** It refers to the state of market in which quantity demanded of a commodity equals the quantity supplied of the commodity.
- **Market Structure:** It refers to the number of firms in an industry and the degree of competition among firms.

4.7 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short Answer Questions

1. Briefly mention some of the important concepts of market in economics.
2. Write a short note on market equilibrium and determination of market price.
3. What is the categorization of market structure based on?

Long Answer Questions

1. Examine, in detail, the demand and supply side of the market system.
2. Discuss how the market structure affects pricing decisions.

4.8 FURTHER READINGS

- Dwivedi, D. N. 2008. *Principles of Economics*, Seventh Edition. New Delhi: Vikas Publishing House.
- Weil, David N. 2004. *Economic Growth*. London: Addison Wesley.
- Thomas, Christopher R. and Maurice S. Charles. 2005. *Managerial Economics: Concepts and Applications*, Eighth Edition. New Delhi: Tata McGraw-Hill Publishing Company Limited.
- Mankiw, Gregory N. 2002. *Principles of Economics*, Second Edition. India: Thomson Press.

UNIT 5 PERFECT COMPETITION AND MARKET EQUILIBRIUM

*Perfect Competition and
Market Equilibrium*

NOTES

Structure

- 5.0 Introduction
- 5.1 Objectives
- 5.2 Meaning and Features of Perfect Competition
- 5.3 Price Determination Under Perfect Competition
- 5.4 Equilibrium of the Firm in Short-Run
- 5.5 Derivation of Supply Curve
 - 5.5.1 Derivation of Firm's Supply Curve
 - 5.5.2 Derivation of Industry's Supply Curve
- 5.6 Equilibrium of the Industry and Firm in Short-Run
 - 5.6.1 Link between Short-run Equilibrium of the Industry and of the Firm
- 5.7 Equilibrium of the Industry and Firm in Long-Run
 - 5.7.1 Equilibrium of the Firm
 - 5.7.2 Equilibrium of the Industry
- 5.8 Long-Run Supply Curve of the Industry
 - 5.8.1 Constant Cost Industry
 - 5.8.2 Increasing Cost Industry
 - 5.8.3 Decreasing Cost Industry
- 5.9 Answers to Check Your Progress Questions
- 5.10 Summary
- 5.11 Key Words
- 5.12 Self Assessment Questions and Exercises
- 5.13 Further Readings

5.0 INTRODUCTION

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

5.1 OBJECTIVES

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

- Discuss the meaning and characteristics of perfect competition

*Self-Instructional
Material*

- Explain cost and revenue conditions faced by the firms under perfect competition
- Describe how a firm attains its equilibrium in a perfectly competitive market

NOTES

5.2 MEANING AND FEATURES OF PERFECT COMPETITION

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

Features of Perfect Competition

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

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

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

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

industrial input. This assumption guarantees that factors of production—land, labour, capital, and entrepreneurship—can enter or quit a firm or the industry at will.

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

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

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

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

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

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

NOTES

5.3 PRICE DETERMINATION UNDER PERFECT COMPETITION

NOTES

Under perfect competition, an individual firm *does not* determine the price of its product. Price for its product is determined by the market demand and market supply. In Figure 5.1 (a) the demand curve, DD' , represents the market demand for the commodity of an industry as a whole. Likewise, the supply curve, SS' , represents the total supply created by all the firms of the industry (derivation of industry's supply curve has been shown in a following section). As Figure 5.1 (a) shows, market price for the industry as a whole is determined at OP . This price is given for all the firms of the industry. No firm has power to change this price. At this price, a firm can sell any quantity. It implies that the demand curve for an individual firm is a straight horizontal line, as shown by the line dd' in Figure 5.1 (b), with infinite elasticity.

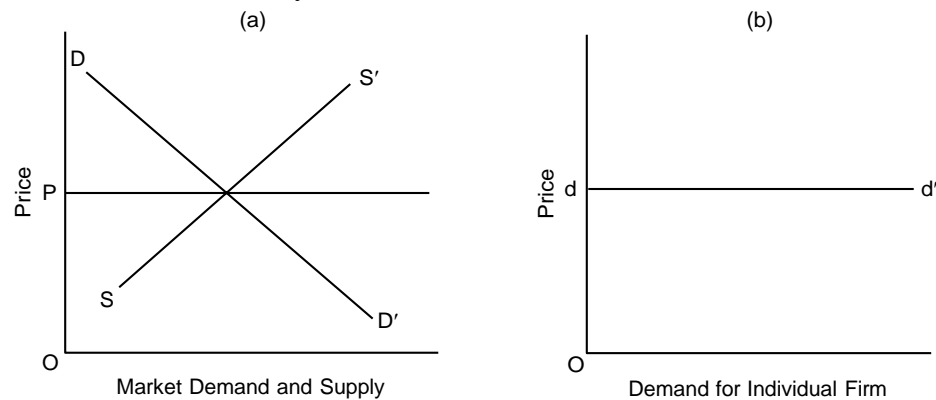


Fig. 5.1 Determination of Market Price and Demand for Individual Firms

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

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

5.4 EQUILIBRIUM OF THE FIRM IN SHORT-RUN

A profit maximising firm is in equilibrium at the level of output which equates its $MC = MR$. However, the level of output which meets the equilibrium condition

NOTES

for a firm varies depending on cost and revenue functions. The nature of cost and revenue functions depends on whether one is considering *short-run* or *long-run*. While the revenue function is generally assumed to be given in both short and long runs, the short-run cost function is not the same in the short and long-runs. The short-run cost function is different from the long-run cost function because in the short run, some inputs (e.g., capital) are held constant while all factors are variable in the long-run. In this section, we will discuss *firm's short-run equilibrium*. Long-run equilibrium of the firm will be discussed in the forthcoming section.

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

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

The firm's equilibrium in the short-run is illustrated in Figure 5.2. Price of a commodity is determined by the market forces—demand and supply—in a perfectly competitive market at OP . The firms, therefore, face a straight-line, horizontal demand curve, as shown by the line $P = MR$. The straight horizontal demand line implies that price equals marginal revenue, i.e., $AR = MR$. The short-run average and marginal cost curves are shown by SAC and SMC , respectively.

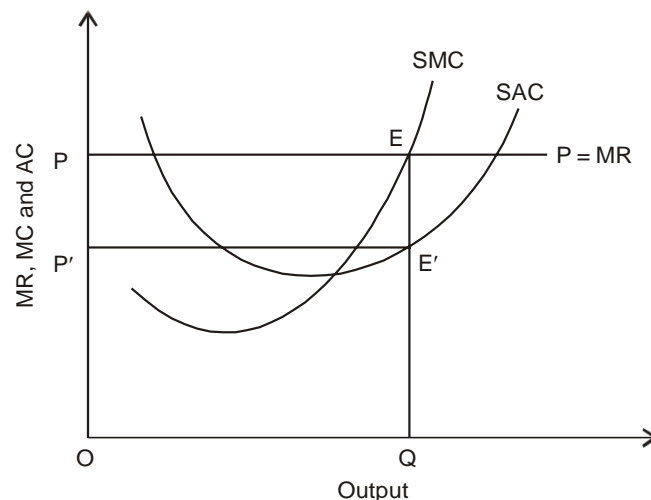


Fig. 5.2 Short-run Equilibrium of the Firm

It can be seen in Figure 5.2 that SMC curve intersects the $P = MR$ line at point E , from below. At point E , therefore, $SMC = MR$. A perpendicular drawn from point E to the output axis determines the equilibrium output at OQ . It can be seen in the figure that output OQ meets both the *first* and the *second order conditions* of profit maximisation. At output OQ , therefore, profit is maximum.

NOTES

The output OQ is, thus, the equilibrium output. At this output, the firm is in equilibrium and is making maximum profit. Firm's total pure profit is shown by the area $PEE'P'$ which equals $PP' \times OQ$ where PP' is the per unit super normal profit at output OQ .

Does a firm always make profit in the short run?: Figure 5.2 shows that a firm makes supernormal profit in the short run. A question arises here: Does a firm make always a supernormal profit in the short run? The answer is 'not necessarily'. As a matter of fact, in the short run, a firm may make a supernormal profit or a normal profit or even make losses. Whether a firm makes abnormal profits, normal profits or makes losses depends on its cost and revenue conditions. If its short-run average cost (SAC) is below the price ($P = MR$) at equilibrium (Figure 5.2), the firm makes abnormal or pure profits. If its SAC is tangent to $P = MR$, as shown in Figure 5.3 (a), the firm makes only normal profit as it covers only its SAC which includes *normal profit*. But, if its SAC falls above the price ($P = MR$), the firm makes losses. As shown in Figure 5.3 (b), the total loss equals the area $PP'E'E (= PP' \times OQ)$, while per unit loss is $PP' = EE'$.

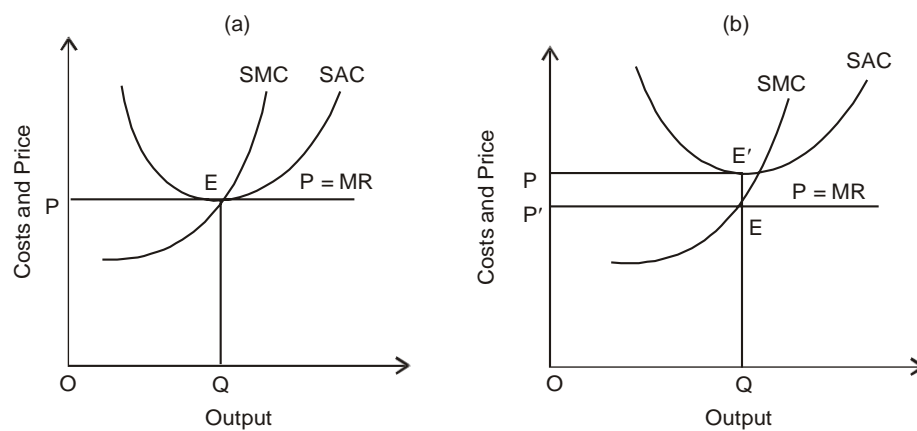


Fig. 5.3 Short-run Equilibrium of Firm with Normal and Losses

Shut-down or close-down point: In case a firm is making loss in the short run, it must minimise its losses. In order to minimise its losses, it must cover its *short-run average variable cost (SAVC)*. The behaviour of *short-run average variable cost* is shown by the curve *SAVC* in Figure 5.4. A firm unable to recover its minimum *SAVC* will have to close down. The firm's *SAVC* is minimum at point E where it equals the *MC*. Note that *SMC* intersects *SAVC* at its minimum level as shown in Figure 5.4.

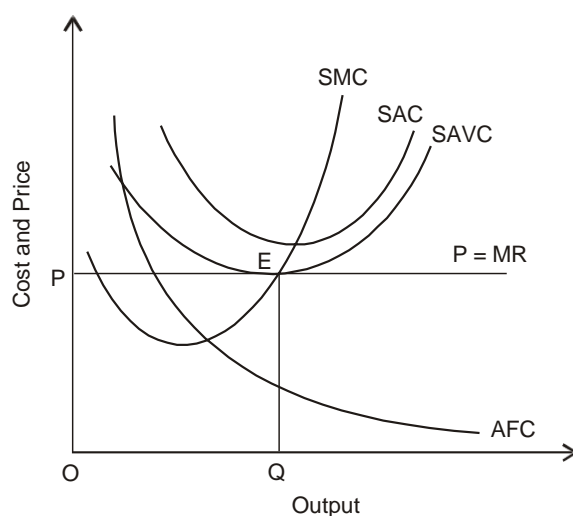


Fig. 5.4 Shut-down Point

Another condition which must be fulfilled is $P = MR = SMC$. That is, for loss to be minimum, $P = MR = SMC = SAVC$. This condition is fulfilled at point E in Figure 5.4. At point E , the firm covers only its fixed cost and variable cost. It does not make any profit—rather it makes losses. The firm may survive for a short period but not for long. Therefore, point E denotes the ‘shut-down point’ or ‘break-down point’, because at any price below OP , it pays the firm to close down as it minimises its loss.

Check Your Progress

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

5.5 DERIVATION OF SUPPLY CURVE

In this section, you will study the derivation of supply curve of both industry and firm.

5.5.1 Derivation of Firm's Supply Curve

The supply curve of an individual firm is derived on the basis of its equilibrium output. The equilibrium output, determined by the intersection of MR and MC curves, is the optimum supply by a profit maximising (or cost minimising) firm. Under the condition of increasing MC , a firm will increase supply only when price increases. This forms the basis of a firm's supply curve. The derivation of supply

NOTES

NOTES

curve of a firm is illustrated in Figure 5.5 (a) and (b). As the figure shows, the firm's SMC passes through point M on its $SAVC$. The point M marks the minimum of firm's $SAVC$ which equals MQ_1 . The firm must recover its $SAVC = MQ_1$ to remain in business in the short run. Point M is the shut-down point in the sense that if price falls below OP_1 , it is advisable for the firm to close down. However, if price increases to OP_2 , the equilibrium point shifts to R and output increases to OQ_2 . Note that at output OQ_2 , the firm covers its SAC and makes normal profit. Let the price increase further to OP_3 so that equilibrium output rises to OQ_3 . When price rises to OP_4 , the equilibrium output rises to OQ_4 and the firm makes abnormal profit. By plotting this information, we get a *supply curve* (SS') as shown in Figure 5.5 (b).

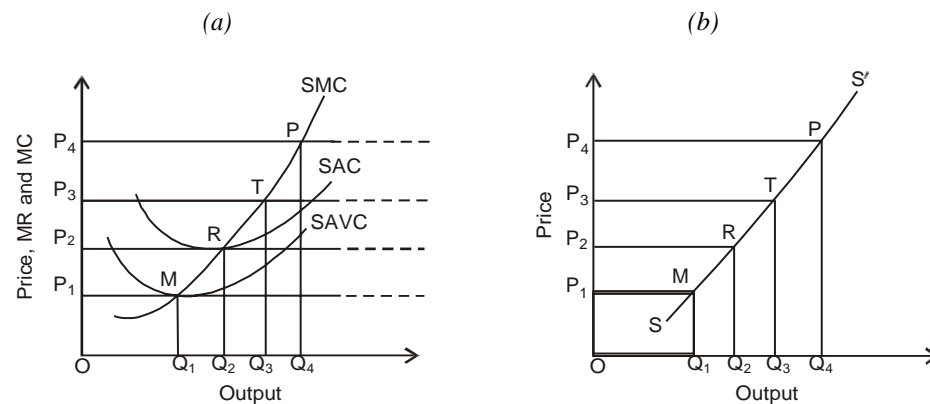


Fig. 5.5 Derivation of a Firm's Supply Curve

5.5.2 Derivation of Industry's Supply Curve

The industry supply curve, or what is also called market supply curve, is the horizontal summation of the supply curve of the individual firms. If cost curves of the individual firms of an industry are identical, their individual supply curves are also identical. In that case, industry supply curve can be obtained by multiplying the individual supply at various prices by the number of firms. In the short run, however, the individual supply curves may not be identical. If so, the market supply curve can be obtained by summing horizontally the individual supply curves. Let us consider only two firms having their individual supply curves as S_1 and S_2 as shown in Figure 5.6 (a). At price OP_1 , the market supply equals $P_1A + P_1B$. Suppose $P_1A + P_1B$ equals P_1M as shown in Figure 5.6 (b). [Note that output scale in part (b) is different from that in part (a).] Similarly, at price OP_2 , the industry supply equals $P_2C + P_2C$ or $2(P_2C) = P_2N$ as shown in Figure 5.6 (b). In the same way, point T is located. By joining the points M , N and T , we get the market or industry supply curve, SS' .

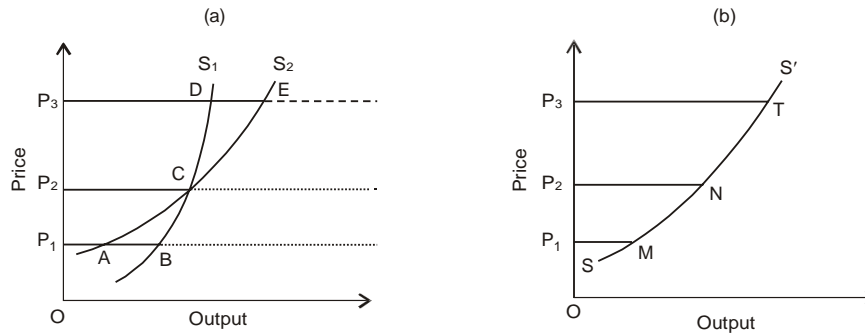


Fig. 5.6 Derivation of the Industry Supply Curve

NOTES

5.6 EQUILIBRIUM OF THE INDUSTRY AND FIRM IN SHORT-RUN

In the previous section you learned about the equilibrium of the firm in the short-run in a perfectly competitive market. To complete the discussion on short-run price and output determination, now we discuss the short-run equilibrium of the industry.

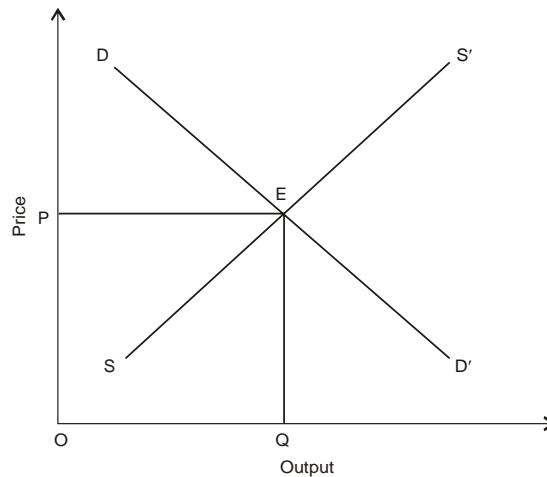


Fig. 5.7 Equilibrium of the Industry

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 called equilibrium price. When an industry reaches its equilibrium, there is no tendency to expand or to contract the output. The equilibrium of industry is shown at point *E* in Figure 5.7. The industry demand curve *DD'* and supply curve *SS'* intersect at point *E*, determining equilibrium price *OP*. At this price, $D = S$. The industry is supplying as much as consumers demand. In the short-run equilibrium of the industry, some individual firms may make pure profits, some normal profits and some may make even losses, depending on their cost and revenue conditions. As we have explained below, this situation will, however, not continue in the long-run.

NOTES

5.6.1 Link between Short-run Equilibrium of the Industry and of the Firm

The short-run equilibrium of the firm and industry have been analysed separately in the previous sections. There exists, however, a link between the equilibrium of a firm and that of the firm industry. In a perfectly competitive market, change in the equilibrium of an individual firm does not affect the industry's equilibrium, simply because the total output of a single firm constitutes a small fraction of the industry's output. But, a change in the industry's equilibrium does alter the equilibrium of an individual firm. In this section, we show how individual firms move from one equilibrium position to another, when industry's equilibrium changes. For the sake of simplicity, we assume that all the firms of an industry have *identical cost conditions*.

The link between industry's and firm's equilibrium is illustrated in Figure 5.8. Suppose industry's initial demand and supply curves are given as DD and SS , respectively, in Figure 5.8 (a). As shown in panel (a) of the figure, industry's demand and supply curves intersect at point P , determining the market price at $PQ = OP_1$ and industry's equilibrium output is OQ . Thus, the price PQ is given to all the firms of the industry. Given the price PQ and firm's cost curves, an individual firm finds its equilibrium at point E in panel (b), where its $MC = MR$. Firm's equilibrium output is OM [see Figure 5.8 (b)]. At price $EM = PQ$, the firm is making an abnormal profit in the short run to the extent of EN per unit of output. The firm's total pure profit is shown by the shaded rectangle.

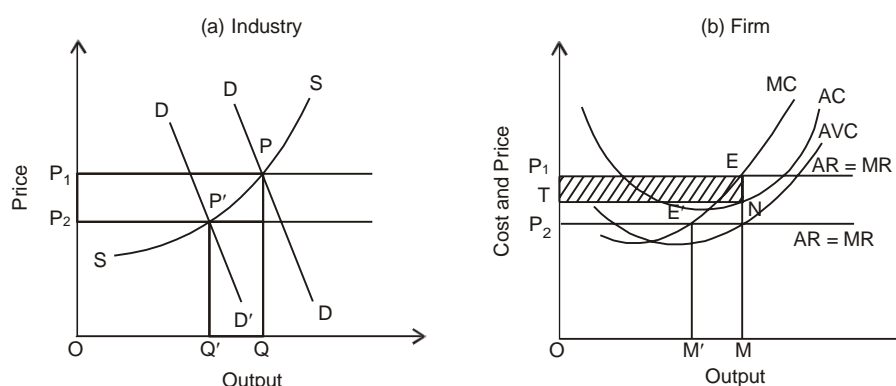


Fig. 5.8 Industry's Vs Firm's Equilibrium in the Short Run

Now let the industry demand curve DD shift downward for some reason to DD' , supply curve remaining unchanged. As a result, market price falls to $OP_2 = P'Q'$ and industry's equilibrium output falls to OQ' . With the fall in price, firm's equilibrium shifts from point E to E' where its $MC = MR$ at $F'M'$. Thus, change in industry's equilibrium changes firm's equilibrium. Although firms are making loss because their AR is less than their AC , this is only a short-run situation. In the long run, loss-making firms will quit the industry. As a result, supply curve SS will shift leftward causing price to go up. Losses will therefore disappear in the long run.

5.7 EQUILIBRIUM OF THE INDUSTRY AND FIRM IN LONG-RUN

The long-run refers to a time period in which short-run conditions do not exist. Recall that short-run is, by definition, a period in which (i) firm's cost and revenue curves are given, (ii) firms cannot change their size—their capital is fixed, (iii) existing firms do not have the opportunity to leave the industry, and (iv) new firms do not have the opportunity to enter the industry. In contrast, long-run is a period in which these constraints disappear. Long-run permits change in technology and employment of both, labour and capital, i.e., firms can change their size. Some of the existing firms may leave and new firms may enter the industry. In the long-run, supply curve not only shifts rightward but also becomes more elastic.

In this section, we will analyse the equilibrium of the firm and industry in the long-run. It should be noted that the process by which firms and industry reach their respective long-run equilibrium, is a continuous process of adjustment and readjustment of price and output with the changing conditions in the long-run. The process of equilibrium adjustment by the firm and industry is discussed below.

5.7.1 Equilibrium of the Firm

The long-run equilibrium of the firm is illustrated in Figure 5.9. To explain it further, let us begin with a short-run situation. Suppose (i) that short-run price is given at OP_1 , in panel (a) of Figure 5.9, and (ii) that firms' short-run cost curves are identical and are given by SAC_1 and SMC_1 , as shown in panel (b). Given the price OP_1 , firms are in equilibrium at point E_1 . It can be seen in part (b) of Figure 5.9 that the firms are making an abnormal profit to the extent of $AR - SAC$, i.e., $E_1M = E_1Q_1 - MQ_1$ per unit of output. The abnormal profit brings two major changes in the industry.

One, existing firms get incentive to increase the scale of their production. Their average and marginal costs go down caused by the economies of scale. This phenomenon is shown by SAC_2 and SMC_2 . When we draw the LAC and LMC curves, these curves show decreasing costs in the long-run.

Two, attracted by the abnormal profit, new firms enter the industry increasing the total supply causing a rightward shift in the industry supply curve.

NOTES

NOTES

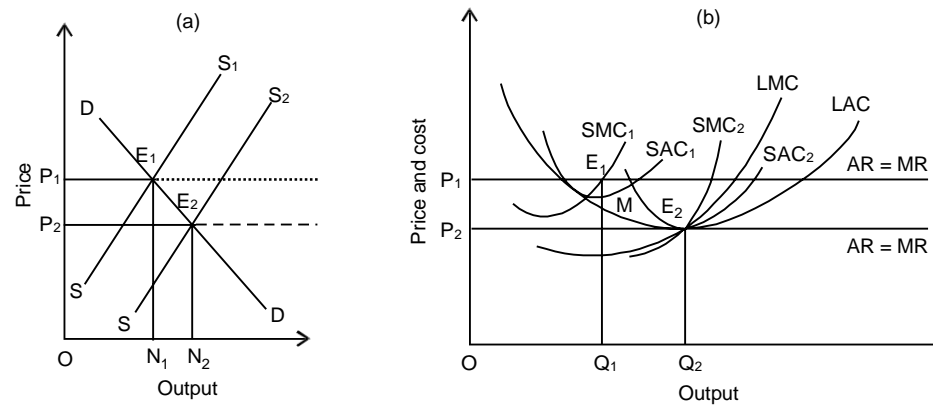


Fig. 5.9 Long-run Equilibrium of the Firm and Industry

For these reasons, the industry supply curve, SS_1 , shifts rightward to SS_2 [Figure 5.9(a)]. The shift in supply curve (demand curve, DD , remaining the same) brings down the market price to OP_2 which is the long-run equilibrium price. Thus, equilibrium price is once again determined for the industry OP_2 .

Given the new market price, OP_2 , firms attain their equilibrium in the long-run where $AR = MR = LMC = LAC = SMC = SAC$. This point is shown at point E_2 in Figure 5.9 (b). As the figure shows, the firms of an industry reach their equilibrium in the long-run where both short-run and long-run equilibrium conditions coincide. In a perfectly competitive market, the cost and revenue conditions are given for the firms. Therefore, when price goes down to OP_2 , what firms are required to do is to adjust their output to the given revenue and cost conditions in order to maximise their profit. Let us now illustrate the process of adjustment of output so as to reach the equilibrium in the long-run.

So long as economies of scale are available to the firms, the LAC tends to decrease and it pays firms to expand their plant-size. This process of output adjustment continues until industry reaches its equilibrium. The industry reaches its equilibrium where LAC is tangent to $P = AR = MR$ for each firm in the industry. This position is shown at point E_2 , in Figure 5.9 (b). At point E_2 , the point of equilibrium, $P = MR = LMC = LAC = SMC = SAC$. Since $P = LAC$, the firms make only normal profits in the long-run. If firms deviate from point E_2 , due to some short-run disturbances, the market forces of demand and supply come into operation and restore the equilibrium.

5.7.2 Equilibrium of the Industry

An industry is in equilibrium at a price and output at which market is cleared, i.e., where market demand equals market supply. The equilibrium of the industry is illustrated in Figure 5.9 (a). When an industry is in equilibrium, all its firms are supposed to be in equilibrium [as shown in Figure 5.9 (b)], and earn only normal profits. This is so because under the conditions of perfect competition, all the firms are assumed to achieve the same level of efficiency in the long-run. Since industry

yields only normal profits, there is no incentive for new firms to enter the industry. These conditions are fulfilled at price OP_2 in Figure 5.9 (a) and (b). At price OP_2 , all the firms are in equilibrium, where each firm has its $LMC = LMR = SMC = SAC = P = LAC$.

Since $P = LAC$, all the firms are earning only normal profit. At industry's equilibrium output ON_2 , market demand equals market supply [Figure 5.9 (a)]. At price OP_2 , therefore, market is cleared. The output ON_2 may remain stable in the long-run. For, there is no incentive for new firms to enter the industry and no reason for the existing ones to leave the industry. The industry is, therefore, in equilibrium.

5.8 LONG-RUN SUPPLY CURVE OF THE INDUSTRY

We have earlier derived the short-run supply curve of the industry by summing up horizontally the supply curves of the individual firms (see Figure 5.6). The long-run supply curve of a competitive industry, however, has nothing to do with the LMC curves. The shape of the long-run supply curve of an industry, under perfect competition, depends on whether factor prices remain constant or increase, when demand for inputs increases in the long-run as a result of expansion of the output of the industry. Depending on whether industry's cost is constant, increasing or decreasing, industries are classified as *constant cost*, *increasing cost* or *decreasing cost industry*. Let us, now, derive the long-run supply curve of industries conforming to their cost behaviour.

5.8.1 Constant Cost Industry

An industry for which factor prices remain constant is referred to as constant cost industry. In other words, when the expansion of output in an industry does not entail a change in factor prices, the industry is said to be a constant cost industry. The shape of the supply curve of such an industry is illustrated by the line LRS in Figure 5.10 (a) and (b). To explain the horizontal shape of the supply curve, let us suppose that the industry is in equilibrium at point P where demand curve DD_1 and supply curve SS_1 intersect. The industry is in equilibrium at price OP_1 and output OQ_1 . At price OP_1 , all firms are in equilibrium as their $LMC = P = MR = SMC = SAC$.

NOTES

NOTES

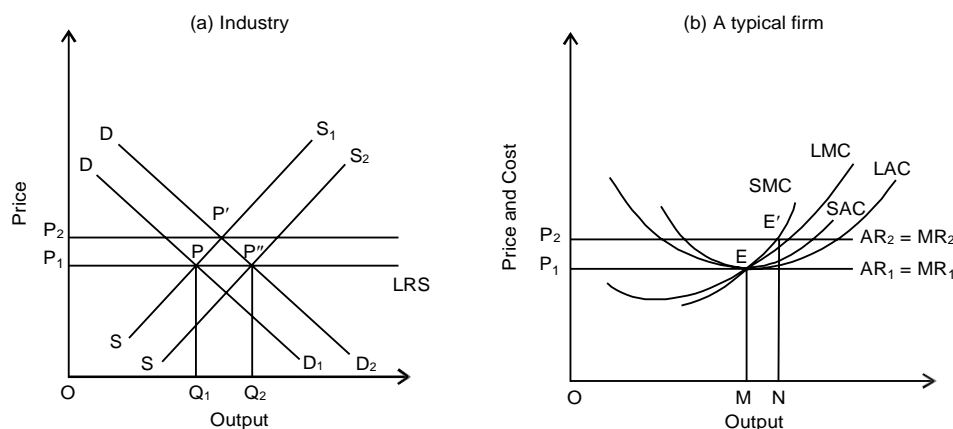


Fig. 5.10 Long-run Supply Curve of the Constant Cost Industry

Now, let the demand curve DD_1 shift to DD_2 due to, say, increase in consumers' income or increase in population or due to both, supply curve remaining the same. As a result, market price increases to OP_2 . In the short-run, this increase in price can induce an increase in supply by the firms only by MN as shown in Figure 5.10 (b), as determined by the point of intersection between firms' SMC and new price line through P_2 . The firms enjoy abnormal profit to the extent of $AR_2 - LAC$.

The abnormal profits attract new firms into the industry. The entry of new firms leads to increase in demand for inputs. However, the industry being a constant cost industry, factor prices do not increase—maybe due to simultaneous increase in input supply. Cost of production for all the firms remains constant at the previous level. But, due to the entry of new firms, market supply increases and market supply curve shifts from SS_1 to SS_2 [Figure 5.10 (a)]. Consequently, in the long-run, market price falls to its previous level, OP_1 , and firms return to their previous equilibrium point E . But the industry output increases from OQ_1 to OQ_2 and industry moves from equilibrium point P to P'' . By joining the two points of industry equilibrium P and P'' , and extending it further, we get long-run supply curve (LRS) of the constant cost industry. Obviously, the long-run supply curve (LRS) of a constant cost industry is a horizontal straight line, as given by the line LRS .

5.8.2 Increasing Cost Industry

An increasing cost industry is one which faces increasing input prices. The increase in input prices may be caused by increase in demand for inputs but their supply does not keep pace with rising demand. The long-run supply curve of an **increasing cost industry** has a positive slope as illustrated in Figure 5.11 (a).

The derivation of long-run market supply curve under increasing cost condition is demonstrated in Figure 5.11. Let the original demand and supply curves of the industry be represented, respectively, by DD_1 and SS_1 , and industry be in equilibrium of point A . Let us also suppose that for some reason demand curve DD_1 shifts to DD_2 , supply curve remaining the same. As a result, short-run

NOTES

market price increases from OP_1 to OP_3 given the supply curve S_1 . With this increase in price, the demand curve for the firms shifts upward to $AR_3 = MR_3$ [Figure 5.11 (b)]. The firms, therefore, enjoy a super normal or economic profit to the extent of P_1P_3 . This supernormal profit attracts new firms to the industry and demand for inputs increases. Since by assumption, the supply of inputs is less than infinitely elastic, the entry of new firms causes an increase in demand for inputs and, therefore, an increase in the input prices. Consequently, cost curves, both short-run and long-run, shift upward from LAC_1 to LAC_2 . In this process of adjustments, however, industry supply increases and market supply curve SS_1 shifts rightwards to SS_2 . With this shift in the supply curve, the industry reaches another equilibrium position at point C where new demand and supply curves intersect. A new market price is determined at OP_2 . At price OP_2 [Figure 5.11 (b)], the long-run and short-run cost curves are tangent to the price line $OP_2 = AR_2 = MR_2$. The individual firms shift to a new long-run equilibrium point E_2 , their individual output remaining the same. Whether equilibrium output of the firms remains constant, increases or decreases, depends, respectively, on whether cost curves shift upward vertically, upward to the right or upward to the left.

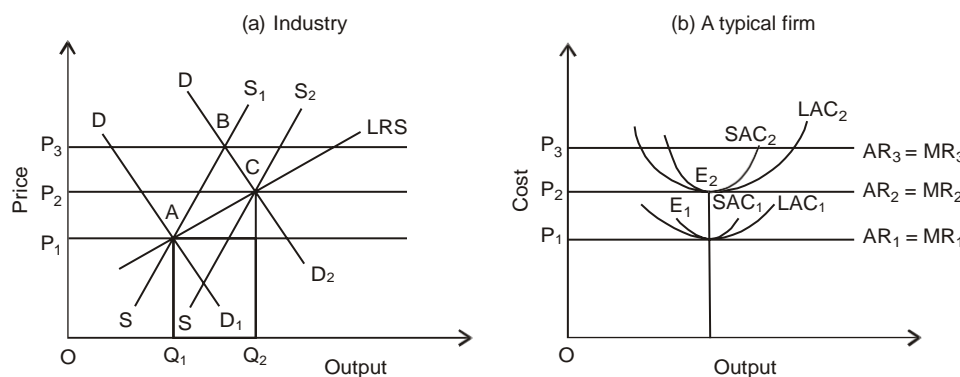


Fig. 5.11 Long-run Supply Curve of an Increasing Cost Industry

Notethat at price OP_2 , both industry and individual firms are in equilibrium. In the absence of any further disturbance, the equilibrium of both firms and industry will remain stable. Thus, at the new equilibrium price OP_2 , the industry-output increases from OQ_1 to OQ_2 and corresponding equilibrium points are A and C , respectively. By joining the long-run equilibrium points A and C , we get the long-run supply curve for the industry, as shown by the curve LRS . Obviously, the LRS has a positive slope in an increasing cost industry.

5.8.3 Decreasing Cost Industry

If expansion of output of an industry is associated with decrease in the input prices, the industry is referred to as a *decreasing cost industry*. A decreasing cost industry has a long-run supply curve with a negative slope.

The derivation of long-run industry supply curve (LRS) under decreasing cost condition is illustrated in Figure 5.12(a) and (b). Let the industry be initially in

NOTES

equilibrium at point A [Figure 5.12 (a)] and firms at E_2 [Figure 5.12 (b)]. Now suppose that demand curve shifts from DD_1 to DD_2 and, consequently, price rises from OP_2 to OP_3 . The short-run equilibrium of firms at price OP_1 [Figure 5.12 (b)] moves upward on the SMC_2 where the firms make abnormal profits. The abnormal profits attract new firms to the industry causing increase in demand for inputs. In case industries are enjoying increasing returns to scale due to economies of scale, the increase in demand for inputs would encourage increased supply of inputs. Increase in the supply of inputs causes input prices to fall. The industry, therefore, enjoys the external economies to scale. As a result, their long-run and short-run cost curves shift downward, from LAC_2 to LAC_1 [Figure 5.12 (b)].

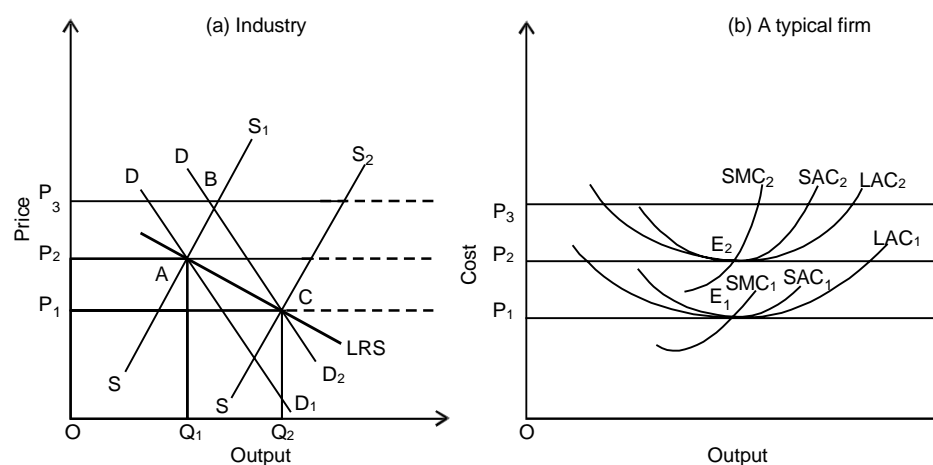


Fig. 5.12 Long-run Supply Curve of a Decreasing Cost Industry

From the industry's point of view, industry supply increases due to the entry of new firms, even if the existing firms maintain their old level of output. Therefore, the industry supply curve shifts from SS_1 to SS_2 which intersects the new demand curve DD_2 at point C. Thus, the equilibrium of the industry shifts from A to C. Industry output increases from OQ_1 to OQ_2 . In the absence of any external disturbance, the industry equilibrium point C, would tend to stabilise. By joining the two equilibrium points A and C and extending it further, we get the long-run supply curve (LRS) of the decreasing cost industry. The LRS has a negative slope.

Issues with decreasing cost: Some authors argue that the phenomenon of decreasing cost... is not consistent with all the requirements of perfect competition. However, the possibility of a decreasing cost industry cannot be ruled out in a very long period. One reason for this is the likelihood of the existence of large external economies of scale, particularly in case of young industries in the undeveloped areas. An increase in the number of industries and the consequent growth of transportation, marketing facilities and financial institutions may reduce the industry's cost of production. Nevertheless, it depends on how substantial are the external economies of scale. R.G. Lipsey has cited the car industry of England as an example of decreasing cost industry. In his own word, "As the output of cars increased, the industry's demand for tyres grew greatly. This ... would have

increased the demand for rubber and tended to raise its price, but it also provided the opportunity for tyre manufacturers to build large modern plants and reap the benefits of increasing returns in tyre production. At first, these economies were large enough to offset any factor price increases and tyre price charged to car manufacturers all. Thus car costs fell because of lower prices of an important input.’

To conclude, whether costs of an industry remain constant or decrease due to increase in the price of some of its inputs, depends also on what proportion of the total input supply is consumed by the industry. For example, output of pencil industry can be increased without substantially affecting the lumber prices as pencil industry uses a small proportion of lumber output.

But a large increase in the output of furniture industry will not leave lumber prices unaffected. Similarly, output of a pin industry can be substantially increased without affecting the steel price. But a substantial increase in car output cannot leave steel prices unaffected.

Another factor which may cause a rise in input prices is whether or not input industries enjoy economies of scale.

Moreover, the most common cases are of the constant and increasing cost industries. Decreasing, cost industries are most unlikely to exist for a long time. The constant and decreasing cost industries, tend over time to become increasing cost industries because external economies have a limit.

Check Your Progress

5. How is the supply curve of an individual firm derived?
6. Under perfect competition, how does the change in the equilibrium of an individual firm affect the industry’s equilibrium?
7. When is an industry in equilibrium?
8. What do you understand by constant cost industry?

5.9 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. Under perfect competition, the number of sellers is assumed to be so large that the share of each seller in the total supply of a product is very small or insignificant. Therefore, no single seller can influence the market price by changing his supply or can charge a higher price.
2. Under perfect competition, a government does not interfere in any way with the functioning of the market. There are no discriminatory taxes or subsidies; no licensing system, no allocation of inputs by the government, or any other kind of direct or indirect control. That is, the government follows the free enterprise policy.

NOTES

NOTES

3. Under perfect competition, an individual firm does not determine the price of its product. Price for its product is determined by the market demand and market supply.
4. A profitmaximising firm is in equilibrium at the level of output which equates its $MC = MR$. However, the level of output which meets the equilibrium condition for a firm varies depending on cost and revenue functions.
5. The supply curve of an individual firm is derived on the basis of its equilibrium output.
6. In a perfectly competitive market, change in the equilibrium of an individual firm does not affect the industry's equilibrium, simply because the total output of a single firm constitutes a small fraction of the industry's output. But, a change in the industry's equilibrium does alter the equilibrium of an individual firm.
7. An industry is in equilibrium at a price and output at which market is cleared, i.e., where market demand equals market supply.
8. An industry for which factor prices remain constant is referred to as constant cost industry. In other words, when the expansion of output in an industry does not entail a change in factor prices, the industry is said to be a constant cost industry.

5.10 SUMMARY

- Perfect competition refers to a market condition in which a very large number of buyers and sellers enjoy full freedom to buy and to sell a homogenous good and service and they have perfect knowledge about the market conditions, and factors of production have full freedom of mobility.
- Under perfect competition, the number of sellers is assumed to be so large that the share of each seller in the total supply of a product is very small or insignificant. Therefore, no single seller can influence the market price by changing his supply or can charge a higher price. Therefore, firms are price-takers, not price-makers.
- Government does not interfere in any way with the functioning of the market. There are no discriminatory taxes or subsidies; no licencing system, no allocation of inputs by the government, or any other kind of direct or indirect control. That is, the government follows the free enterprise policy.
- A profitmaximising firm is in equilibrium at the level of output which equates its $MC = MR$. However, the level of output which meets the equilibrium condition for a firm varies depending on cost and revenue functions.
- The supply curve of an individual firm is derived on the basis of its equilibrium output. The equilibrium output, determined by the intersection of MR and MC curves, is the optimum supply by a profitmaximising (or cost minimising) firm.

- The industry supply curve, or what is also called market supply curve, is the horizontal summation of the supply curve of the individual firms. If cost curves of the individual firms of an industry are identical, their individual supply curves are also identical. In that case, industry supply curve can be obtained by multiplying the individual supply at various prices by the number of firms.
- An industry is in equilibrium at a price and output at which market is cleared, i.e., where market demand equals market supply.
- An industry for which factor prices remain constant is referred to as constant cost industry. In other words, when the expansion of output in an industry does not entail a change in factor prices, the industry is said to be a constant cost industry.

NOTES

5.11 KEY WORDS

- **Perfect Competition:** Perfect competition refers to a market condition in which a very large number of buyers and sellers enjoy full freedom to buy and to sell a homogenous good and service and they have perfect knowledge about the market conditions, and factors of production have full freedom of mobility.
- **Industry Supply Curve:** It is the horizontal summation of the supply curve of the individual firms.

5.12 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short Answers Question

1. What are the features of perfect competition?
2. Distinguish between perfect and pure competition.
3. Explain the short-run equilibrium of a competitive firm.
4. Do you agree that perfect competition leads to optimum size of the firm? Give reasons for your answer.
5. Under perfect competition average revenue equals average cost in the long-run equilibrium. Yet why do firms produce under such a condition?
6. Bring out the essential difference in the nature of equilibrium of a firm under perfect competition in the short-run and in the long-run.
7. Write a short note on the relationship between firm's short-run cost curves and supply curve.

NOTES

Long Answers Question

1. What is the relative position of a firm in a perfectly competitive industry? How does it choose its price and output? Under what market conditions a firm is a price taker?
2. Analyse the equilibrium of a firm under the conditions of perfect competition in the short-run? Discuss in this regard the importance of AR, AC, MR and MC under perfect competition.
3. Show how under the condition of perfect competition in the long-run, the price of a commodity is equal to its average and marginal cost.
4. Distinguish between short-run and long-run equilibrium of a firm under perfect competition. What differences, if any, are there in conditions of equilibrium in the two cases?
5. How is short-run supply curve of a firm derived under perfect competition? Why can't it be downward sloping?
6. Show graphically long-run supply curves of an industry is drawn under perfect competition? Also illustrate graphically the derivation of the long-run supply curve of a firm under perfect competition.

5.13 FURTHER READINGS

- Dwivedi, D. N. 2002. *Managerial Economics*, 6th Edition. New Delhi: Vikas Publishing House.
- Keat Paul G. and K. Y. Philip. 2003. *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th Edition. Singapore: Pearson Education Inc.
- Keating, B. and J. H. Wilson. 2003. *Managerial Economics: An Economic Foundation for Business Decisions*, 2nd Edition. New Delhi: Biztantra.
- Mansfield, E., W. B. Allen, N. A. Doherty and K. Weigelt. 2002. *Managerial Economics: Theory, Applications and Cases*, 5th Edition. NY: W. Orton & Co.
- Peterson, H. C. and W. C. Lewis. 1999. *Managerial Economics*, 4th Edition. Singapore: Pearson Education, Inc.
- Salvantore, Dominick. 2001. *Managerial Economics in a Global Economy*, 4th Edition. Australia: Thomson-South Western.
- Thomas, Christopher R. and Maurice S. Charles. 2005. *Managerial Economics: Concepts and Applications*, 8th Edition. New Delhi: Tata McGraw-Hill.

UNIT 6 MONOPOLY

Structure

- 6.0 Introduction
- 6.1 Objectives
- 6.2 Meaning and Feature of Monopoly
 - 6.2.1 Revenue Curves Under Monopoly
- 6.3 Monopoly—Short-run and Long-run Equilibrium
 - 6.3.1 Price and Output Determination in Short-run
 - 6.3.2 Monopoly Equilibrium in the Long-run
- 6.4 Price Discrimination Under Monopoly
 - 6.4.1 Necessary Conditions for Price Discrimination
 - 6.4.2 Degrees of Price Discrimination
 - 6.4.3 Welfare Aspects
- 6.5 Monopoly Control and Regulation
 - 6.5.1 Government Regulation of Monopoly Prices
- 6.6 Answers to Check Your Progress Questions
- 6.7 Summary
- 6.8 Key Words
- 6.9 Self Assessment Questions and Exercises
- 6.10 Further Readings

NOTES

6.0 INTRODUCTION

In the preceding unit, you learnt about the theory of price and output determination under perfect competition. In this unit, you will study about the theory of price and output determination under monopoly—another extreme case of market and recall that under perfect competition, a firm is a *price-taker*. In contrast, under monopoly, a firm is a *price-maker*. The following aspects of monopoly theory are discussed here.

- (i) Meaning and sources of monopoly power
- (ii) Price and output determination in the short and long runs
- (iii) Price discrimination by a monopoly firm

6.1 OBJECTIVES

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

- Discuss the meaning and features of monopoly
- Describe how monopoly firms determine their equilibrium level of price and output
- Examine why monopoly firms discriminate between consumers and often charge different prices from different kinds of consumers
- Describe government control and regulations of monopolies

NOTES

6.2 MEANING AND FEATURE OF MONOPOLY

The word monopoly has been derived from Greek word *monos*, meaning ‘alone’ and *polein* meaning ‘seller’. By definition, *monopoly is a market situation in which there is a single seller of a commodity of lasting distinction without close substitutes*. A monopoly firm enjoys an absolute power to produce and sell a commodity. This, however, does not mean that a monopoly firm is absolutely free from any kind of competition. Monopoly firms too, have to face *indirect competition*, especially in regard to setting the price of the product. There are at least *two potential sources of indirect competition*.

First potential source of indirect competition is the rivalry between monopoly goods and other goods produced by other monopolies and competitive firms. Therefore, a monopolist cannot charge any price for its product. For example, North Delhi Power Limited, a public sector electricity producing and supplying company is at present a monopolist in Delhi. Pricing of electricity is under the control and regulation of the government. Therefore, while fixing the electricity price, it will have to take into account the government’s pricing policy. Besides, it will have to take into account what people can afford after meeting such essential needs as food, clothing, shelter, education and medicine.

The **second** source of potential indirect competition comes from the availability and price of *inferior substitutes*.

Given these problems confronting the monopolies, one can hardly find many cases of a **pure** or **absolute monopolies**. However, the theoretical discussion on price and output determination under monopoly is based on the case of a *pure monopoly*, i.e., a monopoly firm enjoying absolute power in determining the price and output of its product.

Finally, an **important feature** of a pure monopoly is that a monopolized industry is a single-firm industry, i.e., there is no distinction between the firm and the industry. Therefore, there is no distinction between market demand curve and monopoly firm’s own demand curve, i.e., the demand curve for the monopoly firm’s product is same as the market demand curve.

Sources and Kinds of Monopolies

The emergence and survival of monopoly is attributed to the factors which prevent the entry of other firms into the industry. The barriers to entry are, therefore, the sources of monopoly power. The major sources of barriers to entry are: (i) legal restrictions, (ii) sole control over the supply of certain scarce and key raw materials; (iii) efficiency; (iv) economies of scale; and (v) patent rights. How these factors prevent the entry of new firms is discussed below.

(i) Legal restrictions. Some monopolies are created by law in public interest. Such monopolies may be created in both public and private sectors. Most of the state monopolies in the public utility sector, including postal, telegraph, and telephone services, generation and distribution of electricity, railways, airlines and state roadways, etc., are public monopolies. Such monopolies, wherever they exist, are created by the public law. The state may create monopolies in the private sector also through licence or patent. Such monopolies are intended to reduce cost of production by the economies of scale and investment in technical innovations. Such monopolies are also known as *franchise monopolies*.

(ii) Control over key raw materials. Some firms acquire monopoly power from their overtime control over certain scarce and key raw materials that 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 the 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 technical knowledge or techniques 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, i.e., the most efficient scale of production, coincides almost with the size of the market. In such industries or products, a large-size firm finds it profitable, in the long-run, to eliminate competition by cutting down its price for a short period. Once a monopoly is established, it becomes almost impossible for the new firms to enter the industry and survive. Monopolies born out of efficiency are known as *natural monopolies*. A natural monopoly may emerge out of the technical conditions of efficiency or may be created by the law on efficiency grounds in public interest.

(iv) Economies of scale. Under free economic system, a large firm enjoying large economies of scale is capable of cutting down the price low enough to eliminate the competitors. This method is adopted in the absence of Monopoly Control Act in the country. Once competing firms are eliminated, the large firm enjoys the powers of a monopoly firm.

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

NOTES

NOTES

6.2.1 Revenue Curves Under Monopoly

Tools that are used to analyze price and output determination include cost curves and revenue curves. It is therefore essential that we acquaint ourselves with cost and revenue curves faced by monopoly firms. Let it be noted at the first instance that cost curves—the AC and MC curves—faced by the monopoly firm are U-shaped, just as those faced by the firms under perfect competition. But a monopoly firm has option to charge a high or a low price at its discretion. Therefore, monopoly firm has a downward sloping demand curve. As a result, the AR and MR curves that a monopoly firm faces are different from those faced by the firms under perfect competition. Therefore, before we discuss price and output determination and firm's equilibrium under monopoly, it will be useful to look into the nature of revenue curves faced by a monopoly firm.

As we have noted earlier, in a perfectly competitive market, there is dichotomy between the firm and the industry—firms face a horizontal, straight-line demand curve and industry faces a downward sloping demand curve. Since under monopoly, there is a single firm in the industry, *the industry is a single-firm-industry and industry demand curve has a negative slope*. A monopoly firm faces, therefore, a downward sloping demand curve—it may be a linear or a non-linear demand curve. It is **important** to note here that, given the demand curve, a monopoly firm has the option to choose between the prices to be charged and output to be sold. Once it chooses a price, the demand for its output is fixed. Similarly, if the firm decides to sell a certain quantity of output, then its price is fixed. It cannot charge any other price inconsistent with the demand curve.

For example, suppose that the demand curve for a monopolised industry is given as DM in Figure 6.1. Demand curve, DM , shows the quantities that can be sold at different prices. For instance, if monopoly firm chooses price OP , the quantity that it can sell at this price is fixed at OQ —no other quantity can be sold at this price. Similarly, if it decides to sell quantity ON , its price has to be fixed at OP' —it cannot sell ON output at a higher price. This means that if demand curve is given, the options of monopoly firm becomes limited—it can choose either a price or a quantity at a time. The firm cannot choose a price and a quantity inconsistent with the demand curve. This implies that a monopoly firm has the option of charging a high price and selling a lower quantity or charging a lower price and selling a larger quantity.

Relationship between AR and MR curves: Another point that needs to be noted is the relationship between monopoly's AR and MR curves. As noted earlier, the AR curve for a firm is the same as its demand curve. So is the case with a monopoly firm. Since a monopoly firm faces a downward sloping demand curve, its AR also slopes downward to the right. For example, the demand curve DM in Figure 6.1 is the same as the firm's AR curve.

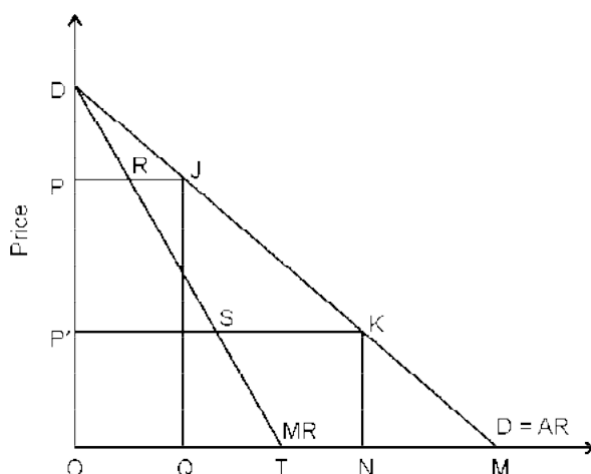


Fig. 6.1 AR and MR Curves for Monopoly

What is more *important* in the analysis of equilibrium of a monopoly firm is the *relationship between the AR and MR curves*. As we have seen in the previous chapter, when price is fixed, as in case of perfect competition, firm's demand curve takes the form of a horizontal line. In that case, $AR = MR$ and MR is a straight line too. But, in case of a monopoly firm, demand curve has a negative slope. Therefore, its MR curve too has a negative slope. There is, however, a *specific relationship* between AR and MR , i.e., *the slope of MR curve is twice that of that AR curve*. That is, given the linear demand function, marginal revenue curve is twice as steep as the average revenue curve. This relationship can be proved as follows. Let us assume that a monopoly firm is faced with a price function given as

$$P = a - bQ \quad \dots(6.1)$$

We know that $TR = Q \cdot P$

By substituting Equation 6.1 for P , we get

$$\begin{aligned} TR &= Q(a - bQ) \\ &= aQ - bQ^2 \end{aligned} \quad \dots(6.2)$$

Since MR equals the first derivative of the TR function,

$$\begin{aligned} MR &= \frac{\partial TR}{\partial Q} = \frac{\partial(aQ - bQ^2)}{\partial Q} \\ &= a - 2bQ \end{aligned} \quad \dots(6.3)$$

Note that the slope of the price function (6.1) equals b whereas the slope of the MR -function (6.3) equals $2b$, it means that the slope of the MR -function is *twice* that of the AR -function. It implies that MR curve is always to the left of AR curve and MR bisects the demand at all levels of price. For example, in Figure 6.1, if price is OP , demand is PJ , and MR passes through point R which divides PJ in two equal parts. Geometrically, $PR = RJ$. Similarly, at price OM , demand equals $P'K$, $P'K = P'S + SK$ where $P'S = SK$, i.e., $P'S = P'K/2$.

NOTES

6.3 MONOPOLY—SHORT-RUN AND LONG-RUN EQUILIBRIUM

NOTES

Let us now discuss price determination under monopoly in the short and the long run.

6.3.1 Price and Output Determination in Short-run

Having defined monopoly and discussed the nature of revenue curves it faces, let us now discuss the price and output determination under monopoly. According to the traditional theory of firm, a monopoly firm (or otherwise) is said to be in equilibrium where it maximises its profit. Maximisation of total profit is a matter of time required to adjust output to the price. Therefore, as in case of perfect competition, equilibrium of a monopoly is studied under both *short-run* and *long-run* conditions. In this section, we explain price and output determination under monopoly in the *short-run*. The equilibrium of monopoly in the *long-run* will be discussed in the next section.

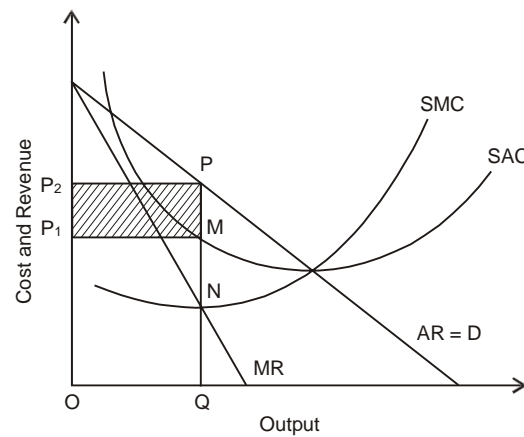


Fig. 6.2 Monopoly Equilibrium: MR-MC Approach

The short-run equilibrium of the monopoly firm is illustrated in Figure 6.2. The short-run revenue curves of the monopoly firm are shown by the *AR* and *MR* curves and its short-run cost curves are shown by the *SAC* and *SMC* curves. Given the revenue and cost conditions and the profit maximisation rule, the equilibrium of the monopoly firm can easily be traced. Recall once again that profit is maximum where $MR = MC$. It can be seen in the figure that *MR* and *SMC* curves intersect at point *N*. Note that point *N* satisfies both the conditions of profit maximisation: (i) $MR = MC$, and (ii) *SMC* curve intersects *MR* curve from below. Point *N*, therefore, determines the equilibrium output and price. An ordinate drawn from point *N* to X-axis determines the profit maximising output at *OQ*. The ordinate *NQ* extended upward to the *AR* curve gives the price *PQ* at which output *OQ* can be sold, given the demand function. Thus, at monopoly equilibrium determines, both equilibrium output and price are determined simultaneously.

Once equilibrium price and output are determined, given the revenue and cost curve, the maximum monopoly profit can be easily determined as follows.

$$\text{Per unit monopoly profit} = AR - SAC = PQ - MQ = PM$$

Given the equilibrium output OQ , total monopoly profit can be obtained by multiplying per unit profit PM by the equilibrium output, OQ . That is, total monopoly profit = $OQ \cdot PM$.

Since $OQ = P_1M$, total monopoly profit at equilibrium can be written as

$$P_1M \cdot PM = P_1P_2PM$$

The total monopoly profit is shown by the shaded area in the Figure 6.2. Since cost and revenue conditions of the monopoly firm are supposed to be given, the monopoly equilibrium is supposed to be stable.

6.3.2 Monopoly Equilibrium in the Long-run

The long-run equilibrium conditions of a monopolist differ from those faced by the competitive firms in another important respect, i.e., the entry of new firms into the industry. While in a competitive market, there is free entry to the industry, a monopoly firm is protected by the *barriers to entry*. The barriers to entry may be in the form of patent rights, legal protection, economies of scale and the well established longstanding of the monopolist.

In the long run, a monopolist gets an opportunity to expand the size of its plant with a view to maximising its long-run profits provided the size of the market is fairly large. The expansion of the plant size may, however, be subject to such conditions as: (a) size of the markets; (b) expected economic profits; and (c) risk of inviting legal restrictions, like *MRTP*.

A general case of monopoly equilibrium in the long-run is presented in Figure 6.3, assuming none of the above conditions limit the expansion of monopoly firm. The AR and MR curves show the market demand and marginal revenue conditions, respectively, faced by the monopoly. Let us begin the analysis with short-run case. The short-run average and marginal cost conditions are shown by SAC and SMC curves. The SMC and MR curves intersect at point A determining the short-run, equilibrium output at OQ_1 and price at P_1Q_1 . The firm is making super-normal profit. Given the demand curve, $AR = D$, the firm has a wide scope for expansion of production. Therefore, the firm adds new plants which results in decrease in long-run production cost. The LAC and LMC curves show the long-term cost conditions. Given the revenue and cost conditions as shown in the figure, the point of intersection between LMC and MR curves determine the equilibrium output at OQ_2 . Given the AR curve, price is determined at P_2Q_2 . Thus the long-run equilibrium output is OQ_2 and equilibrium price is P_2Q_2 . This price-output combination maximises the monopolist's long-run profits. The total long-run profit has been shown by the area $LMSP_2$. Here the monopoly firm is in the long-run equilibrium.

NOTES

NOTES

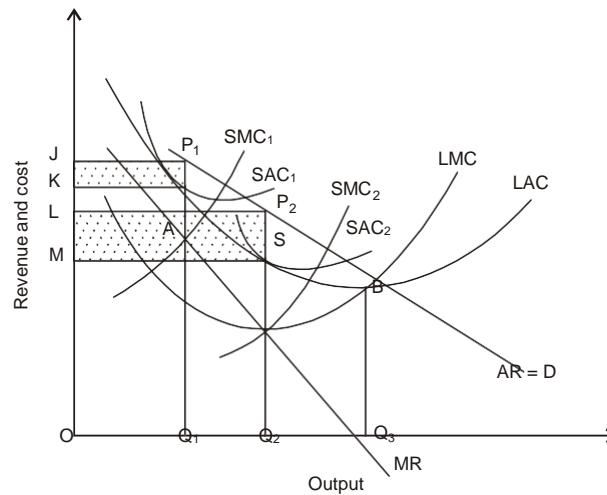


Fig. 6.3 Long-run Equilibrium of the Monopoly Firm

Check Your Progress

1. What are the sources of indirect competition under a monopoly?
2. In a monopoly, what is the difference between the market demand curve and the monopoly firm's demand curve?
3. What are the various barriers to entry for another firm in a monopoly condition?

6.4 PRICE DISCRIMINATION UNDER MONOPOLY

The theory of pricing under monopoly, as discussed above, gives the impression that once a monopolist fixes the price of its product, the same price will be charged from all the consumers. This is, however, not the case generally. A monopolist, simply by virtue of its monopoly position, is capable of charging different prices from different consumers or groups of consumers. When the same (or somewhat differentiated) product is sold at different prices to different set of consumers, it is called price discrimination. When a monopolist sells an identical product at different prices to different buyers, it is called a discriminatory monopoly.

Consumers are discriminated in respect of price on the basis of their income or purchasing power, geographical location, age, sex, quantity they purchase, their association with the seller, frequency of purchase, purpose of the use of the commodity or service, and on several other grounds which the seller may find suitable.

A common example of consumers being discriminated on the basis of their incomes is found in medical professions. Consulting physicians charge different

fees from different clients on the basis of their paying capacity even if quantity and quality of service rendered is the same. Price discrimination on the basis of age is found in railways, roadways and airways: children between 3 and 12 years are charged only half the adult rates. Price discrimination on the basis of quantity purchased is very common. It is generally found that private businessmen charge lower price (or give discount) when bulk purchase is made. In case of public utility services, however, lower rates are charged when commodity or service is consumed in smaller quantity. The most common practice of price discrimination is found in entertainment business, e.g., cinema shows, musical concerts and game-shows.

The product or service in question may be identical or slightly differentiated. For example, services of consulting physician and lawyer are identical. The services of railways, roadways and entertainment shows may be slightly differentiated by providing more comfortable seats, sleepers, security and air conditioning, etc. for the purpose of price discrimination. The modification in service may involve some additional cost. But price differentials are much more than is justified by cost differentials.

Although price discrimination is a common practice under monopoly, it should not mean that this practice exists only under monopoly. Price discrimination is quite common also in other kinds of market structures, particularly when market imperfection exists.

6.4.1 Necessary Conditions for Price Discrimination

As noted above, monopoly firms discriminate between consumers in regard to price charged from them. However, there are certain necessary conditions for price discrimination. The conditions are described here briefly.

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

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

Third, *there must be imperfect competition in the market.* The seller must possess some monopoly power over the supply of the product to be able to distinguish between different classes of consumers and to charge different prices.

NOTES

NOTES

6.4.2 Degrees of Price Discrimination

The degree of price discrimination refers to the extent to which a monopolist can divide the market and can take advantage of market division in extracting the consumer's surplus. According to Pigou, there are three degrees of price discrimination practiced by the monopolists: (i) first degree price discrimination; (ii) second degree price discrimination; and (iii) third degree price discrimination.

(i) First degree price discrimination: The discriminatory pricing that aims at taking away the entire consumer surplus is called **first degree price discrimination**. The first degree discrimination is possible only when a seller is in a position to know the price each buyer is willing to pay. That is, the monopolist knows buyer's demand curve for the product. In that case the seller first sets a price at the highest possible level at which all those who are willing to buy, purchase at least one unit each of the commodity. When the consumer surplus of this section of consumers is exhausted, he gradually lowers down the price so that the consumer surplus of the users of the subsequent units can be extracted. This procedure is continued until the whole consumer surplus available at the price where $MR = MC$ is extracted. Also, consider the case of services of exclusive use, e.g., medical services. A doctor who knows or can guess the paying capacity of his patients, can charge the highest possible fee from visibly the rich patients and the lowest fee from the poorest one. The first degree of price discrimination is the limit of discriminatory pricing.

(ii) Second degree price discrimination: Under the second degree of discriminatory pricing, consumers are classified under different categories and a different price is charged from the different categories of consumer, e.g., consumers belonging to high, middle and low income categories. The second degree price discrimination is also called 'block pricing system'. A monopolist adopting the **second degree price discrimination** intends to siphon off only the major part of the consumer surplus, rather than the entire of it.

This second degree price discrimination is feasible where:

- (i) The number of consumers is large and price rationing can be effective, as in case of utilities like electricity and natural gas and also consumer durabilities;
- (ii) Demand curves of all the consumers are identical; and
- (iii) A single rate is applicable for a large number of buyers.

The second degree price discrimination is illustrated in Figure 6.4. The monopolist fixes the price of its product first at OP_1 for the high income class of consumers and sells OQ_1 quantity. Once this market segment is exploited, the monopolist reduces the price to OP_2 and sells Q_1Q_2 to the second category of consumers, say, the middle income class. Finally, it cuts down the price to OP_3 and exploits the third category of consumers. This process is used over a period of time. Thus, by adopting a **block-pricing system**, the monopolist maximises his total revenues (TR) as

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

Monopoly

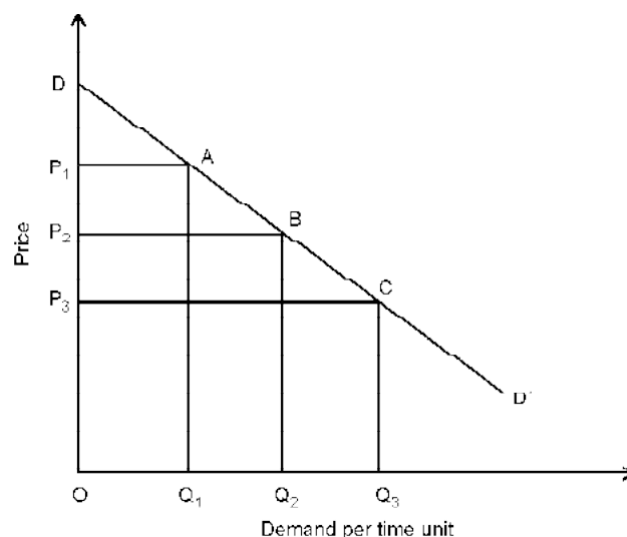


Fig. 6.4 Second Degree Price Discrimination

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

(iii) Third degree price discrimination: When a profit maximising monopoly firm sets different prices in different markets having demand curves with different elasticities, it is using third degree price discrimination. A monopolist is often faced with two or more markets, completely separated from each other—each having a demand curve with different elasticity. Therefore, a uniform price cannot be set for all the markets without losing the possible profits. The monopolist is, therefore, required to allocate total output between the different markets so that profit can be maximised in each market. Profit in each market would be maximum only when $MR = MC$ in each market. The monopolist, therefore, allocates its total output between the markets in such proportions that in all the market his $MR = MC$.

The process of output allocation and determination of price for different markets is illustrated in Figure 6.5. Suppose that a monopolist has to sell goods in only two markets, A and B. The two markets are so separated that resale of commodity is not feasible. The demand curve (D) and marginal revenue curve (MR) given in Figure 6.5 (a) represent the AR and MR curves in market A and the curves D_b and MR_b in Figure 6.5 (b) represent the AR and MR curves, respectively, in market B. The horizontal summation of demand curves D_a and D_b gives the total demand curve for the two markets, as shown by the curve $AR = D$, and the horizontal summation of MR_a and MR_b is given by the curve MR in Figure 6.5 (c). The firm's marginal cost is shown by MC which intersects the aggregated MR at point T . Thus, optimum level of output for the firm is determined at OQ . The whole of OQ cannot be profitably sold in anyone market because of their limited size.

NOTES

NOTES

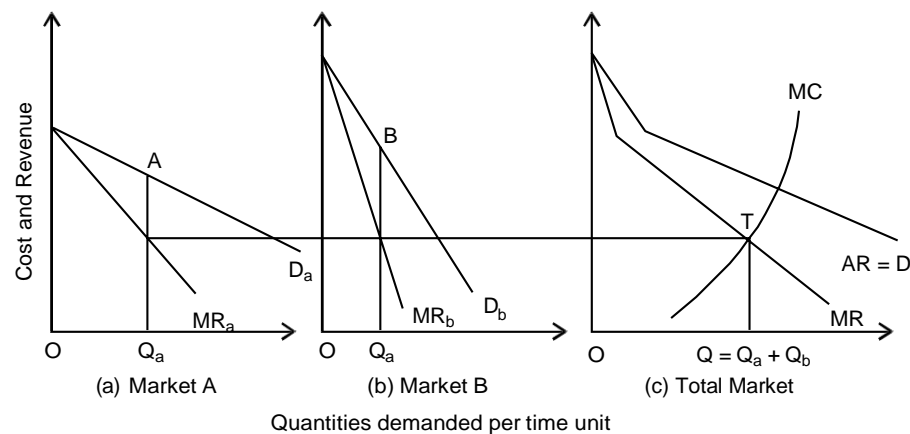


Fig. 6.5 Third Degree Price Discrimination

Therefore, the monopolist would allocate output OQ between the two markets in such proportions that the necessary condition of profit maximisation is satisfied in both the markets (i.e., $MC = MR$). The profit maximising output for each market can be obtained by drawing a line from point T , parallel to X-axis, through MR_b and MR_a . The points of intersection on curved MR_a and MR_b determine the optimum share for each market. As shown in Figure 6.5, the monopolist maximises profit in market A by selling OQ_a units at price AQ_a and by selling OQ_b units in market B at price BQ_b . The firm's total equilibrium output is $OQ = OQ_a + OQ_b$.

Since at OQ_a , $MR_a = MC$ in market A, and at OQ_b , $MR_b = MC$ in market B,
 $MC = TQ = MR_a = MR_b$

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

The third degree method of price discrimination is most suitable where the total market is divided between the home and foreign markets. This practice is called **dumping**—not necessarily by monopoly firms. However, it may be suitably practised between any two or more markets separated from each other by any two or more of such factors as geographical distance, transport barriers, cost of transportation, legal restrictions on the inter-regional or interstate transfer of commodities by individuals, etc.

6.4.3 Welfare Aspects

Price discrimination has had a bad reputation and condemned as illegal and immoral. The objection is: Why charge a higher price from some consumer's and a lower price from others while there is no extra advantage to those who pay a higher price, or why benefit some at the cost of the some others? In the United Kingdom and the United States, railways were prohibited to charge discriminatory rates. Discriminatory pricing has also been criticised as a

destructive tool in the hands of monopoly. For, in the past, large corporations had sought to use price discrimination to prevent the growth of competition. Besides, price discrimination may cause mal-allocation of resources and, hence, may be a deterrent to social welfare.

This is, however, not the case always. In some cases price discrimination is socially advantageous. In fact, as Lipsey has observed, ‘Whether an individual judges price discrimination to be good or bad is likely to depend upon the details of the case as well as upon his own personal value judgements.’ He adds, ‘Certainly there is nothing in economic theory to suggest that price discrimination is always in some sense worse than non-discrimination under conditions of monopoly or oligopoly.’

Price discrimination is however considered to be *desirable* when it adds to total social welfare. Price discrimination is generally justified on the following grounds.

First, price discrimination is socially desirable in case of goods and services which are essential for the society as a whole but their production is unprofitable because long-run average cost curve (LAC) lies much above the aggregated market demand curve. In that case, such goods and services cannot be produced. But, production of such goods and services can be possible provided price discrimination is permitted. Price discrimination, thus, becomes essential for the survival of the industry, on the one hand, and availability of such goods and services, on the other.

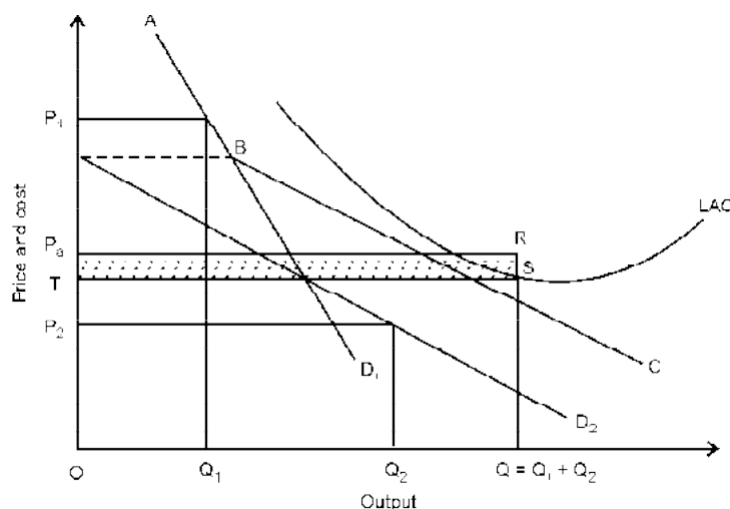


Fig. 6.6 Price Discrimination for Industry's Survival

The desirability of price discrimination is illustrated in Figure 6.6. Suppose (i) there are two markets, I and II with their individual demand curves given as D_1 and D_2 , respectively, (ii) market demand curve is given by ABC , and (iii) the long-run average cost curve is given by curve LAC . Note that LAC lies throughout above the total demand curve ABC . Therefore, a profitable production is not

NOTES

NOTES

possible if a price along the market demand curve ABC is to be charged from all its consumers. However, if price discrimination is adopted and prices are so charged in the two markets that the total revenue exceeds LAC at some level of output, then monopoly may profitably survive to the advantage of the society. Let us suppose that the monopoly firm sets price OP_1 in the market I in which demand is less elastic and price OP_2 in market II in which demand is highly elastic. It would sell OQ_1 units at price OP_1 in market I and OQ_2 at price OP_2 in market II. The total output would then be $OQ = OQ_1 + OQ_2$. His total revenue (TR) would be

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

and
$$AR = \frac{(OP_1 \times OQ_1 + OP_2 \times OQ_2)}{OQ}$$

Suppose AR is estimated to be OP_a as shown in Figure 6.6.

At output OQ , the LAC is $OT = QS$. Thus, the total cost,

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

and its total revenue,

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

Since $OQRP_a > OQST$, the monopoly firm not only covers its cost but also makes a pure profit. Its total profit (Π) can be expressed as

$$\Pi = OQRP_a - OQST = P_a RST$$

This kind of situation arises mostly in public utility services like railways, roadways, post and telegraph services, etc., in which high-paying sections of the market subsidise the low-paying sections. In other words, rich subsidise the consumption of the poor.

Second, discriminatory pricing can be adopted with justification where a uniform, single profitable price is likely to restrict the output and deprive many (particularly the people of lower income groups) of the essential good or service. For example, if specialist doctors in private practice, who charge a discriminatory price for their services, are asked to charge a uniform fee from all the patients, they would charge a fee high enough to maintain the level of their income. But the high fee deprives the poor of the doctor's services and may force them to opt for an inferior or inadequate treatment. This leads to a socially undesirable situation because, on the one hand, rich patients who could pay a higher fee pay a price lower than what they could afford, and could, on the other hand, poor patients are deprived of proper medical care. What is worse, doctor's services remain underutilized.

Third, there is a section of consumers which gains more than the people of other sections from the use of a product. For example, factory-owners gain more from the use of electricity than the households; irrigation projects benefit big farmers more than the marginal farmers; motorists benefit from roads more than pedestrians, and so on. In such cases, uniform price would be unjustified

from a normative point of view, provided the objective is not to restrain the household consumption of certain goods like electricity and spare it for productive purposes. There is, on the other hand, full justification for discriminatory pricing of electricity.

Check Your Progress

4. What are the three degrees of price discrimination practiced by monopolists?
5. When is price discrimination considered desirable?
6. What is discriminatory monopoly?

NOTES

6.5 MONOPOLY CONTROL AND REGULATION

The degree of monopoly power matters a great deal in pricing and output decisions of a monopolist and in respect of control and regulation of monopolies. We discuss here the various measures of monopoly power. It must be borne in mind here that the question of measuring monopoly power arises only under monopolistic competition—not in case of *pure monopoly*. In case of pure monopoly, the degree of monopoly power is 1 and it is known to the policy makers.

Measuring monopoly power has been a very difficult problem. The efforts to devise a measure of monopoly power have not yielded any universal or non-controversial measure. As Hunter has observed, ‘The idea of devising a measure of monopoly power, with reference both to its general incidence and to particular situation, has been and probably always will remain an attractive prospect for economists who wish to probe in this field. If not for any other reason, then for ‘sheer intellectual curiosity’, economic theorists feel compelled to work on this problem, for they could not with good conscience go on talking about ‘great’ or ‘little’ monopoly power or about various degrees of monopoly without trying to ascertain the meaning of these words.

Therefore, to devise at least a ‘conceivable’ measure of monopoly, even if ‘practical’ measurement is impossible, continues to interest the economists, for at least *two reasons*.

First, apart from intellectual curiosity, people would like to know about the economy in which they live, its industrial structure, and the industries from which they get their supplies.

Second, growth of monopolies have forced governments of many countries to formulate policies and devise legislative measures to control and regulate monopolies. If the government is to succeed in its policy of restraining monopoly, it must have at least some practicable measure of monopoly power and monopolistic trade practices.

NOTES

Measuring Monopoly Power

The economist has suggested several devices to measure the degree of monopoly power, though none of the measures is free from flaws. Yet, the various measures suggested by the economists do provide an insight into the monopoly power and its impact on the market structure. Besides, they also help in formulating an appropriate public policy to control and regulate the existing monopolies. We have briefly discussed here some important measures of monopoly power.

(a) Number-of-firms criterion: One of the simplest measures of degree of monopoly power is to count the number of firms in an industry. The smaller the number of firms, the greater the degree of monopoly power of each firm in the industry, and conversely, the larger the number of firms, the greater the possibility of absence of monopoly power. A corollary of this criterion is that if there is a single firm in an industry, the firm has an absolute monopoly power. This criterion seems to have been derived from the characteristics of the perfect competition in which the number of firms is so large that no firm has any monopoly power.

This criterion has, however, a serious *drawback*. The number of firms alone does not reveal much about the relative position of the firms within the industry because (i) 'firms are not of equal size' and (ii) their number does not indicate the degree of control each firm exercises in the industry if an industry with a large number of firms is dominated by a large firm or by a few large firms. Therefore, the numerical criterion of measuring monopoly power is of little practical use.

(b) Concentration ratio: The **concentration ratio** is one of the widely used criterion for measuring monopoly power. The concentration ratio is obtained by calculating the percentage share of the largest group of the firms in the total output of an industry. 'The number of firms chosen for calculating the ratio usually depends on some fortuitous element—normally the census of production arrangements of the country concerned.' In Britain, the share of the largest three firms of a census industry, and in the USA, the share of the largest four firms is the basis of calculating concentration ratio. Apart from the share of the largest firms in the industry-output, 'the size of the firms and the concentration of control in the industry may be measured ... in terms of production capacity, value of assets, number of employees or some other characteristics.'

These measures too are, however, not free from drawbacks. They have *three major drawbacks*. **First**, they involve statistical and conceptual problems. For example, production capacity may not be used straightaway as it may include 'unused obsolete or excess capacity' and the value of assets involves valuation problem as accounting method of valuation and market valuation of assets may differ. Employment figure may not be relevant in case of capital-intensive industries. The use of such figures may be misleading. The two other convenient measures are 'gross output value' or 'net output' (value added). But the former involves the risk of double counting and the latter, the omission of inter-establishment transfers.

Second, an important objection to these measures of degree of monopoly power is that these measures do not take into account the size of the market. Size of the market may be national or local. A large number of firms supplying the national market may be much less competitive than the small number of firms supplying the local market. For, it is quite likely that the national market is divided among thousands of sellers so that each seller has the status of a monopolist in his own area.

Third, the most serious defect of concentration ratio as an index of monopoly power is that it does not reflect the competition from other industries. The degree of competition is measured by the elasticity of substitution between the products of different industries. The elasticity of substitution may be different under different classification of industries. Therefore, an industry with concentration ratio under one classification of industries may have a very low elasticity of substitution and hence a high degree of monopoly. But, if classification of industries is altered, the same industry with a high concentration ratio may have a very low elasticity of substitution, and hence may show a low degree of monopoly.

(c) Excess profitability criterion: J.S. Bain and, following him, many other economists have used *excess profit* as a measure of monopoly power. If profit rate of a firm continues to remain sufficiently higher than all opportunity costs required to remain in the industry, it implies that neither competition among sellers nor entry of new firms prevents the firm from making a pure or monopoly profit. While calculating the excess profit, the opportunity cost of owner's capital and margin for the risk must be deducted from the actual profit made by the firm. Assuming no risk, the degree of monopoly may be obtained by calculating the divergence between the opportunity costs (O) and the actual profit (P) as $(P - O)/P$. If $[(P - O)/P] = 0$, there exists no monopoly, and if $[(P - O)/P] > 0$, there is monopoly. The higher the value of $(P - O)/P$, the greater the degree of monopoly.

(d) Lerner's index of monopoly power: Another measure of degree of monopoly based on excess profitability has been suggested by A.P. Lerner. According to Lerner, for a competitive firm, equilibrium price equals marginal cost but for a monopoly price exceeds MC . Therefore, a logical measure of monopoly power is the ratio of $P - MC$ to P , where P is equilibrium price. Thus, the degree of monopoly power (M^P) may be measured as

$$M^P = \frac{P - MC}{P}$$

where P = price, MC = marginal cost.

Since for a profit maximising firm, $MR = MC$, Lerner's measure of monopoly power may also be expressed as,

$$M^P = \frac{P - MR}{P}$$

NOTES

NOTES

Since $P/(P - MR) = e$ (elasticity), $(P - MR)/P = 1/e$. It means that M^P equals the reciprocal of elasticity. Thus, Lerner's measure of monopoly power may also be expressed as $M^P = 1/e$. It may thus be inferred that lower the elasticity, the greater the degree of monopoly, and *viceversa*. According to Lerner's formula, monopoly power may exist even if firm's $AR = AC$ and it earns only normal profit.

Lerner's formula of measuring the degree of monopoly power is considered to be theoretical most sound. Nevertheless, it has been criticised on the following grounds.

First, it is suggested that any formula devised to measure degree of monopoly power should bring out the difference between the monopoly output and competitive output or the 'ideal' output under the optimum allocation of resources. The divergence between P and MC used in Lerner's formula does not indicate the divergence between the monopoly and the 'ideal' output. This substitution of a price-cost discrepancy for a difference between actual and 'ideal' output is probably the greatest weakness of formula which is supposed to measure deviation from the optimum allocation of resources.

Second, price-cost discrepancy may arise for reasons other than monopoly, and price and cost may be equal or close to each other in spite of monopoly power.

Third, since data on MC are hardly available, this formula is of little practical use.

(e) Triffin's cross-elasticity criterion: Triffin's criterion is said to have been derived from the definition of monopoly itself—monopolist is a single seller of a product without close substitutes. According to Triffin's criterion, cross-elasticity is taken as the measure of degree of monopoly—the lower the cross-elasticity of the product of a firm, the greater the degree of its monopoly power. But, this criterion is based on the inter-relationships between the individual firms and indicates only the relative power of each firm—not the measure of its absolute power. It does not furnish a single index of monopoly power.

6.5.1 Government Regulation of Monopoly Prices

The monopolies are, in general, alleged to restrict production, consumption and employment, widen income and wealth disparities, exploit consumers and employees, distort resource allocation and reduce the social welfare. In most countries, therefore, government intervenes and makes laws to control and regulate monopolies to the advantage of the society. There are various measures—direct, indirect, price, non-price, legal and otherwise—to control and regulate the monopolies. We have discussed below price regulation of natural monopolies.

Price regulation is a common feature in case of natural monopolies. When the size of the market is small relative to the optimum size of the firm, market size cannot support more than one firm of optimal size. The monopoly in such a market

is a natural monopoly, protected by market size itself. The government may either nationalise such monopolies or regulate their prices to eliminate the excess profits.

If government intends to regulate the monopoly price, the question arises: what price should be fixed for the monopolist to charge? The two controlled-price systems have been suggested: one that allows some *excess profit* to the monopolist, and the second that allows only *normal profit* to the monopolist. Both the systems of controlled prices, along with their repercussion on output, are illustrated in Figure 6.7.

NOTES

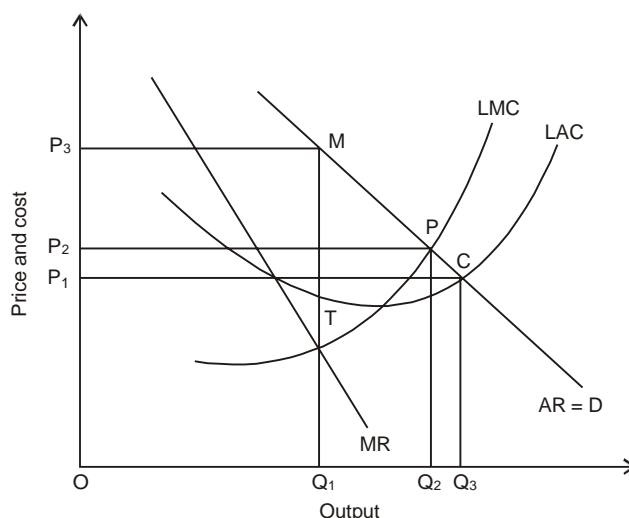


Fig. 6.7 Government Regulated Monopoly

An unregulated monopoly would produce OQ_1 units, charge price OP_3 , and make excess profit of MT per unit. If government intends to regulate monopoly price allowing some super normal profit, then one reasonable price is $OP_2 = PQ_2$, where $LMC = AR$. Alternatively, if government intends to regulate monopoly price with only normal profit, it will fix price at $OP_1 = CQ_3$ at which $AR = LAC$. When OP_1 is the price set for the monopolist, the firm is allowed only normal profit, but the output is maximum possible under the given cost and revenue conditions. On the other hand, if price is fixed at OP_2 , then the monopolist gets some excess profit, but the output is less than that at price OP_1 . In both the cases, however, the total output under regulated monopoly is much greater than that under unregulated monopoly. Which of the two alternative prices (OP_1 and OP_2) is more appropriate is a matter of debate and policy.

Check Your Progress

7. How is concentration ratio obtained?
8. What is Triffin's cross-elasticity criterion?

6.6 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

NOTES

1. Under a monopoly, there are at least two potential sources of indirect competition. First potential source of indirect competition is the rivalry between monopoly good and other goods produced by other monopolies and competitive firms. Therefore, a monopolist cannot charge any price for its product. The second source of potential indirect competition comes from the availability and price of inferior substitutes.
2. Under a monopoly, there is no distinction between market demand curve and monopoly firm's own demand curve, i.e., the demand curve for the monopoly firm's product is same as the market demand curve.
3. The barriers to entry for a firm in a monopoly may be in the form of patent rights, legal protection, economies of scale and the well established long standing of the monopolist.
4. According to Pigou, there are three degrees of price discrimination practiced by the monopolists: (i) first degree price discrimination; (ii) second degree price discrimination; and (iii) third degree price discrimination.
5. Price discrimination is considered to be desirable when it adds to total social welfare.
6. When a monopolist sells an identical product at different prices to different buyers, it is called a discriminatory monopoly.
7. The concentration ratio is obtained by calculating the percentage share of the largest group of the firms in the total output of an industry.
8. According to Triffin's criterion, cross-elasticity is taken as the measure of degree of monopoly—the lower the cross-elasticity of the product of a firm, the greater the degree of its monopoly power.

6.7 SUMMARY

- Monopoly is a market situation in which there is a single seller of a commodity of 'lasting distinction' without close substitutes.
- A monopoly firm enjoys an absolute power to produce and sell a commodity.
- The emergence and survival of monopoly is attributed to the factors which prevent the entry of other firms into the industry. The barriers to entry are, therefore, the sources of monopoly power.
- The major sources of barriers to entry are: (i) legal restrictions, (ii) sole control over the supply of certain scarce and key raw materials; (iii) efficiency; (iv) economies of scale; and (v) patent rights.

- Since under monopoly, there is a single firm in the industry, the industry is a single-firm industry and industry demand curve has a negative slope.
- In the long run, a monopolist gets an opportunity to expand the size of its plant with a view to maximising its long-run profits provided the size of the market is fairly large. The expansion of the plant size may, however, be subject to such conditions as: (a) size of the markets; (b) expected economic profits; and (c) risk of inviting legal restrictions, like MRTP.
- When the same (or somewhat differentiated) product is sold at different prices to different set of consumers, it is called price discrimination.
- When a monopolist sells an identical product at different prices to different buyers, it is called a discriminatory monopoly.
- According to Pigou, there are three degrees of price discrimination practiced by the monopolists: (i) first degree price discrimination; (ii) second degree price discrimination; and (iii) third degree price discrimination.
- The monopolies are, in general, alleged to restrict production, consumption and employment, widen income and wealth disparities, exploit consumers and employees, distort resource allocation and reduce the social welfare.
- In most countries, therefore, government intervenes and makes laws to control and regulate monopolies to the advantage of the society.
- There are various measures—direct, indirect, price, non-price, legal and otherwise—to control and regulate the monopolies.

NOTES

6.8 KEY WORDS

- **Monopoly:** A monopoly is a market situation in which there is a single seller of a commodity of lasting distinction without close substitutes.
- **Natural Monopoly:** Monopolies born out of efficiency are known as natural monopolies.
- **Price Discrimination:** When the same (or somewhat differentiated) product is sold at different prices to different set of consumers, it is called price discrimination.

6.9 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short Answers Question

1. What is monopoly? How does existence of a close substitute affect the monopoly power?
2. What are the sources of monopoly?

NOTES

3. For a profit maximising monopoly, price is greater than marginal cost and it remains so over a large range of output. Why does then a monopolist not produce more than an output at which its $MC = MR$?
4. State the conditions under which price discrimination is possible and profitable as well.
5. Why should the government control a monopoly price? Under what conditions is monopoly economically desirable?

Long Answers Question

1. How is pricing under monopoly different from that under perfect competition? Can a monopoly firm fix any price for its product?
2. Explain the equilibrium of a monopoly firm in the short-run. Why is monopoly price always higher than the competitive price?
3. What is price discrimination? Explain and distinguish between the first, second and the third degrees of price discrimination. Which one makes a general case for price discrimination and why?
4. What is a discriminating monopoly? What are the conditions which help the monopolist in practising price discrimination?
5. What is meant by monopoly power? How is the degree of monopoly power measured?

6.10 FURTHER READINGS

- Dwivedi, D. N. 2002. *Managerial Economics*, 6th Edition. New Delhi: Vikas Publishing House.
- Keat Paul G. and K. Y. Philip. 2003. *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th Edition. Singapore: Pearson Education Inc.
- Keating, B. and J. H. Wilson. 2003. *Managerial Economics: An Economic Foundation for Business Decisions*, 2nd Edition. New Delhi: Biztantra.
- Mansfield, E., W. B. Allen, N. A. Doherty and K. Weigelt. 2002. *Managerial Economics: Theory, Applications and Cases*, 5th Edition. NY: W. Orton & Co.
- Peterson, H. C. and W. C. Lewis. 1999. *Managerial Economics*, 4th Edition. Singapore: Pearson Education, Inc.
- Salvantore, Dominick. 2001. *Managerial Economics in a Global Economy*, 4th Edition. Australia: Thomson-South Western.
- Thomas, Christopher R. and Maurice S. Charles. 2005. *Managerial Economics: Concepts and Applications*, 8th Edition. New Delhi: Tata McGraw-Hill.

UNIT 7 MONOPOLISTIC COMPETITION

*Monopolistic
Competition*

NOTES

Structure

- 7.0 Introduction
- 7.1 Objectives
- 7.2 Features of Monopolistic Competition
- 7.3 Chamberlin Approaches to Equilibrium
 - 7.3.1 Product Differentiation and the Firm's Demand Curve
 - 7.3.2 Selling Costs and the Firm's Cost Structure
 - 7.3.3 Concept of Industry and Product Groups
- 7.4 Equilibrium of the Firm and Group with Product Differentiation and Selling Costs
 - 7.4.1 Short-run Equilibrium of the Firm
 - 7.4.2 Long-run Equilibrium of Firms in Monopolistic Competition
- 7.5 Excess Capacity under Monopolistic and Imperfect Competition
 - 7.5.1 Ideal Output and Excess Capacity
 - 7.5.2 Non-price Competition and Excess Capacity
 - 7.5.3 Selling Cost and the Firm's Equilibrium: Non-price Competition
 - 7.5.4 Monopolistic Competition vs. Perfect Competition
- 7.6 Criticism of Monopolistic Competition
- 7.7 Answers to Check Your Progress Questions
- 7.8 Summary
- 7.9 Key Words
- 7.10 Self Assessment Questions and Exercises
- 7.11 Further Readings

7.0 INTRODUCTION

In the preceding units, you have learned about the theory of price and output determination under perfect (or pure) competition and monopoly—the two limiting and uncommon kinds of market structure. The theories of perfect competition and monopoly constituted the 'classical' microeconomic theory from Marshall to Knight and dominated the theory of value till the early 1920s. In the late 1920s and the early 1930s, however, economists expressed their dissatisfaction with perfect competition and pure monopoly models, as these models did not represent the real world business behaviour. In reality, there were very few monopolies because there were very few commodities for which there were no close substitutes, and there were very few perfectly competitive markets because there were a very few homogeneous products. In today's business world, therefore, perfect competition and monopoly represent only two extreme and uncommon market structures. Such as it is, theories of perfect competition and monopoly could be applied to only two small segments of the markets at two opposite extremes.

NOTES

Piero Shraffa was one of the first to point out the limitations of perfect competition and pure monopoly models. He was followed by Hotelling and Zeuthen who claimed (i) that neither perfect competition nor monopoly represent the real business world, and (ii) that most common markets fall between the perfect competition and the monopoly. It was in this background that Edward H. Chamberlin of Harvard University made a path-breaking contribution to the theory of value in 1933. Another important contribution to the theory of value, as an alternative to Chamberlin's theory, was made by Joan Robinson of the Cambridge University—six months later the same year. It is said that Chamberlin spent a lot of time and labour in pointing out the difference between his own and Joan Robinson's work. Chamberlin's theory is, however, considered to be superior for its, at least, two significant contributions: (i) he introduced firms' practice of 'product differentiation' and its effect on the demand curve and the theory of pricing, and (ii) he included selling cost in his analysis as an element of competition and its effect on firm's cost curves and equilibrium.

In this unit, you will learn about the basic elements of the theory of price and output determination under monopolistic competition as expounded by Chamberlin.

7.1 OBJECTIVES

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

- Discuss the nature and characteristics of monopolistic competition
- Describe how monopolistic competition differs from other kinds of markets
- Explain how price and output are determined by this category of firms in the short run and in the long run
- Examine how firms find their equilibrium when they incur selling cost, i.e., the advertisement expenditure

7.2 FEATURES OF MONOPOLISTIC COMPETITION

By definition, monopolistic competition refers to a market structure in which a large number of sellers sell differentiated products which are close substitutes for one another. Incidentally, a close substitute is one whose cross-elasticity is close to unity or greater. Monopolistic competition combines the basic elements of both perfect competition and monopoly.

The *element of monopoly* in monopolistic competition arises from the fact that each firm has an absolute right to produce and sell a branded or patented product. Other firms are prevented by laws from producing and selling a 'branded product' of other firms. This gives a firm *monopoly power* over production, pricing and sale of its own branded product. For example, consider toilet soap industry. There are a number of brand names available in the market—Lux, Liril, Hamam, Palmolive, Fairglow, Pears, Fa, Rexona, Lifebouy, Carmel, Godrej, Cinthol, Ponds,

OK, Dettol, etc. Each of these branded toilet soaps is produced and sold by a company having monopoly power over its product and price. Similarly, Maruti Udyog Ltd has monopoly power for producing and selling cars under the brand name Maruti and Maruti-Suzuki. No other car manufacturing company can produce and sell cars under this brand name.

The *element of competition* comes from the fact that each branded product is a close substitute for another and firms selling branded products of the same generic category have to compete for the market share. Considering again our example of toilet soaps. All the companies producing and selling these branded toilet soaps are in intensive competition for capturing the largest possible market share. One index of the competition between them is the amount that they spend on advertising their product. These features of the toilet soap industry make it monopolistically competitive. Toothpaste industry with a number of branded product names (Colgate, Close-up, Pepsodent, Forhans, Cibaca, Neem, Meswak, Signal, Promise, Prestige, etc.) is another example of monopolistic competition. So is the case with major industrial products in India, e.g., electrical gadgets, TV sets, refrigerators, air-conditioners, personal computers, textile goods, tea, coffee, cigarettes, soft drinks, cold creams, shampoos, detergents, shaving blades, shaving cream, hair oils, hair dyes, shoes, wrist watches, steel, cement, mobile phones, and so on.

Some of the industries looking monopolistically competitive may be oligopolistic in which there are only a few sellers. The question as to what makes a market monopolistically competitive or oligopolistic will be taken up later in the next unit which deals with *oligopoly markets*. Let us now look at the general characteristics of monopolistic competitions.

General Characteristics of Monopolistic Competition

As mentioned above, monopolistic competition combines the elements of both perfect competition and monopoly power with, of course, product differentiation. Therefore, the main characteristics of monopolistic competition are a blend of perfect competition and monopoly. The main features of monopolistic competition vis-à-vis perfect competition and monopoly are described below.

1. Product differentiation: Under monopolistic competition, the firms differentiate their products from one another in respect of their shape, size, colour, design, minor qualitative differences, efficiency in use, some extra facility, packaging, after-sale service, guarantee and warranty, etc. Product differentiation may be real or fanciful and spurious. Product differentiation is, in fact, the basis of and the main distinctive characteristic of monopolistic competition that distinguishes it from monopoly and perfect competition. In case of monopoly, there is only one product and only one seller, and under perfect competition, a large number of sellers sell a *homogeneous product*. The basic purpose of product differentiation by monopolistic firms is to make the consumers believe that a product is different from others and, thereby, to create brand loyalty of the consumers. Product

NOTES

NOTES

differentiation affects firm's demand curve in a significant way, as discussed ahead in Section 7.4.

2. Large number of sellers: Under monopolistic competition, the number of sellers is *large*. How large? It is difficult to specify number of firms: it may be anywhere 10-20 or even more depending on the size of the market. However, the question 'how large' can be answered in conceptual terms with reference to perfect competition. Under perfect competition, the number of sellers is so large that a firm becomes a *price taker*. In contrast, under monopolistic competition, the number of firms is only so large that a firm retains its power to be a *price maker*. The monopolistically competitive firms have the power to set the price of their product depending on the objective of the firm and the level of competition.

3. Free entry and free exit: As in case of perfect competition, there is no barrier on the entry of new firms and exit of old ones from the industry. New firms are free to enter the monopolistically competitive industry and to quit at will. Entry of new firms reduces the market share of the existing ones and exit of firms does the opposite. These consequences of free entry and free exit lead to intensive competition among the firms for both retaining and increasing their market share.

4. Selling costs: Unlike firms under perfect competition and monopolies, firms under monopolistic competition make heavy expenditure on advertisement and other sales promotion schemes for their product. This is an important feature that distinguishes monopolistic competition from perfect competition and monopoly. *Selling costs* include all the expenditure on advertisement, sales promotion schemes, and salaries of sales personnel. Selling costs and their effect on firm's equilibrium will be discussed later in detail.

5. Downward sloping demand curve: As in case of monopoly, a monopolistically competitive firm faces a downward sloping demand curve. The reason is that a monopolistically competitive firm can, by exercising its monopoly power, increase its price and still retain some buyers with brand loyalty and can increase the demand for its product by decreasing the price because of a relatively higher cross-elasticity of the competitive product.

7.3 CHAMBERLIN APPROACHES TO EQUILIBRIUM

Let us now discuss Chamberlin's theory of monopolistic competition. For a better comprehension of his theory, it is essential to understand the implications of the following three basic elements of monopolistic competition as defined by Chamberlin.

- (i) Product differentiation and firm's demand curve;
- (ii) Selling costs and firm's cost structure, and
- (iii) Product differentiation and the concept of industry.

These aspects of monopolistic competition are briefly discussed below.

7.3.1 Product Differentiation and the Firm's Demand Curve

Product differentiation is the basis of competition among the monopolistically competitive firms. 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 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, of 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.'

Thus, the basic purpose of product differentiation is to make customers distinguish the product of a firm from those of other firms of the industry and to develop a preference or brand loyalty. Once brand loyalty is developed, it alters the course of the demand curve for the product. In the ultimate analysis, *product differentiation aims at changing the slope and position of the demand curve for the product and from converting it from a horizontal demand line (as under perfect competition) to a downward sloping demand curve*. The downward sloping demand curve gives the firm power to use his discretion in changing the price of his product.

What is more important in regard to inter-firm competition, with product differentiation, is that each firm perceives that the demand curve for its own product is more elastic than that of the rival firms. This aspect is illustrated in Figure 7.1. Suppose industry's demand curve (i.e., market demand curve) is given by the curve D_M . If industry demand is proportionately divided between the firms, each firm is supposed to have a demand curve shown by D_A .

However, firms under monopolistic competition do not take it to be the demand curve for their individual product. Each firm perceives that the demand curve for its own product is *more elastic* than that of the other firms. Its *perceived demand curve* is shown by the demand curve D_p . Note that the *perceived demand curve*, D_p , is more elastic than the actual demand curve D_A . The basis of the perceived demand curve is the firm's belief that if it changes the price of its own product, it will go unnoticed by the other firms and they will not react to change the price of their products. The reason for this assumption is the firm's belief that the number of firms is so large that price changes made by a single firm is very much likely to go unnoticed by the rival firms.

NOTES

NOTES

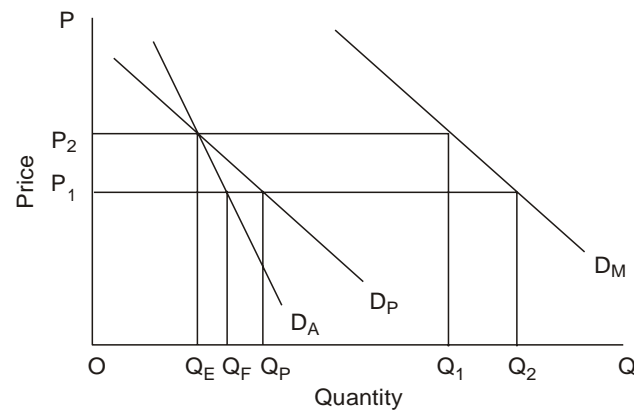


Fig. 7.1 Demand Curves for the Industry and the Firms

The concept of the *perceived demand curve* can be explained as follows. Given the industry demand curve D_M in Figure 7.1, if price for the industry (the ‘product group’) is given at OP_2 , the demand for the industry as a whole will be OQ_1 , each firm selling an equal quantity OQ_E . Given this price-quantity combination, an individual firm perceives that if it cuts down the price of its own product to OP_1 , the other firms will not change their price, then the demand for its product will increase by Q_EQ_P , not by Q_EQ_F . The additional increase in the demand for its product is the result of cross elasticity, i.e., when one firm decreases its price and other firms do not, its product becomes relatively cheaper. Since products of all the firms are close substitutes for one another, some customers switch over from the constant-price products to the product whose price is cut down. Similarly, if only one firm increases its price, it loses its customers to other firms. As a result, demand for its product decreases more than indicated by the demand curve D_A . This holds for all price changes and gives rise to a perceived demand curve D_P . The perceived demand curve plays a significant role in price and output determination in Chamberlin’s theory of monopolistic competition with price competition. This aspect is discussed further in a following section.

7.3.2 Selling Costs and the Firm’s Cost Structure

Introduction of selling costs in the theory of price and output determination under monopolistic competition is another innovative contribution made by Chamberlin. He defines ‘selling costs’ as ‘costs incurred in order to alter the position or the slope of the demand curve for a product.’ Chamberlin’s concept of selling costs is not exactly the same as advertisement cost: it is advertisement cost plus. By Chamberlin’s definition, selling costs include:

- (i) Cost of advertisement
- (ii) Expenditure on sales promotion schemes (including gifts and discounts to buyers)
- (iii) Salary and commission paid to sales personnel

- (iv) Allowance to retailers for displays
- (v) Cost of after-sale-services

Also, Chamberlin distinguishes *selling costs* and *production cost* on the basis of their basic purpose and functions. According to Chamberlin, costs that are incurred to create a product or service of utility and making it available to the consumers are *production costs*. In Chamberlin's perception, production cost includes also the cost of transportation. The basic function of the production cost is to create a commodity and to make it available to the consumers. The selling costs, on the other hand, is aimed at the following functions:

- (i) Informing potential buyers about the availability of the product
- (ii) Increasing demand for the product by attracting customers of the rival products
- (iii) To make the demand curve shift upward

What is more important in price and output determination is the effect of selling costs on the total cost that figure in pricing decisions. In his model of monopolistic competition, Chamberlin assumes the traditional U-shaped cost curves—*AC*, *AVC* and *MC*—and also a U-shaped *average selling cost (ASC)* curve. *ASC* is defined as SC/S (where *SC* = *selling costs* and *S* = *quantity sold*). The U-shaped *ASC* curve is illustrated in Figure 7.2.



Fig. 7.2 The Average Selling Cost Curve

As the figure shows, the *ASC* first decreases until it reaches its minimum and then begins to increase. In the beginning, it is very high because a little selling cost proves to be ineffective. With increase in selling costs, sales increase at a rate higher than the rate of increase in selling costs, i.e., $\Delta SC/\Delta S < 1$. As a result, *ASC* decreases. The decrease in *ASC* is attributed to 'increasing returns' to advertising and *economies of scale* in advertisement cost. It must, however, be noted that returns to selling cost is determined by the following factors:

- (i) *Price of the product*—a high price makes selling cost less effective;
- (ii) *Price of the substitute*—a lower price of the substitutes makes selling cost less productive;

NOTES

NOTES

- (iii) *Buyers' income*—advertising a costly product in low-income society has no pay-off; and
- (iv) *Buyers' loyalty to rival brands*—the stronger the loyalty, the lower the cross-elasticity and the less effective the selling cost.

However, even if all these factors are favourable, a stage is finally reached when returns to selling costs, especially to advertisement cost, tends to become constant. This is the stage of saturation. The stage of saturation may be marked by the size of the market and/or competitive advertisement by the rival firms. Increasing selling costs or advertisement cost at the stage of saturation tends to become less and less effective in attracting more buyers. Therefore, sales increase at a much lower rate than the increase in the selling costs, i.e., $\Delta SC/\Delta S > 1$. Consequently, ASC begins to increase and goes on increasing. This is how ASC gets its U-shape. The ASC curve is added to the AC curve in determining the profit maximising level of output and price. This aspect is discussed in the next section.

What is the optimum level of advertising cost? Given the nature of the ASC curve (in Figure 7.2), a question arises: What is the optimum level of advertising cost? The optimum level of advertising cost is determined by the objective of the firm whether it is profit maximisation, retaining market share or countervailing the advertisement by the rival firms. Optimisation of advertising expenditure is illustrated in Figure 7.3 under the following assumptions.

- (i) Objective of the firm is to maximise its profit;
- (ii) Price of the product is given;
- (iii) Average production cost (APC) and MPC curves are given; and
- (iv) Average selling cost (ASC) is also known.

Under these conditions, the optimum level of selling cost is determined where the firm's overall marginal cost (MC) including marginal cost of production (MCP) and marginal cost of advertising equals the price. Since price is given by assumptions, $price = MR$. This is illustrated in Figure 7.3.

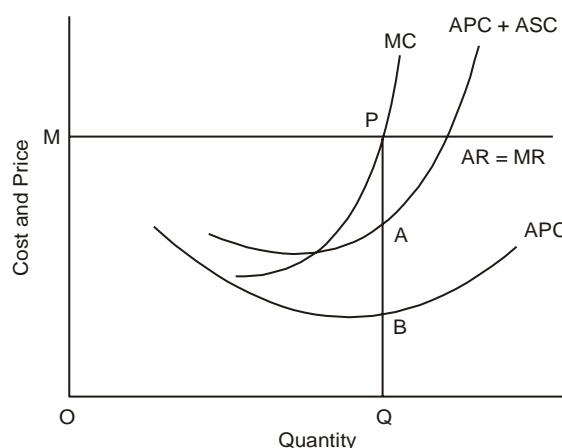


Fig. 7.3 Optimization of Selling Cost

The price of the product is given at OM and firm's APC is shown by the curve APC . The firm's APC curve added vertically with ASC curve (not given in the figure) is shown by the curve labelled $APC + ASC$. The vertical distance between the APC and $APC + ASC$ measures the average selling cost (ASC). Finally, firm's overall MC (associated with $APC + ASC$) is shown by the MC curve.

As Figure 7.3 shows, price line, $AR = MR$, and MC curve intersect at point P . An ordinate drawn from point P to the quantity axis determines the profit maximizing output at OQ . Once profit-maximizing output is determined, the optimum level of all costs (given the cost curves) is automatically determined. It can be seen in Figure 7.3 that at profit maximizing output OQ , price equals $PQ = BQ + AB + AP$. Note that at output OQ , average selling cost (ASC) equals AB . That is, at the profit maximizing level of output, average advertising cost equals AB . Therefore, AB is the *optimum average selling cost (ASC)*. The total optimum selling cost or advertisement expenditure can be obtained by multiplying ASC with output OQ . That is, *total optimum advertisement expenditure* = $OQ \times AB$.

7.3.3 Concept of Industry and Product Groups

An industry in a perfectly competitive market is defined as a group of firms producing a homogeneous product. But, this concept of industry cannot be applied to the cases where products are differentiated. Where products are differentiated—slightly or substantially—each firm is, in a sense, an industry in itself, exactly as a monopoly firm is an industry in itself. The heterogeneity of the products, therefore, causes a problem in analytical treatment of the industry. Problem arises because *industry demand curve under monopolistic competition is not the same as under perfect competition*. Since under perfect competition, product is homogeneous, demand curve for an industry can be obtained by adding demand curves of individual firms. But, under monopolistic competition, product is made heterogeneous through product differentiation techniques and in case of heterogeneous products, the demand for individual products cannot be added to obtain market demand and supply curves.

For the reason given above, Chamberlin defines the monopolistically competitive industry as a *group* of firms producing a *closely related* commodity, called *product group*. The products of the *group* must be *close, technological and economic substitutes*. The two products are *technological substitutes* for each other if they technically satisfy the same want, e.g., soaps, cigarettes, toothpastes, automobiles, TV sets, etc. The two products are considered as *economic substitutes* for each other when they satisfy the same want and have *more or less* the same price. For example, all non-AC cars, of different brands with comparable price are economic substitutes for one another. But non-AC cars are not economic substitutes for AC cars since their prices are widely different but they are technological substitutes.

NOTES

NOTES

Operationally, the product group may be defined as the group of firms whose products have between themselves high *price elasticity* and high *cross-elasticities*. This definition, although theoretically plausible, involves the problems of measuring cross-elasticities and determining its degree that can make a commodity admissible to the group. Determining the product group, therefore, involves subjective judgement.

Check Your Progress

1. From where does the element of monopoly in monopolistic competition arise?
2. From where does the element of competition in monopolistic competition arise?
3. What kind of demand curve does a monopolistically competitive firm face?
4. What does product differentiation aim at?

7.4 EQUILIBRIUM OF THE FIRM AND GROUP WITH PRODUCT DIFFERENTIATION AND SELLING COSTS

In this section, we discuss Chamberlin's theory of price and output determination under monopolistic competition and show how firms in monopolistic competition find their equilibrium. Chamberlin prefaces his theory with a comparison of *price and output choices* under pure competition and monopoly and under monopolistic competition. He points out that under pure competition, firms selling homogeneous products are given a price at which they can sell any quantity they desire. Advertising by the individual firm is of no consequence: it is rather a waste. A monopoly facing a downward sloping demand curve has the discretionary power in respect of price and quantity to be sold. However, given the demand curve, it can choose either price or output, not both. Advertising by a pure monopoly is of little consequence. Under monopolistic competition, however, a firm can alter its sales prospects by the following *three methods*:

- (i) By changing the price of its product
- (ii) By changing the nature of the product
- (iii) By incurring the advertisement outlays

As to changing price, since a firm under monopolistic competition faces a downward sloping demand curve with elasticity less than infinity, it has the option to change the price. In regard to changing the nature of the product, a firm can do it by changing the quality of its product by making technical changes, introducing a

new design, use of superior material, by a new style of packaging, by establishing a close link with buyers, and so on. Besides, a firm can increase its sales by prompt and courteous service, credit facilities, and by enhancing expenditure on advertisement. While making changes in price and output is a short-run phenomenon, changing the quality of the product and attracting larger number of buyers are long-run phenomena. Therefore, Chamberlin's theory of price and output determination is discussed under *short-run* and *long-run* conditions.

Assumptions: Chamberlin has made the following *explicit* and *implicit* assumptions to develop his theory of monopolistic competition.

- (i) There is a large number of firms selling somewhat differentiated products which are close substitutes for one another.
- (ii) The number of firms in a product group is so large that their activities, especially, manoeuvring of price and output, go unnoticed by the rival firms.
- (iii) Demand and cost curves for all the products and for all the firms of the group are uniform, i.e., firms face identical demand (including *perceived* one) and cost curves.
- (iv) Consumer's preferences are evenly distributed among the different products and difference in products is not such that they make a difference in cost.

The last two assumptions are called rather 'heroic' in the sense that these are somewhat unrealistic assumptions. However, it can be assumed, for theoretical convenience, that the differences, wherever they are, are not significant enough to influence the price and output decisions of the rival firms. Given the assumptions, let us discuss first the *short-run equilibrium of the firm*. Firm's *long-run equilibrium* will be discussed in the next section.

7.4.1 Short-run Equilibrium of the Firm

The *final short-run equilibrium* position of the monopolistically competitive is illustrated in Figure 7.4. In this figure, the firm's perceived demand curve and the corresponding marginal revenue curve are shown by OP and MR curves, respectively, and its short-run average and marginal cost curves are shown by the SAC and SMC curves, respectively.

As Figure 7.4 shows, the MR and SMC curves intersect at point E which determines the profit maximising output at OQ and price at P_e . Each firm produces OQ and sells at price PQ . At this price and output, the firms maximise their short-run profit. The firms are, therefore, in short-run equilibrium at point E . Their maximum pure profit equals $BC \times PC = P_e PCB$ and is shown by the shaded area. Any other price and output will reduce their total profit. Therefore, firms have no incentive to change their price and output. Note that the final price is determined at the point of intersection between D_p and D_A curves.

NOTES

NOTES

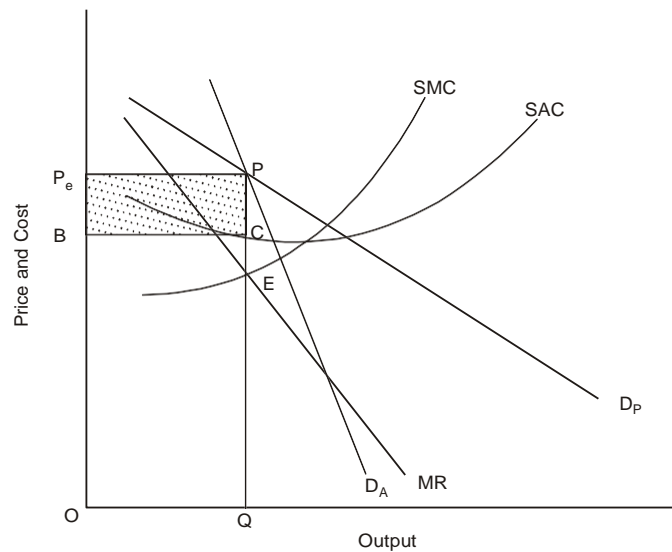


Fig. 7.4 Short-run Equilibrium of the Firm under Monopolistic Competition

This, however, should not mean that all the firms in monopolistic competition make pure profits in the short-run. Chamberlin does not rule out the possibility of some firms making losses.

7.4.2 Long-run Equilibrium of Firms in Monopolistic Competition

In the long run, the following changes take place in monopolistic competition.

- (i) New firms enter the industry
- (ii) Firms begin to indulge in price competition
- (iii) Firms make vigorous advertisement with a view to enhancing their market share or to retain their market share especially when there is entry of new firms into the industry or the product group

These changes affect the market share of the rival firms which leads to intensive competition. Chamberlin has used three models to analyse the effects of these changes on firm's equilibrium. His three models present, in fact, the three different stages of analysis. In this *first model*, he analyses firm's equilibrium with free entry of new firms into the industry, all other things given. His *second model* analyses the effect of price competition between the firms on their equilibrium, with no entry or exit. His *third model* presents a combined analysis of free entry and price competition on the firm's equilibrium. These models are discussed below in the same sequence.

Model I: Long-run equilibrium with free entry of new firms: The long-run equilibrium of firms with free entry of new firms, other things remaining the same, is illustrated in Figure 7.5. The *LAC* and *LMC* curves represent the firms' cost curves. The initial *AR* and *MR* curves, (i.e., prior to the entry of new firms) are given by $AR = D_2$ and MR_2 , respectively. Given the cost and revenue curves, the

firms will be in short-run equilibrium at price OP_2 and output OQ_2 . The firms are making super-normal profit to the extent of EB per unit of output.

The existence of super-normal profit 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 (i.e., AR curve) for the old firms shifts downward. Therefore, firms adjust their price and output to the new AR and MR conditions, cost conditions remaining the same. This process continues until the demand curve is tangent to the LAC , as shown by the demand curve $AR = D_1$ 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, thus, be stable at point A .

NOTES

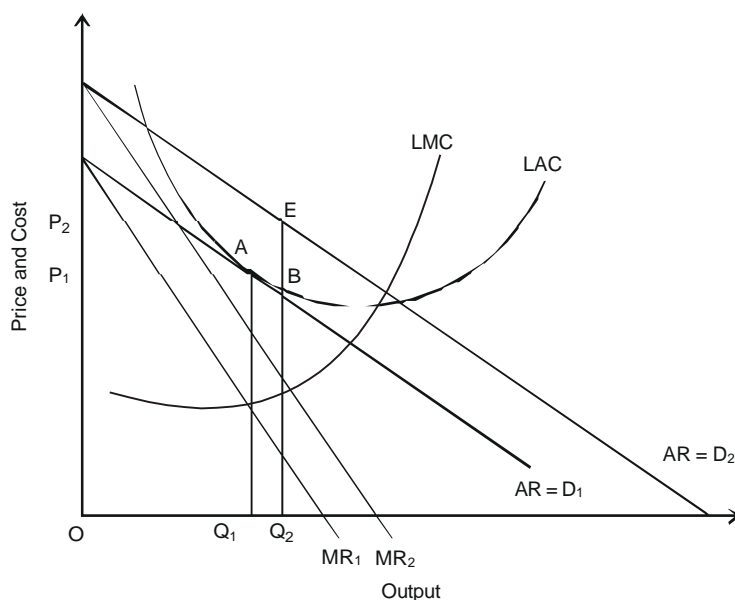


Fig. 7.5 Long-run Equilibrium with Free Entry

Model II: Long-run equilibrium with price competition: In his model of long-run equilibrium with price competition, Chamberlin assumes (i) that the number of existing firms in the product group is optimal, i.e., the number of firms is compatible with long-run equilibrium of the industry; and (ii) there is no entry or exit of the firms. Given these assumptions, he analyses the long-run equilibrium with price competition in two stages. In the first stage, he traces the expected demand curve resulting from the change in price, assuming that rival firms do not react to the price changes. In the second stage, he traces the long-run equilibrium under the conditions of optimal number of firms and price competition.

The derivation of the perceived demand curve is shown in Figure 7.6. Let us assume that demand curve for the product of all the firms is identical and is given by DD_1 and that all the firms are in equilibrium at point E , with price OP and

NOTES

output OQ . Now suppose, that one of the firms contemplates a price reduction and assumes that the rival firms will not react to its price reduction. By reducing its price, the firm can expect 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. This is *substitution effect*. Thus, the firm can expect a substantial increase in its total sales. For example, if the firm *reduces* its price by EM , it expects the demand for its product to increase by MB . The MN part of the incremental demand MB is attributed to the elasticity of its original demand DD_1 and NB is due to the substitution effect. Similarly, if the firm *raises* its price, and the rival firms do not raise their prices, it loses its market partly because its demand curve is elastic and partly because its customers switch over to other products which become automatically cheaper. Thus, the firm expects a different and more elastic demand curve for its product, as shown by dd_2 . This is called as the *perceived 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 .

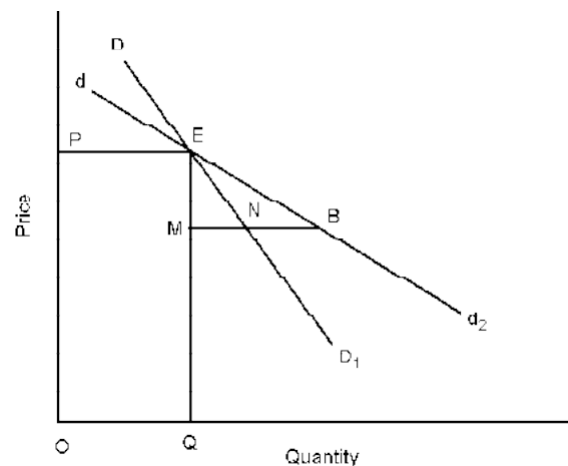


Fig. 7.6 Introduction of Perceived Demand Curve

Having introduced the perceived demand curve, let us now replicate Chamberlin's analysis of long-run equilibrium of a firm with price competition. The long-run equilibrium with price competition is presented in Figure, 7.7. The curves DD' and dd_1 are the firm's normal and perceived demand curves, respectively, 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 where price is OP_2 and output is OQ_1 . At this price and output, the firm makes the total supernormal profit represented by the area $CMPP_2$.

Let us now introduce price competition and analyse its effect on firm's equilibrium. Competition begins with the firms' belief that DD' is market (or group) demand curve and dd_1 is the demand curve for their own product. Note that dd_1 is more elastic than DD' . Therefore, each firm thinks that it can increase its profit

by cutting down its price and, thereby, increasing the demand for its product. In an attempt to increase profits, each firm reduces its price expecting to move along the demand curve dd_1 . But, instead of moving along dd_1 , the firms move along the market demand curve DD' because all of them reduce their prices simultaneously.

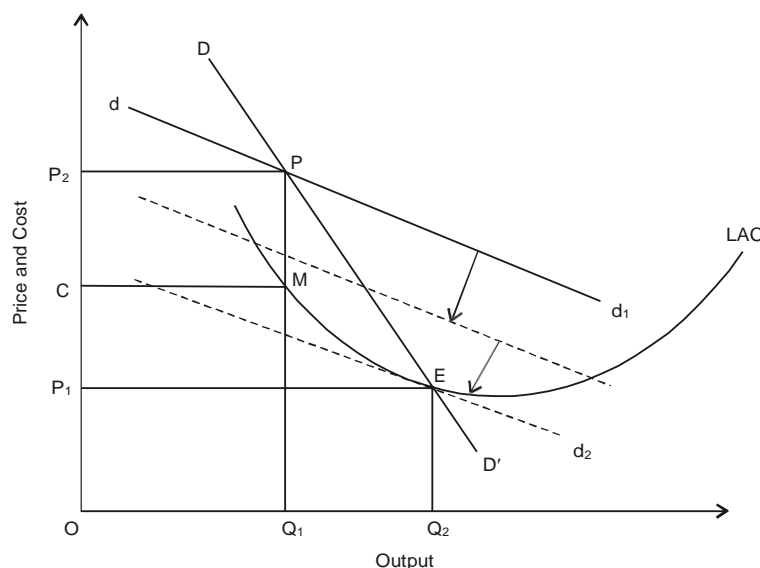


Fig. 7.7 Long-run Equilibrium: Price Competition

Why do firms not realise this phenomenon? According to Chamberlin, the firms do not learn from their past experience and each firm sticks to its own belief that price-reduction will make the demand curve for its product more elastic than the market demand curve (DD'). Therefore, the firms go on reducing their prices independently and their assumed demand curve (dd_1) slides downward as shown by the arrow. This process of downward shift in dd_1 continues until it is tangent to the LAC curve at point E , as shown by dd_2 in Figure 7.7. Here equilibrium price is OP_1 . A further reduction in price will make firms incur losses. Therefore, reduction in price below OP_1 is not desirable. Thus, long-run equilibrium of firms takes place at E , where each firm produces OQ_2 and fixes its price at OP_1 .

Model III: Long-run equilibrium with free entry and price competition:

We have explained above separately the equilibrium of the firm with free entry and with price competition. In this section, we combine the two equilibrium analyses and explain Chamberlin's third and the final model of equilibrium of monopolistic competition.

According to Chamberlin, the ultimate equilibrium under monopolistic competition can be achieved through an integrated analysis of effects of free entry and price adjustments. The integrated analysis of equilibrium is presented in Figure 7.8. In this figure, 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 , where output is OQ_5 and price is BQ_5 . Here firms make abnormal profits per unit of sale to the extent of BF . Since entry to the product

NOTES

NOTES

group' is free, new firms are attracted to the industry by the abnormal profit. When new firms enter the product group with slightly differentiated products, the share of each existing firm in the market 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 at point A , as shown by DD_3 in Figure 7.8. For, till this point is reached, firms will make abnormal profits and new firms will continue to enter the 'product group'.

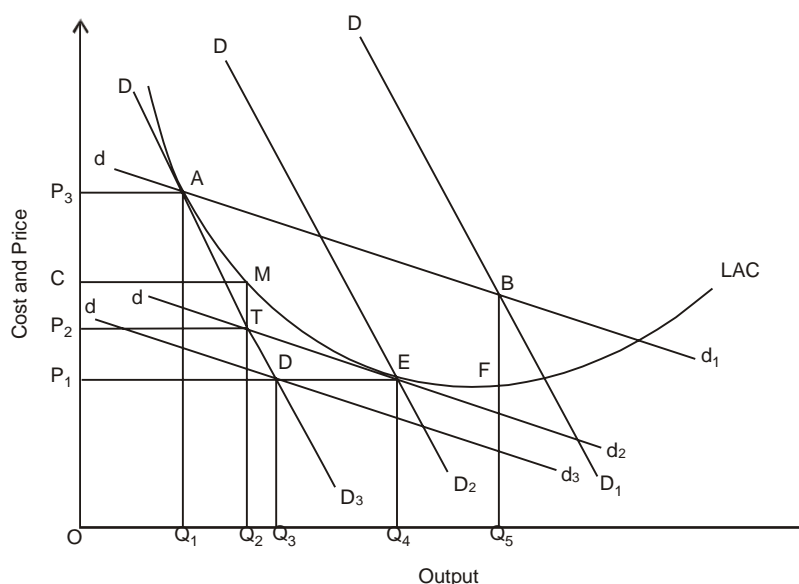


Fig. 7.8 Long-run Equilibrium with Free Entry and Price Competition

It appears from this part of analysis that the long-run equilibrium is attained at point A with output OQ_1 and price OP_3 . This is however not the case. This is only half of the story, confined to the influence of free entry on firms' equilibrium. Let us now consider the competitive manoeuvring of price and its role in determining the long-run equilibrium. When abnormal profits disappear, firms get levelled up at normal profits. Here begins the price-war. As shown above, 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 one after another because each firm has the same incentive to do so. As a result, firms move down along DD_3 whereas they had expected to move down along dd_1 . In this process, 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 , their subjective demand curve dd_2 slides further down to dd_3 , the dotted line. As a result, the firms make ever-increasing losses. A temporary equilibrium will be attained at point D with output OQ_3 . Here all firms incur heavy losses. Consequently, the firms which are unable to sustain losses, eventually leave the industry. The remaining firms find

their market share increasing. Therefore, demand curve DD_3 moves to the right and dd_3 shifts upward. This continues until DD_3 shifts to DD_2 and dd_3 shifts upward to the position of dd_2 and intersects DD_2 at point E . Thus, the long-run equilibrium is attained at point E which, according to Chamberlin, is stable because all firms earn only normal profits and, therefore, there is no entry or exit of the firms.

The final version of Chamberlin's model given in Figure 7.8 is highly complicated. A simplified version of the model is presented in the following subsection-Model IV.

Model IV: Equilibrium of firms in the long-run: The final version: The final version of Chamberlin's model of long-run equilibrium of firms under monopolistic competition is presented in Figure 7.9. In fact, Figure 7.9 is a reproduction of Figure 7.8 eliminating the process related to shifts in demand curves. Recall that with the entry of new firms, the existing firms lose a part of their markets to the new entrants. As a result, the actual demand curve (D_1) of the firms shifts leftward and perceived demand curve (d_1) shifts downward. This shift continues until pure profit disappears. This status of firms is shown by point P in Figure 7.9. At point P , demand curve D_P is tangent to LAC curve and demand curve D_F intersects with them. Since price equals LAC at point P , no firm is making pure profit. There is, therefore, no incentive for new firms to enter the industry nor for the existing ones to quit the industry. The firms are, therefore, in the state of their long-run equilibrium at point P . It is important to note that point P is not coincidental. It is determined by the intersection of the MR and MC curves determining profit maximising output at OQ_A and price at PQ_A . In equilibrium, firms produce and sell quantity OQ_A at price $PQ_A = OP_E$. This is how, according to Chamberlin, monopolistically competitive firms reach their equilibrium in the long-run. An important outcome of this analysis is that, in the long-run, all the firms in monopolistic competition make only normal profit.

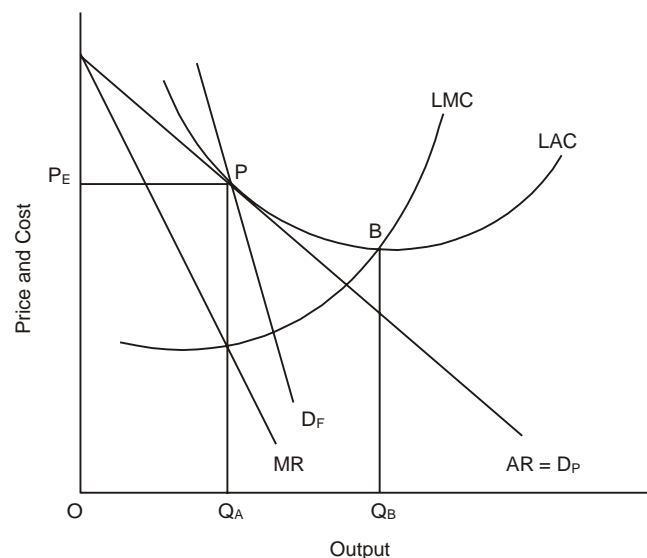


Fig. 7.9 Long-run Equilibrium of Firms in Monopolistic Competition

NOTES

7.5 EXCESS CAPACITY UNDER MONOPOLISTIC AND IMPERFECT COMPETITION

NOTES

As in case of monopoly, the existence of excess capacity is an important feature of monopolistic competition. The excess capacity or unutilized capacity is a social waste which has attracted the economists' attention. In this section, we will discuss

- (a) The concept and measure of 'ideal output'
- (b) Non-price competition and excess capacity

7.5.1 Ideal Output and Excess Capacity

The 'excess capacity' of a firm is defined as the difference between the 'ideal output' and the 'actual output' attained in the long-run. What is 'ideal output'? The economists from Marshall down to Kahn, Harrod, and Cassels defined *ideal output* as the output that can be produced at the minimum *long-run average cost (LAC)*. This concept of 'ideal output' is linked to social optimality of production. *Excess capacity* is also called as 'idle capacity' and 'unused capacity'.

The existence of excess capacity under monopolistic competition can be seen in Figure 7.9. The firm is in equilibrium at point *P*. Its *actual output* at equilibrium is OQ_A . As regards the 'ideal output', note that *LMC* intersects *LAC* at point *B* which marks, as a matter of rule, the minimum point at the *LAC* curve. Point *B* determines the 'ideal output' at OQ_B . Thus 'actual output' is OQ_A and 'ideal output' is OQ_B . The difference between the actual output OQ_A and the ideal output OQ_B equals Q_AQ_B . Thus, Q_AQ_B is the 'excess capacity' under monopolistic competition.

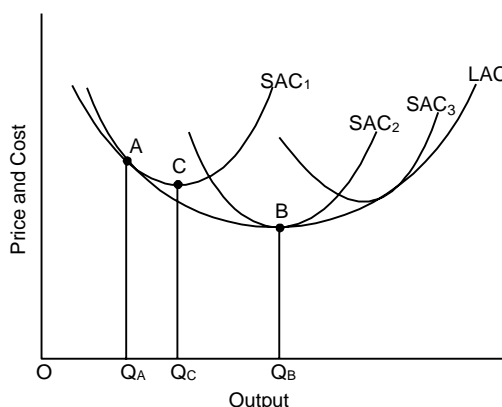


Fig. 7.10 Ideal Output and Excess Capacity

Cassels has divided excess capacity in two parts on the basis of what is optimum from individual firm's and society's points of view. The two parts of the excess capacity are illustrated in Figure 7.10. Suppose a firm is optimally placed at point *A* (where SAC_1 is tangent to its *LAC*) whereas its ideal output is given by

point C . The firm produces OQ_A whereas ideal output is OQ_C . The difference between OQ_A and OQ_C , i.e., $Q_A Q_C$ is the *first part* of the excess capacity.

The **second part** of excess capacity is associated with what is socially optimum size of the firm. A firm in monopolistic competition can expand its scale of production to point B which marks the minimum of both SAC (i.e., SAC_2) and LAC . That is, socially optimum output is given by point B at OQ_B . But a firm in monopolistic competition does not use enough resources of the society to produce at point B . Assuming that the firm produces at point A , the firm leaves out the prospects producing an additional output $Q_C Q_B$. This is the *second part* of the excess capacity.

NOTES

7.5.2 Non-price Competition and Excess Capacity

According to Chamberlin, excess capacity in monopolistic competition arises in the long-run when there is free entry but *absence of active price-competition* between the firms. It implies that if monopolistic competition is characterised by free entry and active price competition, there would be no excess capacity. Chamberlin attributed the excess capacity to the absence of active competition. Firms may abstain from active price competition for such policy reasons as — formal or tacit agreements, open price association, trade association activities building up an *esprit decorps*, — ‘price maintenance’, the imposition of uniform prices on dealers by manufacturers, and excessive differentiation of the product in the attempt to turn attention away from price’. The other factors he added are — ‘customary prices’, — ‘professional ethics’, fear that price cutting might give a wrong signal to the consumers about quality of the product, and avoiding problems arising out of price competition.

Chamberlin’s view that excess capacity is the result of non-price competition is illustrated in Figure 7.11. As usual, the LAC curve represents the *long-run average cost* curve which is the same for all the firms and their *perceived* and *proportional* demand curves are shown by D_p and D_F curves. If there is *free entry with price competition*, the firms will be in equilibrium at point P where the firms’ perceived demand curve D_p is tangent to the LAC curve and their proportional demand curve D_F intersects both the curves D_p and LAC . At equilibrium point P , firms produce OQ_C at price OP_C . The reason for the firms being in equilibrium at point P is given in Figure 7.8. If firms adopt an aggressive price competition, point P will shift closer to point M , the point of perfectly competitive equilibrium.

Let the *price competition now be eliminated from the monopolistic competition*. If there is no price competition, the *perceived demand curve* D_p has no existence. The firms are left with their *proportional demand curve* D'_F . In the absence of active price competition, entry of new firms makes the demand curve D_F shift leftward because market share of each firm decreases. New firms continue to enter the industry until their demand curve D_F shifts to the position of D'_F which is tangent to the LAC curve. Any further entry of firms will make the D'_F

NOTES

shift further to the left and away from the LAC curve. In that case, all the firms will make losses forcing some of the firms to quit the industry.

Therefore, entry of new firms stops at point E . Therefore, point E marks the point of long-run equilibrium where long-run output is OQ_A and price is $EQ_A = OP_E$.

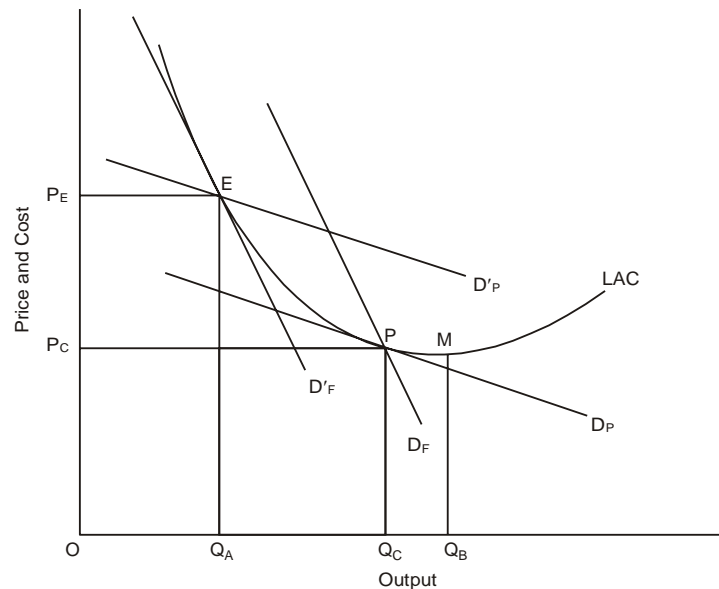


Fig. 7.11 Long-run Equilibrium: Non-price Competition and Excess Capacity

Now, by comparing equilibrium output OQ_C corresponding to *free entry with active competition*, and equilibrium output OQ_A corresponding to *free entry with no active price competition*, we get Chamberlin's measure of excess capacity: it is obviously $Q_A Q_C$. Alternatively, Chamberlin's excess capacity can be obtained as $OQ_C - OQ_A = Q_A Q_C$.

According to Chamberlin, 'excess capacity' may exist even 'under pure competition, owing to mis-calculation on the part of the producers or to sudden fluctuations in demand or cost conditions'. But in the long-run, excess capacity 'is the peculiarity of the monopolistic competition' which is a 'waste' caused by the 'monopoly element in monopolistic competition' and for it 'there is no automatic corrective'.

7.5.3 Selling Cost and the Firm's Equilibrium: Non-price Competition

The various aspects of the selling cost including meaning and purpose of selling costs, returns to the selling costs, the nature of average selling cost (ASC) curve, its effect on the total cost (i.e., average production cost *plus* selling costs), and optimization of selling cost, as enunciated by Chamberlin, have already been discussed in detail in section 7.4.2 of this unit. Here, we discuss the effect of

selling costs on the equilibrium of the firms in monopolistic competition. Chamberlin has developed his theory of value under monopolistic competition with selling costs at two levels: (i) the case of ‘Individual Equilibrium’, i.e., the equilibrium of an individual firm assuming other firms do not advertise their products, and (ii) the case of ‘Group equilibrium’, i.e., when all the firms adopt a competitive advertising. We present here a brief description of his analysis of ‘Individual Equilibrium’ and ‘Group Equilibrium’.

Individual Equilibrium

Chamberlin has enunciated equilibrium of an individual firm under the following three conditions: (a) determination of the optimum selling costs with price given; (b) determination of equilibrium price with a given amount of selling cost, and (c) determination of price and selling costs simultaneously. Case (a) is already illustrated in Figure 7.3. Given our limited purpose here, we will confine to case (c) in which a firm facing a downward sloping demand curve and U-shaped ASC curve has to find equilibrium price and selling costs. The equilibrium of the firm under these conditions is explained below under the following assumptions:

- (i) The firm seeks to maximise its profit
- (ii) It is free to adjust its price, output, and selling cost to this end
- (iii) ‘Price, products, and selling policies of all competition’ are given
- (iv) There is no interdependence between the firms

The equilibrium of the individual firm is illustrated in Figure 7.12. The firm’s downward sloping demand curve is given by the curve $D (= AR)$ and its marginal revenue curve by the curve MR . Its average production cost curve is given by the curve APC and its overall cost is shown by the curve labelled $APC + ASC$ and its combined marginal cost by the curve MC_c . Firm’s MR and MC_c curve intersect at point M determining the equilibrium output at OQ . Given the demand curve, equilibrium price is determined at $OP = EQ$. That is, firm’s profit maximizing output is OQ and price is OP . At the equilibrium level of output, its average selling cost (ASC) equals ED and its total selling cost equals $ED \times OQ$. Since $OQ = SD$, firm’s total selling cost equals $ED \times SD = SDEP$. This is optimum selling cost. The total revenue of the firm equals output \times price = $OQ \times OP = OPEQ$. Since firm’s demand curve and overall average cost curves are tangent to one another at point E , selling price equals the overall average cost, i.e., $APC + ASC$. The firm is therefore making only normal profit. This is the general case of firm’s individual equilibrium in isolation. However, if firm’s demand curve lies above the $APC + ASC$ curve, the firm will make pure profit, but only in the short-run.

NOTES

NOTES

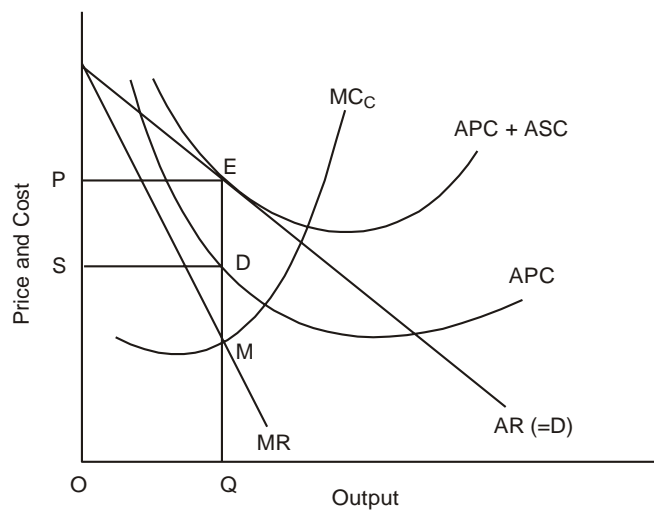


Fig. 7.12 Firm's Equilibrium with Variable Price and Selling Costs

Group Equilibrium

Chamberlin has developed his 'group equilibrium' model taking into account, as in case of individual equilibrium, a variety of conditions pertaining to product, price, selling cost, price and selling cost taken together, action and reaction of competitors, and constant and variable price and selling costs. We present here only the final form of his model of group equilibrium. His group equilibrium model is based on the following explicit and implicit assumptions:

- (i) There is a group of firms all aiming at maximisation of profits
- (ii) Price for the group product is given, say by custom or otherwise
- (iii) Demand curves, production cost curves, and selling cost curves are uniform for all the products in the group — 'the drastic assumption'

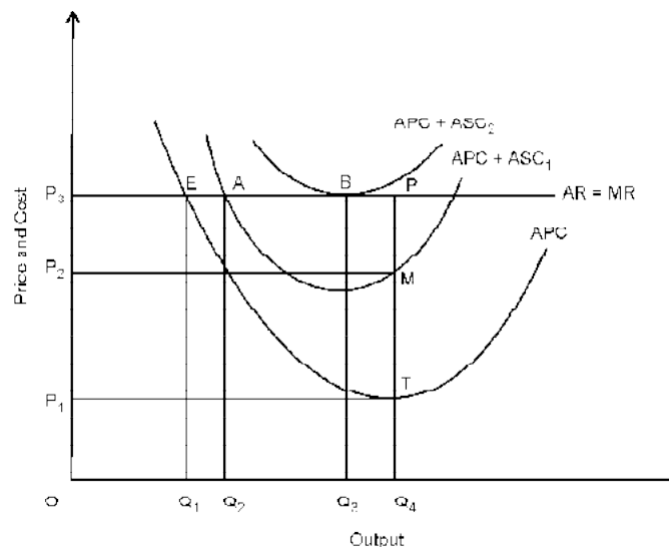


Fig. 7.13 Selling Costs and Group Equilibrium

- (iv) The firms make competitive advertisement either to increase the demand for their product or to recapture their lost market share
- (v) All of them face the identical demand and cost curves

The group equilibrium of the firms in monopolistic competition is illustrated in Figure 7.13 under the assumptions given above. The curve APC represents the average production cost of the firms. Price of the group product is given at OP_3 and, at the given price, the price line for all the firms is given by the line labelled $AR = MR$. None of the firm is incurring any selling cost. Under these conditions, all the firms are shown to be in equilibrium at point E where they all make only normal profits.

Now suppose that one of the firms 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 sale increases to OQ_4 . At output OQ_4 , the firm makes supernormal profits of P_3PMP_2 . This profit is, however, possible only so long as other firms do not 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 fall to OQ_2 . In fact, all the firms reach equilibrium at point A and produce 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 $APC + ASC_1$ curve shifts further upward. This process continues until $APC + ASC_2$ becomes tangent to the $AR = MR$ line, as shown by point B. Beyond point B, advertising is of no avail to any firm. The equilibrium will be stable at point B where each firm produces OQ_3 and makes only normal profit.

7.5.4 Monopolistic Competition vs. Perfect Competition

As mentioned earlier, the monopolistic competition as perceived by Chamberlin, is characteristically closer to perfect competition. There are, however, significant differences between the two kinds of markets. This section presents a comparison of monopolistic competition and perfect competition in respect of (i) the number of firms, (ii) the nature of products, (iii) the nature of competition, (iv) efficiency in production, and (v) capacity utilization.

1. The number of firms: The number of firms in both monopolistic competition and perfect competition is very large. Though numbers cannot be specified numerically, there is a significant conceptual difference. The number of firms in perfect competition is so large that an individual firm has absolutely no control on the price of its product: price is determined by the market forces and is given to the firm. A firm cannot change its price without incurring losses. Under monopolistic competition, however, the number of firms is only so large that an individual firm does have power to change price of its product. A firm can increase the price of its product and still retain some of its buyers (which is not possible in perfect competition) and if a firm cuts down the price of its product, it captures a part of

NOTES

NOTES

the market of the rival firms. On the contrary, if a firm in perfect competition cuts down the price, it goes out of the market itself.

2. The nature of the product: Under perfect competition, product is *homogeneous* and, therefore, the product of each seller is a *perfect substitute* for that of the others. In monopolistic competition, on the other hand, there is *product differentiation* and the product of each firm is a *close substitute* for that of the others. As discussed earlier, *product differentiation* adds the *element of monopoly* and the scope for competition, making the market monopolistically competitive.

3. The nature of competition: Under perfect competition with homogeneity of products, there is *virtually no competition*. Each firm facing a horizontal demand curve, can sell any quantity without affecting the market share of other firms. Under monopolistic competition, on the contrary, there is competition among the firms as they face a downward sloping demand curve due to product differentiation. Competition may take the form of *price competition* or *non-price competition*. The basis of *price competition* is the firm's perception that the demand curve for its product (i.e., the *perceived demand curve*) is more elastic than the market demand curve. So the firms can cut down the price of its product and increase its sales. This leads to price competition.

More important is the *non-price competition*. *Non-price competition* takes the form of *competitive advertising* of the product by the firms—it may be informative, or persuasive. *Informative advertising* aims at providing information to the potential buyers about the product, its quality and price. *Persuasive advertising* aims at attracting customers of the other firms which leads to intensive advertising of the product. For example, Pepsi and Coca-Cola companies have been involved in aggressive advertising of their product over a decade without making profits. This is a non-price competition for market share. On the contrary, advertising by a firm in perfect competition is a waste.

4. Efficiency in production: Efficiency in production in monopolistic competition and perfect competition is compared on the basis of their equilibrium output. Although rules for profit maximisation are the same for the firms in both the kinds of the markets (i.e., $MR = MC$ with MC rising), equilibrium output under perfect competition is higher than that under monopolistic competition. For illustration, refer to Figure 7.9. The equilibrium of a firm in monopolistic competition is determined at point P with output OQ_A and equilibrium of a firm in perfect competition will be at point B where equilibrium output is OQ_B . Recall that the firms under perfect competition face a horizontal demand curve passing through the minimum point of the LAC curve, i.e., point B . Thus, the equilibrium output under monopolistic competition is OQ_A and equilibrium output under perfect competition is OQ_B . It can be seen in Figure 7.9 that $OQ_B > OQ_A$. It means that production under perfect competition is more efficient than under monopolistic competition.

5. Capacity utilization. Related with equilibrium output is the issue of capacity utilization. In continuation with the issue of efficiency in production, it can be shown that capacity utilization under monopolistic competition is lower than that under perfect competition. Referring again to Figure 7.9, full capacity utilization lies at point B conforming to equilibrium output OQ_B in perfect competition. Under monopolistic competition, however, firms produce OQ_A which is lower than full-capacity output. It means that in monopolistic competition, there is underutilization of capacity. That is, there is *excess capacity* in monopolistic competition whereas there is none in perfect competition.

NOTES

7.6 CRITICISM OF MONOPOLISTIC COMPETITION

Chamberlin's theory of monopolistic competition was received enthusiastically as a path-breaking contribution to the theory of value as it sought to analyse the determination of price and output in the market structure other than monopoly and pure competition. In later years, however, Chamberlin's theory was criticized by the economists on both theoretical and empirical grounds. Major criticisms of Chamberlin's theory are briefly given below.

1. It has low predicting power: The test of the validity of an economic theory lies in its power to predict the effect of changes in economic conditions. On this test, Chamberlin's theory has been found to have low predicting power. Critics claim that this model of monopolistic competition is not significantly different from traditional models of perfect competition and pure monopoly. Where models of perfect competition and pure monopoly do not apply in predicting the price and output, *oligopoly models* apply better than Chamberlin's model of monopolistic competition. Therefore, his model has not been found useful for making predictions.

2. His model has theoretical flaws: In monopolist competition, product differentiation makes the product, in reality, heterogeneous in terms of quality, price, use, distinctive appearance, consumers' liking, etc. Chamberlin has, therefore, devised the concept of 'product group', i.e., the group of firms selling products which are very close substitutes for one another. Stigler, a Nobel Laureate, has, however, criticized Chamberlin's definition of the 'product group'. He argues that products are so heterogeneous that grouping them under a category of very close substitutes is not reasonable and dependable. In fact, the concept of 'product group' is *ambiguous*. For example, look at the so-called car industry. The number of brand names and variety of cars with significantly different prices and use is so large that it is difficult to group certain brands as a 'product group'. An attempt to group them under different product groups will violate the concept of industry: each brand name of certain size will make a different industry. Under these conditions, it is extremely difficult to find a demand curve for the product group.

NOTES

3. Chamberlin's model makes unrealistic assumptions: Chamberlin's theory has also been criticized for making unrealistic assumptions. The arguments against three of his basic assumptions are as follows.

(i) *Assuming identical cost and revenue curves are not justified.* Critics argue that Chamberlin's *heroic assumption* that firms have *identical* or *'uniform'* cost and revenue curves is questionable. In monopolistic competition, products are so widely differentiated that each product makes an industry and has a different demand curve. Also, where product differentiation is very significant, it does make a difference in the cost of production. Therefore, assuming identical demand and cost curves is unrealistic.

(ii) *Assuming no interdependence is not reasonable.* Chamberlin assumes that the number of firms in monopolistic competition is so large that pricing and selling strategies adopted by an individual firm go unnoticed by the competitors and, therefore, there is no *interdependence* between the firms. This assumption has been questioned on the ground that rival firms are bound to be affected by the pricing and selling strategies adopted by a firm because their products are deemed to be close substitutes. Therefore, rival firms are bound to react and adopt a countervailing policy. It is, therefore, not reasonable to assume *absence of interdependence*.

(iii) *Assuming that firms do not learn is not correct.* Chamberlin's model of price output determination assumes that firms do not learn from their mistakes and experience. Cohen and Cyert question the validity of this assumption. They argue that it is puzzling to accept that firms repeat the mistake of cutting down the price of their products time and again even if price reduction results in no gain. Therefore, firms cannot be supposed to continue to stick to their belief that their *perceived demand curve* provides them with real opportunities. If this argument is accepted, Chamberlin's model of price competition breaks down.

4. Measure of excess capacity is logically inconsistent: Harrod has questioned the logic of Chamberlin's measures of the excess capacity. He argues that it is logically *inconsistent* because Chamberlin uses *long-run MC* curve with *short-run MR* curve to determine the actual output. Consistency requires that *with long-run MC curve, only long-run MR curve should be used*. If long-run *MR* curve is used, the actual output will be larger than Chamberlin's output because the long-run demand and *MR* curves will be more elastic. In that case, the measure of excess capacity will be lower than one suggested by Chamberlin. Also, Chamberlin's measure of excess capacity is not corroborated by empirical evidence.

5. Chamberlin's model lacks empirical validity: Cohen and Cyert claim that *it is difficult to find any example in the real world to which Chamberlin's model of monopolistic competition is relevant*. According to them, most markets that exist in the real world may be classified as *perfect competition*, *pure monopoly*, *oligopoly* or *monopoly* and none under monopolistic competition.

These critics hold the view that Chamberlin's model of monopolistic competition is not a useful addition to economic theory as it does not pertain to any real market.

Monopolistic
Competition

Despite these damaging criticisms, Chamberlin's theory of monopolistic competition is regarded as a significant contribution to the theory of value and remains a subject matter of microeconomics for analytical rigour and insight that it provides in analysing a monopolistically competitive market.

NOTES

Check Your Progress

5. Under monopolistic competition, how does a firm alter its sales prospects?
6. According to Chamberlin, what are the changes that take place in monopolistic competition?
7. Define excess capacity of a firm.
8. What are the assumptions on which Chamberlin's group equilibrium model is based?

7.7 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. The *element of monopoly* in monopolistic competition arises from the fact that each firm has an absolute right to produce and sell a branded or patented product. Other firms are prevented by laws from producing and selling a 'branded product' of other firms. This gives a firm *monopoly power* over production, pricing and sale of its own branded product.
2. The *element of competition* in monopolistic competition comes from the fact that each branded product is a close substitute for another and firms selling branded products of the same generic category have to compete for the market share.
3. As in case of monopoly, a monopolistically competitive firm faces a downward sloping demand curve. The reason is that a monopolistically competitive firm can, by exercising its monopoly power, increase its price and still retain some buyers with brand loyalty and can increase the demand for its product by decreasing the price because of a relatively higher cross-elasticity of the competitive product.
4. Product differentiation aims at changing the slope and position of the demand curve for the product and from converting it from a horizontal demand line (as under perfect competition) to a downward sloping demand curve.
5. Under monopolistic competition, a firm can alter its sales prospects by the following *three methods*:
 - (i) By changing the price of its product;

NOTES

- (ii) By changing the nature of the product; and
 - (iii) By incurring the advertisement outlays.
6. In the long run, the following changes take place in monopolistic competition.
- (i) New firms enter the industry,
 - (ii) Firms begin to indulge in price competition, and
 - (iii) Firms make vigorous advertisement with a view to enhancing their market share or to retain their market share especially when there is entry of new firms into the industry or the product group.
7. The excess capacity of a firm is defined as the difference between the ideal output and the actual output attained in the long-run.
8. Group equilibrium model is based on the following explicit and implicit assumptions:
- (i) There is a group of firms all aiming at maximisation of profits;
 - (ii) Price for the group product is given, say by custom or otherwise;
 - (iii) Demand curves, production cost curves, and selling cost curves are uniform for all the products in the group — the drastic assumption.

7.8 SUMMARY

- Monopolistic competition refers to a market structure in which a large number of sellers sell differentiated products which are close substitutes for one another
- Monopolistic competition combines the basic elements of both perfect competition and monopoly.
- The main features of monopolistic competition vis-à-vis perfect competition and monopoly are: (i) Product differentiation; (ii) Large number of sellers; (iii) Free entry and free exit; (iv) Selling costs; (v) Downward sloping demand curve.
- There are three basic elements of monopolistic competition as defined by Chamberlin:
 - (i) Product differentiation and firm's demand curve;
 - (ii) Selling costs and firm's cost structure, and
 - (iii) Product differentiation and the concept of industry.
- Under monopolistic competition, a firm can alter its sales prospects by the following *three methods*:
 - (i) By changing the price of its product;
 - (ii) By changing the nature of the product; and
 - (iii) By incurring the advertisement outlays.

- According to Chamberlin, excess capacity in monopolistic competition arises in the long-run when there is free entry but *absence of active price-competition* between the firms. It implies that if monopolistic competition is characterised by free entry and active price competition, there would be no excess capacity.
- The major criticisms of Chamberlin's theory are as follows: (i) It has low predicting power; (ii) His model has theoretical flaws; (iii) Chamberlin's model makes unrealistic assumptions; (iv) Measure of excess capacity is logically inconsistent; (v) Chamberlin's model lacks empirical validity.

NOTES

7.9 KEY WORDS

- **Monopolistic Competition:** monopolistic competition refers to a market structure in which a large number of sellers sell differentiated products which are close substitutes for one another.
- **Product Differentiation:** It is the process of distinguishing a product or service from others, to make it more attractive to a particular target market. This involves differentiating it from competitors' products as well as a firm's own products.
- **Selling Costs:** According to Chamberlin, selling costs are costs incurred in order to alter the position or the slope of the demand curve for a product.
- **Ideal Output:** Ideal output is the output that can be produced at the minimum long-run average cost (LAC).

7.10 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short Answer Questions

1. What is meant by product differentiation? What is the purpose of product differentiation? How does it affect a firm's demand curve?
2. How is the concept of 'product group' different from the traditional meaning of industry?

Long Answer Questions

1. 'Theories of monopolistic competition and monopoly do not represent the real business world whereas theory of monopolistic competition offers an explanation to price and output determination in a real business world'. Explain and justify the statement.
2. According to Chamberlin, average selling cost (ASC) curve is U-shaped. What factors determine the shape of the ASC? Illustrate graphically how selling costs affect the overall cost structure of the firm.

NOTES

3. The purpose of selling cost is to increase the demand for the product and to make demand curve more elastic. But there is a limit to it. How can a firm find the optimum level of selling cost? Explain using appropriate diagrams.
4. How does a monopolistically competitive firm adjust its price and output to arrive at its equilibrium? Explain and illustrate how a firm in monopolistic competition reaches its equilibrium in the short-run. Does a firm in equilibrium in monopolistic competition always make a super-normal profit?
5. What does Chamberlin mean by 'group equilibrium'? How does selling cost affect the group equilibrium? Illustrate firm's equilibrium with selling costs.
6. Write a note on the critical evaluation of Chamberlin's theory of monopolistic competition? Why is his theory not in use even though it is analytical and sound?

7.11 FURTHER READINGS

- Dwivedi, D. N. 2002. *Managerial Economics*, 6th Edition. New Delhi: Vikas Publishing House.
- Keat Paul G. and K. Y. Philip. 2003. *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th Edition. Singapore: Pearson Education Inc.
- Keating, B. and J. H. Wilson. 2003. *Managerial Economics: An Economic Foundation for Business Decisions*, 2nd Edition. New Delhi: Biztantra.
- Mansfield, E., W. B. Allen, N. A. Doherty and K. Weigelt. 2002. *Managerial Economics: Theory, Applications and Cases*, 5th Edition. NY: W. Orton & Co.
- Peterson, H. C. and W. C. Lewis. 1999. *Managerial Economics*, 4th Edition. Singapore: Pearson Education, Inc.
- Salvantore, Dominick. 2001. *Managerial Economics in a Global Economy*, 4th Edition. Australia: Thomson-South Western.
- Thomas, Christopher R. and Maurice S. Charles. 2005. *Managerial Economics: Concepts and Applications*, 8th Edition. New Delhi: Tata McGraw-Hill.

UNIT 8 OLIGOPOLY

Structure

- 8.0 Introduction
- 8.1 Objectives
- 8.2 Meaning and Features of Oligopoly
 - 8.2.1 Factors Causing Oligopoly
 - 8.2.2 Features of Oligopoly
- 8.3 Oligopoly Models: An Overview
- 8.4 Kinked Demand Curve
 - 8.4.1 Criticism of Sweezy's Model
- 8.5 Answers to Check Your Progress Questions
- 8.6 Summary
- 8.7 Key Words
- 8.8 Self Assessment Questions and Exercises
- 8.9 Further Readings

NOTES

8.0 INTRODUCTION

As noted in earlier units, perfect competition and monopoly are uncommon cases of market structure. During the early 20th century monopolistic competition emerged as a major kind of market structure. Over time, however, monopolistic competition too gave way to oligopoly kind of market structure. Therefore, the attention of economists was drawn towards the analysis of price and output determination under oligopolistic market structure in which there are 'a few sellers of a product', although the first oligopoly model (in the form of a duopoly model) was developed by a classical economist, Augustin Cournot, as early as 1838.

In this unit, we will discuss theories of price and output determination in oligopoly. A number of complex theories were developed over time to analyse price and output determination in an oligopolistic market. This, however, added more complication and confusion to the subject rather than offering a non-controversial theory.

8.1 OBJECTIVES

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

- Discuss the nature and characteristics of oligopoly
- Describe the major oligopolistic models provided by various thinkers
- Explain Sweezy's kinked demand curve model

NOTES

8.2 MEANING AND FEATURES OF OLIGOPOLY

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

In any case, if oligopoly firms sell a homogeneous product, it is called *pure* or *homogeneous oligopoly*. For example, industries producing bread, cement, steel, petrol, cooking gas, chemicals, aluminium and sugar are industries characterised by *homogeneous* oligopoly. And, if firms of an oligopoly industry sell *differentiated products*, it is called *differentiated* or *heterogeneous oligopoly*. Automobiles, television sets, soaps and detergents, refrigerators, soft drinks, computers, cigarettes, etc. are some examples of industries characterized by *differentiated* or *heterogeneous oligopoly*.

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

Market share (%) of 4-10 firms	No. of industries
1 – 24.9	8
25 – 49.9	11
50 – 74.9	15
75 – 100	50
Total	84

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

8.2.1 Factors Causing Oligopoly

The main sources of oligopoly are described here briefly:

1. Huge capital investment: Some industries are by nature capital-intensive, e.g., manufacturing automobiles, aircraft, ships, TV sets, refrigerators, steel and aluminium goods, etc., and hence require huge investment. Therefore, only a few big firms can enter these kinds of industries. In fact, a huge investment requirement works as a natural barrier to entry to the oligopolistic industries.

2. Economies of scale: By virtue of huge investment and large scale production, large units enjoy *absolute cost advantage* due to economies of scale in purchase

of industrial inputs, market financing, and sales organization. This gives the existing firms a comparative advantage over the new firms, especially in price competition. This works as a deterrent for the entry of new firms.

3. Patent rights: In case of *differentiated oligopoly*, firms get their differentiated product patented which gives them monopoly power, i.e., an exclusive right to produce and market the patented commodity. This prevents other firms from producing the patented commodity. Therefore, unless new firms have something new to offer and can match the existing products in respect of quality and cost, they cannot enter the industry. This keeps the number of firms limited.

4. Control over certain raw materials: Where a few firms acquire control over almost the entire supply of important inputs required to produce a certain commodity, new firms find it extremely difficult to enter the industry. For example, if a few firms acquire the right from the government to import certain raw materials, they control the entire input supply.

5. Merger and takeover: Merger of rival firms or takeover of rival firms by the bigger ones with a view to protecting their joint market share or to put an end to waste of competition is working, in modern times, as an important factor that gives rise to oligopolies and strengthens the oligopolistic tendency in modern industries.

8.2.2 Features of Oligopoly

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

1. Small number of sellers: As already mentioned, there is a small number of *sellers* under oligopoly. How small is the number of sellers? There is no precise answer to this question: it depends largely on the size of the market. Conceptually, however, the number of sellers is so small and the market share of each firm is so large that a single firm can influence the market price and the business strategy of its rival firms. The number may vary from industry to industry.

2. Interdependence of decision-making: The *most striking feature* of an oligopolistic market structure is the *interdependence* of oligopoly firms. The characteristics fewness of firms under oligopoly brings the firms in keen competition with each other. The competition between the firms takes the form of action, reaction and counteraction in the absence of collusion between the firms. Since the number of firms in the industry is small, the business strategy of each firm in respect of pricing, advertising, product modification is closely watched by the rival firms and it evokes imitation and retaliation. What is equally important in strategic business decisions is that firms initiating a new business strategy anticipate and take into account the counteraction by the rival firms. This is called *interdependence* of oligopoly firms.

However, *price competition* is not the major form of competition among the oligopoly firms as price war destroys the profits. A more common form of competition is *non-price competition* on the basis of product differentiation, vigorous advertising and provision of services.

NOTES

NOTES

3. Barriers to entry: Barriers to entry to an oligopolistic industry arise due to such market conditions as (i) huge investment requirement to match the production capacity of the existing ones, (ii) economies of scale and absolute cost advantage enjoyed by the existing firms, (iii) strong consumer loyalty to the products of the established firms based on their quality and service, and (iv) resistance by the established firms by price cutting. However, the new entrants that can cross these barriers can and do enter the industry, though only a few, that too mostly the branches of MNCs.

4. Indeterminate price and output: Another important feature, though controversial, of the oligopolistic market structure is the indeterminateness of price and output. The characteristic fewness and interdependence of oligopoly firms make derivation of the demand curve a difficult proposition. Therefore, price and output are said to be indeterminate. However, price and output are said to be determinate under collusive oligopoly. But, collusion may last or it may breakdown. *An opposite view is that price under oligopoly is sticky, i.e., if price is once determined, it tends to stabilize.*

Types of Economy in Oligopoly

There are two types of economy in Oligopoly.

- Collusive
- Non-collusive

Collusive oligopoly is when the companies come together and work as a group. (Change the price of the goods, in effect acting as a monopoly but dividing any profits that they make.)

Non-collusive oligopoly exists when the firms in an oligopoly do not collude and so have to be very aware of the reactions of other firms while taking price decisions.

Check Your Progress

1. Define oligopoly.
2. What is the most striking feature of an oligopolistic market?
3. Why do barriers to entry to oligopolistic industry arise?

8.3 OLIGOPOLY MODELS: AN OVERVIEW

As already mentioned, under oligopolistic conditions, rival firms adopt an intricate pattern of actions, reactions and counteractions showing a variety of behaviour patterns. The uncertainty arising out of unpredictable behaviour, actions and reactions of oligopoly firms makes systematic analysis of oligopoly an extremely difficult task. As Baumol puts it, 'Under [these] circumstances, a very wide variety of behaviour pattern becomes possible. Rivals may decide to get together and

cooperate in the pursuit of their objectives, ... or, at the other extreme, may try to fight each other to the death. Even if they enter an agreement, it may last or it may breakdown. Economists have, therefore, found it extremely difficult to make a systematic analysis of price and output determination under oligopoly. This has, however, not deterred the economists from their efforts to find a reasonable explanation to the problem of price determination in an oligopolistic market.

In accordance with the wide variety of behaviour patterns, economists have developed a variety of analytical models based on different behavioural assumptions. The famous models of oligopoly include Cournot's duopoly model (1838), Bertrand's leadership model (1880), Edgeworth's duopoly model (1897), Stackelberg's model (1933), Sweezy's kinked demand curve model (1939), Neumann and Morgenstern Game Theory model (1944), Baumol's sales maximization model (1959). None of these models, however, provides a universally acceptable analysis of oligopoly, though these models do provide an insight into firms' oligopolistic behaviour.

Here, we will discuss some selected classical and modern oligopoly models with the purpose of showing the behaviour of oligopoly firms and working of the oligopolistic markets. The analytical models discussed here are selected on the basis of how price and output are determined under price competition, cartel system and the dilemma that oligopoly firms face in their price and output decisions. Specifically, we will discuss the following oligopoly models.

- (i) Cournot's model of duopoly: A limiting case of oligopoly
- (ii) Chamberlin's oligopoly model
- (iii) Sweezy's kinked-demand curve model
- (iv) Price leadership models:
 - (a) Price leadership by low-cost firm
 - (b) Price leadership by dominant firm
 - (c) Price leadership by barometric firm
- (v) Collusive model: The cartel arrangement
- (vi) Baumol's sales revenue maximization model
- (vii) The game theory model of oligopoly
- (viii) Prisoner's dilemma model

Let us discuss some of the major models.

1. A Classical Model of Duopoly: Cournot's Model

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

- (a) There are two firms, each owning an artesian mineral water well
- (b) Both the firms operate their wells at zero marginal cost

NOTES

- (c) Both of them face a demand curve with constant negative slope
- (d) Each seller acts on the assumption that his competitor will not react to his decision to change his *output*—Cournot's behavioural assumption

NOTES

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

Cournot's duopoly model is presented in Figure 8.1. The demand curve for mineral water is given by the AR curve and firm's MR by the MR curve. To begin with, let us suppose that there are only two sellers A and B , but initially, A is the only seller of mineral water in the market. By assumption, his $MC = 0$. Following the profit maximizing rule, he sells quantity OQ where his $MC = 0 = MR$, at price OP_2 . His total profit is OP_2PQ .

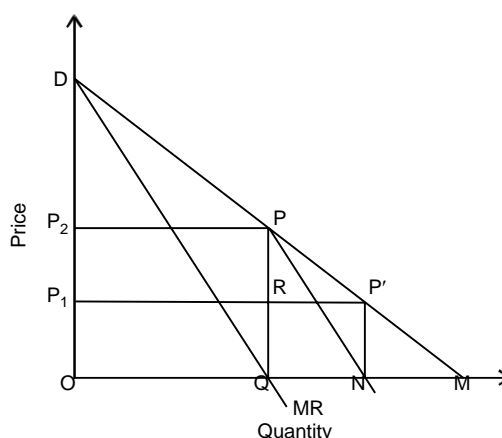


Fig. 8.1 Price and Output Determination under Duopoly: Cournot's Model

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

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






A supplies 1/2 of the unsupplied market (3/4), i.e., 3/8 of the market. It is noteworthy that A's market share has fallen from 1/2 to 3/8.

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

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

The actions and reactions and equilibrium of the sellers A and B, according to Cournot's model, are presented in Table 8.1.

Table 8.1 Determination of Market Share

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

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

Cournot's equilibrium solution is stable. For, given the action and reaction, it is not possible for any of the two sellers to increase their market share as shown in the last row of the table.

Cournot's model of duopoly can be extended to a general oligopoly model. For example, if there are three sellers in the industry, each one of them will be in equilibrium when each firm supplies 1/4 of the market. The three sellers together supply 3/4 of the total market, 1/4 of the market remaining unsupplied. Similarly, when there are four firms each one of them supply 1/5th of the market and 1/5th of the market remains unsupplied. The formula for determining the share of each seller in an oligopolistic market is: $Q \div (n + 1)$ where Q = market size, and n = number of sellers.

NOTES

Algebraic solution of duopoly: Cournot's model can also be presented algebraically. Let us suppose that the market demand function is given by linear function as

$$Q = 90 - P \quad \dots(8.1)$$

NOTES

As noted above, under zero cost condition, profit is maximum where $MC = MR = 0$ and when $MR = 0$, the profit maximizing output is $1/2 (Q)$.

Let us suppose that when A is the only seller in the market, his profit-maximising output is Q_A which is determined by the profit maximising rule under zero cost condition. A 's market share can be written as

$$Q_A = 1/2 (90 - P) \quad \dots(8.2)$$

When seller B enters the market, his profit maximising output is determined as follows.

$$Q_B = 1/2 [(1/2(90 - P))] \quad \dots(8.3)$$

Thus, the respective shares of sellers A and B are fixed at Q_A and Q_B . The division of market output may be expressed as

$$Q = Q_A + Q_B = 90 - P \quad \dots(8.4)$$

The demand function for A may now be expressed as

$$Q_A = (90 - Q_B) - P \quad \dots(8.5)$$

and for B as

$$Q_B = (90 - Q_A) - P \quad \dots(8.6)$$

Given the demand function (8.5), the market open to A (at $P = 0$) is $90 - Q_B$. The profit maximising output for A will be

$$Q_A = \frac{90 - Q_B}{2} \quad \dots(8.7)$$

and for B , it will be

$$Q_B = \frac{90 - Q_A}{2} \quad \dots(8.8)$$

The equations (8.7) and (8.8) represent the reaction functions of sellers A and B , respectively. For example, consider equation (8.7). The profit maximising output of A depends on the value of Q_B , i.e., the output which B is assumed to produce. If B chooses to produce 30 units (i.e., $Q_B = 30$), then A 's profit maximizing output = $[(90 - 30)/2] = 30$. If B chooses to produce 60 units, A 's profit maximizing output = $(90 - 60)/2 = 15$. Thus, equation (8.7) is A 's reaction function. It can similarly be shown that equation (8.8) is B 's reaction function.

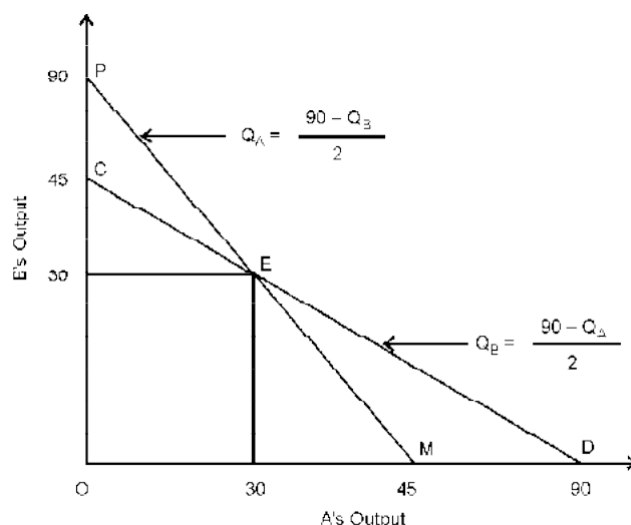


Fig. 8.2 Reaction Functions and Equilibrium: Cournot's Model

The reaction functions of A and B are graphed in Figure 8.2. The reaction function PM shows how A will react on the assumptions that B will not react to changes in his output once B's output is fixed. The reaction function CD shows a similar reaction of B. The two reaction functions intersect at point E . It means that the assumptions of A and B coincide at point E and here ends their action and reaction. Point E is, therefore, the point of stable equilibrium. At this point, each seller sells only 30 units.

The same result can be obtained by equating the two reaction equations (8.7) and (8.8). The market slope of A and B can be obtained by equating A's and B's reaction functions (8.7) and (8.8), respectively. That is, market equilibrium lies where

$$\frac{90 - Q_B}{2} = \frac{90 - Q_A}{2}$$

Since, $Q_B = (90 - Q_A)/2$, by substitution, we get first term as

$$Q_A = \frac{90(90 - Q_A)/2}{2}$$

$$Q_A = 30$$

Thus, both the sellers are in equilibrium at their respective output of 30. The market output will be 60 units. Given the market demand curve, market price will be $P = 90 - Q = 90 - 60 = \text{Rs } 30$.

As mentioned above, the duopoly model can be extended to oligopoly market.

Criticism of Cournot's model: As we have seen above, Cournot's model is logically sound and yields a stable equilibrium solution. His model has, however, been criticized on the following grounds.

NOTES

NOTES

First, Cournot's behavioural assumption, specifically assumption (d) above, is said to be naive as it implies that firms continue to make wrong calculations about the behaviour of the rival firms even though their calculations are proved wrong. For example, each seller continues to assume that his rival will not change his output even though he finds frequently that his rival does change his output.

Second, Cournot assumed zero cost of production, which is not realistic. However, even if this assumption is ignored, Cournot's results are not affected.

2. Bertrand Model of Non-collusive Oligopoly

Bertrand, a French mathematician, criticised Cournot's model and developed his own model of duopoly in 1883. Bertrand's model differs from Cournot's model in respect of its behavioural assumption. While under Cournot's model, each seller assumes his rival's output to remain constant, under Bertrand's model each seller determines his price on the assumption that his rival's price, rather than his output, remains constant.

Bertrand's model concentrates on price-competition. His analytical tools are reaction functions of the duopolists. Reaction functions of the duopolists are derived on the basis of isoprofit curves. An isoprofit curve, for a given level of profit, is drawn on the basis of various combinations of prices charged by rival firms. Assuming two firms A and B, the two axes of the plane on which isoprofit curves are drawn measure one each the prices of the two firms. Isoprofit curves of the two firms are convex to their respective price axis, as shown in Figures 8.3 and 8.4. Isoprofit curves of firm A are convex to its price-axis P_A (Figure 8.3) and those of firm B are convex to P_B (Figure 8.4).

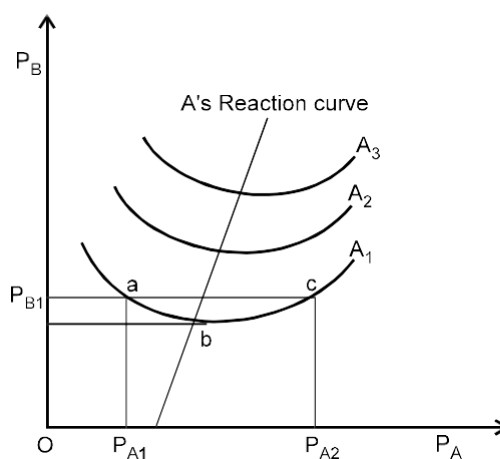


Fig. 8.3 A's Reaction Curve

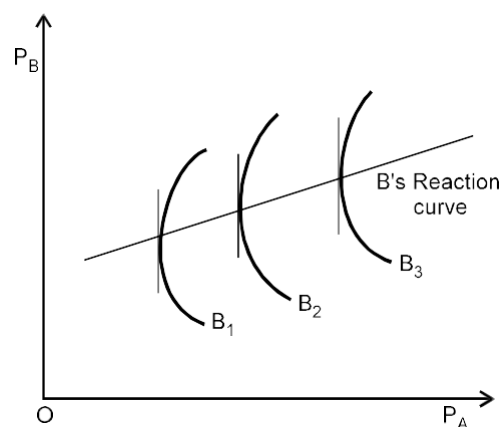


Fig. 8.4 *B's Reaction Curve*

To explain the implication of an isoprofit curve, consider curve A in Figure 8.3. It shows that A can earn a given profit from the various combinations of its own and its rival's price. For example, price combinations at points a , b and c on isoprofit curve A_1 , yield the same level of profit. If firm B fixes its price P_{B1} , firm A has two alternative prices, P_{A1} and P_{A2} , to make the same level of profits. When B reduces its price, A may either raise its price or reduce it. A will reduce its price when he is at point c and raise its price when he is at point a . But there is a limit to which this price adjustment is possible. This point is given by point b . So there is a unique price for A to maximize its profits. This unique price lies at the lowest point of the isoprofit curve. The same analysis applies to all other isoprofit curves. If we join the lowest points of the isoprofit curves A_1 , A_2 and A_3 , we get A 's reaction curve. Note that A 's reaction curve has a rightward slant. This is so because, isoprofit curves tend to shift rightward when A gains market from its rival B .

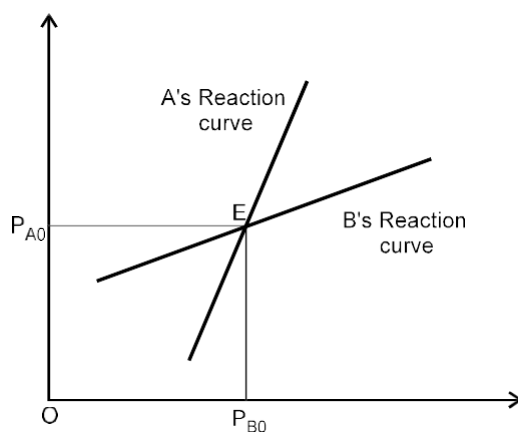


Fig. 8.5 *Duopoly Equilibrium: Bertrand's Model*

NOTES

NOTES

Following the same process, B 's reaction curve may be drawn as shown in Figure 8.4. The equilibrium of duopolists suggested by Bertrand's model may be obtained by putting together the reaction curves of the firms A and B as shown in Figure 8.5. The reaction curves of A and B intersect at point E where their expectations materialise. Point E is therefore equilibrium point. This equilibrium is stable. For, if any one of the firms deviates from the equilibrium point, it will generate a series of actions and reactions between the firms which will lead them back to point E .

Criticism

Bertrand's model has however been criticised on the same grounds as Cournot's model. Bertrand's implicit behavioural assumption that firms never learn from their past experience is naive. Furthermore, if cost is assumed to be zero, price will fluctuate between zero and the upper limit of the price, instead of stabilizing at a point.

3. Stackelberg Model of Non-collusive Oligopoly

Stackelberg, a German economist, developed, his leadership model of duopoly in 1930. His model is an extension of Cournot's model. Stackelberg assumes that one of the duopolists (say A) is sophisticated enough to play the role of a leader and the other (say B) acts as a follower. The leading duopolist A recognizes that his rival firm B has a definite reaction function which A uses into his own profit function and maximizes his profits.

Suppose market demand function is $Q = 90 - P$ and B 's reaction function is given as in Equation (8.9), i.e.,

$$Q_B = \frac{90 - Q_A}{2} \quad \dots(8.9)$$

Now, let A incorporate B 's reaction function into the market function and formulate his own demand function as

$$Q_A = 90 - Q_B - P \quad \dots(8.10)$$

Since $Q_B = (90 - Q_A)/2$, Equation (8.10) may be written as

$$Q_A = 90 - \frac{90 - Q_A}{2} - P$$

$$\text{or} \quad Q_A = 45 + \frac{Q_A}{2} - P$$

$$\begin{aligned} \text{or} \quad 2Q_A &= 90 + Q_A - 2P \\ Q_A &= 90 - 2P \end{aligned} \quad \dots(8.11)$$

Thus, by knowing B 's reaction function, A is able to determine his own demand function. Following the profit-maximization rule, A will fix his output at 45 units ($= 90/2$), i.e., half of the total demand at zero price.

Now, if seller A produces 45 units and seller B sticks to his own reaction function, he will produce

$$Q_B = \frac{90 - 45}{2} = 22.5 \text{ units} \quad \dots(8.12)$$

Thus, the industry output will be

$$45 + 22.5 = 67.5.$$

The problem with Stackelberg's model is that it does not decide as to which of the firms will act as leader (or follower). If each firm assumes itself to be the leader and the other to be the follower then Stackelberg's model will be indeterminate with unstable equilibrium.

8.4 KINKED DEMAND CURVE

The origin of kinked-demand curve can be traced into Chamberlin's theory of monopolistic competition. Later, Hall and Hitch used kinked-demand curve to explain rigidity of prices in oligopolistic market. But, neither Chamberlin nor Hall and Hitch used kinked-demand curve as a tool of analysis in their respective theories. It was Paul M. Sweezy who used the kinked-demand curve in his model of price stability in oligopolistic market. Sweezy's Model is described below.

The kinked-demand curve model developed by Paul M. Sweezy has features common to most oligopoly pricing models. This is the best known model to explain, relatively more satisfactorily, the behaviour of the oligopolistic firms. It must, however, be noted at the outset that kinked-demand curve analysis does not deal with price and output determination. Rather, it seeks to establish that once a price-quantity combination is determined, an oligopoly firm does not find it profitable to change its price even when there is a considerable change in the cost of production and change in demand for the product.

The logic behind the proposition that price once determined remains stable runs as follows. An oligopoly firm believes that if it reduces the price of its product, the rival firms would follow and neutralise the expected gain from price reduction. But, if it raises the price, the firms would either maintain their prices or even go for price-cutting, so that the price-raising firm loses a part of its market to the rival firms. This behaviour is true of all the firms. The oligopoly firms would, therefore, find it more desirable to maintain the prevailing price and output. This is the basic theme of Sweezy's theoretical model. This model is explained and illustrated below.

NOTES

NOTES

In order to analyse the effects of possible reactions of the rival firms on the demand for the product of the firm initiating the change in price, let us make the following assumptions.

- (i) There are four oligopoly firm—A, B, C and D
- (ii) Market demand curve is given by dd' in Figure 8.6
- (iii) All the firms are in equilibrium at point P

Let us suppose that firm A takes lead in changing its price and examine the effect of various kinds of reactions of the rival firms on demand for A's product.

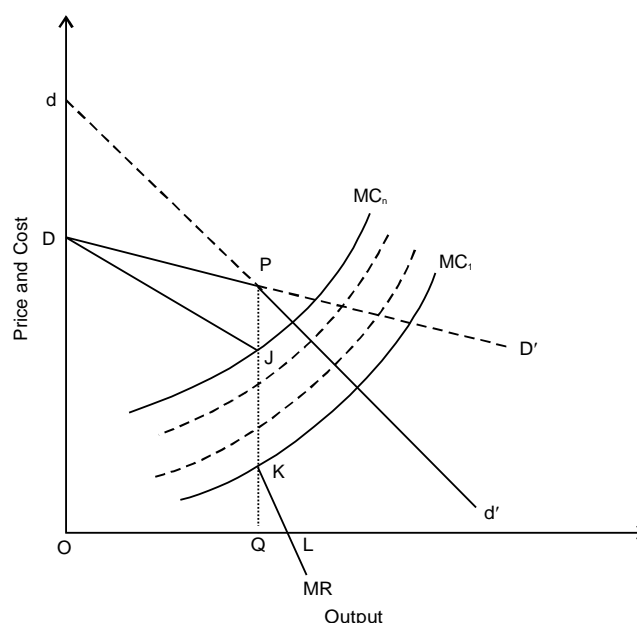


Fig. 8.6 Kinked-Demand Curve Analysis

Reaction (i) When firm A increases or decreases its price, the rival firms follow the suit. Then firm A finds itself moving along the demand curve dd' . It does not gain nor does it lose.

Reaction (ii) When the rival firms do not react to price changes made by the firm A, its demand curve becomes highly elastic as shown by the curve DD' in Figure 8.6. To explain it further, when firm A raises its price and rival firms do not follow, firm A loses a part of its market to the rival firms and moves along PD part of the demand curve. But, when firm A cuts its price and rival firms do not follow, then it captures a part of the rival's market share and finds itself moving along the PD' part of the demand curve. This is what firm A would like to achieve. Note that PD' part of demand curve is more elastic than PD .

Reaction (iii) When firm A raises its price and rival firms do not follow, then firm A loses a part of its market share to the rival firms. Then the relevant demand curve for firm A is DP . But, when firm A decreases its price, rival firms

react by cutting down their own prices by an equal amount or even more. This is a more realistic reaction. This countermove by the rival firms prevents firm A from taking any advantage of price cut. Therefore, the relevant segment of demand curve for firm A (below point P) is Pd' . If the two relevant segments of the two demand curves are put together, the demand curve for A's product takes the form of the curve DPd' . Note that this demand curve has a **kink** at point P . It is, therefore, called a **kinked-demand curve**.

Let us now derive MR curve. We know that given the demand function as $D = a - bP$, marginal revenue (MR) function is given as $MR = a - 2bP$. The derivation of the MR curve on the basis of this MR function is shown in Figure 8.6 under the condition of kinked demand (or AR) curve. The segment DJ of the MR curve corresponds to DP segment of the demand curve and KL segment of MR curve corresponds to Pd' segment of the demand curve. By joining the two segments of the MR curves, we get the full MR curve as $DJKL$.

Let us suppose that the marginal cost curve is given as MC_1 which intersects MR at point K . Point K satisfies the necessary condition for profit maximization ($MR = MC$). Therefore, oligopoly firms are in equilibrium at output OQ and they are making maximum profit. Now, if marginal cost curve shifts upwards to MC_n or to any level between points J and K , their profit would not be affected because profit maximization condition remains undisturbed. Therefore, they have no motivation for increasing or decreasing their price. It is always beneficial for them to stick to the price PQ and output OQ . Thus, both price and output are *stable*. The oligopoly firms would think of changing their price and output only if MC rises beyond point J . The same analysis applies to decrease in MC below point K . The firms would not cut their prices down unless MC decreases below point K (Figure 8.6).

8.4.1 Criticism of Sweezy's Model

As mentioned earlier, Sweezy's model is considered to be the best known model that explains relatively more satisfactorily, the behaviour of the firms in oligopoly. On the face of it, it appears to be logically sound and realistic. However, economists have criticized this model on both theoretical and empirical grounds as follows.

1. Sweezy's model does not explain price determination: The basic function of price theory is to explain price and output determination in a particular kind of market. Sweezy's model, however, does not explain price and output determination. His model only assumes the price to be given at a point of time. It explains only why price once determined tends to be sticky even if there are changes in cost conditions to a certain extent. Sweezy's model is, therefore, regarded as an *ex-post rationalization* rather than *ex-ante* explanation of market equilibrium.

2. This model does not determine the point of kink: This is a criticism related to non-determination of price. The kinked demand curve analysis explains why

NOTES

NOTES

‘kink’ appears on the demand curve. It does not explain how and at what level of price and output, the point of kink is determined. George Stigler doubts even the existence of the kinked-demand curve. Stigler’s view is supported by Julian Simon. This makes the model a purely hypothetical one, not as realistic as it appears on the face of it. However, Cohen and Cyert argue that kink in the demand curve and price rigidity may exist for a short period, for lack of inter-firm information, especially when new and unknown rivals enter the market. They are of the opinion that kink is clearly not a stable long-run equilibrium.

3. Price rigidity is not supported by empirical facts: Sweezy’s claim of price rigidity in oligopoly does not stand the test of empirical verification. Empirical facts reveal a surprising lack of price stability in oligopoly markets. Empirically, monopoly prices have been found to be more stable than oligopoly prices. Economists’ opinion is, however, divided on the issue of price rigidity in oligopoly. While Stigler has questioned price rigidity in oligopoly market, Liebhafsky finds considerable evidence of price rigidity in oligopolistic industries of the US.

4. Sweezy’s conclusion conflicts with marginal productivity theory: In Sweezy’s model, *MC* curve can shift up and down (say, between finite points *J* and *K* in Figure 8.6), while *MR* remains the same. This argument is in conflict with *marginal productivity theory* of factor pricing as this means that factor prices do not necessarily equal the marginal revenue productivity.

Check Your Progress

- 4. What were the assumptions made by Cournot to formulate his oligopoly theory?
- 5. What is the problem with Stackelberg’s model?
- 6. From where did the kinked-demand curve originate?

8.5 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

- 1. Oligopoly is a form of market in which there are a few sellers selling homogeneous or differentiated products. Economists do not specify how few are the sellers in an oligopolistic market. However, two sellers is the limiting case of oligopoly. When there are only two sellers, the market is called duopoly.
- 2. The most striking feature of an oligopolistic market structure is the interdependence of oligopoly firms. The characteristics fewness of firms under oligopoly brings the firms in keen competition with each other. The

competition between the firms takes the form of action, reaction and counteraction in the absence of collusion between the firms. Since the number of firms in the industry is small, the business strategy of each firm in respect of pricing, advertising, product modification is closely watched by the rival firms and it evokes imitation and retaliation. What is equally important in strategic business decisions is that firms initiating a new business strategy anticipate and take into account the counteraction by the rival firms. This is called interdependence of oligopoly firms.

3. Barriers to entry to an oligopolistic industry arise due to such market conditions as (i) huge investment requirement to match the production capacity of the existing ones, (ii) economies of scale and absolute cost advantage enjoyed by the existing firms, (iii) strong consumer loyalty to the products of the established firms based on their quality and service, and (iv) resistance by the established firms by price cutting.
4. To illustrate his oligopolistic model, Cournot made the following assumptions:
 - (a) There are two firms, each owning an artesian mineral water well
 - (b) Both the firms operate their wells at zero marginal cost
 - (c) Both of them face a demand curve with constant negative slope
 - (d) Each seller acts on the assumption that his competitor will not react to his decision to change his output—Cournot's behavioural assumption.
5. The problem with Stackelberg's models is that it does not decide as to which of the firms will act as leader or follower.
6. The origin of kinked-demand curve can be traced into Chamberlin's theory of monopolistic competition.

NOTES

8.6 SUMMARY

- Oligopoly is a form of market in which there are a few sellers selling homogeneous or differentiated products.
- Economists do not specify how few are the sellers in an oligopolistic market. However, two sellers is the limiting case of oligopoly. When there are only two sellers, the market is called duopoly.
- The main sources of oligopoly are (i) Huge capital investment; (ii) Economies of scale; (iii) Patent rights; (iv) Control over certain raw materials; (v) Merger and takeover.
- The important characteristics of oligopolistic industries are (i) Small number of sellers; (ii) Interdependence of decision-making; (iii) Barriers to entry; (iv) Indeterminate price and output.

NOTES

- Augustine Cournot was the first to develop a formal oligopoly model in 1838. He formulated his oligopoly theory in the form of a duopoly model which can be extended to oligopoly model.
- To illustrate his model, Cournot made the following assumptions:
 - (a) There are two firms, each owning an artesian mineral water well;
 - (b) Both the firms operate their wells at zero marginal cost;
 - (c) Both of them face a demand curve with constant negative slope;
 - (d) Each seller acts on the assumption that his competitor will not react to his decision to change his output—Cournot's behavioural assumption.
- On the basis of his model, Cournot has concluded that each seller ultimately supplies one-third of the market and both the firms charge the same price. And, one-third of the market remains unsupplied.
- The classical models assume, as shown in case of Cournot's model, that the actions taken by the firms in order to maximise their profits are independent of possible reactions of the rival firms. Chamberlin rejected the assumption of independent action by competing firms. He has developed a model of oligopoly assuming interdependence between the competitors. He argues that firms do not act independently. They do recognise their mutual interdependence.
- Chamberlin suggests in his oligopoly model that, if rival firms are assumed to recognise their mutual interdependence and act accordingly, a stable equilibrium can be reached where each firm charges a monopoly price and shares equally the monopoly equilibrium output. When all firms are in equilibrium, industry profit is maximized.
- The kinked-demand curve model developed by Paul M. Sweezy has features common to most oligopoly pricing models.
- The kinked-demand curve analysis does not deal with price and output determination. Rather, it seeks to establish that once a price-quantity combination is determined, an oligopoly firm does not find it profitable to change its price even when there is a considerable change in the cost of production and change in demand for the product.

8.7 KEY WORDS

- **Oligopoly:** It is a form of market in which there are a few sellers selling homogeneous or differentiated products.
- **Duopoly:** It is a type of market where there are only two sellers, the market is called duopoly.
- **Heterogeneous Oligopoly:** It is a type of oligopoly in which the industry sell differentiated products.

8.8 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short Answer Questions

1. What are the assumptions of theory of oligopoly?
2. Suppose a demand function is given as $Q = 100 - 2P$. Using Cournot's model, find the equilibrium output and price. Is equilibrium of the firm stable in Cournot's model?

Long Answer Questions

1. What are the characteristics of oligopoly? How is an oligopoly market different from monopolistically competitive market in respect of price and output determination? Is price determinate in an oligopoly market?
2. Explain Cournot's model of duopoly. Illustrate graphically that price is determinate and stable in Cournot's model of duopoly. Show that this model of duopoly can be applied to oligopoly?
3. What are the assumptions behind the existence of kinked demand curve? How does it reflect behaviour of the oligopoly firms? Why do rival firms react to a price cut but not to price rise in oligopoly markets?
4. Kinked demand curve model establishes that price once determined in oligopoly does not change even if there is change in cost of production. Using kinked demand curve model show that change in cost of production does not lead to change in price in oligopoly.
5. Explain and illustrate graphically Sweezy's kinked demand curve theory of rigidity in oligopoly. Are prices in oligopoly really sticky? What are the weaknesses of this model?

8.9 FURTHER READINGS

- Dwivedi, D. N. 2002. *Managerial Economics*, 6th Edition. New Delhi: Vikas Publishing House.
- Keat Paul G. and K. Y. Philip. 2003. *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th Edition. Singapore: Pearson Education Inc.
- Keating, B. and J. H. Wilson. 2003. *Managerial Economics: An Economic Foundation for Business Decisions*, 2nd Edition. New Delhi: Biztantra.
- Mansfield, E., W. B. Allen, N. A. Doherty and K. Weigelt. 2002. *Managerial Economics: Theory, Applications and Cases*, 5th Edition. NY: W. Orton & Co.

NOTES

Oligopoly

NOTES

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

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

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

BLOCK - III
MARGINAL PRODUCTIVITY AND
ADDING UP THEOREM

*Marginal Productivity
Theory of Distribution*

NOTES

UNIT 9 MARGINAL PRODUCTIVITY
THEORY OF DISTRIBUTION

Structure

- 9.0 Introduction
- 9.1 Objectives
- 9.2 Concept of Marginal Productivity Theory
- 9.3 Answers to Check Your Progress Questions
- 9.4 Summary
- 9.5 Key Words
- 9.6 Self Assessment Questions and Exercises
- 9.7 Further Readings

9.0 INTRODUCTION

For any businessman it is important to know all the costs associated with the production, so that decisions regarding further employment of production or termination of employment can be taken. When dealing with the factors of production, productivity turns out to be one of the most decisive factors of input and this is where the marginal productivity theory of distribution comes in. The general equation that is kept in mind is that the businessmen will keep increasing the factors of production up until the marginal productivity equals the marginal cost of production. In this unit, you will learn about the marginal productivity theory of distribution.

9.1 OBJECTIVES

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

- Discuss the concept of marginal productivity of distribution
- Explain the demand for firm's factor
- Describe factor price determination in perfect market

9.2 CONCEPT OF MARGINAL PRODUCTIVITY THEORY

NOTES

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

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

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

Marginal Productivity and Factor Demand

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

Demand for a Single Factor: Labour

The demand for a variable factor depends on the value of its marginal productivity. Therefore, we shall first derive the *value of marginal productivity* (VMP) curve of labour. The VMP_L for labour is drawn from the marginal productivity curve (MP_L). The MP_L curve is shown in Figure 9.1. The curve MP_L shows diminishing

returns to the variable factor—labour. If we multiply the MP_L at each level of employment a *constant* price P_x , we get the *value of marginal physical product curve*, as shown by the curve $VMP_L = MP_L \cdot P_x$. It is this curve which is the basis of demand curve for labour. The derivation of labour demand curve is illustrated in the following section.

NOTES

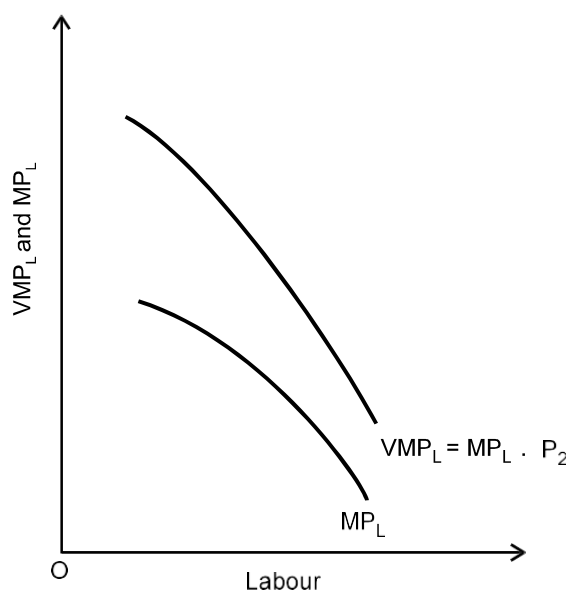


Fig. 9.1 MP_L and VMP_L Curves

Derivation of a Firm's Labour

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

- (i) Firm's objective is to maximise profit and profit condition is $MR=MC=w$.
- (ii) The firm uses a single variable factor, labour and the price of labour, wages (w), is constant.
- (iii) The firm produces a single commodity whose price is constant at P_x .

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

The short-run equilibrium of the profit-maximising firm is illustrated in Figure 9.2. The VMP_L curve shows the value of marginal product of labour, the only

NOTES

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

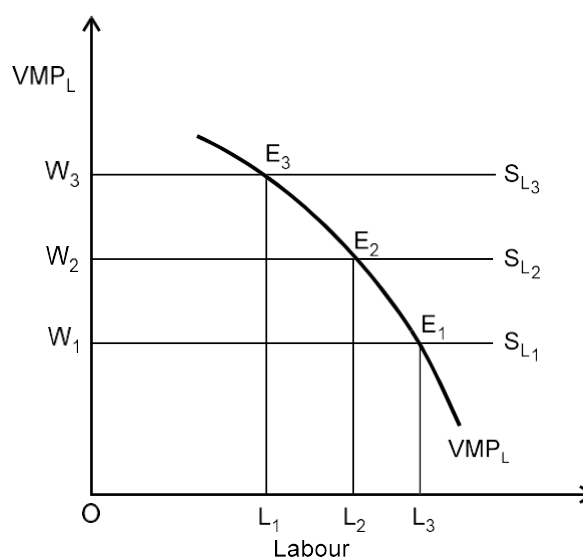


Fig. 9.2 MP_L and VMP_L Curves

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

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

*Marginal Productivity
Theory of Distribution*

Factor Price Determination in Perfect Market

We have derived above the market demand curve for labour, as shown by curve D_L in Figure 9.3. The labour supply curve is shown through the curve S_L . The labour supply curve (S_L) shows that labour supply increases in wage rate. The tools may now be applied to illustrate the factor price (wage) determination in perfectly competitive markets. Figure 9.3 shows the determination of wage in a competitive market. As shown in the figure, the demand curve for and supply curve of labour intersect each other at point P , where demand for and supply of labour are equal at OL , and wage rate is determined at OW . This wage rate will remain stable in a competitive market so long as demand and supply conditions do not change.

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

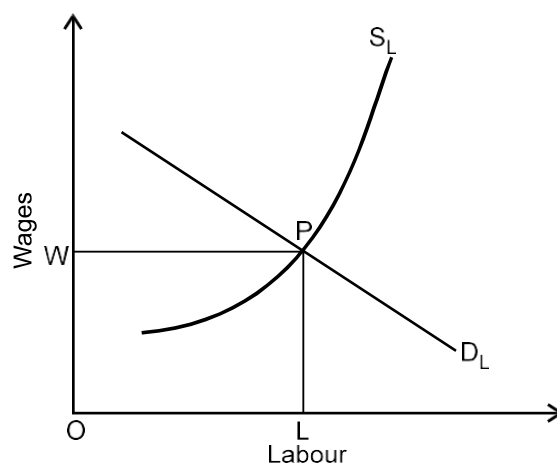


Fig. 9.3 *Determination of Wages in a Perfectly Competitive Market*

Check Your Progress

1. What are the elements whose combination with the factor demand curve gives the factor price determination?
2. What are the factors of production that a firm demands?
3. Mention the curve which is the basis of demand curve for labour.

NOTES

NOTES

9.3 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. The factor demand curve, derived on the basis of marginal productivity, combined with factor supply curve gives the factor price determination.
2. Firms demand factors of production like land, labour and capital.
3. The value of marginal physical product curve is the basis of demand curve for labour.

9.4 SUMMARY

- The neo-classical approach to factor price determination is based on marginal productivity theory of factor. Marginal productivity theory is regarded as the general micro-theory of factor price determination. It provides an analytical framework for the analysis of determination of factor prices.
- The origin of marginal productivity concept can be traced into the writings of economic thinkers of the nineteenth century. It was, in fact, John Bates Clark who had developed the marginal productivity theory as an analytical tool of analysing wage determination.
- According to Clark, the marginal productivity principle is a complete theory of wages, which could be well applied to other factors of production also. Although many theorists, including Marshall and Hicks, have objected to the marginal productivity theory being regarded as theory of wages or as theory of distribution, it is regarded as a sound theory of factor price determination.
- The marginal productivity theory provides an analytical framework for deriving the demand for a factor which is widely used in modern economic analysis. The factor demand curve, derived on the basis of its marginal productivity, combined with factor supply curve, gives the factor price determination.
- Demand for a factor of production depends on the existence of demand for the goods and services that a factor of production can create.
- The demand for a variable factor depends on the value of its marginal productivity which is made through the value of marginal productivity curve.
- The 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_1 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_1) or its marginal revenue product curve (MRP_1).

*Marginal Productivity
Theory of Distribution*

9.5 KEY WORDS

- **Marginal Productivity Theory:** It provides an analytical framework for the analysis of determination of factor prices.
- **Value of Marginal Product:** It is a measure of a firm's revenue contributed by the last unit of a productive factor employed.

NOTES

9.6 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short Answer Questions

1. Write a short note on the origin of marginal productivity concept.
2. Why is the demand for a factor a derived demand?
3. Briefly explain factor price determination in perfect market.

Long Answer Questions

1. Explain the derivation of a firm's labour.
2. Describe using diagrams the demand for single factor: Labour.

9.7 FURTHER READINGS

- Dwivedi, D. N. 2002. *Managerial Economics*, 6th Edition. New Delhi: Vikas Publishing House.
- Keat Paul G. and K. Y. Philip. 2003. *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th Edition. Singapore: Pearson Education Inc.
- Keating, B. and J. H. Wilson. 2003. *Managerial Economics: An Economic Foundation for Business Decisions*, 2nd Edition. New Delhi: Biztantra.
- Mansfield, E., W. B. Allen, N. A. Doherty and K. Weigelt. 2002. *Managerial Economics: Theory, Applications and Cases*, 5th Edition. NY: W. Orton & Co.

NOTES

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

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

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

UNIT 10 THEORY OF ADDING UP THEOREM

Theory of Adding
up Theorem

NOTES

Structure

- 10.0 Introduction
- 10.1 Objectives
- 10.2 Production Exhaustion Theorem
 - 10.2.1 Clark-Wicksteed-Walras Product Exhaustion Theorem
- 10.3 Relative Factor Shares and Income Distribution
 - 10.3.1 Technological Progress and Income Distribution
- 10.4 Answers to Check Your Progress Questions
- 10.5 Summary
- 10.6 Key Words
- 10.7 Self Assessment Questions and Exercises
- 10.8 Further Readings

10.0 INTRODUCTION

In the previous unit, you learn about the concept of marginal productivity. The theory stated that the factors involved in the production are paid as per their marginal product. In this unit, you will learn how this payment of marginal product affects the value of total product. The adding up theorem comes forth due to the proposition of the Euler's product exhaustion theorem. The product exhaustion theorem explains the exhaustion of the total product on account of the factors of production getting paid as per their marginal product.

10.1 OBJECTIVES

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

- Describe Euler's production exhaustion theorem
- Discuss the adding up controversy
- Explain Clark, Wickseed and Walrus' contribution

10.2 PRODUCTION EXHAUSTION THEOREM

The ultimate aim of the distribution theory is to explain how the share of factors of production in total output is determined.

According to the marginal productivity theory, the share of each factor in national income is determined by the marginal productivity of a factor and the number of units of the factor employed, i.e., national income = $VMP_L \cdot L + VMP_K \cdot K$. However,

NOTES

there has been a controversy on as to how is the share of each factor in the national income determined. The controversy is known as 'adding-up controversy'. This controversy is discussed first to focus on the nature of the problem. This is followed by Euler's theorem of distribution and then 'relative factor share'.

Adding-up Controversy and Solution

When the marginal productivity theory first gained acceptance by the end of the 19th century, a controversy arose whether distribution of national income among the various factors of production according to their marginal productivity was morally justifiable. In the course of the debate, another question came up, i.e., whether the sum of total labour income and of capital income equals the total product.

In other words, the controversy was, if each factor is paid the value of its marginal product (*VMP*), does this mean that the entire output is exhausted and nothing is left that falls into the hands of exploiting capitalists? Some economists attempted to show that, if each factor is paid its *VMP*, the distribution of income under free enterprise or capitalist system must be equitable. Precisely, they attempted to demonstrate that

$$Q = (MP_l)L + (MP_k)K \quad \dots(10.1)$$

In terms of value, national income is equal to $P \cdot Q$, and:

$$P \cdot Q = (MP_l \cdot P) L + (MP_k \cdot P) K \quad \dots(10.2)$$

where P is the average price of the products.

Since, $MP_l \cdot P = VMP_l$ and $MP_k \cdot P = VMP_k$

$$P \cdot Q = VMP_l + VMP_k \quad \dots(10.3)$$

It is, thus, proved that national income is distributed between the factors of production according to their marginal productivity.

Euler's Product Exhaustion Theorem

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

$$Q = \frac{\partial Q}{\partial L} \cdot L + \frac{\partial Q}{\partial K} \cdot K \quad \dots(10.4)$$

Since $\partial Q / \partial L = MP_l$ and $\partial Q / \partial K = MP_k$, Eq. (10.4) takes the form of Eq. (10.1), i.e.,

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

This may be proved as follows.

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

$$f(\lambda L, \lambda K) = \lambda^v \cdot f(L, K) \quad \dots(10.5)$$

By differentiating Eq. (10.5) with respect to λ , we get

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

When return to scale is constant, $v = 1$, and then Eq. (10.5) may be written as

$$Q = L (MP_L) + K (MP_K) = f(L, K)$$

$$\text{Thus, } Q = MP_L \cdot L + MP_K \cdot K$$

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

$$\begin{aligned} P \cdot Q &= (MP_L \cdot P) L + (MP_K \cdot P) K \\ &= VMP_L \cdot L + VMP_K \cdot K \end{aligned}$$

If $VMP_L = w$ and $VMP_K = r$, then

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

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

10.2.1 Clark-Wicksteed-Walras Product Exhaustion Theorem

Euler's product exhaustion theorem assumes a homogeneous production function, i.e., constant returns to scale. Clark, Wicksteed and Walras have, however, shown that the assumption of homogeneous production function is not necessary for the product exhaustion theorem. It holds for all types of production functions. That is, according to Clark-Wicksteed-Walras theorem, if each factor is paid its VMP , then the total factor payments will exhaust the value of total output. A graphical proof of Clark-Wicksteed-Walras theorem of product exhaustion is given in Figure 10.1.

Let us assume (i) an economy consists of n identical firms, (ii) each firm employs the same number of homogeneous labour, (iii) the marginal physical product of labour is given by the curve MPL in Figure 10.1, and (iv) each firm employs OL number of workers. The total output of each firm will then be represented by the area $OMBL$. Suppose also that each labour is paid a real wage of $OQ = BL$ and that the total wages equal the area $OQBL$. That is, the share of labour in total output $OMBL$ is $OQBL$. The residual ($OMBL - OQBL = QMB$) goes to land as *rent*. The rent so computed is merely a residual. But, Clark and others proved that QMB is not merely a residual: it is also the marginal physical product of land. By proving this, they had established the product exhaustion theorem. Note that, given n firms, the total output of the industry is $n \times OMBL$. Now suppose that the number of firms increases to $n + 1$, the number of workers remaining the same, i.e., $n \times OL$. The new firm gets its labour supply from the old firms. Suppose that $n \times OL$ workers are so distributed between $n + 1$ farms that each farm again has the same number of workers, say OL' . Note that number of workers employed by each firm decreases from OL to OL' so that

$$(n + 1) OL' = n \cdot OL$$

NOTES

NOTES

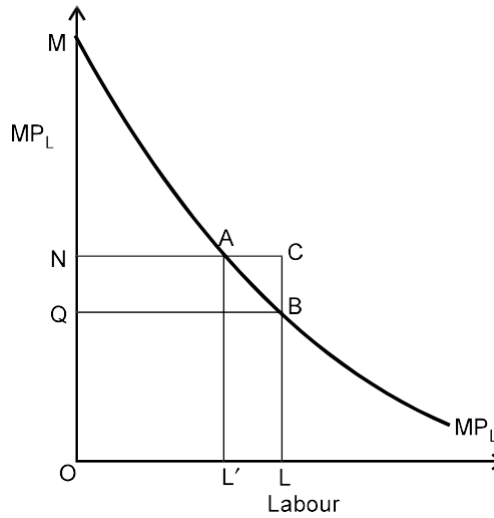


Fig. 10.1 Product Exhaustion Theorem

When each firm employs OL' workers, output per firm is $OMAL'$, and the total output of the industry is

$$(n + 1) \cdot OMAL' = n \cdot OMAL' + OMAL' \quad \dots(10.6)$$

The initial output of the industry with n firms can be written as

$$n \cdot OMBL = n \cdot OMAL' + n \cdot L'ABL \quad \dots(10.7)$$

The difference between Eqs. (10.6) and (10.7) is the marginal product of land (MP_{LD}) That is,

$$\begin{aligned} MP_{LD} &= (n \cdot OMAL' + OMAL') - (n \cdot OMAL' + n \cdot L'ABL) \\ &= OMAL' - n \cdot L'ABL \end{aligned} \quad \dots(10.8)$$

As can be seen from Figure 10.1,

$$OMAL' = NMA + ONAL' \quad \dots(10.9)$$

$$\text{and } L'ABL = L'ACL - ABC \quad \dots(10.10)$$

By substituting Eqs. (10.9) and (10.10) in Eq. (10.8), we get

$$\begin{aligned} MP_{LD} &= NMA + ONAL' - n(L'ACL - ABC) \\ &= NMA + ONAL' - n \cdot L'ACL + n \cdot ABC \end{aligned} \quad \dots(10.11)$$

Since $n \cdot L'ACL = ONAL'$, by substitution, we can write Eq. (10.11) as

$$\begin{aligned} MP_{LD} &= NMA + ONAL' - ONAL' + n \cdot ABC \\ &= NMA + n \cdot ABC \end{aligned} \quad \dots(10.12)$$

Consider the last term, $n \cdot ABC$. As $n \rightarrow \infty$, the share of each firm in the given supply of labour tends to be zero. Therefore the last term $n \cdot ABC \rightarrow 0$. Consequently,

$$MP_{LD} = NMA = \text{rent of land}$$

It is the same residual, for all firms with OL' number of workers, calculated earlier as rent. Thus, Clark-Wicksteed-Walras theorem is proved.

Check Your Progress

1. State the ultimate aim of the distribution theory.
2. What does Euler's production exhaustion theorem assume?

NOTES

10.3 RELATIVE FACTOR SHARES AND INCOME DISTRIBUTION

Now we will discuss how a *change* in relative factor prices affect the relative factor shares and income distribution. When relative factor prices change, one factor becomes relatively cheaper and the other becomes relatively costlier. This impels the profit maximizing firms to substitute the cheaper factor for the costlier one. As a result, factor ratio changes. For example, suppose there are only two variable factors, labour (L) and capital (K) and factor ratio is given as K/L . This factor ratio changes, at margin when one factor is substituted for another. When factor ratio changes, relative factor-share changes.

Elasticity of factor substitution and relative factor shares

The extent to which relative **factor shares** in income are affected by the change in relative factor prices depends on the *elasticity of factor substitution*. The concept of the elasticity of factor substitution was developed by J.R. Hicks. It is regarded as the foundation of the modern neo-classical theory of distribution and relative factor shares. Ferguson remarks that the concept of *elasticity of substitution* lies at the heart of the neoclassical theory of distribution. The *elasticity of substitution* (σ) is defined as

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

Recall that, in a perfectly competitive input market, a firm is in equilibrium when it chooses a labour-capital combination at which $MRTS$ is equal to the ratio of factor prices (w/r). That is, under perfectly competitive conditions, a firm is in equilibrium when

$$MRTS_{L,K} = \frac{w}{r} \quad \dots(10.13)$$

where w = price of labour (wage rate) and r = price of capital (interest). Thus, in a perfectly competitive factor market, the firm's equilibrium condition given in Eq. (10.13), may also be written as

$$\sigma = \frac{\partial(K/L)/(K/L)}{\partial(w/r)/(w/r)} \quad \dots(10.14)$$

The elasticity of substitution (i.e., the value of σ) is always positive, though in some cases, $\sigma = 0$. Thus, *the value of σ ranges from zero to infinity*. The value of σ yields useful information regarding the degree of substitutability between the factors. If $\sigma = 0$, it means that substitution between factors, say labour and capital, is

NOTES

impossible; the two factors can be used only in a fixed proportion; and that isoquant is L -shaped.

The positive range of σ may be classified and interpreted as follows:

$\sigma < 1$: **Inelastic substitutability:** The degree of substitutability between the two factors is very low.

$\sigma = 1$: **Unit elasticity of substitution:** The two factors can be proportionately substituted for one another (see properties of Cobb-Douglas production function).

$\sigma > 1$: **Highly elastic substitutability:** One factor can substitute another to a large extent.

$\sigma = \infty$: **Perfect substitutability:** One factor can substitute another to any extent.

Let us now examine the relationship between the value of σ and the relative shares of factors in the total output. Consider a two-factor model in which the total income $P \cdot Q$ is the sum of labour-share ($w \cdot L$) and capital-share ($r \cdot K$). That is,

$$P \cdot Q = w \cdot L + r \cdot K \quad \dots(10.15)$$

The relative share of *labour* in the total value of output is then given by

$$\frac{w \cdot L}{P \cdot Q}$$

Similarly, the relative share of *capital* in the total value of the output is given by

$$\frac{r \cdot K}{P \cdot Q}$$

Thus, the *ratio* of relative share of L and K

$$\begin{aligned} &= \frac{wL}{PQ} + \frac{rK}{PQ} \\ &= \frac{wL}{rK} \end{aligned} \quad \dots(10.16)$$

$$= \frac{w/r}{K/L} \quad \dots(10.17)$$

Eq. (10.17) can be used to show the effect of change in relative factor price (w/r) on the relative shares of L and K in the value of total output.

Suppose w/r increases by 10 per cent, i.e., labour becomes costlier by 10 per cent. This will lead to a substitution of capital (the relatively cheaper factor) for the labour (the relatively costlier factor). The extent of substitution depends on the value of σ (i.e., the elasticity of substitution). Suppose $\sigma = 0.5$, i.e., the elasticity of substitution is less than one. Then a 10 per cent increase in w/r will result in a 5 per cent increase in the capital-labour ratio (K/L). This will alter the relative shares of K and L .

There new relative shares can be obtained as

$$\begin{aligned} \frac{F_{KL}}{F_{LL}} &= \frac{(w/r)(1+0.10)}{(K/L)(1+0.05)} \\ &= \frac{1.10}{1.05} \cdot \frac{(w/r)}{(K/L)} \end{aligned}$$

where * denotes the new relative factor share.

Obviously,

$$\frac{F_{KL}}{F_{LL}}^* > \frac{F_{KL}}{F_{LL}}$$

That is, new relative factor-share ratio is greater than the initial ratio. One may show by the same reasoning that if $\sigma > 1$, and other things remain the same, then

$$\left(\frac{wL}{rK} \right)^* < \left(\frac{wL}{rK} \right)$$

There is a two-way causation in the change in relative factor shares caused by a change in relative factor prices. Changes in the relative supply position of factors and changes the relative factor prices which, in turn, changes the factor-ratios depending on the value of σ . This leads to a change in the relative shares of the factors in the total output.

It is clear that the concept of elasticity of substitution plays a very important role in neo-classical theory of income distribution.

10.3.1 Technological Progress and Income Distribution

We have so far analysed factor price determination and relative factor shares on the basis of a tacit assumption that production function is given and technology of production remains unchanged over the reference period. In the real world, however, technological progress does take place. Technological progress means a given quantity of output can be produced with less quantity of inputs or a given quantity of inputs can produce a greater quantity of output. This means a downward shift in the production function (the isoquant) towards the point of origin (O).

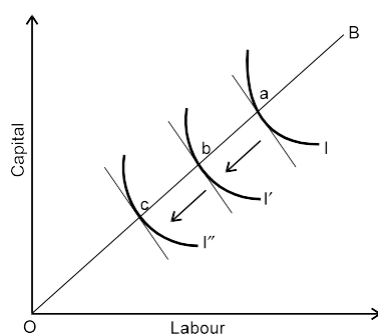


Fig. 10.2 Technological Progress-Neutral
Technological Progress

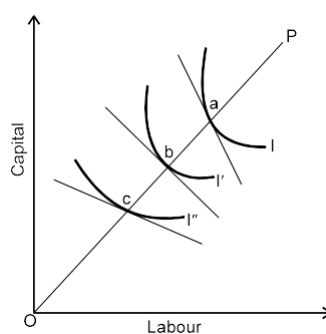


Fig. 10.3 Capital Deepening

Technological progress is graphically shown in Figure 10.2. A given level of output is shown by isoquants I , I' and I'' . That is, all three isoquants, I , I' , I'' represent the same level of output.

NOTES

NOTES

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

According to J. R. Hicks, technological progress may be classified as neutral, capital-deepening and labour-deepening. Technological progress is *neutral* if, at constant K/L , the marginal rate of technical substitution of capital for labour i.e., $MRTS_{L,K}$ remains constant. The neutral technological progress is illustrated in Figure 10.2. At each equilibrium point, $MRTS_{L,K} = w/r$. When technological progress is neutral, both K/L and w/r remain unchanged.⁶ It follows that relative factor share remains unchanged when technological progress is neutral.

Capital-deepening technological progress is illustrated in Figure 10.3. Technological progress is capital-deepening when, at a constant capital/labour ratio (K/L), $MRTS_{L,K}$ declines. It implies that, at constant K/L , MP_K increases relative to MP_L . Therefore, at equilibrium w/r declines, as r increases relative to w , because $w = VMP_L$. Consequently, the relative factor share changes in favour of K . That is, share of capital in the total output increases while that of labour decreases.

Technological progress is labour-deepening when, at a given K/L , the $MRTS_{L,K}$ increases. Labour-deepening technological progress is illustrated in Figure 10.4. It can be shown, following the above reasoning, that under labour-deepening technological progress, the share of labour in the total output increases while that of capital increases.

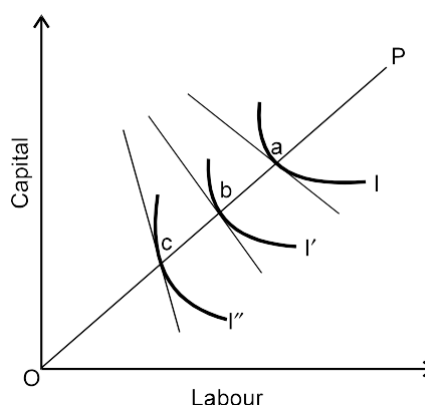


Fig. 10.4 Labour Deepening Technological Progress

Check Your Progress

3. Who developed the concept of elasticity of factor substitution?
4. What is technological progress in economics?

10.4 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. The ultimate aim of the distribution theory is to explain how the share of factors of production in total output is determined.
2. Euler's production exhaustion theorem assumes a homogenous production theorem i.e., constant returns to scale.
3. The concept of the elasticity of factor substitution was developed by J R Hicks.
4. Technological progress means a given quantity of output can be produced with less quantity of inputs or a given quantity of inputs can produce a greater quantity of output. This means a downward shift in the production function (the isoquant) towards the point of origin (O).

NOTES

10.5 SUMMARY

- According to the marginal productivity theory, the share of each factor in national income is determined by the marginal productivity of a factor and the number of units of the factor employed, i.e., national income $= \sum VMP_i \cdot L_i + \sum VMP_i \cdot K_i$.
- The adding up controversy was that if each factor is paid the value of its marginal product (VMP), does this mean that the entire output is exhausted and nothing is left that falls into the hands of exploiting capitalists? Some economists attempted to show that, if each factor is paid its VMP, the distribution of income under free enterprise or capitalist system must be equitable.
- As per Euler's product exhaustion theorem, if each factor is paid a sum equal to its VMP, the total value of product is exhausted.
- Clark, Wicksteed and Walras have shown that the assumption of homogeneous production function is not necessary for the product exhaustion theorem. It holds for all types of production functions. That is, according to Clark-Wicksteed-Walras theorem, if each factor is paid its VMP, then the total factor payments will exhaust the value of total output.
- When relative factor prices change, one factor becomes relatively cheaper and the other becomes relatively costlier. This impels the profit-maximizing firms to substitute the cheaper factor for the costlier one. As a result, factor ratio changes.
- The extent to which relative factor shares in income are affected by the change in relative factor prices depends on the elasticity of factor substitution. The concept of the elasticity of factor substitution was developed by J.R. Hicks. It is regarded as the foundation of the modern neo-classical theory of distribution and relative factor shares.

NOTES

10.6 KEY WORDS

- **Euler’s theorem:** It demonstrates that if each factor is paid a sum equal to its VMP, the total value of the product is exhausted.
- **Relative factor shares:** It refers to the condition where in when relative factor prices change, one factor becomes relatively cheaper and the other relatively costlier.
- **Elasticity of factor substitution:** It refers to the extent to which relative factor shares in income are affected by the change in relative factor prices.

10.7 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short Answer Questions

1. What is the adding-up controversy?
2. Briefly explain Euler’s product exhaustion theorem.
3. Write a short note on technological progress and income distribution.

Long Answer Questions

1. Discuss, with a graphical proof, the Clark-Wicksteed Walrus theorem of product exhaustion.
2. Examine the concept of elasticity of factor substitution and relative factor shares.

10.8 FURTHER READINGS

- Dwivedi, D. N. 2002. *Managerial Economics*, 6th Edition. New Delhi: Vikas Publishing House.
- Keat Paul G. and K. Y. Philip. 2003. *Managerial Economics: Economic Tools for Today’s Decision Makers*, 4th Edition. Singapore: Pearson Education Inc.
- Keating, B. and J. H. Wilson. 2003. *Managerial Economics: An Economic Foundation for Business Decisions*, 2nd Edition. New Delhi: Biztantra.
- Mansfield, E., W. B. Allen, N. A. Doherty and K. Weigelt. 2002. *Managerial Economics: Theory, Applications and Cases*, 5th Edition. NY: W. Orton & Co.
- Peterson, H. C. and W. C. Lewis. 1999. *Managerial Economics*, 4th Edition. Singapore: Pearson Education, Inc.
- Salvantore, Dominick. 2001. *Managerial Economics in a Global Economy*, 4th Edition. Australia: Thomson-South Western.
- Thomas, Christopher R. and Maurice S. Charles. 2005. *Managerial Economics: Concepts and Applications*, 8th Edition. New Delhi: Tata McGraw-Hill.

BLOCK - IV

THEORIES OF INTEREST, PROFIT AND INNOVATION

NOTES

UNIT 11 THEORIES OF RENT

Structure

- 11.0 Introduction
- 11.1 Objectives
- 11.2 Ricardian Theory of Rent
- 11.3 Modern Theory of Rent
- 11.4 Quasi Rent
- 11.5 Answers to Check Your Progress Questions
- 11.6 Summary
- 11.7 Key Words
- 11.8 Self Assessment Questions and Exercises
- 11.9 Further Readings

11.0 INTRODUCTION

The concept of economic rent is different from the common day usage of the term. In normal day usage, rent denotes the periodic payment for essentially hiring a particular asset (in most cases land or house). In economics, however, rent refers to price of the asset in excess of the cost involved in bringing that factor into production. Many theorists have provided their conceptual analysis of this rent. In this unit, you will learn about these different theories of rent in economics. The unit will discuss the Ricardian theory, Modern theory of Rent and Quasi – Rent.

11.1 OBJECTIVES

After going through this unit you will be able to:

- Describe Ricardian theory of rent
- Explain modern theory of rent
- Discuss quasi theory of rent

11.2 RICARDIAN THEORY OF RENT

The marginal productivity theory applies to all factor prices, including land rent. The Ricardian theory of rent is the earliest known rent theory and is generally known as the classical theory of rent. The point of distinction between Ricardian

NOTES

and modern theories of rent is that while Ricardo considered rent as ‘surplus produce’ attributable solely to land as a factor of production, modern economists consider rent as ‘economic surplus’ which accrues as well to all other factors in fixed supply in the short-run.

Antecedents of Rent Theory

Ricardian theory of rent has an interesting antecedent. In the early 19th century, food prices in Britain had considerably increased partly due to Napoleonic War and partly due to increase in population. This caused a great deal of anxiety to the British government. So, both the House of Lords and the House of Commons appointed committees to find the cause of rise in food price. The committees reported that ‘food prices were high because rents were high’. Contemporary economists, namely, Robert Torrens, Thomas Robert Malthus-David Ricardo reacted to this suggestion and offered, separately, an alternative explanation to the problem. In their opinion, food prices were high not because rents were high, rather, rents were high because food prices were high. According to them, food prices had gone up due to Napoleonic War and increase in population causing increase in demand for food. Scarcity of food led to increase in food prices which, in turn, increased profitability of cultivation. This resulted in increase in demand for land, which caused rise in rents. Ricardo, who was said to be a new bourgeoisie, added that the landed aristocracy (the landlords) was thriving on the misfortune of the rest of the society and causing misery to the tenant farmers. For holding this view, Ricardo was criticised as being anti-landed aristocracy. However, Ricardo’s theory of rent emerged out of his effort to establish his argument.

Ricardian Theory of Rent

Ricardo defined rent as ‘that portion of the produce of earth which is paid to the landlord for the use of the *original* and indestructible powers of soil’. Ricardo considered payment of rent as an indication of niggardliness of nature. 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 arises 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*.

NOTES

NOTES

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 accrues 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.

Check Your Progress

1. Which theory of rent is known as the classical theory of rent?
2. Why does rent arise as per Ricardo?

11.3 MODERN THEORY OF RENT

The modern theory of rent has been developed by using two modern concepts—transfer earning and economic rent. According to the modern theory of rent, the equilibrium price of a factor service can be divided into two components:

- (i) Transfer earning
- (ii) Economic rent

Transfer earning or what is also known as opportunity cost, may be defined as the amount that a factor must earn to remain in its present occupation. Or, in other words, transfer earning is the amount that a factor expects to earn if transferred to its second best use. For example, suppose a doctor earns ₹ 50,000 per month from his private practice. The alternative available to him is to serve in a hospital as an employee where he expects to earn ₹ 40,000 per month. Thus, the doctor's transfer earning is ₹ 40,000 per month. He must earn a minimum of ₹ 40,000 per month to remain in his private practice. So long as he earns ₹ 40,000 per month from his private practice, he has no incentive to join a hospital as an employee.

Economic rent is the excess of actual earning of a factor over its transfer earning. Economic rent may thus be defined as factor's actual earning minus its transfer

earning. Consider the factor supply curve, S_f in Figure 11.1, which has a positive slope. It implies that more and more units of factor shall be supplied to an industry if factor payments increase, and, conversely, less and less units will be supplied to the industry if factor payments decrease. That is, when factor payment decreases, factors are transferred to their alternative uses. For example, given the demand curve D_f , the market factor price is determined at OP_3 , where equilibrium supply of factor is OM . Note that, given the supply curve, S_f , all but one of OM units (i.e., $OM - 1$) of factor are willing to remain in this industry at factor prices lower than OP_3 . That is, minimum payments that must be made to all but the last factor unit, in order to prevent transfer of factors to alternative uses, is less than the equilibrium price OP_3 .

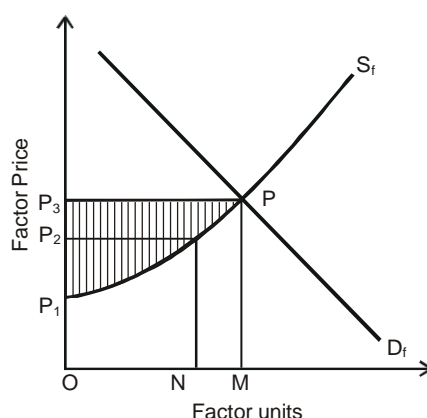


Fig. 11.1 Economic Rent

In other words, the transfer earning of all factor units, *excepting the last one*, is less than their actual earning. For example, the *transfer earning* of ON th unit of the factors is only OP_2 , whereas the actual earning, i.e., the market factor price, is OP_3 . Thus, economic rent earned by the ON th unit is $OP_3 - OP_2 = P_2O_3$. The same exercise may be performed for all the OM factor units, and economic rent computed. The shaded area, $P_1 P P_3$, represents the *total economic rent* of OM units. Note that the OM th unit, i.e., the last unit, of the factor does not earn economic rent because its actual earning equals its transfer earning. The total transfer earning is shown by the area below the shaded area.

Note that the terms ‘economic rent’ has a different meaning from ‘rent’ in common parlance. In its common usage, the term ‘rent’ means the *actual* payment to the landlord, much of which is transfer earning. But, when an economist speaks of ‘rent’ he means ‘economic rent’, i.e., the excess of payment over transfer earning.

Elasticity of factor supply and economic rent: The existence of economic rent depends on the elasticity of factor supply. Economic rent may be zero or equal to transfer earning depending on whether factor supply is perfectly elastic or perfectly inelastic. These are the two limiting cases of economic rent.

When factor supply is perfectly elastic, economic rent is zero. Perfectly elastic factor supply (i.e., $e_s = \infty$) means that an individual factor-owner can supply

NOTES

NOTES

his factor as much as he wishes, and an individual firm or industry can buy as many units of the factor as it wants to, *at a given price*. In such a case, *the whole price paid to the factor*, i.e., its actual earning, equals its transfer earning. There is no excess payment over the transfer earning. Hence, economic rent is zero. This case is illustrated in Figure 11.2. The market factor price is determined at OP , i.e., the actual earning of the factor. No factor owner can charge a price in excess of OP . Hence economic rent is zero.

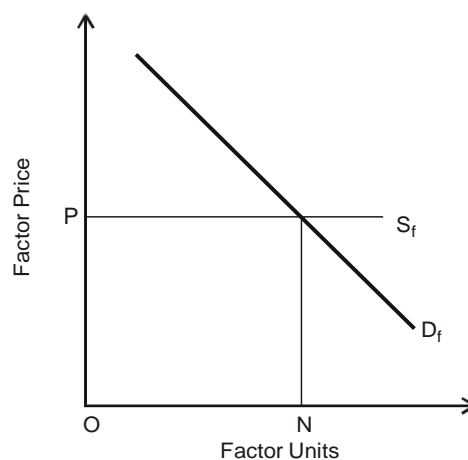


Fig. 11.2 Zero Economic Rent

Factor supply is perfectly inelastic, economic rent equals actual earning. If factor supply is fixed and factor has only one use, the factor owners would have to put their factors on the market for whatever they can earn. Even if factor owners are not satisfied with what the market offers, they cannot transfer their factors to other uses, since there is none. Therefore, in such cases *transfer earning* is zero. Thus, the whole factor price is *economic rent*. This case is illustrated in Figure 11.3. The market price is fixed at OM . The whole of which is economic rent. When factor supply curve has a positive slope, economic rent equals factor price.

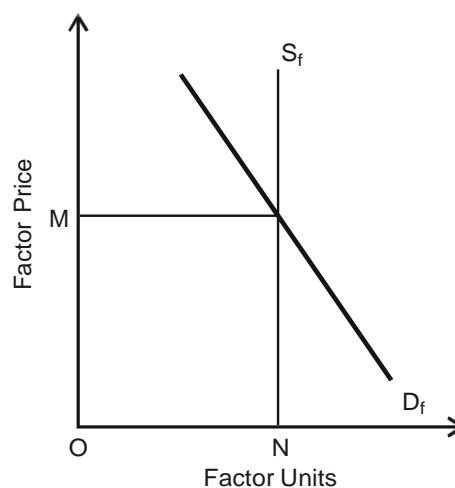


Fig. 11.3 Zero Transfer Earning

11.4 QUASI RENT

Quasi rent, a concept introduced by Alfred Marshall, refers to the short-term earnings of factors which are in fixed supply in the short-run. To explain the concept of *quasi rent*, let us make a distinction between the short-run and the long-run. In the long-run, all inputs are variable in large quantities as their supply is elastic. In the short-run, however, the supply of certain inputs is fixed. For example, the supply of plant and machinery in the short-run is inelastic.

In the short-run, variable factors can be transferred to their alternative uses if they are paid in their current use an amount less than their transfer earning (or opportunity cost). Therefore, if variable factors are to be retained in their current use in the short-run, they must be paid equal to their transfer earning. Otherwise, variable factors shall be transferred to their alternative uses. On the contrary, the fixed factors cannot be transferred to their alternative uses in the short-run. Therefore, in the short-run, fixed factors are paid what is left after the variable factors are paid their opportunity cost. That is, fixed factors are paid, in the short-run, the *residual* of the total revenue. This residual payment to a factor fixed in the short-run is called quasi rent. The quasi rent may thus be defined as $TR - TVC$.

The determination of quasi rent is illustrated in Figure 11.4. Suppose, given the AVC , AC and MC curves, price is OP , and the firm is in equilibrium at point E .

At equilibrium, firm's total revenue is

$$OP \times OQ = OPEQ$$

and

$$TVC = OB \times OQ = OBMQ$$

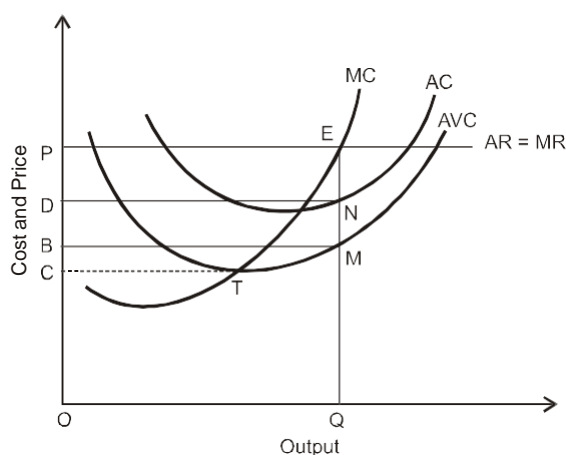


Fig. 11.4 Determination of Quasi Rent

The firm must pay a total sum of $OBMQ$ to retain the variable factors. Under perfectly competitive conditions, this sum equals their transfer earnings, i.e., the earning that a factor expects from its second best use. The quasi rent may be obtained as:

$$\text{Quasi rent} = OPEQ - OBMQ = PBME$$

NOTES

NOTES

The quasi rent will always be a non-negative quantity. For example, so long as price is greater than *OC*, the *quasi rent* will be greater than zero. When price is *OC*, total revenue (*TR*) equals total variable cost (*TVC*), i.e.,

$$TR = OC \times CT \text{ and } TVC = OC \times CT$$

Since $TR - TVC = 0$, quasi rent = 0. When price falls below *OC*, there will be no production. Therefore there is no question.

The quasi rent can be divided into two components: (i) opportunity cost; and (ii) economic profits. We have seen that when prices is *OP*, quasi rent is represented by the area *PBME*. Of this, the area *DPEN* represents the difference between the *TR* and *TC* ($= OQ \times OD$). 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 conditions. Therefore, the area *BDNM* is the *opportunity cost* of fixed factors. Thus

$$\text{Quasi rent} = TFC + \text{Economic Profit}$$

Check Your Progress

- 3. What is transfer learning also known as?
- 4. What does the existence of economic rent depend upon?
- 5. Who introduced the concept of quasi rent?

11.5 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

- 1. The Ricardian theory of rent is known as the classical theory of rent.
- 2. According to Ricardo, rent arises due to differential in surplus accruing to cultivators resulting from the differences in fertility of soil of different grades of land.
- 3. Transfer earning is what is also known as opportunity cost.
- 4. The existence of economic rent depends upon the elasticity of factor supply.
- 5. The concept of quasi rent was introduced by Alfred Marshall.

11.6 SUMMARY

- The Ricardian theory of rent is the earliest known rent theory and is generally known as the classical theory of rent.

- 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.
- The modern theory of rent has been developed by using two modern concepts—transfer earning and economic rent. According to the modern theory of rent, the equilibrium price of a factor service can be divided into two components: (i) Transfer earning (ii) Economic rent.
- The existence of economic rent depends on the elasticity of factor supply. Economic rent may be zero or equal to transfer earning depending on whether factor supply is perfectly elastic or perfectly inelastic.
- *Quasi rent*, a concept introduced by Alfred Marshall, refers to the short-term earnings of factors which are in fixed supply in the short-run.
- 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 $TR - TVC$.

NOTES

11.7 KEY WORDS

- **Transfer Earning:** It is defined as the amount that a factor must earn to remain in its present occupation.
- **Economic Rent:** It is the excess of actual earning of a factor over its transfer earning.
- **Quasi Rent:** It refers to the short-term earnings of factors which are in fixed supply in the short-run.

11.8 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short Answer Questions

1. Briefly state the difference between Ricardian and Modern theory of rent.
2. Give an account of the antecedents of the rent theory.
3. What are the two components into which quasi rent can be divided?

NOTES

Long Answer Questions

1. Explain the Ricardian theory of rent. Mention its criticisms.
2. Describe the modern theory of rent.
3. Examine the concept of quasi rent.

11.9 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 12 THEORIES OF WAGES

Structure

- 12.0 Introduction
- 12.1 Objectives
- 12.2 The Subsistence Theory of Wages and Wage Fund Theory
- 12.3 Marginal Productivity Theory of Wages
- 12.4 Answers to Check Your Progress Questions
- 12.5 Summary
- 12.6 Key Words
- 12.7 Self Assessment Questions and Exercises
- 12.8 Further Readings

NOTES

12.0 INTRODUCTION

You have learnt about the theories of rent in the previous unit. In this unit, we turn towards the theories on wages. There are many different theories propounded by authors about the manner of determination of wages. Amongst the popular theories are the Subsistence Theory of Wages, Standard of Living Theory, Wage Fund Theory, Residual Claimant Theory, Marginal Productivity Theory and Discounted Marginal Productivity Theory. In this unit, you will study three of these popular theories: the subsistence theory of wage, wage fund and the marginal productivity theory. Bear in mind that the basics of marginal productivity of distribution has already been dealt with in Unit 9 and here the theory will be a recapitulation of the same.

12.1 OBJECTIVES

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

- Discuss the subsistence theory of wage
- Describe the wage fund theory
- Explain the marginal productivity theory of wages

12.2 THE SUBSISTENCE THEORY OF WAGES AND WAGE FUND THEORY

In this section, you will study about the subsistence and wage fund theory. Let us first discuss the subsistence theory. This theory is known to have been developed by physiocrats. These theories neglect the demand side of the labour market and concentrate on the supply aspects. As per this theory, the driver for real wages to the level of minimum required subsistence is the supply of workers. The minimum

NOTES

subsistence here constitutes as the fundamental needs such as shelter and food. The earliest mention of the concept of subsistence can be found in Smith's work *The Wealth of Nations*, but his successors like David Ricardo and Thomas Malthus are credited with its origin. While Smith's was a more positive outlook, Ricardo's and Malthus's was more pessimistic. The theory of subsistence level of wages was inspired by Theory of Population Control by Malthus. He held the view that as per the theory of population, the means of supporting adjust the level of population. Ricardo, in a similar vein, considered that the price needed to perpetuate the race and subsist was the natural price of labour.

As per the theorists of this view, the natural prices will be like the guiding forces of market price of labour. This is to say market prices and natural prices will not vary much. Therefore, with the increase in wages about the rate of subsistence, the number of workers in the market would also increase and this will then lead to bringing the rate of wages down. Similarly, if the wages fall below the rate of subsistence, the number of workers in the market will decrease and this will consequently lead to the pushing the rate of wages up. This view when was being formulated was supported by the contemporary economic conditions, meaning that the labour rates were nearing the surpassing of means of subsistence. As per Ricardo, the natural price of labour was not fixed. But contrary to his views, many authors considered that the wages would also be a factor that would be driven down and this is why the theory got its name: the iron law of wages.

Wage Fund Theory

As per Smith, the demand for labour would increase only in proportion to the increase of the funds set out for the payment of wages. Further, Ricardo, too maintained that a rise in the capital would drive the demand for labour upwards. These postulations led to the considerations that there is always a pre-decided 'fund' for the payment of wages. This wage fund was defined by Smith as the disposable income of the wealthy which could be utilised for the employment of labour. Ricardo, differing from Smith, held the view that wage fund was actually the capital constituting the requirements for employment of labour including factors like raw material, food, clothing, tools, etc. The nature of this wage fund is considered to be fixed at a particular point of time, but it may fluctuate over a period of time. The average of such a fund would then logically be calculated as the total divided by the number of workers.

This then concludes, that if the volume of the fund, whatever be its constituents, was large, the wages would be high and vice versa. And if the population increases more than the food and other subsistence related items, the wages would be directed towards the subsistence level. This led to the belief that a contribution of sorts to the fund from the labourers would result positively for them and consequently any irrational or over-the-top demands would put them at a disadvantage as the capital would decrease and so would their wages.

The theory received acceptance for over 50 years by economists like Nassau William Senior and John Stuart Mill until it was discredited by W.T. Thornton, F.D. Longe, and Francis A Walker after 1865. The general point of disagreement was that demand for labour was actually connected with the consumer demand for products and not by a 'wages fund'. Additionally, there is also lack of evidence proving the relation of pre-existing relationship between capital and labour demand. There was also lack of clarity of the portion of the labour's contribution actually being used for payment of wages. Further, it was also considered that there were several other factors like the bargaining power of the labourers which affected the level of wages paid to the labourers. Even so, the wages-fund theory was prominent until the end of the 19th century.

NOTES

12.3 MARGINAL PRODUCTIVITY THEORY OF WAGES

We can extend the neo-classical marginal productivity theory to explain macro-distribution of income that is to explain the share of labour and capital.

It should be noted that all the factors of production are regarded as substitutes of each other in neo-classical marginal productivity theory. To calculate the marginal product of factors we have to find the absolute and the relative shares of the factor which is equal to which real rewards for factors are determined.

Now we can take the example of labour, we represent the labour by L . Let's assume that Q stands for level of output when L units of labour is employed. The marginal product of the labour can be written as,

$$\Delta Q / \Delta L.$$

This means change in output resulting from a marginal unit change in labour. If the total labour is L and the marginal product of the labour is $\Delta Q / \Delta L$, the absolute share of labour will be $\Delta Q / \Delta L \times L$.

We can get the relative share of labour, by dividing the absolute share by the total national product (Q), which we represent by λ . Therefore, Relative share of labour in national product,

$$\Delta = \Delta Q / \Delta L \times L / Q$$

or

$$\Delta = \Delta Q / Q \times L / \Delta L$$

$$= \Delta Q / Q \div \Delta L / L$$

$$\Delta Q / Q \div \Delta L / L,$$

means the relative change in output as a result of relative change in the factor L . In other words, $\Delta Q / Q \div \Delta L / L$ represents elasticity of production in respect of labour.

NOTES

The share of the labour tells us that if 1% change in the quantity of labour, brings about 50% changes in the national product, then the labour's share in national product will be 50%.

Similarly, in case of capital the relative share of capital will be,

$$K = \Delta Q / \Delta k \times k / Q$$

or

$$= \Delta Q / Q \times K / \Delta K$$

$$= \Delta Q / Q \div \Delta K / K$$

It shows the elasticity of total production with regards to capital. According to this theory the distribution of national income should be in the share of marginal product of the factors of production.

Check Your Progress

1. Name the inspiration for the theory of the subsistence level of wages.
2. Who were the critics of the wage fund theory?
3. What is the assumption of factors of production in the neo-classical marginal productivity theory?

12.4 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. The theory of subsistence level of wages was inspired by Theory of Population Control by Malthus.
2. T. Thornton, F.D. Longe, and Francis A Walker were the prominent critics of the wage fund theory.
3. All the factors of production are considered as substitutes for each other in neo-classical marginal productivity theory.

12.5 SUMMARY

- The subsistence theory is known to have been developed by physiocrats. These theories neglect the demand side of the labour market and concentrate on the supply aspects. As per this theory, the driver for real wages to the level of minimum required subsistence is the supply of workers.
- The minimum subsistence here constitutes as the fundamental needs such as shelter and food.
- The earliest mention of the concept of subsistence can be found in Smith's work *The Wealth of Nations*, but his successors like David Ricardo and

Thomas Malthus are the credited with its origin. While Smith's was a more positive outlook, Ricardo's and Malthus's was more pessimistic. The theory of subsistence level of wages was inspired by Theory of Population Control by Malthus.

- As per the proponents of theory of subsistence, the natural prices will be like the guiding forces of market price of labour. This is to say market prices and natural prices will not vary much. Therefore, with the increase in wages about the rate of subsistence, the number of workers in the market would also increase and this will then lead to bringing the rate of wages down. Similarly, if the wages fall below the rate of subsistence, the number of workers in the market will decrease and this will consequently lead to the pushing the rate of wages up.
- As per Smith, the demand for labour would increase only in proportion to the increase of the fund sort out for the payment of wages. Further, Ricardo, too maintained that a rise in the capital would drive the demand for labour upwards. These postulations led to the considerations that there is always a pre-decided 'fund' for the payment of wages.
- The nature of this wage fund is considered to be fixed at a particular point of time, but it may fluctuate over a period of time. The average of such a fund would then logically be calculated as the total divided by the number of workers.
- If the volume of the fund, whatever be its constituents, was large, the wages would be high and vice versa. And if the population increases more than the food and other subsistence related items, the wages would be directed towards the subsistence level.
- The theory of wage fund received acceptance for over 50 years by economists like Nassau William Senior and John Stuart Mill until it was discredited by W.T. Thornton, F.D. Longe, and Francis A Walker after 1865.
- It was considered that there were several other factors like the bargaining power of the labourers which affected the level of wages paid to the labourers. Even so, the wages-fund theory was prominent until the end of the 19th century.

NOTES

12.6 KEY WORDS

- **Subsistence:** It refers to the state of having what you need in order to stay alive, but no more.
- **Disposable Income:** It refers to the income remaining after deduction of taxes and social security charges, available to be spent or saved as one wishes.

12.7 SELF ASSESSMENT QUESTIONS AND EXERCISES

NOTES

Short Answer Questions

1. Write a short note on the origin of the wage fund theory.
2. Briefly explain the marginal productivity theory of wages.

Long Answer Questions

1. Explain the subsistence theory of wages.
2. Discuss the wage fund theory.

12.8 FURTHER READINGS

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

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

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

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

UNIT 13 THEORY OF INTEREST AND PROFIT

*Theory of Interest
and Profit*

NOTES

Structure

- 13.0 Introduction
- 13.1 Objectives
- 13.2 Classical Theory of Interest
- 13.3 Neo-Classical Loanable Funds Theory
 - 13.3.1 Keynes Liquidity Preference Theory
- 13.4 Modern Theory of Interest
- 13.5 Answers to Check Your Progress Questions
- 13.6 Summary
- 13.7 Key Words
- 13.8 Self Assessment Questions and Exercises
- 13.9 Further Readings

13.0 INTRODUCTION

Now that you have learnt about the theories of rent and wages, we move on to another important area economics, the theories of interest and profit. The theories of interest deal with the manner in which rate of interests are ascertained. There are four main theory of interest: Classical theory of interest propounded by the likes of Marshall, Pigou, Walrus and Knight, etc; the neo-classical or loanable funds theory introduced by Wicksell, Ohlin Robinson and A.C. Pigou; the Keynesian theory or interest (also known as the monetary theory of interest) and the modern theory of interest. In this unit, you will learn about these theories of interest. The next unit will discuss the theory of profit.

13.1 OBJECTIVES

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

- Discuss the classical theory of interest
- Examine the neo-classical theory of interest
- Describe the liquidity preference theory of interest
- Explain the modern theory of interest

13.2 CLASSICAL THEORY OF INTEREST

NOTES

To the classical economists, money was only a 'veil' that partially concealed but did not really alter any of the basic 'real' forces in the economy. This was as true of their interest theory as of other elements of their elegant structure. Their reasoning ran as follows:

The amount of borrowed money demanded by business firms depends essentially on the marginal productivity of capital. If the productivity of any added machine is expected to be high enough to more than cover the interest cost financing its purchase, it must be bought to maximize profits. However, additional capital goods, when combined with the other factors of production, will produce diminishing marginal productivity. Therefore, more such goods will be bought only at lower rates of interest. This shows that demand curve for capital goods is downwards sloping.

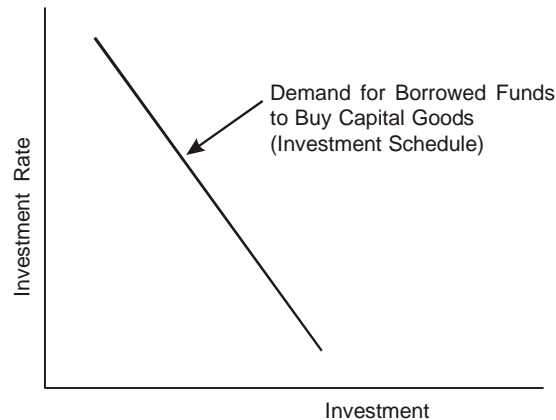


Fig. 13.1 Demands for Funds for Investment

On the other side of market, the supply of funds made available by savers is motivated by receiving interest on funds. A rational person would prefer some interest return on his savings to none. Therefore, as it is the reward of interest that motivates saving, the higher the rate of interest, the greater the quantity saved. The supply of savings, then, is the rising function of rate of interest.

Interest is what must be paid for the right to borrow money. But this amount is a price that, like all other prices, is determined by demand and supply. The demand for borrowed money comes almost entirely from business managers, who desire to use the funds to purchase capital goods. The supply, on the other hand, comes from individuals who, having decided not to spend all their income on consumption, have some left over in the form of saving, available to lend out.

Explanation

According to the classical approach, the rate of interest was a factor which brought into equilibrium the demand for investment and the supply of savings. Investment

represented the demand for investible funds and savings represented the supply of these funds while the rate of interest was that 'price' of the investible funds at which the supply of and the demand for investible funds were equated. Under the play of free market forces, the long-run equilibrium rate of interest rested at that point where the total amount of investment at that rate was equal to the total amount of saving at that rate. The classical interest rate theory, which was developed under the assumption of full employment of labour and capital, is a *flow* analysis in which both investment and saving are *flow* variables directing attention to a *period of time* rather than to a *point of time*. Since both investment and saving are flow variables, these can only be expressed as quantities per time unit. In short, the equilibrium relates to the capital market which must be continuously cleared of saving which represents flow on to the market and investment which represents flow off the market.

The classical theory of interest rate can be diagrammatically explained as shown in Figure 13.2 where the negatively sloping linear investment demand function I is a negative function of the rate of interest while the positively sloping linear saving supply function S is a positive function of the rate of interest. In other words,

$$I = f(r); \text{ and } \frac{dI}{dr} < 0$$

$$S = g(r); \text{ and } \frac{dS}{dr} > 0$$

In Figure 13.2, the positively sloping savings supply curve SS and the negatively sloping investment demand curve II intersect each other at point A corresponding to which the long-run equilibrium rate of interest is Or_e which will come to stay in the capital market since at this interest rate, the demand for and the supply of investible funds are equal leaving no scope for deviation of the rate of interest from the 'norm' as long as the underlying supply and demand schedules do not alter their position. The market rate of interest was regarded as fluctuating around this long-run equilibrium rate of interest.

The classical theory of the rate of interest has three important features. In the *first* place, it is a purely flow theory, i.e., the saving supply and investment demand variables which determine the rate of interest are the *flow* quantities as distinct from the *stock* quantities.

Secondly, both saving and investment are the *real* variables as distinct from the monetary variables. Savings represent the real resources which become available as a result of voluntary saving on the part of people. These are governed by the distribution of income and wealth, the nature of expected future income streams and wealth-owners' time preferences between the present and future goods. Investment represents net addition to economy's total capital stock involving an increase in the economy's total productive capacity. Although both saving and investment are reckoned in money units, money does not matter and it is *neutral*

NOTES

NOTES

in the economy. Expressed differently, operations in the capital market relate to the supply of and the demand for real securities. Those who borrow capital funds are the suppliers or sellers of real securities while the suppliers of capital funds are the purchasers of these securities. Money acts only as a mediating agent—as a mere veil. In the classical theory, the rate of interest is not determined by the quantity of money in circulation. Consequently, it is invariant with respect to changes in the money supply because any change in the money supply is neutralized by equi-proportionate change in the prices leaving the quantity of real money in the economy unchanged. Consequently, the demand and supply curves of money intersect at the same rate of interest.

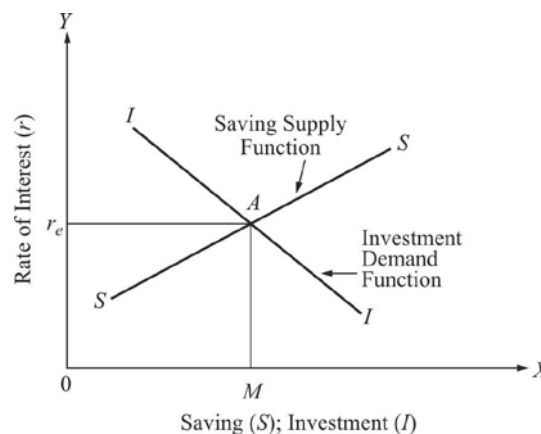


Fig. 13.2 Classical Equilibrium Interest Rate

Thirdly, if the equilibrium between saving and investment is disturbed due to the shift in any one of the functions, it is re-established exclusively through changes in the rate of interest without in anyway affecting any other variable. In other words, the theory is self-contained and is relevant only to the equilibrium of economy's real sector while the monetary sector of the economy is completely left out of the purview of the theory.

Criticism

The classical theory of interest rate has been criticized on several grounds. Besides others, it has been severely criticized by Keynes. In the first place, the classical theory of interest rate is incomplete because it considers only the real as distinct from the monetary and only the flow as distinct from the stock variables. The result is that both the stock as well as the monetary variables which affect the rate of interest are completely left out from the discussion of interest rate determination. Any complete theory worth its name should be broad enough to include both the stock and the money variables. This weakness of the theory which flows from its excluding the consideration of monetary factors was recognized by the neo-classical economists, including Swedish economist Knut Wicksell and noted Cambridge University economist Dennis Holme Robertson. These economists attempted to

synthesize the monetary and non-monetary or real variables by developing the loanable funds theory according to which the equilibrium rate of interest is that rate which equates the supply of and the demand for the loanable funds. It was recognized that the classical theory was faulty because it considered the current voluntary savings as the only source of supply and investment as the only source of demand for the **loanable funds**.

Starting from a more realistic understanding of the operations taking place in the capital market, the exponents of the loanable funds theory recognized that the flow of money on to the market could be increased or decreased by the activities of the monetary authorities—central bank and commercial banks—through credit creation or credit squeeze and through **hoarding or dishoarding** on the part of wealth-owners. Similarly, the flow of securities on to the market did not exclusively represent the issue of new securities to borrow the funds for investment; it was also fed by hoarding (borrowing in order to accumulate the surplus cash balances) on the part of wealth-owners.

Secondly, the classical theory of interest rate has been criticized by Keynes on several grounds. Keynes has questioned the nature of the classical savings supply and investment demand functions which are regarded as interest-elastic. According to Keynes, saving is more a function of income rather than of the rate of interest for at very low levels of income, people will not save at all (they will rather spend) even if they are offered inducement in the form of high rate of interest. Similarly, according to Keynes, investment is largely autonomous and at any rate the investment and interest rate relationship is a very weak and unreliable relationship. Had this relationship been strong, the monetary policy alone would have sufficed to ensure full employment in the economy and there would have been no necessity to resort to fiscal policy measures.

Besides questioning the form of the classical savings supply and investment demand schedules, Keynes has criticized the classical view that rate of interest is the ‘price’ which equates the demand for and the supply of investible resources. According to Keynes, ‘the rate of interest is not the ‘price’ which brings into equilibrium the demand for resources to invest with the readiness to abstain from present consumption.’ Asserting that the rate of interest is a *purely monetary phenomenon* as distinct from the classical real phenomenon Keynes has stated that ‘it is the ‘price’ which equilibrates the desire to hold wealth in the form of cash with the available quantity of cash...’ He has criticized the classical approach of regarding the rate of interest as a return on saving on the ground that a man who hoards his savings in the form of cash earns no interest although he saves. In Keynes’ view ‘the mere definition of the rate of interest tells us in so many words that the rate of interest is the reward for parting with liquidity for a specified period.’

According to Keynes, the classical theory of interest rate was faulty and misleading in another respect also. According to the theory, an upward or a downward shift in the investment demand schedule would result in the new

NOTES

NOTES

equilibrium rate of interest determined by the point of intersection between the new investment demand curve and the given savings supply curve. In other words, when the investment demand curve shifts the saving supply curve does not shift. Keynes' criticism of this approach was that the assumption of a given savings supply curve corresponding to a shifting investment demand curve was faulty and untenable. Consequently, the conclusions that followed were also faulty. Keynes argued that when the aggregate investment outlay changed, the aggregate income also changed. Consequently, the amounts saved at different interest rates also changed resulting in the shift of the saving supply curve simultaneously. If the aggregate **investment outlay** increases (decreases), the aggregate income must also increase (decrease) the multiplier time of the increase in investment outlay. In other words, it was faulty to assume the aggregate income as given when the aggregate investment was changing. And if the aggregate income changed when the aggregate investment changed (represented by shift in the investment demand curve), the saving supply curve would also appropriately shift because people would save different amounts out of the changed income at the different rates of interest. When both the saving supply and the investment demand curves shift simultaneously, the whole position becomes indeterminate.

Keynes' criticism of the classical theory of the rate of interest may be stated in his own words. He writes: 'The independent variables of the classical theory of the rate of interest are the demand curve for capital and the influence of the rate of interest on the amount saved out of a given income; and when, (e.g.,) the demand curve for capital shifts, the new rate of interest, according to this theory, is given by the point of intersection between the new demand curve for capital and the curve relating the rate of interest to the amounts which will be saved out of given income. The classical theory of the rate of interest seems to suppose that if the demand curve of capital shifts or if the curve relating the rate of interest to the amounts 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 positions of the two curves. But this theory holds no ground. For the assumption that income is constant is inconsistent with the assumption that these two curves can shift independently of one another. If either of them shifts, then, in general, income will change; with the result that the whole schematism based on the assumption of given income breaks down. In truth, the classical theory has not been alive to the relevance of changes in the level of income or the possibility of the level of income being actually a function of the rate of the investment.'

Illustrating his above views diagrammatically in Figure. 13.2, Keynes has further, stated that 'the functions used by the classical theory, namely, the response of investment and the response of amount saved out of a given income to change in the rate of interest, do not furnish material for a theory of the rate of interest'

Thirdly, the classical theory of rate of interest is faulty because it ignores the influence which the bank created money (credit) exerts on the rate of interest.

The classical theory is also faulty since it completely ignores the consideration of the asset demand for money and regards money as being demanded exclusively for the transactions purpose in order to remove the difficulties of barter. In short, the classical theory of the rate of interest is at best a half-baked faulty explanation of determination of the long-run equilibrium rate of interest. It cannot be accepted as a complete and scientific explanation of the complicated process through which the interest rate is determined in an economy where, far from being neutral, money plays an important and active role in shaping the entire processes of production and distribution by acting not only as the medium of exchange but also as the permanent store of value.

NOTES

Check Your Progress

1. According to the classical approach, what did the rate of interest do as a factor?
2. What is the role of money in classical theory of interest?

13.3 NEO-CLASSICAL LOANABLE FUNDS THEORY

There is much that is appealing and much that is correct in the classical interest theory. But it has not lasted the test of time—partly because of institutional and structural changes in the economy over the years and partly because of flaws in the underlying assumptions.

Critics' first query was—is saving out of current income really the only source of funds to be lent out? What about banks? Certainly as creators of money, they lend money *in addition* to what savers save. Is it really true that what is saved is *always* lent out at interest? Suppose savers *do* choose to hoard part of what they save in first year and then lend it out (in addition to second year's saving) in the following year? All this affects the interest rate in both years.

Second, what about demand side? True, business investors do borrow much of what is lent, but are they the only borrowers? Surely we know that the government has been a significant borrower in recent decades. And for that matter, consumers are by no means all net savers. They, too, borrow and pay interest.

Questions such as these resulted in the reformulation of the classical interest theory into a more complete explanation called the loanable funds theory of interest. The loanable funds interest theory is a logical extension and modification of the much older classical theory of interest.

The explanation of interest that has perhaps the greatest appeal to common sense is the loanable funds theory. It considers the interest rate, reasonably enough, to be simply *the price paid for the right to borrow and use loanable funds*. With this straightforward definition of interest as the focus, it then becomes necessary

NOTES

to consider the components that determine the demand for and supply of loanable funds.

Components of supply of loanable funds

The components of supply of loanable funds are as follows:

- 1. Gross saving out of current income
- 2. New money created by banks
- 3. Dishoarding

Let us consider a numerical example. Suppose that in first year (year 1), the following events occur:

	₹
Total Income received	800 billion
Total Consumption Spending	700 billion
Therefore, total saving	100 billion
Total amount lent out by savers at interest	75 billion
Therefore, total hoarding	25 billion
Total new money created by banks in making loans	50 billion

What is the total supply of loanable funds made available in year 1? Clearly, it is 125 billion. Banks put in 50 billion of newly created money. Consumers saved 100 billion, but not all of it was lent out; the 25 billion that was hoarded during year 1 must be subtracted to arrive at the correct figure.

Suppose now in second year (year 2) that:

	₹
Total Income received	900 billion
Total Consumption Spending	800 billion
Therefore, total saving	100 billion
Total amount lent out by savers at interest (100 billion from this year's saving, plus the 25 billion that was hoarded in year 1)	125 billion
Therefore, total dishoarding	25 billion
Total new money created by banks in making loans	50 billion

In year 2, the supply of loanable funds 175 billion – 100 billion from year 2 saving, 50 billion from banks, 25 billion from lending out money saved and hoarded last year.

Components of demand for loanable funds

*Theory of Interest
and Profit*

The components of demand for loanable funds are as follows:

1. Business demand to finance the purchase of capital goods (investment)
2. Government demand to finance deficit spending (or minus government tax surplus)
3. Consumer demand to finance consumer purchases on credit

In essence, then, the loanable funds theory of interest concludes that, other things remaining equal, the rate of interest will tend to rise as a result of an increase in business, government or consumer borrowing; consumer hoarding; a decline in the money supply; or a shift to the left in the savings schedule.

Graphically, the loanable funds theory of interest is shown in Figure 13.3, where S equals saving, DH stands for dishoarding, DM represents an increase in the money supply, I equals demand by business for funds to finance investment, G_d stands for the government budget deficit and C_d represents gross consumer borrowing. Note that the diagram assumes that dishoarding will occur at high interest rates and hoarding will be likely at low interest rates. The equilibrium rate of interest, as pointed out, is determined where the demand for loanable funds forthcoming from business, government and consumers equals the supply of loanable funds made available by savers, dishoarders and the banks.

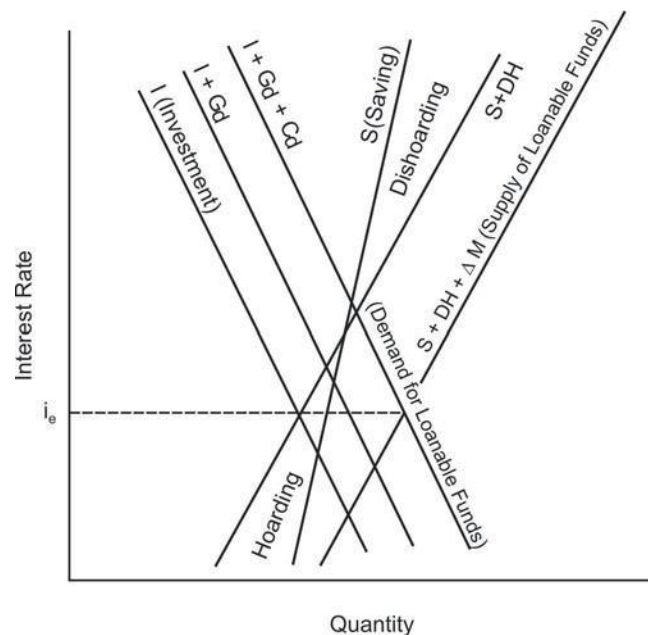


Fig. 13.3 Graphical Representation of Loanable Funds Theory of Interest

Criticism

Like the classical theory, the loanable funds theory of interest rate has also been criticized on various grounds. In the first place, the theory is neither a purely flow

NOTES

NOTES

theory nor a stock theory. It is a curious mixture of stocks and flows which hardly makes any sense. A misleading attempt has been made by the loanable funds theorists to add the stock quantities that are relevant to a *point of time* by showing these as flow schedules to the flow quantities that are relevant to a *period of time*. The successful effort to synthesize the *flows* and *stocks* was made by John R Hicks in 1937 in his general equilibrium formulation developed in his now well-known article published in *Econometrica* and which was subsequently pushed forward by Alvin H. Hansen so as to become now the well-known Hicks-Hansen general equilibrium synthesis of the classical flow theory and the Keynesian stock theory of interest rate.

Secondly, like the classical theory, the loanable funds theory is also indeterminate since the loanable funds' total supply schedule (since it includes voluntary savings which vary with income) varies with income. Not only does the saving portion of the supply schedule of loanable funds vary with income but even the 'new money and activated balances' which form part of the loanable funds rise and fall with the increase or decrease in the current income.

Thirdly, as is evident from Figure 13.4, according to the loanable funds theory, corresponding to the equilibrium rate of interest there is no equilibrium between current voluntary savings and the investment outlay and the disequilibrium between the two flow quantities is bridged by net hoarding. No difficulty arises so far. However, when this conclusion is presented in the form of a diagram through the aid of a set of interrelated schedules, a formidable problem of reconciling the stock and flow variables confronts us which cannot be got over without assuming a *period* too short enough to approach to a *point*. And the moment it is done, the distinction between a *period* and a *point* disappears and with it also disappear the concepts of *flow* and *stock*. In a way, then the entire loanable funds approach crashes like the frail house of cards and the theory is reduced to a mere pedagogic exercise with no bearing on the practical affairs.

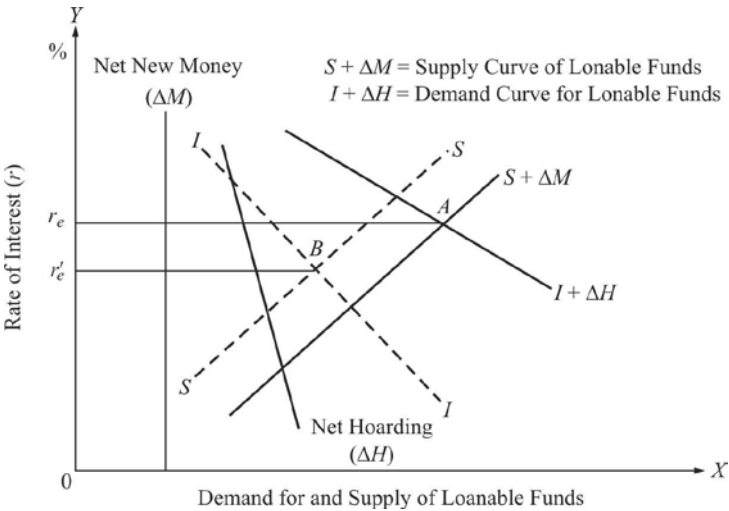


Fig. 13.4 Loanable Funds Theory

NOTES

The *fourth* criticism of the theory is that it is wrong to show the net new money and net hoarding as schedules in the diagram because (i) these changes in stocks are not functions of the rate of interest, and (ii) their interpretation in the schedules' sense clearly implies that at a certain rate of interest the monetary authorities go on adding continuously to the money supply in the economy at a constant or given rate per time period and similarly the people go on adding to their cash balances continuously at constant rate year after year. Such a conclusion hardly makes any sense without incorporating income growth in the analysis. Looking from the practical point of view, it has never been the experience of any real world economy that the central banking authority there ever indulged in such a regular credit creation or destruction nor does the behaviour of the wealth-owners support the argument.

The fact is that like those in the classical theory, in the loanable funds theory too both saving and investment are purely flow concepts while hoarding and new money are, in reality, the changes in stocks. Consequently, these represent 'additions to or deductions from' the market flows within the period during which the change is taking place. But a situation in which such change is taking place cannot be regarded as an equilibrium nor can a change in stocks from one level to a new desired level be commensurate with continuous flows. Voluntary saving and investment expenditure are continuous flows which, at any given rate of interest, will continue throughout the Marshallian 'short-run period' at least. In actual life, we may assume these rates of flows to persist over a year or more; it is otherwise with hoarding and the new money.

It is one thing to say that during a very short interval market conditions will be affected by the creation of a certain additional sum of new money; it is clearly a very different thing to imply that at a certain rate of interest the monetary authorities will go on creating new money continuously at a constant rate throughout a period of years'. The diagram gives this impression because by implication, the flow of the new money is commensurate with savings. Similarly, hoarding is commensurate with investment.

Criticizing the theory Hansen has stated that 'according to the loanable funds analysis the rate of interest is determined by the intersection of the demand schedule for loanable funds with the supply schedule. Now the supply schedule of loanable funds is composed of saving (in the Robertsonian sense) plus net additions to loanable funds from new money and the dishoarding of idle balances. But since the 'savings' portion of the schedule varies with the level of 'disposable' income, it follows that the total supply schedule of loanable funds also varies with income. Thus this theory is also indeterminate.'

13.3.1 Keynes Liquidity Preference Theory

The concept of liquidity preference was introduced by economist John Maynard Keynes in 1936 in his book *The General Theory of Employment, Interest and Money*. Keynes stated that people value money for 'the transaction of current

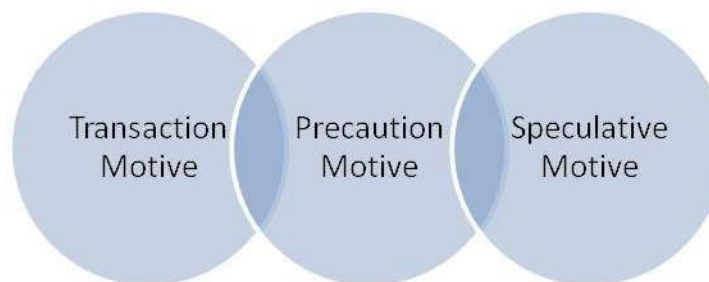
NOTES

business and its use as a store of wealth'. So, they will forego the option of earning interest on money that they want to spend currently, and that they want to have it on hand as a precaution. In contrast when the interest rates go up, people are keen to hold less money for these purposes to secure a profit.

Liquidity Preference Theory of Interest or Keynesian Theory of Interest

People demand money to maintain it as liquid. Keynes describes this demand or preference for money as liquidity preference. Giving loans involves a turn down in the stock of money detained as liquid. Put it differently, lending involves giving up liquidity by the lender.

Interest, as compensation, is made to encourage people to give up their liquidity. Keynes considered interest as the rate of the reward for parting with liquidity for a specific period. The currency money is entitled liquidity and the demand for money is called liquidity preference. According to Keynes, people demand/prefer liquidity because of the following three motives:



Transaction motive: Firms and individuals do many transactions on a daily basis and for that they need money in liquid form. This type of demand for liquidity for meeting day-to-day transaction is called demand for liquidity for transaction purpose. This demand further depends on the size of income, spending habits and time gap in the receipt of the income.

Precaution motive: This type of demand for liquidity is for unforeseen crisis, eventuality and misfortune. Both business firms and individual firms need liquidity for this purpose also and this depends on the size of income and vision and nature of the person.

Speculative motive: In this type of demand, individuals and business firms want to keep money to take advantages of the price fluctuations and earn profits when the price is high in the market.

Liquidity Preference Curve

Transactions balances and precautionary balances are held with the purpose of being used to make purchases as and when necessary, at times, these two objectives jointly called are demand for active balances.

The significant point regarding demand for active balances is that it does not react to changes in the rate of interest. In other words, we can say that the demand for active balances is interest inelastic. In contrast to this, demand for speculative balances (sometimes referred to as the demand for idle balances) is interest elastic.

When the rate of interest decreases, the demand for money will augment and when the rate of interest increases, the demand for money will reduce. Consequently, the demand for money is unenthusiastically associated with the interest rate and demand curve for money will incline down from left to right. There are mainly two sources of supply of money namely (i) government and (ii) the banking system. Money set into flow by the government is called legal tender money. The depositors can pull out their money from the bank and the deposited money is called bank liquidity.

After adding the demand for active (L_a) and idle balances (L_i) together, we can get community's total demand for money or liquidity preference schedule. As shown in the following diagram.

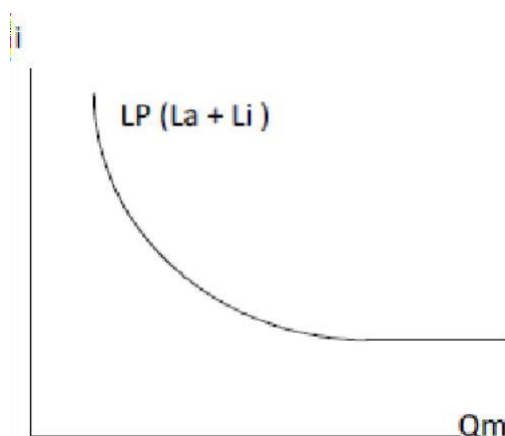


Fig. 13.5 Total Demand for Money or Liquidity Preference Schedule

Economist John Maynard Keynes illustrates liquidity preference theory in the book titled *The General Theory of Employment, Interest and Money*.

As per Keynes, people value money both for the transaction of current business and its utilization as a store of wealth.

Consequently, they will forego the capability to make interest on money that they desire to use in the present, and that they desire to contain it on hand as a safety measure. Alternatively, when interest rates increase, they turn out to be eager to grasp less money for these reasons to lock up profit.

Criticism

The liquidity preference theory of interest rate has not escaped criticisms. *First*, being completely a stock theory, the rate of interest in this theory is as much

NOTES

NOTES

indeterminate as it was in the classical theory. According to Alvin H. Hansen, 'The Keynesian theory, like the classical, is indeterminate. In the Keynesian case the money supply and demand schedules cannot give the rate of interest unless we already know the income level; in the classical case the demand and supply schedules for saving offer no solution until income is known. Keynes' criticism of the classical theory applies to his own theory.'

Secondly, the theory cannot explain the determination of the rate of interest in the long period because it focusses attention only on those factors which are relevant only in the short period.

Thirdly, the theory cannot explain the co-existence of the different interest rates on the basis of liquidity preference because interest rates will have to be perfectly uniform due to the perfect uniformity of money or cash balances.

The *fourth* criticism of this theory is that it is wrong to say that rate of interest is not the reward for saving or waiting as such. 'Keynes blissfully forgets that without saving or waiting, investment funds cannot be obtained. Jacob Viner stated the correct position when he asserted that 'without saving there cannot be liquidity to surrender... the rate of interest is the return for saving without liquidity.'

The *fifth* criticism of the theory is that although in the short period at any given point of time the rate of interest will be such that the community's total holding of cash must be equal to the total existing stock of money but it is equally true that over a long period the rate of interest must tend to that level where the saving and investment flows are in equilibrium.

Lastly, Keynes' basic proposition in his theory is that the rate of interest and the demand for money, more particularly the speculative demand for money, are inversely related. According to Don Patinkin, 'Keynes' analysis of the implication of this dependence is repeatedly marred by a confusion (which characterizes the later Keynesian literature as well) between his basic proposition that the amount of money demanded is inversely dependent upon the rate of interest and the completely different proposition that the equilibrium rate of interest is inversely dependent on the amount of money. His discussion of the liquidity preference in *The General Theory* shifts uninhibitedly from one proposition to the other with never an indication that they are in any way not identical. More specifically, there is never a recognition that, in our terminology, the first of these propositions describes an individual experiment and the second a market experiment, and that the truth of the first does not imply the truth of the second.'

Check Your Progress

3. What are the components of the supply of loanable funds?
4. What did Keynes consider interest as?

13.4 MODERN THEORY OF INTEREST

The modern theory of interest was developed by Hicks and Hansen. The main postulate of the modern theory of interest is that the equilibrium between the savings function, investment function, liquidity preference function and the quantity of money determine the rate of interest as well as the level of income at equilibrium. This theory was developed due to the shortcomings of the other previous theories, namely, that the classical theory of interest does account for any change in the level of income, the loanable fund theory only provides us the IS curve without pointing towards the rate of interest and the liquidity preference theory merely provides an LM curve without again telling the rate of interest. Let's see how the interaction between the IS and LM curve define the modern rate of interest.

The general equilibrium model of the economy comprises of the two parts. The first part brings together the determinants of equilibrium in the real sector or the goods market of the economy. The second part brings together the determinants of equilibrium in the money market or the monetary sector of the economy. The equilibrium in the real sector is defined in terms of the equality between the aggregate saving and aggregate investment. Corresponding to that aggregate real income where aggregate saving equals the aggregate investment ($S = I$), the aggregate demand for goods just equals the aggregate supply of goods in the economy, i.e., $C + I = Y$. Thus, the economy's real sector is in equilibrium at this level of the aggregate real income. The equilibrium of the economy's money market requires equality between the total supply of money and the total demand for money. The equality between the total supply of and demand for money furnishes us with the equilibrium rate of interest. Thus, the monetary sector of the economy will be in equilibrium at that rate of interest corresponding to which the total demand for money equals the total supply of money, i.e., where $M_d = M_s$.

The equilibrium aggregate income corresponding to which the aggregate saving equals the aggregate investment, i.e., $S = I$, partly depends on the conditions in the monetary sector. Similarly, the equilibrium rate of interest at which the total demand for money and the total supply of money are in equilibrium, i.e., $M_d = M_s$, partly depends on the conditions in the real sector or the goods market. To understand what is involved, a review of the simple Keynesian model of income and interest rate determination, as shown in Figure 13.6, is in order. With the given positively sloping savings supply curve SS and the investment demand curve I_1I_1 as shown in Figure 13.6(A), the equilibrium aggregate income is OY_1 because at this income the aggregate demand ($C + I$) and the aggregate supply of goods (Y) are in equilibrium. Expressed differently, in terms of the savings supply and the investment demand approach, corresponding to the OY_1 aggregate income, the aggregate saving and aggregate investment are in equilibrium. If the rate of interest enters as a variable in investment determination, it is then obvious that I_1I_1 investment demand curve has been drawn assuming some given interest rate, say of 6 per cent. A lower rate

NOTES

NOTES

of interest, say of 4 per cent, *ceteris paribus*, would cause an upward shift in the entire investment demand schedule from the position of the original investment demand curve I_1I_1 to the position of new investment demand curve I_2I_2 . This will raise the equilibrium aggregate income from OY_1 to OY_2 . We cannot, however, say anything about the rate of interest from Figure 13.6(A) as the analysis proceeds on the basis of certain assumed given rate of interest.

Figure 13.6(B) shows the equilibrium rate of interest. Given the demand function for money $M_d = L(Y, r)$ and the vertical supply function of money M_sM_s showing the money supply being autonomously determined by the monetary authority, the equilibrium rate of interest is Or_1 because at this rate of interest the demand curve for money and the supply curve of money intersect showing that at this rate of interest the total demand for money equals the total supply of money, i.e., $M_d = M_1 + M_2 = M_s$. Since the total demand for money is composed of the transactions demand (including the precautionary demand) for money (M_1) and the speculative demand for money (M_2) and since the transactions demand for money depends on income, the composite demand curve for money $M_d = L(Y, r)$ has been drawn assuming the aggregate money income to be Y_1 . A higher aggregate money income Y_2 , *ceteris paribus*, would give us a higher demand curve for money $M_d = L(Y_2, r)$ as shown in Figure 13.6(B). Consequently, given the same fixed money supply as represented by the vertical supply curve M_sM_s , the equilibrium rate of interest will rise to Or_2 . Figure 13.6(B) does not, however, show what the equilibrium aggregate income will be; it assumes a certain given level of the equilibrium aggregate income and then proceeds from there.

It is, therefore, obvious that the equilibrium aggregate income and the equilibrium rate of interest are so closely interdependent that one cannot be determined without first knowing the other. Both the aggregate income Y and the rate of interest r have, therefore, to be simultaneously determined. Although this cannot be done through Figure 13.6, there will, however, be a particular combination of income and interest rate corresponding to which the economy's real sector of Figure 13.6(A) and the monetary sector of Figure 13.6(B) would simultaneously be in equilibrium. This chapter develops a general equilibrium model furnishing the simultaneous solution of the equilibrium values for the aggregate income (Y) and the rate of interest (r) and explains some other important problems and the policy matters.

NOTES

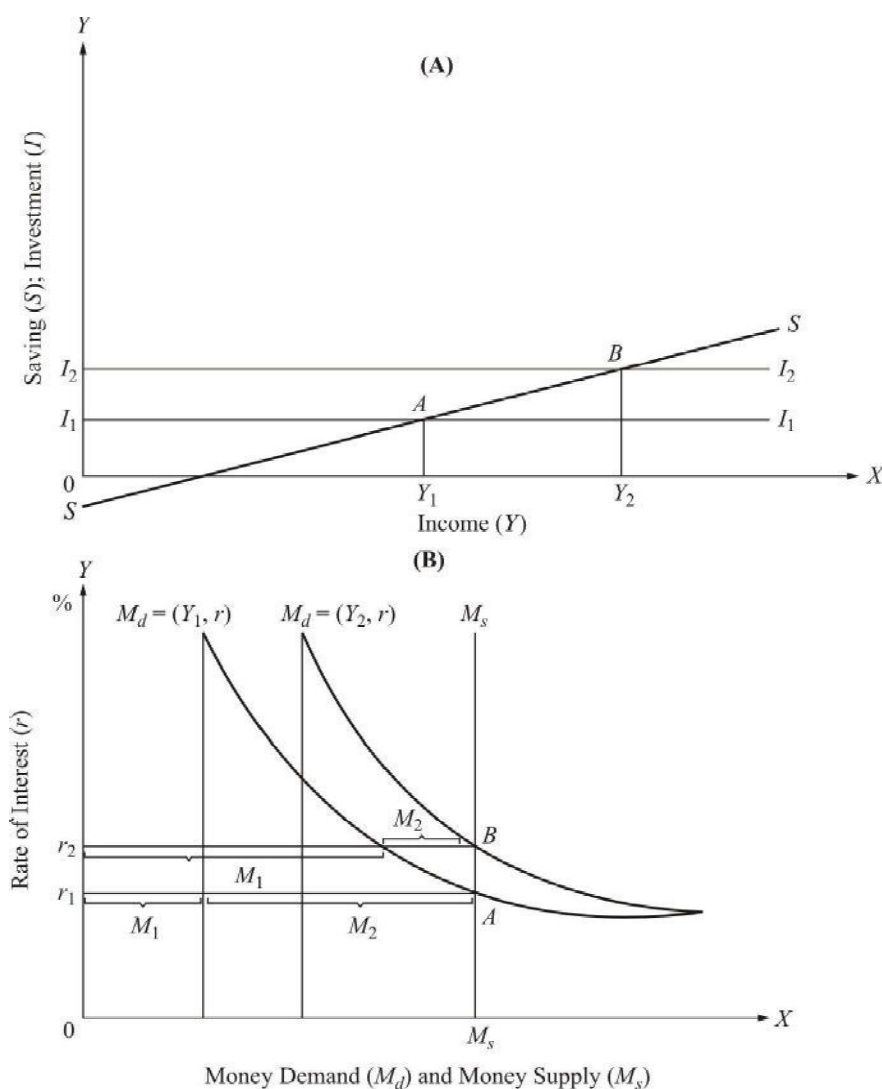


Fig. 13.6 Keynesian Model of Income and Interest Rate Determination

Real Sector (Goods Market) Equilibrium

Equilibrium in the economy's real sector requires equality between the aggregate saving (S) and the aggregate investment (I). Consequently, all those factors which cause changes in the aggregate saving and the aggregate investment will influence the determination of equilibrium in the real sector. For the simplicity of analysis, it is assumed that investment is a function of the rate of interest alone while saving is a function of income alone. The following set of equations represents the equilibrium in the goods market or real sector of the economy.

Savings supply function	$S = f(Y)$
Investment demand function	$I = g(r)$
Equilibrium condition	$f(Y) = g(r)$

NOTES

Figure 13.7, which has been divided into four parts, shows equilibrium in the real sector of the economy. Part *A* merely shows the $r \equiv r$ equality straight-line drawn at the 45° -angle from the origin of the axes. It is simply an equality-line since the rate of interest has been shown on both the axes. Part *B* of the figure shows the aggregate investment demand schedule *I* showing that investment is a negative function of the rate of interest. Part *C* of the figure shows the aggregate savings supply schedule *SS* demonstrating that saving is a positive function of income. Part *D* of the figure shows the *IS* curve which has been derived from other parts of the figure. This *IS* curve is the locus of all those different combinations of the aggregate income and the rate of interest at which the aggregate saving equals the aggregate investment, i.e., those particular income-interest combination pairs where $S = I$ showing that the aggregate supply equals the aggregate demand. Consequently, the *IS* curve shows the equilibrium of the real sector of the economy. To illustrate the derivation of the *IS* curve, we first find the amount of investment at different rates of interest from Part *B* of the figure and thereafter we find out those different levels of the aggregate income at which the aggregate saving equals the aggregate investment. For example, in Part *B* at the 6 per cent rate of interest, the total investment is ₹20 crore. To satisfy the equilibrium condition, the total saving must also be ₹20 crore. Part *C* shows that the aggregate saving will be ₹20 crore when the aggregate income (*Y*) is ₹150 crore. Bringing together the aggregate income *Y* of ₹150 crore from Part *C* and the rate of interest *r* of 6 per cent from Part *B* of the figure we obtain one combination of *Y* (₹150 crore) and *r* (6 per cent) at which $S = I$ or the real sector of the economy is in equilibrium. Plotting this equilibrium combination of *Y* and *r* in Part *D* of the figure gives us point *E*. Similarly, corresponding to the 4 per cent rate of interest the total investment is ₹40 crore. Part *C* of the figure shows that the total saving of ₹40 crore is possible when the total income is ₹200 crore. Accordingly, we have the second combination of *Y* (= ₹200 crore) and *r* (= 4 per cent) at which the equilibrium condition in the real sector $S = I$ is satisfied. Plotting this equilibrium combination of *Y* and *r* in Part *D* of the figure gives us point *F*. At an interest rate of 2 per cent, the total investment is ₹60 crore. To make this aggregate investment equal to aggregate saving, the aggregate income must be ₹250 crore. This gives us the third combination of *Y* and *r* at which the equilibrium condition of $S = I$ in the goods market is satisfied. Plotting this combination in Part *D* gives us point *G*. Similarly, the other equilibrium combinations of *Y* and *r* can be found out by taking the other different rates of interest and finding out those levels of aggregate income at which the aggregate saving equals the aggregate investment against that rate of interest. Joining the points *E*, *F* and *G* points gives us the negatively sloping *IS* curve which shows the equilibrium in the goods market or in the real sector of the economy.

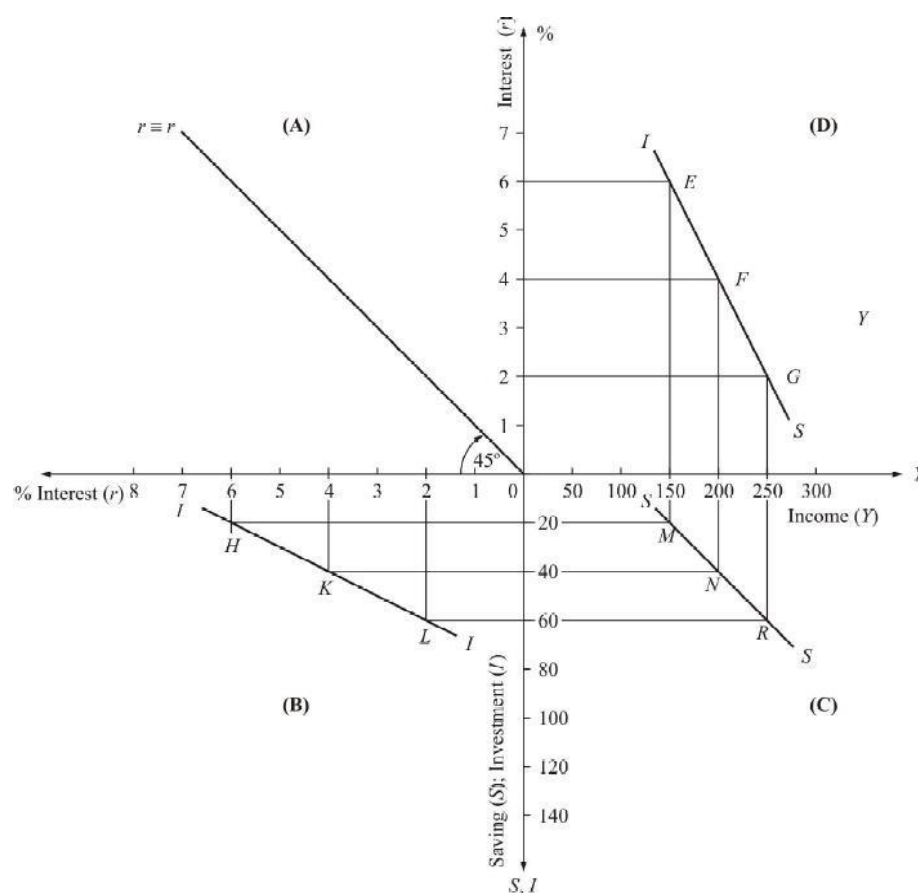


Fig. 13.7 Equilibrium in the Real Sector of the Economy

It is evident from the IS curve that there is a different income level for each different rate of interest at which saving is equal to investment, i.e., $S = I$. It is also evident that higher is the rate of interest, lower is the level of income at which $S = I$. The combination of lower aggregate income with a higher rate of interest follows from the fact that since investment is a negative function of the rate of interest, a higher rate of interest, *ceteris paribus*, means the lower aggregate investment spending which in turn through the multiplier action means lower aggregate income. Explained differently, this follows from the fact that the lower aggregate income means smaller aggregate saving. Since equilibrium requires equality between S and I , a low S means a low I and a low I results from a high r . The IS curve shows that corresponding to a higher rate of interest r , the economy's real sector will be in equilibrium at the lower level of aggregate income Y and *vice versa*. It does not, however, show that particular combination of Y and r which will exist in the market at any given time. The existence of all the different combinations of Y and r located on the IS curve are equally possible.

Monetary Sector (Money Market) Equilibrium

Equilibrium in the money market requires that the total supply of money and the total demand for money should be in equilibrium, i.e., $M_d = M_s$. Following John

NOTES

NOTES

Maynard Keynes we treat the transactions demand for money (including the precautionary demand for money) as a positive function of income alone and the speculative demand for money as a negative function of the rate of interest alone. The equilibrium in the money market is depicted by the following set of equations:

Transactions demand for money	$M_1 = kY$
Speculative demand for money	$M_2 = f(r)$
Equilibrium condition	$M_1 + M_2 = M_d = M_s$

The total supply of money M_s is determined autonomously in the economy by the monetary authority.

The equilibrium of economy's monetary sector and the derivation of the *LM* curve showing the different equilibrium combinations of the aggregate income (Y) and rate of interest (r) show the derivation of and shifts in the *LM* function. The *LM* curve shows that there is a different level of money income for each different rate of interest at which the total demand for money M_d equals the given total supply of money M_s . The *LM* curve slopes upward to the right showing that with the given money supply, the money market equilibrium is possible at those combinations of income and rate of interest at which a higher money income goes along with a higher rate of interest and *vice versa*. This follows from the fact that with the given money supply, at the higher income a larger transactions demand for money can be satisfied only by releasing a part of the speculative cash balances which is possible only at the higher rate of interest. Explained differently, at the higher rates of interest, the speculative demand for money will be low. Thus, with the given money supply, a larger part of the total money supply becomes available which can be absorbed in the transactions demand for money only at the higher income. Although the *LM* curve shows that equilibrium in the money market at the higher rate of interest is possible only corresponding to the higher aggregate income, it does not tell us anything about the particular equilibrium combination of Y and r since the existence of all the different combinations located on the *LM* curve are equally possible in the money market.

General Equilibrium—Integration of the Real and Monetary Sectors

We have seen that the equilibrium between S and I is possible at the different combinations of Y and r . Similarly, equilibrium between the total demand for money M_d and the total supply of money M_s is also possible at the different combinations of aggregate income Y and rate of interest r . There will, however, be only a single equilibrium combination of Y and r which will be unique in the sense that at this particular combination of Y and r both the real sector and the monetary sector of the economy are simultaneously in equilibrium, i.e., at this equilibrium combination of Y and r the double condition of equilibrium, namely $S = I$ and $M_d = M_s$ which is necessary for the real and the monetary sectors of the economy to be simultaneously

NOTES

in equilibrium, is satisfied. The particular combination of Y and r satisfying the double condition of simultaneous equilibrium of the economy's real and monetary sectors is obtained by getting the level of Y and r at the point of intersection of the IS and LM curves as shown in Figure 13.8. In the figure, the general equilibrium, where $I = S$ and $M_d = M_s$, is attained at the combination of $Y = ₹200$ crore and $r = 4$ per cent. Any other combination of Y and r is a disequilibrium combination in the sense that at this combination both the sectors of the economy will not be simultaneously in equilibrium. For example, corresponding to $Y = ₹75$ crore and $r = 8$ per cent, although the real sector of the economy is in equilibrium because $S = I$ but the monetary sector of the economy is in disequilibrium because at this combination of Y and r , the total demand for money is less than the total supply of money, i.e., $M_d < M_s$. This means that at an income of ₹75 crore after satisfying the transactions demand for money M_1 the residual money supply ($M_s - M_1$) exceeds the total speculative demand for money M_2 at 8 per cent interest rate. In other words, looking at the total money supply the combination of ₹75 crore of money income (Y) and 8 per cent rate of interest (r) is not the money market clearing combination.

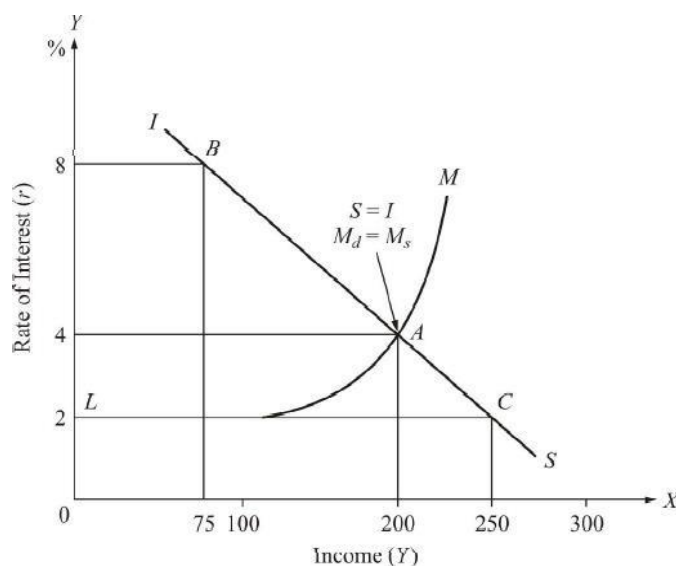


Fig. 13.8 Simultaneous Equilibrium of Economy's Real and Monetary Sectors

Corresponding to $Y = ₹75$ crore, r should be 2 per cent if the money market has to be in equilibrium. However, at $r = 2$ per cent the equilibrium in the goods market requires that Y should be ₹250 crore. Thus, while at $Y = ₹75$ crore and $r = 2$ per cent combination of Y and r although the money market is in equilibrium, the goods market is in disequilibrium because at this combination of Y and r the aggregate saving is less than the aggregate investment, i.e., $S < I$. Consequently, in order to make $S = I$ —a condition which must be satisfied if the goods market has to be in equilibrium—either Y must increase sufficiently (r remaining unchanged) to generate the larger aggregate saving to become equal to the investment or else r must rise sufficiently (Y remaining unchanged) to reduce the aggregate investment

NOTES

sufficiently to become equal to the aggregate saving at $Y = ₹75$ crore. Alternatively, partly Y must rise and partly r must fall simultaneously involving a process by which the equilibrium will be attained at point A.

There are an infinite number of disequilibrium combinations of Y and r at which S is either more or less than I and M_d is either more or less than M_s . Assuming that there takes place no shift in any of the schedules underlying the IS and LM curves, the adjustment process will restore equilibrium at the point of intersection of the IS and the LM curves.

Change in General Equilibrium

A shift either in the LM curve or the IS curve or simultaneously in both the curves will cause a shift in the point of intersection of the IS and LM curves causing a change in the general equilibrium combination of Y and r . Shifts in the IS function are caused by shifts either in the investment demand function or in the saving supply function or simultaneously in both these functions. Shifts in the LM function are caused by shifts either in the aggregate money supply function or the aggregate money demand function or simultaneously in both these functions.

Increase in Aggregate Investment

Let us assume that as a consequence of innovations that have taken place in the economy, the aggregate investment demand function shifts upward so that at each rate of interest the aggregate investment demand is higher than before. Consequent upon an increase in the aggregate investment spending at each rate of interest resulting from an upward shift in the investment demand function, the IS function will also shift upward to the right such that (ignoring the LM function for the moment) the increase in the aggregate income at each rate of interest is equal to the initial increase in the aggregate autonomous investment times the investment multiplier, i.e., $\Delta Y = \Delta I \cdot K$. With an increase in aggregate investment of ₹20 crore and MPS of 0.4, i.e., with the value of investment multiplier of 2.5, the IS function must shift upward to the right so as to give $\Delta Y = \Delta I / \text{MPS} = ₹50$ crore. Figure 13.9 shows the original and the new IS functions IS_1 and IS_2 respectively. The original equilibrium in Figure 13.8 was attained at the aggregate income-interest rate combination of $Y = ₹200$ crore and $r = 4$ per cent. Consequent upon the upward shift in the investment demand schedule, the new equilibrium is now attained at $Y = ₹225$ crore and $r = 5$ per cent corresponding to the intersection of the new higher IS function IS_2 and the original LM function at point B which lies to the right and above the old intersection point A. If the rate of interest had not risen, the equilibrium income Y would have increased by ₹50 crore from ₹200 crore to ₹250 crore and the increase in income (ΔY) would have been equal to the full increase of ₹20 crore in investment outlay (ΔI) times the investment multiplier of 2.5, i.e., $\Delta Y = \Delta I \cdot K = ₹50$ crore. The actual increase in the aggregate income, however, is only ₹25 crore which is less than the multiplier (K) times the given initial increase in the autonomous investment outlay of ₹20 crore. What is the reason for this dampened

increase in equilibrium income? The reason for this smaller increase in income is that the increase in investment spending by raising the level of income also increases the transactions demand for money.

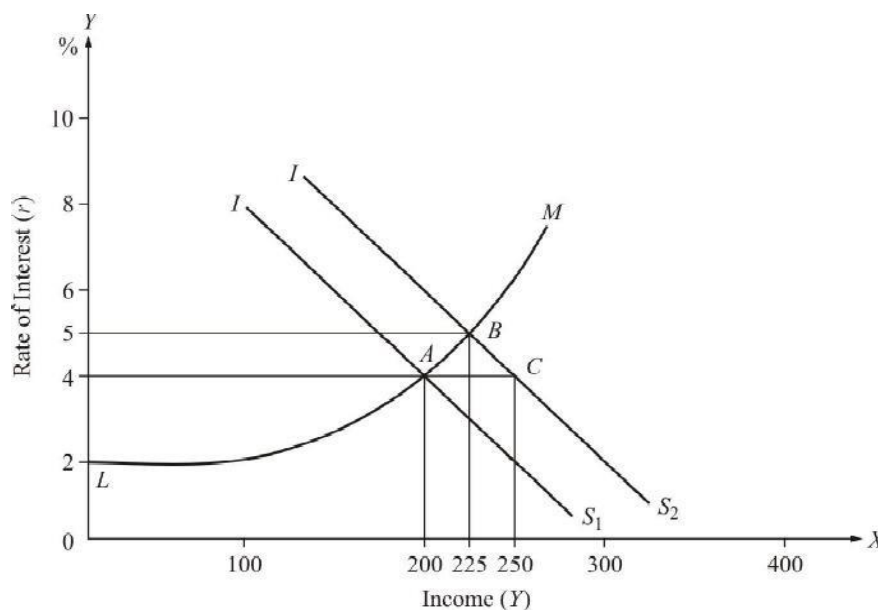


Fig. 13.9 Increase in Aggregate Investment

With the given money supply, the increase in the transactions demand for money corresponding to a higher income (with no decrease in the speculative demand for money) raises the rate of interest which is necessary in order to induce the wealth-holders to release a part of their speculatively held cash balances to meet the increased transactions demand for the cash balances. Since investment is a negative function of the rate of interest, a rise in the rate of interest by recoiling on the increase in investment spending reduces the increase in investment outlay from ₹20 crore to ₹10 crore. Consequently, the total increase in income is reduced from ₹50 crore to ₹25 crore. In this manner, the expansionary income effect of a given increase in the aggregate investment demand is dampened by an increase in the rate of interest brought about by an increase in the aggregate income. The size of the dampening effect will depend partly on the extent of the increase in the rate of interest which in turn will depend on the elasticity of the LM curve and partly on the interest elasticity of the investment demand.

Increase in Money Supply

We may now consider the effect of a given increase in the total money supply on the equilibrium income and the rate of interest. Let us assume that the aggregate money supply increases by ₹20 crore, i.e., $\Delta M_s = ₹20$ crore. Consequently, the supply function of money M_s will shift to the right. The aggregated demand function for money M_d remaining unchanged, the increase in the total money supply of ₹20 crore will cause a rightward shift in the LM function to the extent of ₹50 crore at each rate of interest so that treating the LM functional alone (ignoring for the moment

NOTES

NOTES

the IS function), the aggregate income will rise by ₹ 50 crore at each rate of interest. Equilibrium between the demand for money and the increased supply of money calls for an increase in the aggregate income sufficient to absorb the entire ₹ 20 crore increase in the supply of money in the transactions demand for money M_1 . Since $M_1 = kY$, $Y = M_1/k$ and $\Delta Y = \Delta M_1/k$. With the value of k —the proportion or fraction of its income which the community keeps in the form of money being given as 0.4, ΔY must be equal to ₹ 50 crore in order to give the new equilibrium between the demand for and the supply of money at each rate of interest. In Figure 13.10, the old and the new LM functions as shown by LM_1 and LM_2 have been combined with the original IS_1 .

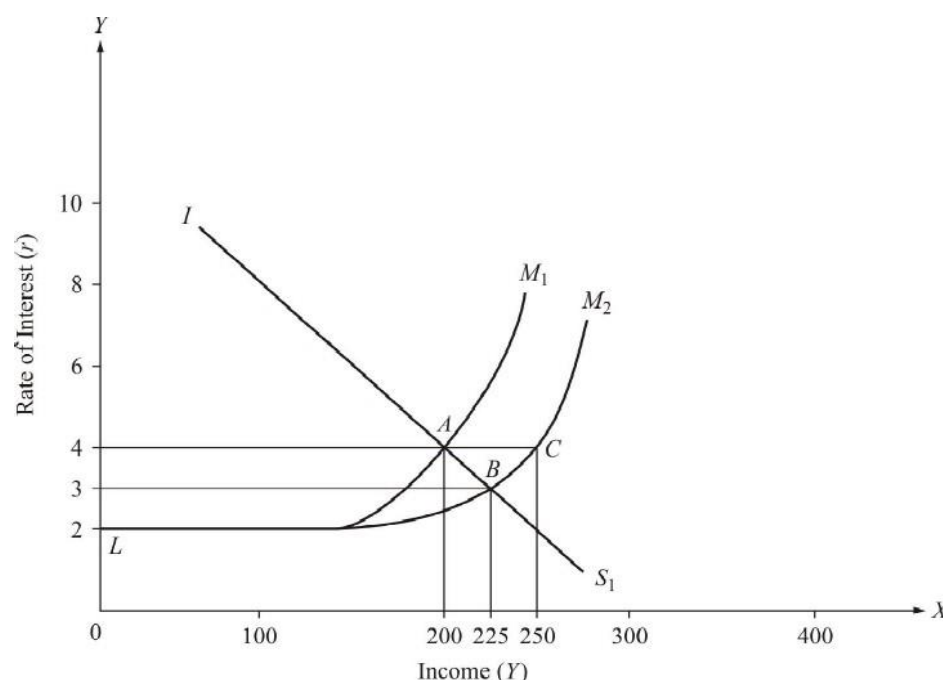


Fig. 13.10 Increase in Money Supply

Originally, the general equilibrium given by the intersection of the IS_1 and LM_1 functions was attained at $Y = ₹ 200$ crore and $r = 4$ per cent. The new general equilibrium resulting from the rightward shift in the LM function gives the new equilibrium combination of income and interest rate at $Y = ₹ 225$ crore and $r = 3$ per cent. Although the increase in the money supply of ₹ 20 crore shifts the LM_1 function rightward to the position of LM_2 function by ₹ 50 crore corresponding to each rate of interest, but with no shift in the IS function the equilibrium income increases by only ₹ 25 crore from ₹ 200 crore to ₹ 225 crore. The reason why the actual increase in the general equilibrium income is less than the shift in the LM function alone shows is simple. With the given increase in the money supply and with the given demand curve for money, the rate of interest will fall which in turn will raise the aggregate income. The fall in the rate of interest, however, also increases the speculative demand for money. Consequently, the effective increase in the total supply of money in the context of an expansionary income effect becomes

less than the actual increase of ₹20 crore in the aggregate money supply. This being so, the increase in the aggregate income required to absorb the smaller effective increase in the total supply of money will be less than ₹50 crore. The increase in aggregate income would be ₹50 crore only if the rate of interest remained unchanged at 4 per cent. But with the fall in the rate of interest from 4 per cent to 3 per cent, a part of the increased money supply is absorbed in the increased demand for speculative cash balances. The fall of 1 per cent in the rate of interest causes the aggregate investment demand to increase by ₹10 crore, which with the investment multiplier of 2.5 causes the aggregate equilibrium income to increase by ₹25 crore. Although the increase in equilibrium income is less than ₹50 crore the increase in income is only ₹25 crore when only the money supply increases, nevertheless the income increases purely due to the pursuit of an expansionary monetary policy. In short, monetary policy can help bring about changes in the aggregate output in the economy. However, the extent of the total increase (decrease) in the equilibrium income which a given increase (decrease) in the money supply can bring about will depend on (i) the extent of the fall in the rate of interest which in turn will depend up on the interest elasticity of the speculative demand for money function; and (ii) the extent of the increase in the aggregate investment demand which in turn will depend up on the interest elasticity of the investment demand. Provided that the interest rate falls consequent upon an increase in the money supply and provided that the investment increases as the rate of interest falls, some increase in the equilibrium aggregate income must take place as a result of the increase in the aggregate money supply.

Simultaneous Increase in Investment and Money Supply

We have discussed the effects of a given increase in the aggregate investment and a given increase in the money supply on the level of equilibrium income and interest rate. We may now consider the effect on the general equilibrium income and interest rate of a simultaneous given increase in the aggregate investment and the money supply. In other words, we will discuss the effect on the equilibrium income of the expansionary monetary and fiscal policies applied simultaneously by the government and the monetary authority. As a result of the simultaneous increase in the aggregate investment spending and the increase in the money supply, both the *IS* and the *LM* functions will shift. The shift in the investment demand function will cause the *IS* function to shift upward to the right from the position of *IS*₁ curve to that of *IS*₂ curve while the increase in the money supply will shift the *LM* function from the position of *LM*₁ curve to the position of *LM*₂ as shown in Figure 13.11. As a consequence of the simultaneous shifts in the *IS* and *LM* functions, the general equilibrium combination of income and rate of interest changes from $Y = ₹200$ crore and $r = 4$ per cent to $Y = ₹250$ crore and $r = 4$ per cent. The increase in the aggregate autonomous investment spending raises the aggregate income.

NOTES

NOTES

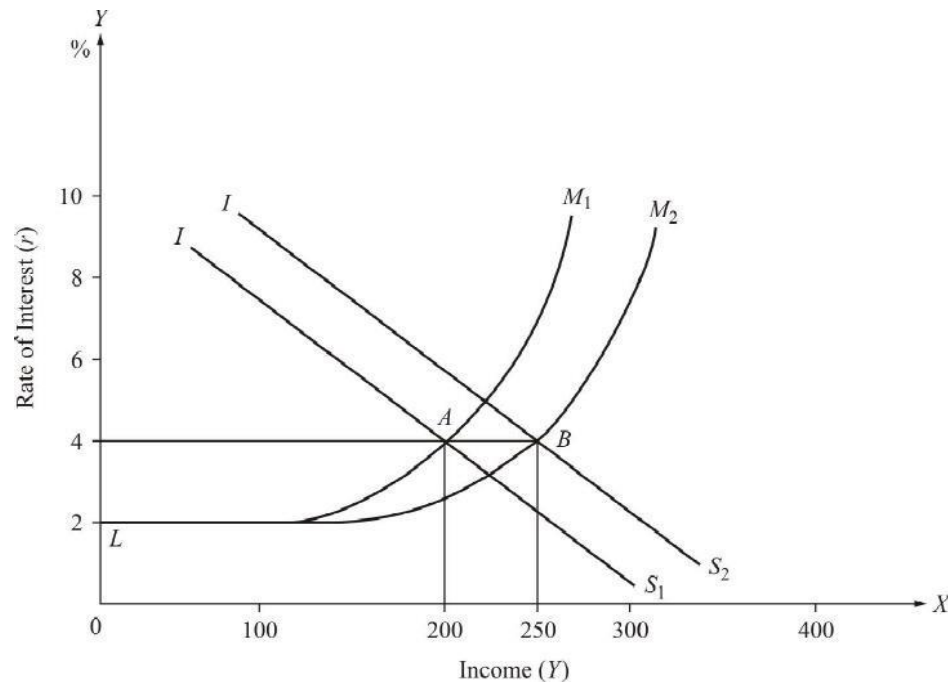


Fig. 13.11 Effect of Investment Demand Function on IS Curve

But with the total money supply remaining unchanged, the increase in income is dampened by the increase in the rate of interest resulting from an increase in the transactions demand for money arising from an increase in the level of income. If the supply of money increases just enough to prevent the increase in the rate of interest which would have occurred consequent upon an increase in the level of income, the full income expansionary effect of the given initial increase in the aggregate investment will be materialized. The increase in the equilibrium income from ₹200 crore to ₹250 crore as a consequence of a ₹20 crore increase in the investment spending with a multiplier of 2.5 is just equal to the increase in aggregate income which would result in the simple Keynesian theory of income determination. This analysis shows that the equilibrium income will increase by the full amount of the increase in the aggregate investment spending times the investment multiplier only if the increase in the total money supply exactly matches the increase in the investment spending so as to prevent any increase in the rate of interest.

The effects of shifts in the other functions on the equilibrium income can be analysed in a similar manner. For example, an increase in the aggregate consumption spending which would be shown by the downward shift in the aggregate saving supply function in Figure 13.7(C), will cause the IS function to shift to the right raising both the equilibrium income and the equilibrium rate of interest. An increase in the speculative demand for money (M_2) expressed as a shift to the right in the demand function for money will, by causing a shift to the left in the LM function, raise the equilibrium rate of interest and reduce the equilibrium income. On the other hand, a decrease in the transactions demand for money M_1 resulting from a

given fall in the value of k occasioned by a change in the payments practices or in the community's habits reflected in the practice of economy in the use of money for the transactional purpose will, by shifting the LM function to the right, cause the equilibrium rate of interest to fall while raising the level of equilibrium aggregate income.

Government Spending, Taxes and General Equilibrium

When the government spending and taxes are included in the analysis, the equilibrium condition of $S = I$ in the goods market for a two-sector economy becomes $C + S + T = C + I + G$ for a three-sector economy which includes the household, business and government sectors. The equilibrium condition $S + T = I + G$ means that the aggregate demand for goods will be equal to the aggregate supply of goods when the diversions or leakages in the form of personal savings and government taxes from the real income stream just equal the injections inserted into the real income flow stream in the form of private business investment and government spending. The equilibrium condition for the money market is the same as before, i.e., $M_d = M_s$. The effect of a given increase in government spending, it was stated that the equilibrium aggregate income will increase by the full amount of the deficit-financed increase in government spending times the multiplier. However, this conclusion has to be qualified when we include the money market in our analysis. An expansionary fiscal policy aiming at increasing the equilibrium income through a deficit-financed increase in government spending will fail to produce the maximum possible increase in the level of equilibrium income if the increase in government spending is not accompanied by an appropriate increase in the money supply.

In balanced budget multiplier it is assumed that both government spending (G) and taxes (T) are independent of the level of income, being determined autonomously, the multiplier was found to be one. In that simple model, the increase in income was equal to the increase in government budgetary spending. With the inclusion of money market in the analysis, the rate of interest also becomes the equilibrium income determining variable. The actual multiplier will now be less than the earlier simple balanced budget multiplier having the value of one because the increase in government spending with the balanced budget while raising the aggregate income will also tend to raise the rate of interest by raising the transactions demand for money, assuming no increase in the aggregate money supply and no decrease in the speculative demand for money. The rise in the rate of interest will dampen the increase in equilibrium income. Thus, a fiscal policy aiming at raising the income while maintaining a balanced budget will not produce the maximum possible increase in income if it is not accompanied by an expansionary monetary policy to prevent the rise in the rate of interest and the consequent fall in private investment spending. The $IS-LM$ function analysis may be elaborated by introducing

NOTES

the more realistic fiscal policy assumptions. For example, by treating taxes (T) as a function of income (Y), the effect of an increase in the balanced budget on the equilibrium income (Y) and the rate of interest (r) can be analysed.

NOTES

Check Your Progress

- 5. When will the monetary sector of the economy be in equilibrium?
- 6. What causes a shift in the LM function?

13.5 ANSWERS TO CHECK YOUR PROGRESS
QUESTIONS

- 1. According to classical approach, the rate of interest was a factor which brought into equilibrium the demand for investment and supply of savings.
- 2. In classical theory of interest, money is only a mediating agent. Herein, the rate of interest is not determined by the money in circulation.
- 3. The components of the supply of loanable funds are as follows: gross saving out of current income, new money created by banks and dishoarding.
- 4. Keynes considered interest as the rate of reward for parting with liquidity for a specific period.
- 5. The monetary sector of the economy will be in equilibrium at that rate of interest corresponding to which the total demand for money equals the total supply of money.
- 6. Shifts in the LM function are caused by shifts either in the aggregate money supply function or the aggregate money demand function or simultaneously in both these functions.

13.6 SUMMARY

- To the classical economists, money was only a ‘veil’ that partially concealed but did not really alter any of the basic ‘real’ forces in the economy.
- According to the classical approach, the rate of interest was a factor which brought into equilibrium the demand for investment and the supply of savings.
- The classical interest rate theory, which was developed under the assumption of full employment of labour and capital, is a *flow* analysis in which both investment and saving are *flow* variables directing attention to a *period of time* rather than to a *point of time*.
- The loanable funds interest theory is a logical extension and modification of the much older classical theory of interest.

NOTES

- The explanation of interest that has perhaps the greatest appeal to common sense is the loanable funds theory. It considers the interest rate, reasonably enough, to be simply *the price paid for the right to borrow and use loanable funds*. With this straightforward definition of interest as the focus, it then becomes necessary to consider the components that determine the demand for and supply of loanable funds.
- People demand money to maintain it as liquid. Keynes describes this demand or preference for money as liquidity preference. Giving loans involves a turn down in the stock of money detained as liquid. Put it differently, lending involves giving up liquidity by the lender.
- Interest, as compensation, is made to encourage people to give up their liquidity. Keynes considered interest as the rate of the reward for parting with liquidity for a specific period. The currency money is entitled liquidity and the demand for money is called liquidity preference.
- The modern theory of interest was developed by Hicks and Hansen. The main postulate of the modern theory of interest is that the equilibrium between the savings function, investment function, liquidity preference function and the quantity of money determine the rate of interest as well as the level of income at equilibrium.

13.7 KEY WORDS

- **Interest:** It refers to the amount which must be paid for the right to borrow money.
- **Liquidity Preference:** As per Keynes, the demand or preference for money to maintain it as liquid is known as liquidity preference.

13.8 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short Answer Questions

1. As per the classical theory of interest, why is the demand curve for capital goods downward sloping?
2. What are the three important features of the classical theory of interest?
3. What were the important questions over classical theory of interest which led to the development of neoclassical theory of interest?
4. What are the motives for which people demand/prefer liquidity?
5. Briefly discuss the money sector equilibrium.

NOTES

Long Answer Questions

1. Explain the classical theory of interest. Mention its criticism.
2. Discuss the loanable funds theory of interest.
3. Describe Keynes' liquidity preference theory of interest.
4. Examine the general equilibrium in the real and monetary sectors. Also discuss the increase and decrease in money and investment.

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 PROFIT AND INNOVATION THEORY

*Theories of Profit and
Innovation Theory*

NOTES

Structure

- 14.0 Introduction
- 14.1 Objectives
- 14.2 Major Theories of Profit
- 14.3 Schumpeter's Innovation Theory
- 14.4 Answers to Check Your Progress Questions
- 14.5 Summary
- 14.6 Key Words
- 14.7 Self Assessment Questions and Exercises
- 14.8 Further Readings

14.0 INTRODUCTION

While discussing the 'marginal productivity theory', we noted that, in the long-run competitive equilibrium, the reward for each factor, including the reward for 'entrepreneurship' equals the value of its marginal product. It implies that, according to the marginal productivity theory, 'profit', which is the reward for 'entrepreneurship' equals the value of its marginal product. There are, however, a number of other profit theories developed by various economists over time. Here, we briefly review some important profit theories. The unit will conclude with a discussion on Schumpeter's innovation theory.

14.1 OBJECTIVES

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

- Discuss the important theories of profit
- Explain Schumpeter's innovation theory

14.2 MAJOR THEORIES OF PROFIT

The meaning and source of 'profit' have always been a centre of controversy. –The word 'profit' has different meanings to businessmen, accountants, tax collectors, workers and economists... For example, 'profit' to a layman means all incomes that go to the capitalist class'. To an accountant, profit means the excess of revenue over all paid-out costs including both manufacturing and overhead expenses. For all accounting purposes, businessmen use accountants' definition of profit.

NOTES

For all practical purposes, profit or business income means profit in accountancy sense plus non-allowable expenses. Economist's concept of profit is of 'pure profit' called 'economic profit' or 'just profit'. Pure profit is a return over and above the opportunity cost, i.e., the income which a businessman might expect from the second best alternative use of his resources. The two concepts of profit are discussed below in detail.

Accounting Profit vs. Economic Profit

The two important concepts of profit in business decisions are 'economic profit' and 'accounting profit'. It will be useful to understand the difference between the two concepts of profit. As already mentioned, in accounting, profit is surplus of revenue over and above all paid-out costs, including both manufacturing and overhead expenses. Accounting profit may be calculated as:

$$\text{Accounting profit} = TR - (W + R + I + M + OC)$$

where W = wages and salaries, R = rent, I = interest, M = cost of materials, OC = other paid out costs like electricity, transportation, etc.

While calculating accounting profit, only explicit or book costs, i.e., the cost recorded in the books of accounts, are considered.

The concept of 'economic profit' differs from that of 'accounting profit'. Economic profit takes into account also the implicit or imputed costs. The implicit cost is opportunity cost. Opportunity cost is defined as the payment that would be 'necessary to draw forth the factors of production from their most remunerative alternative employment'. Alternatively, opportunity cost is the income foregone which a businessman could expect from the second best alternative use of his resources. For example, if an entrepreneur uses his capital in his own business, he foregoes interest which he might earn by purchasing debentures of other companies or by depositing his money with joint stock companies for a period. Furthermore, if an entrepreneur uses his labour in his own business, he foregoes his income (salary) which he might earn by working as a manager in another firm. Similarly, by using productive assets (land and building) in his own business, he sacrifices his market rent. These foregone incomes—interest, salary and rent—are called *opportunity costs* or *transfer costs*. Accounting profit does not take into account the opportunity cost whereas all these costs are taken into account while working out the *economic profit*.

In addition, it should also be noted that in working out *economic* or *pure profit* a provision is also made for (a) insurable risks, (b) depreciation, and (c) necessary minimum payment to shareholders to prevent them from withdrawing their capital. Pure profit may thus be defined as 'a residual left after all contractual costs have been met, including the transfer costs of management, insurable risks, depreciation and payments to shareholders sufficient to maintain investment at its current level'. Thus,

$$\text{Economic profit} = \text{Total Revenue} - (\text{Explicit Cost} + \text{Implicit Cost})$$

In real life situation, if economic profit is greater than zero, one thing is certain that people will invest their money in business. However, in case economic profit is zero or insignificantly different from zero, people prefer to invest their money in business because it has a better future prospect in spite of risk involved.

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.

NOTES

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 economist 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 visualized 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 organization 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.

NOTES

2. Clark's theory: Distribution of wealth: A theory of wages, interest and profits

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 visualized 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 world, 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 organization
- (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

*Theories of Profit and
Innovation Theory*

With the passage of time, profits resulting from the inventions and improvements in production methods disappear. What happens, in fact, is that competition forces other entrepreneurs to imitate or innovate the new technology. This leads to rise in demand for labour and capital. Consequently, wages and interest rise and cost of production increases. On the other hand, with larger employment of labour and capital, production increases leading to fall in product prices. The ultimate result is that profits disappear. In Clark's own words, 'profit is an elusive sum which entrepreneurs grasp but cannot hold. It slips through their 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 for ever. In fact, under dynamic conditions, the generic changes continue to take place: it is a continuous process. The process of dynamic change gives entrepreneurs opportunities time and again to adjust their business to the changing conditions, make inventions and improve production methods, with a view to make pure profit.

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 conditions. Therefore, profit does not result from risk-bearings.

To sum up, according to J.B. Clark, profit is a reward for coordinating managerial functions of entrepreneurs under dynamic conditions. It is a reward for dynamism. It is not a reward for risk bearing. Pure profit, according to him, is a residue that remains after interest and wages are paid. That is, the difference between the gross receipts and payments for wages and interest represents profit.

Criticism of Clark's theory: Clark's theory, though impressive, has failed to win unqualified acceptance and has been criticized 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

NOTES

NOTES

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 is settled down to almost routine conditions and yet profits are 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.

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 add-on 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 actuarial 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 actuarial 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 actuarial value of risk. According to Hawley profit consists of two parts: **first**, represents compensation for actuarial 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 criticized on the following grounds:

In his reaction to the risk theory of profit, Clark remarked that the profit visualized 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 losing of capital) is interest, it is not profit.

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.

Hawley's theory of profit concentrates only on risk-bearing element, and ignores other entrepreneurial functions, viz., organization and coordination, which also lead to organization profit.

It is also argued that Hawley failed to make a distinction between predictable and unpredictable risks. Predictable (or foreseeable) 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 puts 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.

Carver observed that profits are a reward for avoiding risk and not for bearing risk, because only those entrepreneurs who are able to avoid risk make profits.

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.

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

NOTES

NOTES

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 visualized 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; (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.

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 uncertainty, many are not. For example, suppose that a person is betting in a horse race. If he has the knowledge of age, training, rearing, etc., of different horses and their jockeys, he would be operating in the region of *risk*. And, if he does not have the knowledge about the horses and jockeys participating in the race, he would be regarded as operating in the area of *uncertainty*. But, if he has some knowledge about the horses and/or jockeys, it will be difficult to decide whether the person is operating in the area of risk or in the area of uncertainty.

By considering profit as a reward exclusively for bearing uncertainty, 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.

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.

14.3 SCHUMPETER’S INNOVATION THEORY

The innovation theory of profit was developed by Joseph A. Schumpeter. His theory of profit is, in fact, the constituent of his *theory of economic development*. According to him, economic development takes place only when there are innovations in goods and service, manufacturing techniques, and methods of supply. Innovations are made and introduced by the business firms to make *pure profit*—profit in excess of normal profit. Introduction of innovations creates conditions for additional investment and labour employment and this leads to economic development and generate new business opportunities and also for profit.

Schumpeter developed his theory of economic development by assuming a stationary economic condition in the country. Under the condition of stationary equilibrium, demand is equal to supply, prices are equal to cost, and total revenue of firms is exactly equal to total cost. Under these conditions, firms make only normal profit, i.e., *managerial wages*—there is no *pure profit*. Pure profit, i.e., profit in excess of *management wages*, can be made only by making innovations in goods and service, manufacturing techniques and in the methods of supply goods. According to him, innovations may include:

- (i) introduction of new products and/or better quality goods and services,
- (ii) introduction of a new production technology,
- (iii) creating or finding new sources of raw materials,
- (iv) opening new markets for the innovated products, and
- (v) introduction of a more efficient and innovative management.

These kinds of innovations provide opportunities to innovative firms to fix a price of their product higher than the static equilibrium price. In simple words, innovative firms charge a price higher than production cost and hence make a *net profit*. Thus, according to Schumpeter, innovation is the source of profit.

Theory of Innovation

According to Schumpeter, –business cycles are almost exclusively the result of innovations in the industrial and commercial organizations. By innovations he means –... such changes of the combination of the factors of production as cannot be effected by infinitesimal steps or variations on the margin. [Innovation] consists primarily in changes in methods of production and transportation, or changes in industrial organization, or in the production of a new article, or opening of a new market or of new sources of material...». Innovations do not mean inventions. Innovations are simply the commercial application of new techniques, new materials, new means of transportation and new sources of energy. According to Schumpeterian theory, innovations are the cause of cyclical fluctuations.

In his formal approach to the business cycles theory, Schumpeter has developed a model in two stages which he calls the *first approximation* and the *second approximation*. The first approximation deals with the initial impact of the innovative ideas, and the second approximation deals with the subsequent waves that are created by the application of innovations.

The *first approximation* of Schumpeter's model starts with the economic system in equilibrium in which there is no involuntary unemployment; each firm has its $MC = MR$ and P (price) $= AC$. Under the conditions of complete equilibrium in the economy, if an innovation in the form of a new technique or a new process of production is introduced, it will have to be financed through bank credit. For, the economy being in equilibrium, there is no surplus fund to finance the new adventure. With additional funds available from the banking system, the innovating firms go on bidding higher prices for other inputs with a view to withdraw them from other uses. Due to increased spending in the economy, prices begin to rise. This process is further accelerated when other firms imitate the innovation and

NOTES

NOTES

acquire additional funds from the banks. With the widespread adaptation of the innovation, output begins to flow into the market. The phase of expansion gets underway. But beyond a certain level, increased output causes a decrease in price and profitability. Since further innovations do not come by quickly, there would not be additional demand for funds. Rather, the firms which had earlier borrowed from the banks start paying back. This leads to contraction in money supply. Hence prices fall further. The process of recession begins and continues until equilibrium is once again restored.

The *second approximation* of Schumpeter's model analyzes the secondary waves that are created by the first approximation. The main element in the secondary wave is *speculation*. When the primary wave of expansion begins, investors in capital goods industries, expect the upswing to be permanent. With this expectation, existing firms borrow heavily. Even consumers anticipating higher price in future go into debt to acquire durable consumer goods. This heavy indebtedness causes a problem when prices begin to fall. Debtors, both investors and consumers, find it extremely difficult to meet their obligations. This situation leads to a panic and then to depression.

Critical Evaluation

Schumpeter's theory has been criticised on the ground that it does not offer a sound theory of trade cycle. As M.W. Lee puts it, –An objective evaluation of Schumpeter's theory of the cycle is not only difficult but also unavailing because most of his arguments are based on _sociological rather than economic factors'. Hence, this theory can hardly be put to test. Besides, Schumpeter's theory is not basically different from over-investment theory: it differs only in respect of the cause of variation in the investment when the economy is in state of equilibrium. Not only that, this theory too, like many other theories, leaves out many other important factors causing fluctuations. Innovation is only one of the factors and not the sole factor.

Check Your Progress

1. What does the economic profit being zero or insignificantly different from zero indicate?
2. What do profit theories reveal?
3. Name the economist whose theory is called the residual theory of profit.
4. What does the second approximation of Schumpeter's theory deal with?

14.4 ANSWERS TO CHECK YOUR PROGRESS
QUESTIONS

1. In case economic profit is zero or insignificantly different from zero, people prefer to invest their money in business because it has a better future prospect in spite of risk involved.

2. The profit theories reveal only the source of profit and not the determination of profit rate.
3. F B Hawley's theory of profit is also called as the residual theory of profit.
4. The second approximation of Schumpeter's theory deals with the subsequent waves that are created by the application of innovations.

NOTES

14.5 SUMMARY

- While discussing the 'marginal productivity theory', we noted that, in the long-run competitive equilibrium, the reward for each factor, including the reward for 'entrepreneurship' equals the value of its marginal product. It implies that, according to the marginal productivity theory, 'profit', which is the reward for 'entrepreneurship' equals the value of its marginal product.
- For all practical purposes, profit or business income means profit in accountancy sense plus non-allowable expenses. Economist's concept of profit is of 'pure profit' called 'economic profit' or 'just profit'. Pure profit is a return over and above the opportunity cost, i.e., the income which a businessman might expect from the second best alternative use of his resources.
- One of the most widely known theories advanced to explain the nature of profit was formulated by economist 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 add-on 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.
- Frank H. Knight treated profit as a residual return to uncertainty bearing—not risk bearing. Obviously, Knight made a distinction between risk and uncertainty. He divided risks into calculable and non-calculable risks.
- The innovation theory of profit was developed by Joseph A. Schumpeter. His theory of profit is, in fact, the constituent of his theory of economic development. According to him, economic development takes place only when there are innovations in goods and services.
- Schumpeter developed his theory of economic development by assuming a stationary economic condition in the country. Under the condition of stationary equilibrium, demand is equal to supply, prices are equal to cost,

NOTES

and total revenue of firms is exactly equal to total cost. Under these conditions, firms make only normal profit, i.e., managerial wages—there is no pure profit. Pure profit, i.e., profit in excess of management wages, can be made only by making innovations in goods and service, manufacturing techniques and in the methods of supply goods.

14.6 KEY WORDS

- **Economic Profit:** In the economic terms, it refers to pure profit which is a return over and above the opportunity cost, i.e., the income which a businessman might expect from the second best alternative use of his resources.
- **Innovations:** As per Schumpeter, innovations are simply the commercial application of new techniques, new materials, new means of transportation and new sources of energy. They are the cause of cyclical fluctuations.

14.7 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short Answer Questions

1. Write a short note on profit and pure profit.
2. Differentiate between accounting and economic profit.
3. Briefly discuss Walker's theory of profit.
4. What are the criticisms against Knight's theory of profit?

Long Answer Questions

1. Examine the Clark's theory of profit.
2. Discuss Schumpeter's innovation theory. Mention the criticism against it.

14.8 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.