

---

## UNIT 9 RETURNS TO SCALE

---

### Structure

- 9.0 Objectives
- 9.1 Introduction
- 9.2 Concept of Returns to Scale
  - 9.2.1 Increasing Returns to Scale
  - 9.2.2 Constant Returns to Scale
  - 9.2.3 Diminishing Returns to Scale
- 9.3 Economies and Diseconomies of Scale
  - 9.3.1 Internal Economies of Scale
    - 9.3.1.1 Real Internal Economies of Scale
    - 9.3.1.2 Pecuniary Internal Economies if Scale
  - 9.3.2 Internal Diseconomies of Scale
  - 9.3.3 External Economies
  - 9.3.4 External Diseconomies
- 9.4 Let Us Sum Up
- 9.5 References
- 9.6 Answers or Hints to Check Your Progress Exercises

---

### 9.0 OBJECTIVES

---

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

- state the concept of returns to scale;
- distinguish between the stage of increasing, constant and diminishing returns to scale; and
- explain the concepts of economies and diseconomies of scale (both internal and external).

---

### 9.1 INTRODUCTION

---

Sometimes to increase the level of output, all factors are increased simultaneously and factor proportions are held constant. This is known as expansion in scale. In this context, three phases of production are discussed: increasing returns to scale, constant returns to scale, and diminishing returns to scale. Expansion of scale confers a number of economies i.e. advantages on the firm – both internal and external. Internal economies, in turn, can be divided into real internal economies of scale and pecuniary internal economies. If the scale of production is continuously expanded, a stage of internal diseconomies of scale sets in i.e. after a certain point, increase in production is less than proportionate increase in the factors of production. In this unit, we propose to discuss all these issues. We shall also explain the concept of external economies and external diseconomies.

## 9.2 CONCEPT OF RETURNS TO SCALE

The concept of returns to scale is associated with the tendency of production that is observed when the ratio between the factors is kept constant but the scale is expanded, that is use of all the factors is changed in same proportion.

When all the factors of production (labour, capital, etc.) are increased in the conditions of constant techniques, three possibilities arise:

- 1) Output increases in a greater proportion as compared to the increase in the factors of production. This is the case of **increasing returns to scale**.
- 2) Output increases in the same proportion as the increase in the amount of the factors of production. This is the case of **constant returns to scale**.
- 3) Output increases in a smaller proportion as compared to the increase in the amounts of the factors of production. This is the case of **diminishing returns to scale**.

We can illustrate these three situations with the help of numerical examples as follows:

OUTPUT SCHEDULE-1				
Input-X	Input -Y	% Change in Inputs	Output	% Change in Output
2	4	100	1000	-
4	8	100	3000	200
8	16	100	10000	233
16	32	100	35000	250

It can be observed from the given output schedule that:

- 1) At all the stages, we are increasing the quantity of inputs by 100%.
- 2) With increase in the quantity of inputs, the quantity of output is increasing by more than 100% at all stages. In other words output is increasing proportionately more than the increase in input.

Compare this situation with the illustrations given below:

OUTPUT SCHEDULE-2				
Input-X	Input -Y	% Change in Inputs	Output	% Change in Output
2	4	100	1000	-
4	8	100	2000	100
8	16	100	4000	100
16	32	100	8000	100

Output shedule-2 indicates that:

- 1) Inputs increase by 100% at each stage.
- 2) Output also increased by 100% at each stage.

In this illustration, output increased by the same proportion in which inputs have been increased.

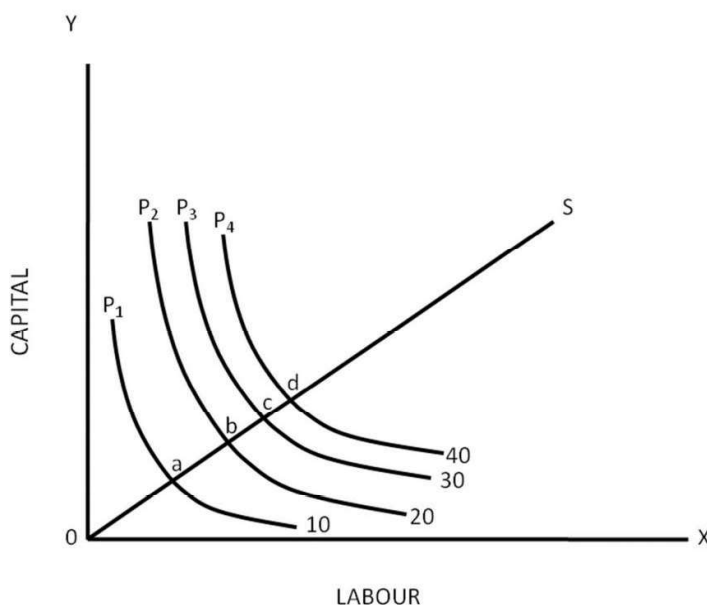
Compare this situation with the one given below:

OUTPUT SCHEDULE-3				
Input-X	Input -Y	% Change in Inputs	Output	% Change in Output
2	4	100	1000	-
4	8	100	1800	80
8	16	100	2500	39
16	32	100	3000	20

It would be observed that the total output at all stages increases less proportionately than the increase in inputs.

### 9.2.1 Increasing Returns to Scale

**When the ratio between the factors of production is kept fixed and the scale is expanded, initially output increases in a greater proportion than the increase in the factors of production.**



**Fig. 9.1: Increasing returns to scale output increases in a greater proportion than the increase in the factors of production**

For example, if factors are doubled the output is more than doubled. In other words, to double the quantity of the output, it is not necessary to double the quantity of the factors of production. This can be understood with the help of Fig. 9.1. In this figure  $P_1$ ,  $P_2$ ,  $P_3$ ,  $P_4$  are isoquants. They show 10, 20, 30, 40 units of output respectively.  $OS$  is the scale line which is cut by the isoquants at unequal distances. In the figure, it can be seen that  $cd < bc < ab < oa$ . This means that to enable the firm to rise from isoquant  $P_1$  to  $P_2$  (so that production increases from 10 to 20 units), the amount of factors of production required is less than the amount required to produce the initial 10 units of output.

Similarly, to increase the output further by 10 units so as to reach isoquant  $P_3$ , the amount of factors of production required is less than the amount required to produce the earlier 10 units of output as  $bc < ab$ . This position seems to hold true till isoquant  $P_4$ . There are three main factors which account for increasing returns to scale:

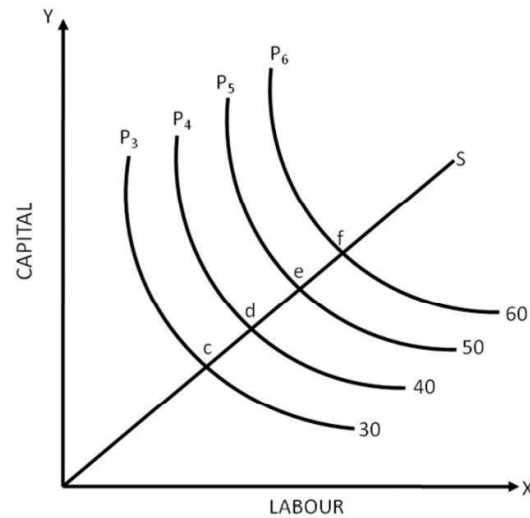
- 1) **Indivisibility:** The most important reason of increasing returns to scale is the ‘technical and managerial indivisibilities’. The meaning of an indivisible factor of production is that there is a certain minimum size of the factor and even if it is large in relation to the size of the output, it has to be used (i.e., it cannot be divided). For example, even if only 10-15 letters are to be despatched from an office, it would be necessary to keep a typewriter. It is not possible to purchase only half the typewriter since only a small number of letters have to be typed daily. We would, therefore, say that typewriter is not divisible. In a similar way, plants and managerial services in modern factories are not divisible. Accordingly, when the scale of production is enlarged initially, there is no equi-proportionate increase in the demand for the factors of the production.
- 2) **Specialisation:** Chamberlin does not regard indivisibility as an important cause of ‘increasing returns to scale’. According to him, the main reason of increasing returns to scale is specialisation. When due to division of labour, workers are given jobs according to their ability, their productivity increases while cost declines. According to Donald S. Watson, acknowledgement of this fact contradicts the assumption that the ratio of different factors of production remains constant. Accordingly, he casts doubts whether specialisation can be regarded as leading to increasing returns to scale. The importance of specialisation can be accepted only if we assume that although an increase by an equal amount in quantity of labour and capital employed is necessary for an expansion in scale, this increase does not mean the doubling or trebling their units employed but it does mean an increase in their fixed money cost. But this can lead to technical changes and it is very much possible that increasing returns emerge not due to an expansion in scale but due to technical reasons.

### 9.2.2 Constant Returns to Scale

**Increasing returns to scale can be obtained only upto a point. After this point is reached, expansion of scale only leads to equal proportionate change in output.**

Empirical evidence suggests that the phase of constant returns is a fairly long one and is observed in the case of a number of commodities. In a scientific sense, constant returns to scale implies that when the quantity of the factors of production is increased in such a way that the ratio of the factors remains unchanged, output increases in the same proportion in which the factors are increased. In other words, when the quantity of the factors is doubled, the output also doubles. Such a production function is often called linear homogeneous production function or homogeneous production function of the first degree. The phase of constant returns to scale can be understood with the help of Fig. 9.2. In this figure, when the firm goes from isoquant  $P_3$  to  $P_4$ , or

from isoquant  $P_4$  to  $P_5$  or from isoquant  $P_5$  to  $P_6$ , constant returns to scale are obtained. The fact  $cd = de = ef$  on the scale line indicates this phenomenon.



**Fig. 9.2: Constant returns to scale-output increases in the same proportion in which inputs are increased**

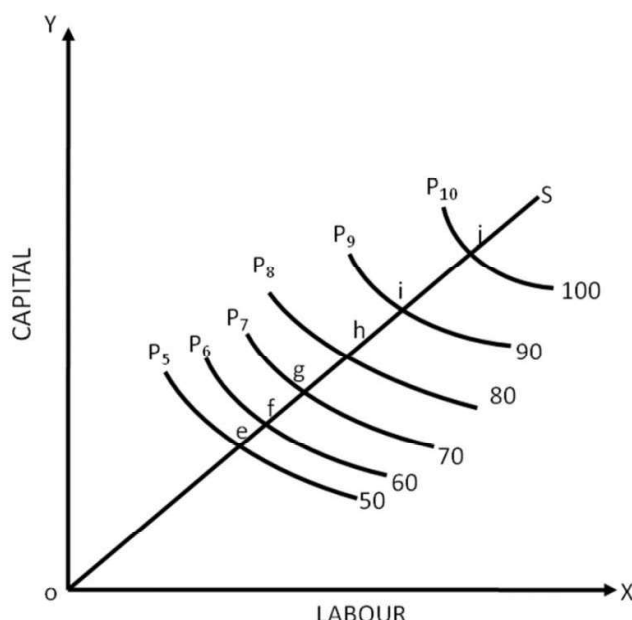
The question that now arises is what are the reasons which account for constant returns to scale. Generally when inefficiencies of production on a small scale are overcome and no problems regarding technical and managerial indivisibilities remain, expansion in scale leads to a situation where returns increase in the same proportion as the factors of production. Some economists are of the view that when benefits of specialisation of a factor in the unit of production are small or when such benefits have already been reaped at a small level of production, then for a considerable period of time, production increases according to the law of constant returns to scale.

Economists have argued that if the factors of production are perfectly divisible, the production function must exhibit constant returns to scale. In their opinion, if constant returns to scale does not prevail in some industries, it is because in these industries either due to scarcity or indivisibility of some factors, it is not possible to vary all them in the same proportion. Indivisibility of a factor often results in its under-utilisation at lower levels of output. When a producer for obtaining a larger output increases quantities of other factors, the amount of the lumpy factor which had not been fully utilised at lower levels of output, will not be increased. These economists do not think that economies of scale will be available when the factors of production are perfectly divisible. They however, stress the role of optimum factor proportionality in production. When factors of production are perfectly divisible, they can be increased or decreased in such amounts that an optimum proportion between factors is achieved. The output can be increased or decreased by increasing, or decreasing the amounts of the factors in the optimum proportion without any economies or diseconomies of scale which means that constant returns to scale will necessarily prevail.

### **9.2.3 Diminishing Returns to Scale**

Diminishing returns to scale ensure that the size of the productive firms cannot be infinitely large. Generally after a limit when the quantity of the factors of production is increased in such a way that the proportion of the factors remains unchanged, output increases in a smaller proportion as compared to increases

in the amounts of the factors of production. For example, it may happen that an increase in amount of labour and capital by 100 per cent leads to an increase in output by only 75 per cent. In other words, if output has to be doubled, the factors of production will have to be more than doubled. We can understand this phenomenon with the help of Fig. 9.3. In this figure, when the firm is at isoquant  $P_6$ , the tendency of constant returns to scale has come to an end. From here, the increasing distance between two consecutive isoquants is an indication that to obtain the same increase in output, factors of production will have to be increased at a higher and higher rate. On the scale line  $OS$ ,  $ij > hi > gh > fg > ef$  indicate this phenomenon very explicitly.



**Fig. 9.3: Diminishing returns to scale – output increases proportionally less than inputs**

Economists do not agree on the causes which leads to operation of diminishing returns to scale. Nevertheless, the two causes that are often mentioned are as follows:

- 1) **Enterprise:** Some economists emphasise that enterprise is a constant and indivisible factor of production and its supply cannot be increased even in the long run. Accordingly, when the quantity of other factors is increased and the scale of production expanded in a bid to boost up production, the proportion of other factors in relation to enterprise increases. Beyond a certain point, this results in diminishing returns as enterprise becomes scarce in relation to other factors.
- 2) **Managerial difficulties:** According to some other economists, the main reason for the operation of diminishing returns to scale is managerial difficulties. When the scale of production expands, the co-ordination and control on different factors of production tend to become weak and therefore output fails to increase in the same proportion as the factors of production increase. This results in diminishing returns to scale.

---

## 9.3 ECONOMIES AND DISECONOMIES OF SCALE

---

Expansion of the scale confers a number of economies on the firm. Some of these are in 'real terms' while others are in 'pecuniary terms'. Economies that are obtained in production work, marketing, management, transport, etc. are in real terms, while economies that are obtained in terms of, say, purchase of inputs at wholesale rate, availability of finance at lower rate of interest, saving on advertisement costs, etc. are in money terms. Then, there are certain economies that do not accrue to the firm whose scale of operation is large but accrue to certain other firms which benefit from the large scale of this firm.

**In Economics, those economies which accrue to a firm on expansion of its own size are known as internal economies. As against this, those economies which accrue to a firm not due to its own operations but due to the operations of other firms are termed external economies.**

### 9.3.1 Internal Economies of Scale

Generally, when the scale of production is sought to be enlarged, the firm replaces its small plant by a larger plant. This increases the efficiency of production. However, it is not always necessary to change the plant for expanding the scale of production. The firm can keep the old plant in a running condition and either establish a new plant of the same type or a new plant of some new type. In all these alternatives, the firm obtains many different kinds of economies. The fact is that it is the economies of scale that determine the nature of the long-run average cost curve.

#### 9.3.1.1 Real Internal Economies of Scale

When expansion in the scale of production takes place, the firm obtains some real internal economies. These economies accrue in the form of saving in the physical quantities of raw materials, labour, fixed and variable capital, and other inputs. Broadly speaking, real internal economies are of the following four types: (i) production economies, (ii) selling or marketing economies, (iii) managerial economies, and (iv) economies in transport and storage.

- 1) **Production economies:** When the scale of production expands, a number of economies accrue to the firm in the production process itself. First, opportunities for obtaining various types of economies emerge in the workshop of the factory. Production on a large scale enables the firm to carry out extensive division of labour and employ large automatic machines. The capacity of the machines is also fully used on account of the large volume of production. Instead of depending on others for carrying out repairs of machines and machine tools, the firm can itself employ technicians and workmen for the purpose. Techniques of production are changing so rapidly in the modern world that every producer has to remain ever alert. Large size of the firm and large-scale of operations is distinctly better in this regard since a larger firm can easily make use of its big financial resources to conduct research in its laboratories and/or adapt technology discovered elsewhere to suit its own requirements. The firm is, thus, able to discover better and less expensive techniques of production.

Whatever be the scale of operations, some waste material is invariably left out in each factory. If the scale of operations is small, a relatively larger quantity of this material goes unutilised. However, if the scale of operations is large, some useful goods can be prepared even from the waste material. For example, from the syrup left out in the sugar factory, liquor can be prepared. Similarly, in bangles factory, a number of small glass goods can be prepared out of the broken bangles.

When the scale of production is small, the producer generally cannot afford a packaging department. Therefore, he has to depend on others for obtaining packaging material like boxes, labels, etc. This leads to a substantial expenditure on packaging. However, if the scale of production is large, the firm can setup its own packaging department which is economical and also leads to lower per unit packaging costs.

- 2) **Selling or marketing economies:** Every producer produces with the purpose of selling. Therefore, he has to incur some expenditure in making his goods available to the consumer. When the scale of production is large, the per unit expenditure of the producer on marketing of goods is reduced substantially due to a number of reasons. All firms advertise their products in a number of ways. Even very small firms have to spend a certain minimum amount on advertising, though this expenditure of the small firms is considerably less than the expenditure of the large firms, yet the per unit cost of the large firms is smaller due to the fact that advertising cost is not required to be increased proportionately as the volume of production increases. Also, when the scale of production is large, the firm can economise on the expenditure on salesmen, agents, etc. The large firms can also enter into such contracts with the wholesalers and distributors that they take more interest in selling the products of the firm. Naturally, a small firm is deprived of this benefit.
- 3) **Managerial economies:** Managerial costs are partially production costs and partially selling costs. However, they are generally considered separately since it is convenient to do so. Managerial economies are obtained on account of the following two basic reasons: First, benefits of specialisation in the field of management can be obtained only when the scale of operations is considerably large. When the scale of production is small, all managerial responsibilities regarding production, marketing, finance, etc. will have to be borne by one person only. However, as the scale of operations expands, separate managers are appointed to look after these tasks. This raises the level and quality of management. At the same time, cost does not increase in proportion to the increase in the scale of operations. Large firms are in a position to use a number of machines for purposes of management. The use of computers, telephone, fax, etc. can be made only by a sufficiently large firm. If small firms use these machines, the total costs incurred on them would be very much higher in relation to the level of production attained.

The economists are, however, not in complete agreement on the managerial economies. Some economists argue that with the expansion of scale, managerial economies are obtained only upto a limit. After this limit, costs on management increase in a greater proportion. This is due to two reasons. First, the managerial structure in large companies is bureaucratic and when the scale of production expands, delays in decision making creep in. This weakens managerial efficiency. Second,



the degree of uncertainty increases as the size of the firm increases. On account of this reason, various difficulties have to be encountered in decision making leading to an increase in managerial costs.

- 4) **Economies in transport and storage:** When the scale of production expands, economies in transport and storage accrue to the firm. Small firms have usually to depend on public transport and therefore their per unit transport cost is higher. As the scale of operation expands, the firm can purchase its own truck, lorry, etc. This will reduce the per unit transport cost for the firm. If the scale of operation expands still further, the firm can go in for larger trucks and lorries. The railways also give siding facilities to large producers and this reduces their loading costs. In reality, the transport cost is partly production cost and partly selling and marketing cost. When the firm purchases raw material, the loading cost is a part of its production cost. On the other hand, when finished goods are transported to the market, it is a part of selling and marketing cost. However, for convenience in analysis, the economists prefer to treat transport costs separately.

Like transport costs, storage costs are also partly production costs and partly selling and marketing costs. For example, expenditure on storing the raw material is a production cost whereas expenditure on finished and semi-finished goods is a part of marketing costs. From the point of view of the size of the warehouse, an important thing to remember is that larger the size of warehouse, larger will be the economies accruing to the firm. The reason is that the cost of construction of the warehouse does not increase in the same proportion as the increase in the storage capacity of the warehouse.

#### **9.3.1.2 Pecuniary Internal Economies of Scale**

Some pure pecuniary economies accrue to a firm as its scale of operation expands. The more important ones are the following:

- 1) A large sized firm can ask the suppliers of raw materials to give specific concessions and discounts. No raw material supplier usually ignores such requests (or pressures) of the large firm.
- 2) Perfect competition generally does not prevail in the capital market. Since the large companies have greater goodwill in the capital market, they are in a position to obtain loans at lower rates of interest from the banks and financial institutions.
- 3) Transport companies are also willing to provide discounts and concessions if the cargo is substantially large. This enables the firm to obtain monetary economies in transport costs by expanding its scale of operations.
- 4) When production is large, the firm is required to spend a large amount on advertising as well. However, advertising on a large scale attracts discounts and concessions from the media in which the advertisements appear.

#### **9.3.2 Internal Diseconomies of Scale**

If the scale of production is continuously expanded, is it possible that after a certain point, increase in production is less than proportionate than increase in the factors of production? Many economists believe that such a situation can and does arise if production is pushed beyond the point of optimum scale. The reasons that they advance are as follows:

- 1) **Limitations on the availability of factors of production:** The factors of production are always available in limited supply at the place of production. When the scale of production is increased beyond a certain point, it no longer remains possible to meet the requirements of some factors from local sources and, accordingly, factors have to be transported from other regions. This is generally possible only at higher prices. Let us suppose that an engineering factory is to be set up at Rudrapur in the Terai region. When the scale of production is small, it would be possible to meet the demand for some materials from local sources. As the scale of production expands, it will become more and more difficult to get even the labour from local sources and after a certain point, workers will have to be attracted from other regions by offering them higher wages.
- 2) **Problems in management:** When the scale of production is very large, the task of management at the top level becomes increasingly more and more burdensome and some inefficiency is bound to creep in. At times, information vital for taking a decision does not reach the top managers of the company in time. This delay, in turn, leads to a delay in decision making and increases the per unit cost.
- 3) **Technical factors:** When the scale of production is expanded, per unit cost increases due to a number of technical reasons. The establishment cost of large and sophisticated plants and machinery is generally high. The buildings of large factories should also have stronger foundations and the factory itself must be equipped with coolers, air-conditioners, etc. All these factors lead to an increase in per unit cost.

### 9.3.3 External Economies

External economies were discussed first of all by Alfred Marshall. According to him, when a firm enters production, it obtains a number of economies for which the firm's own production strategy, managerial arrangements, etc. are not responsible. In fact, these are economies external to the firm. For example, let us suppose that a firm is established at a place where transport, advertising facilities, etc. are not available. If the size of the firm remains small, it is possible that these facilities are not locally available in the future as well. However, if the size of the firm increases significantly, these facilities will themselves start coming to the firm. These are, in fact, external economies.

When a firm expands its scale of production, other firms also earn many economies. For example, when a large factory attracts various factors of production fairly regularly, many other factories set up in the neighbourhood, that could not have attracted these factors on their own, also stand to gain. They obtain these factors at practically the same prices at which the large factory obtained them.

Because of external economies of large-scale production, there is a gap between private and social returns. When a firm expands its scale of production, it becomes possible for the other firms to reduce their cost of production. However, there is no method available in the prevalent price mechanism to the firm expanding its scale of operations to charge for the benefits it confers on the other firms.

### 9.3.4 External Diseconomies

When the scale of operations is expanded, many such diseconomies emerge that have no particular ill-effect on the firm itself. In fact, their burden falls on

the other firms. On account of this reason, they are termed external diseconomies. The smoke rising from the chimney of a factory pollutes the atmosphere. When the firm is of a small size, the pollution is less and its ill-effects on the people living in colony nearby is limited. However, if the scale of the firm is large, the smoke will be very dense and can cause serious health hazard to the people. Similarly, as the scale of production of the factories increases, employment rises sharply. This creates problems of traffic congestion and overcrowding in the city where these factories are located. In agriculture, increase in the scale of production leads to problems of soil erosion and this reduces the fertility of the adjoining fields as well. From the above illustrations, it is clear that external economies and diseconomies can be both pecuniary and technological.

### **Check Your Progress 1**

- 1) Indicate the following statements as true (T) or false (F):
  - i) When output increases in a greater proportion as compared to the increase in the amount of the factors of productions, we have the stage of increasing returns to scale.
  - ii) Those economies which accrue to a firm an account of the other firms are known as external economies.
  - iii) Production economies are a part of pecuniary internal economies.
  - iv) In the case of linear homogenous production function, we have constant returns to scale.
- 2) Discuss the factors which account for increasing returns to scale.  
.....  
.....
- 3) Explains how decreasing returns to scale arise.  
.....  
.....
- 4) Discuss the internal economies of scale.  
.....  
.....
- 5) What do you mean by external economies and external diseconomies?  
.....  
.....

---

## **9.4 LET US SUM UP**

---

In this unit, the concept of returns to scale has been explained. As noted in the beginning itself, this concept is associated with the tendency of production that is observed when the ratio between the factors is kept constant but the scale is expanded. This, in turn, can give rise to three possibilities – increasing returns to scale, constant returns to scale, and diminishing returns to scale. After

discussing all these possibilities, we shift our focus to a discussion of economies and diseconomies of scale. Economies of scale, in turn, are divided into two parts – internal economies of scale and external economies of scale. Economies which accrue to a firm on expansion of its own size are known as internal economies while economies which accrue to a firm not due to its own operations but due to operations of other firms are termed as external economies. We have discussed in detail all the causes which can result in the generation of such economies. In the end, we have focussed our attention on diseconomies of scale – both internal and external.

---

## 9.5 REFERENCES

---

- 1) Robert S. Pindyck, Daniel L. Rubinfeld and Prem L. Mehta, *Microeconomics* (Pearson Education, Seventh edition, 2009), Chapter 5, Section 5.4.
- 2) Dominick Salvatore, *Principles of Microeconomics* (Oxford University Press, Fifth edition, 2010) Chapter 7, Section 7.5.
- 3) A. Koutsoyianms, *Modern Microeconomics* (The Macmillan Press Ltd, Second edition, 1982) Chapter 3.
- 4) John P. Gould and Edward P. Lazear, *Microeconomic Theory* (All India Traveller Bookseller, Sixth edition, 1996), Chapter 8, Section 8.6.

---

## 9.6 ANSWERS OR HINTS TO CHECK YOUR PROGRESS EXERCISES

---

### Check Your Progress 1

- 1) (i) (T), (ii) (T), (iii) (F), (iv) (T)
- 2) See Sub-section 9.2.1 of Section 9.2
- 3) See Sub-section 9.2.3 of Section 9.2
- 4) See Sub-section 9.3.1 of Section 9.3
- 5) See Sub-section 9.3.3 and 9.3.4 of Section 9.3.