UNIT 4 SUPPLY OF HEALTHCARE SERVICES

Structure

4.0 OBJECTIVES

After reading this unit, you will be able to:

- state the basic features of factor markets of healthcare services;
- specify the equilibrium condition for non-physicians in the input market structure of healthcare sector distinguishing them from that of physicians;
- discuss the consequences of monopsony power in the input market of healthcare sector;
- describe the equilibrium conditions of a physician under: (i) pure monopoly; and (ii) price discriminating monopoly;
- outline the advantages of competition among price discriminating among physicians;
- indicate how monopolistic competition among physicians services lead to long run stability in the price for physicians services;
- explain the features of production function of healthcare services;

- derive the average and marginal cost function for healthcare services both for short-run and long-run; and
- distinguish between the concepts of 'economies of scale' and 'economy of scope' in the context of market for supply of healthcare services.

4.1 INTRODUCTION

Healthcare encompasses a lot more than just medical care. One can avoid medical care (or medical expenses) if basic healthcare services are available (e.g. different immunisation programmes initiated by a government). Healthcare includes a host of social welfare programs contributing to keeping people healthy. Healthcare falls under the category of merit good (or near public good) and therefore calls for government intervention or regulation. The market price of healthcare is difficult to determine due to imperfect market characteristics of healthcare sector. These are: (a) healthcare is a heterogeneous product as a range of outcomes could be there for different patients; (b) patients who are insured by third-party payers would behave differently than those without such coverage; and (c) the feedback mechanism on a standard 'market price' is missing due to which there is scope for malpractice (e.g. a hospital which has invested in machinery and equipments might order treatments with the sole objective of recovering the sunk cost).

Healthcare services follow certain characteristics like (a) intangibility (i.e. healthcare services are incapable of being assessed by the five senses), (b) inseparability (i.e. consumption and production of healthcare services take place simultaneously) and (c) inconsistency (i.e. the composition and quality of healthcare services consumed vary widely across medical events). Thus, unlike normal goods, consumers seeking healthcare services are less informed in general. The production and supply of healthcare is also therefore quite complex as compared to normal goods and services.

4.2 FACTOR MARKET OF HEALTHCARE SERVICES

Supply of healthcare services is multi-product and interdependent. This is because it includes the labour time of multitude of trained professionals like physicians, surgeons, specialists, nurses, medical technicians and pharmacists. Lack of coordination of any one of these critical services places the patient's life in danger making the healthcare services mutually interdependent and complex in its character. In other words, healthcare services are not homogeneous as they are composed of interdependent multiple inputs.

4.2.1 Structure of Input Market

In standard microeconomics, if the product and input market is perfectly competitive, then the firm (i.e. a nursing home or hospital) will use input up

to that level where the value of marginal product is equal to wage rate. This means:

$$\pi = TR - TC \tag{4.1}$$

where π stands for total profit, TR = total revenue, TC = total cost. Since the total cost comprise of labour (L) and capital (K), if we assume that the two input markets are perfectly competitive, then Equation (4.1) becomes:

$$\pi = P.f(X) - wL - rK \tag{4.2}$$

where, P is the unit price of healthcare services, $f(X)$ is the production function of healthcare, w stands for total wage, r is the interest rate on capital, L and K are the amount of total labour and capital invested respectively. Thus, in order to maximise π, the first order condition yields:

$$\frac{\delta\pi}{\delta L} = P.\frac{\delta f}{\delta L} - w = 0 \tag{4.3}$$

Equation (4.3) implies that $P.MP_L = w$ (where MP_L is the Marginal Productivity of Labour given by $\frac{\delta f}{\delta L}$). The healthcare provider will hire physicians up to that level where the 'value of marginal product' (VMP) of labour (physician) is equal to wage rate. Graphically, the equilibrium condition is attained at VMP as in Fig. 4.1. The healthcare provider (hospital) will employ OL^* number of labour units in order to maximise profit. At OL^*, the value of marginal product ($P.MP_L$) is equal to the wage rate OW. Under perfect competition in both the product (healthcare) and factor (labour) market, the exploitation of labour will not occur.

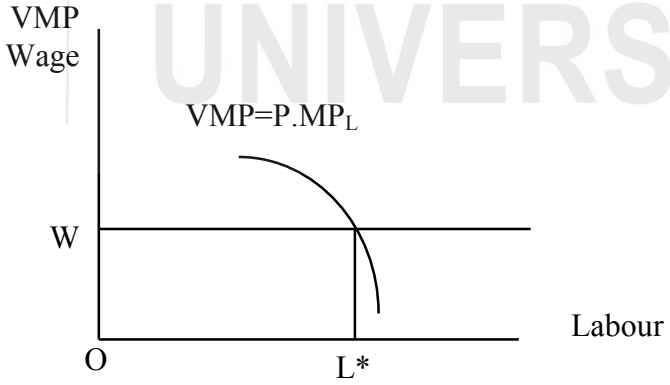

Fig. 4.1: Equilibrium under Perfect Competition

The above equilibrium condition is valid in case of hiring nurses and other medical staff but not in case of doctors or physicians. This is because the input market of physician is more heterogeneous in terms of skill, experience and expertise. Physicians further enjoy market power unlike nurses and other non-medical staff who cannot provide services independently of doctors. However, there may be monopsony power in markets for nurses in case of single buyer hospital.

4.2.2 Monopsony Power in Input Market

Monopsony occurs when there is a single buyer in an input market. This is similar to the concept of monopoly when there is a single seller in a product market. Buyers may possess monopsony power in an input market if the mobility of the input is restricted to that market. Thus, if a hospital enjoys monopsony power in the input market for nurses, then the wage rate of nurses would be lower than what would prevail under perfect competition. This is indicated by the downward sloping curve EF in Fig. 4.2 which represents the 'marginal revenue product' MRP which is obtained as: MR.MPP [i.e. product of marginal revenue (MR) of healthcare services and 'marginal physical product' (MPP)]. Since the price of healthcare (Px) is higher than MR in an imperfectly competitive market, MRP is less than 'value of marginal product' (VMP). But under perfect competition, VMP = MRP.

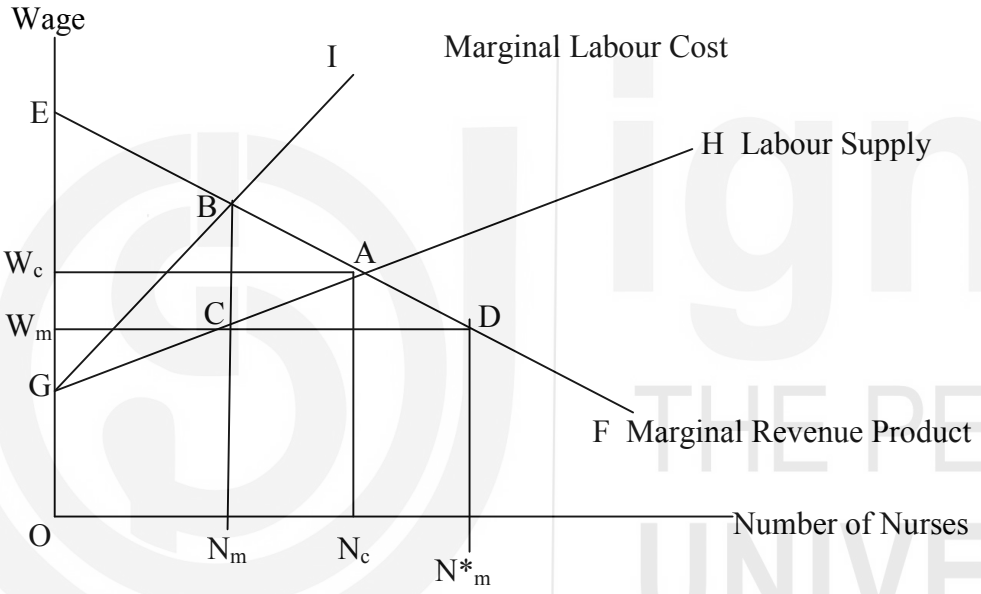

Fig. 4.2: Monopsony in Health Market

The lower upward sloping curve (GH) is the supply curve of labour (viz. nurses). The single employer (hospital, in our case) realises that to attract more nurses, it has to offer a higher wage rate. Thus, the 'marginal labour cost' (MLC) curve viz. GI lies above the supply curve. The MLC (GI) curve determines how many nurses are to be employed by the hospital. Since wage rate *w* is a function of labour L, from Equation (4.2), we can write the profit function as:

$$\pi = P.X - w(L).L - rK \qquad (4.4)$$

To maximise π, with respect to *L*, the first order condition gives us:

$$\frac{\delta\pi}{\delta L} = P.\frac{\delta X}{\delta L} + X.\frac{\delta P}{\delta X}.\frac{\delta X}{\delta L} - \left(\left(w(L) + L.w'(L)\right)\right) = 0$$

$$\Rightarrow P.MP_L + X.\frac{\delta P}{\delta X}.MP_L = MC_L \qquad (4.5)$$

$$\Rightarrow \left[MP_L \left(P + X.\frac{\delta P}{\delta X} \right) \right] = MC_L \qquad (4.6)$$

Since, $P = f(X)$, $X = X(L)$, assuming only L to be the variable cost and identifying $MR = \frac{\delta}{\delta X}(TR) = P + X.\frac{\delta P}{\delta X}$, (4.6) can be written as:

$$MP_L * MR_X = MC_L \qquad (4.7)$$

This means that under equilibrium the firm (hospital) will hire labour (nurse) up to the level where MRP is equal to marginal cost of input for nurses. Therefore, the hospital will employ N_m number of nurses at wage rate W_m. But, at W_m, the demand for nurses is N^*_m, and there is a shortage of labour equal to $(N^*_m - N_m)$. If the monopsonist hospital desires to hire additional nurses, it cannot do so until it raises the wage beyond W_m which reduces its profit. Therefore, in a monopsonist market, the equilibrium is not efficient because: (i) vacancies persist; (ii) employment would be less than that in a competitive market; and (iii) the equilibrium wage W_m would be less than MRP as shown by the vertical distance between points B and C in Fig. 4.2. The gap between MRP and wage under monopsony is:

$$\text{Gap (G)} = (\text{MRP} - W_m)/W_m = 1/\varepsilon \qquad (4.7)$$

where ε stands for elasticity of supply of labour to the hospital with respect to wage W_m. 'G' is thus interpreted as a measure of 'exploitation' or hospital's market power in the input market.

Check Your Progress 1 [answer within the space given in about 50-100 words].

1) What is the basic characteristic of healthcare sector making it both different and complex from other normal goods and services?

.....

.....

.....

.....

.....

2) Why is attainment of equilibrium in numbers hired and wages paid possible for nurses but not for physicians in the healthcare market?

.....

.....

.....

.....

.....

3) Why is it in case of a monopsonist market, equilibrium conditions are not met?

.....

.....

.....

.....

.....

4.3 DETERMINATION OF EQUILIBRIUM PRICE FOR PHYSICIANS SERVICES

Physicians play a dual role: to a hospital as inputs and independently as supplier of healthcare services to patients directly. A physician thus enjoys monopoly power in some respects. In this section, we discuss the following three scenarios pertaining to physicians.

4.3.1 Pure Monopoly

The monopolist can set the market price. The profit function of the monopolist is:

$$\pi = TR - TC \tag{4.8}$$

where Total Revenue $TR = P.Q$ with P as the price per unit of supply of healthcare and Q as the quantity of healthcare supplied. The first order condition of profit (π) maximisation requires that MR = MC. Thus:

$$\frac{\Delta\pi}{\Delta Q} = \frac{\Delta TR}{\Delta Q} - \frac{\Delta TC}{\Delta Q} = 0$$

$$P + Q.\frac{\Delta P}{\Delta Q} = MC \tag{4.9}$$

$$P\left(1 + \frac{Q}{P}.\frac{\Delta P}{\Delta Q}\right) = MC$$

Since the price elasticity of demand is given by:

$(E_P) = (-).\frac{\Delta Q}{\Delta P}.\frac{P}{Q}$, Equation (4.9) can be re-written as:

$$P\left(1 - \frac{1}{-E_P}\right) = MC \text{ or } P\left(1 - \frac{1}{|E_P|}\right) = MC \tag{4.10}$$

Thus the change in total revenue is influenced by two opposing forces when a monopolist physician lowers the fees (price) of his healthcare services. In other words, while on the one hand, total revenue increases because the physician is selling more healthcare services, on the other hand, total revenue (TR) falls because the physician has lowered the price. This happens because

of the downward sloping demand curve. The equilibrium of the physician can be graphically illustrated as in Fig. 4.3.

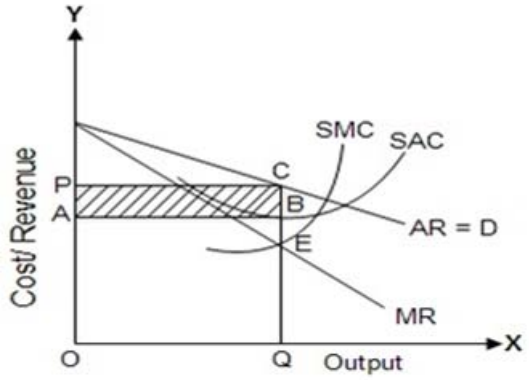

Fig. 4.3: Equilibrium Price and Output of a Monopolist Physician

In Fig. 4.3, the equilibrium condition is at E where MR = MC. The physician charges the price/fee OP and supplies OQ amount of healthcare services, earning ABCP profit (π). Since the equilibrium amount of healthcare services is determined by the intersection of MR and MC, there is no unique supply curve of a monopolist. Further, since π = MR – MC, the second order condition of profit maximisation requires that:

$$\frac{\Delta^2\pi}{\Delta Q^2} = \frac{\Delta MR}{\Delta Q} - \frac{\Delta MC}{\Delta Q} < 0 \qquad (4.11)$$

Thus, slope of MR < Slope of MC or slope of MC > slope of MR. Since at point E, both the first and second order conditions for profit maximisation are satisfied, the rising portion of the marginal cost curve (MC) can be considered as supply curve of a monopolist for healthcare services. In the long-run, the physician earns normal profit when price (P) is equal to the Average Cost (AC). The physician can enjoy absolute monopoly power if he is the sole supplier of healthcare services. This is shown by the Lerner's Index (*LI*) of degree of monopoly power given by:

$$\begin{aligned} LI &= \left(\frac{P - MC}{P}\right) \\ &= \frac{P - MR}{P} \qquad \because MR = P\left(1 - \frac{1}{E_P}\right) \\ &= \frac{1}{E_P} \end{aligned} \qquad (4.12)$$

The Lerner's Index of monopoly power is therefore nothing but the reciprocal of the numerical coefficient of price-elasticity of demand for the product or services. Thus, the less elastic is the demand for the product or services, more would be the degree of monopoly power and vice versa. The economic meaning of this idea is that 'smaller the price-elasticity of demand, smaller would be the response of demand for the product (in response to a

change in its price) and larger would be the power of the monopolist to charge a price in excess of *MC*'.

4.3.2 Price Discriminating Monopoly

We consider here the case of physician as 'price discriminating monopolist'. Price discrimination being the practice of charging a different price for the same good or service, there can be following three types of price discrimination.

- *First-degree* price discrimination, alternatively known as *perfect* price discrimination, which occurs when a physician charges a different price for every unit consumed.
- *Second-degree* price discrimination where a different price is charged for different quantities (such as discounts for bulk purchase of healthcare services). The physician can provide discount to patients in case of multiple medical tests.
- *Third-degree* price discrimination means charging a different price to different consumer groups. For instance, the patients can be divided into two groups on the basis of health insurance. In the same way, patients can also be divided into two groups viz. poor and non-poor. This type of discrimination is the commonest type.

The equilibrium condition of the discriminating monopolist physician can be explained as follows. Suppose, the physician divides the market based on health insurance with TR_1 and TR_2 as the revenues earned by the physician from two markets (patients having health insurance and others not insured). The profit function can be written as:

$$\pi = TR_1 + TR_2 - TC \qquad (4.13)$$

To maximise profit (π), the first order condition yields:

$$\frac{\delta\pi}{\delta Q_1} = \frac{\delta TR_1}{\delta Q_1} + \frac{\delta TR_2}{\delta Q_1} - \frac{\delta TC}{\delta Q_1} = 0 \Rightarrow MR_1 - \frac{\delta TC}{\delta Q}.\frac{\delta Q}{\delta Q_1} = 0 \Rightarrow MR_1 - MC.\left(\frac{\delta(Q_1+Q_2)}{\delta Q_1}\right) = 0$$

$$\Rightarrow MR_1 = MC(1+0) = MC$$

$$\frac{\delta\pi}{\delta Q_2} = \frac{\delta TR_1}{\delta Q_2} + \frac{\delta TR_2}{\delta Q_2} - \frac{\delta TC}{\delta Q_2} = 0 \Rightarrow MR_2 - \frac{\delta TC}{\delta Q}.\frac{\delta Q}{\delta Q_2} = 0 \Rightarrow MR_2 - MC.\left(\frac{\delta(Q_1+Q_2)}{\delta Q_2}\right) = 0$$

$$\Rightarrow MR_2 = MC(0+1) = MC$$

Assuming Q_1 and Q_2 are the quantity of healthcare services supplied by the physician in Markets 1 and 2 respectively so that $Q = Q_1 + Q_2$, the price discriminating physician in equilibrium will equate $MR_1 = MR_2 = MC$. Using the relationship between AR (Price), marginal revenue (MR) and elasticity of demand (E_p) as shown in Equation (4.10), we can write:

$$P_1\left(1-\frac{1}{E_1}\right)=P_2\left(1-\frac{1}{E_2}\right) \quad (4.14)$$

where P_1 and P_2 are the prices charged by the monopolist in the 1st and 2nd markets and E_1 and E_2 are the price elasticities of the 1st and 2nd markets respectively. Noting that if $P_1 > P_2$ then $E_1 < E_2$, we observe that discrimination is only worth undertaking if the profit from keeping the markets separated is greater than from keeping the markets combined. Since this will depend upon the relative elasticities of demand in the sub-markets, consumers in the relatively inelastic sub-market will be charged the higher price and those in the relatively elastic sub-market will be charged the lower price.

Thus, under price discrimination, the physician is able to discriminate between different markets in respect of the price of his healthcare services. From a physician's perspective price discrimination can offer some advantages. These include: profit maximisation, economies of scale and positive externalities. For instance, matching prices to the specific characteristics of the market, and its various segments, is a profit maximising strategy where the physician can extract some of the consumer surplus into producer surplus (i.e. profits). Likewise, given that charging different prices can increase sales volume, especially as a result of new consumers entering into the market attracted by the discounted prices, physicians can benefit from 'economies of scale' which arise from increased output and production. Economies of scale are thus the cost advantages that physicians derive due to their scale of operation with cost per unit of healthcare services decreasing with increasing scale of supply of healthcare services. This line of argument can be extended to consider the role of price discrimination in reducing market failure (e.g. enabling wider consumption of merit goods like healthcare services). For instance, if a private hospital charge relatively high medical fees for those who can afford them where demand is inelastic, the revenue generated allows them to cover their costs and run the hospital. With fixed costs covered, they can then offer services at discounted fees (to cover only the variable costs) to poor patients. Given that the demand for private healthcare by less well-off patients is likely to be price elastic, the lower price will encourage greater demand for healthcare. The benefit to 'society' is therefore that more healthcare services are 'consumed' and more positive externalities are generated.

4.3.3 Competition among Physicians

Monopolistic competition is a market structure which combines elements of monopoly and competitive markets. Essentially, a monopolistic competitive market is one with freedom of entry and exit with physicians providing differentiated services. The physicians therefore have an inelastic demand curve on which they can set prices. However, because there is freedom of

entry, supernormal profits in the short run will encourage more firms to enter into the market leading to normal profits in the long term.

Healthcare service by physicians, in reality, is characterised by monopolistic competition. As a result, each physician's demand curve is negatively sloped. This is mainly because of certain reasons. Firstly, search among doctors is costly for patients. So, if the patient is satisfied with a doctor, the cost of further searching is not worthwhile. Thus, a patient may not leave his or her physician even if the physician's price is slightly above another physician's. A second reason for monopolistic competition among physicians is that people may be willing to pay more for the convenience of visiting a physician who is geographically closer to the patient. Further, there being no barriers to entry, as more sellers enter the market, the demand curve shifts until each seller earns zero profit (Fig. 4.4).

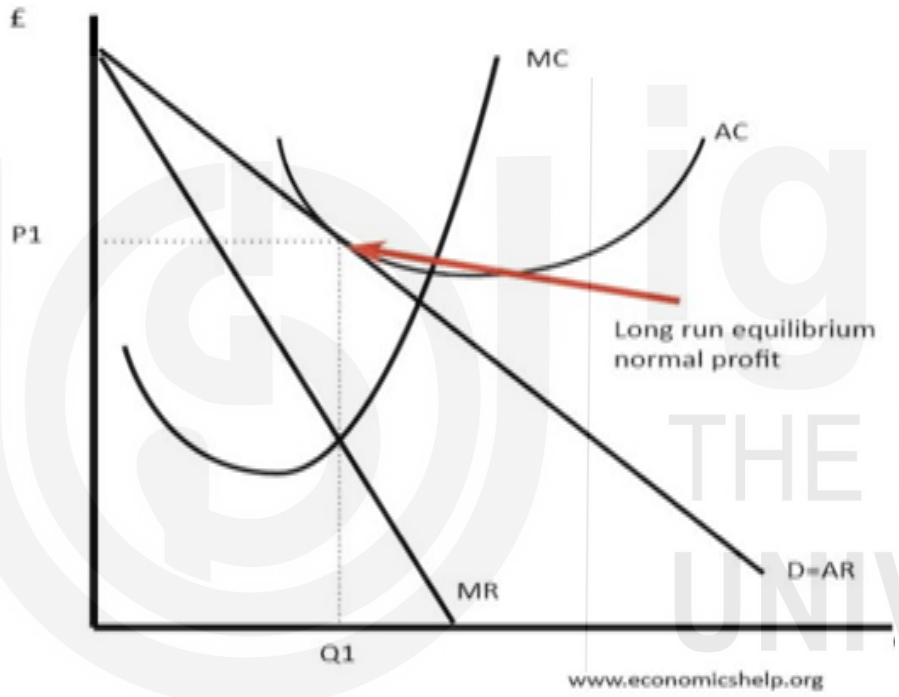

Fig. 4.4: Monopolistic Competition in Physicians Healthcare Market

Thus, under monopolistic competition, since physicians produce differentiated services, they are not at complete liberty to set the prices. In other words, they have inelastic demand. Physicians compete among themselves both for quality of healthcare as also on price. Differentiated healthcare services being a key element of the market, monopolistic competition remains a middle ground in which physicians have the same but relatively low degree of market power even though they are all price makers. In the long run, demand becomes elastic and ensures sensitivity to price changes. In other words, it is only in the short run that economic profit is positive and it approaches zero in the long run.

Check Your Progress 2 [answer within the space given in about 50-100 words].

1) How is the 'total revenue' for a monopolist physician influenced by two opposite forces?

.....

.....

.....

2) What does Lerner's index indicate? On which factor, does this basically depend?

.....

.....

.....

.....

3) What is meant by 'third degree discrimination'?

.....

.....

.....

.....

4) What factor enables the physician to practice 'price discrimination' in different markets? What advantages accrue to physicians by practicing this?

.....

.....

.....

5) How would you distinguish for the short run and long run characteristic of price setting physicians in a market? Why?

.....

.....

.....

4.4 PRODUCTION AND SUPPLY OF HEALTHCARE SERVICES

Medical firms like hospitals earn revenue from producing and selling some type of medical or healthcare output. Production and retailing activities occur regardless of the form of ownership. A comprehensive healthcare production and supply is performed by a well organised hospital.

4.4.1 Production Function

A healthcare production function can be written as:

$$HC = f(P, N, NM, K) \quad (4.15)$$

where HC = Output of healthcare service, P = number of physicians, N = number of nursing staff, NM = number of non-medical staff and K = amount of physical capital. P, N and NM are the inputs referred to as human capital and K is the physical capital (that includes building, medical equipments and machines). If we assume that a given level of output of healthcare (HC) can be produced by using two inputs [human capital (H_1) and physical capital (H_2)], graphically, the efficiency frontier is the lower envelope of the set of production possibilities (Fig. 4.5). This would be associated with a particular type and volume of output of healthcare. The collection of all these input combinations yields the concept of production function.

The efficiency frontier, or isoquant, represents all possible efficient ways to achieve a certain production level. Fig. 4.5 shows two distinct types of isoquants based on two types of production technologies represented by BB*B** and LO*L* (associated with two different types of production functions). If we move along OO*O**, output of healthcare (HC) increases. Therefore, health output produced along the isoquant BB*B** is lower than the output produced along the isoquant LO*L*. The marginal rate of technical substitution (MRTS) is defined as: keeping the total output of health desired to be produced as constant, how much of H_1 has to be decreased if H_2 increases by one additional unit. In other words, it shows the relationship between the two inputs, and the trade-offs amongst them, without changing the level of total output of health. Technically, MRTS amounts to:

$$MRTS_{H_1,H_2} = (-\frac{dH_1}{dH_2}) = \frac{MP_{H_1}}{MP_{H_2}} \quad (4.16)$$

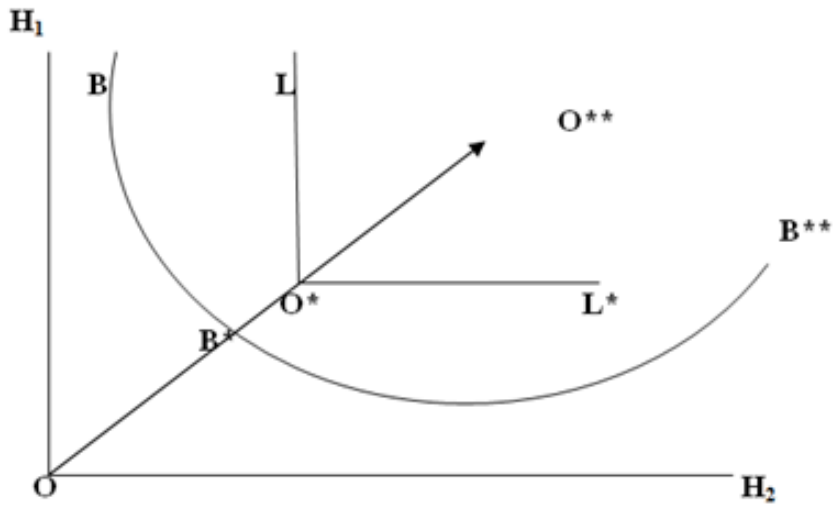

Fig. 4.5: Efficiency Frontier for Differing Production Levels

In Equation (4.16) $\frac{dH_1}{dH_2}$ is the slope of the isoquant and MP_{H_1} and MP_{H_2} are the marginal productivity of H_1 and H_2 respectively. MRTS is thus the negative slope of the isoquant. If $HC = \overline{H}$, the healthcare production function $HC = H(H_1, H_2)$ for fixed level of healthcare output becomes: $\overline{H} = H(H_1, H_2)$. Taking the total differential on $\overline{H}$, we have: $d\overline{H} = \frac{\partial H}{\partial H_1}.dH_1 + \frac{\partial H}{\partial H_2}.dH_2$. Since there is no change in output of health along a fixed isoquant, $d\overline{H} = 0$. Thus, we have:

$$d\overline{H} = \frac{\partial H}{\partial H_1}.dH_1 + \frac{\partial H}{\partial H_2}.dH_2 = 0 \text{ or } \frac{dH_1}{dH_2} = -\frac{\frac{\partial H}{\partial H_2}}{\frac{\partial H}{\partial H_1}} = -\frac{MP_{H_2}}{MP_{H_1}}$$

If H_1 and H_2 are assumed to be perfect substitutes of each other, then the isoquant will be a straight line having a negative slope of –1. This means, if we add one unit of H_1, we must sacrifice the same unit of H_2 (in order to remain on the same isoquant H). In case of the isoquant BB*B**, substitution that takes place between inputs is imperfect. As a result, MRTS tends to zero while diminishing the quantity of H_1 and to infinity while diminishing the quantity of H_2.

If we assume that the healthcare output is produced along the L-shaped isoquant LO*L*, there is no scope of input substitution with the two inputs (H_1 and H_2) being treated as complementary to each other i.e. the two inputs are always used in a fixed proportion. In this case, the horizontal segment of LO*L* has a MRTS = 0 and along the vertical segment of LO*L*, the MRTS = ∞. Since MRTS measures the slope of an isoquant, along LO*, the slope is infinity (since H_1 changes but H_2 remains same along LO*). As a result, $dH_1/dH_2 = \infty$ (infinity) [since $dH_2 = 0$]. Thus, along the isoquant LO*L*, the optimum health supply $H = Min\left(\frac{H_1}{\alpha}, \frac{H_2}{\beta}\right)$, where, α and β are two positive constants. In Fig. 4.5, if the angle L*O*O** is 45º then α = β. In our present case of fixed coefficient type production, the production path is along the locus of line OO*O**. Thus, in case of LO*L* isoquant, no substitution is possible between the two factors H_1 and H_2 i.e. the two factors of production are perfect complement to each other. In other words, if the two factors are used in fixed proportion, we can avoid wastage of inputs either H_1 or H_2.

Practically, the production of healthcare service is fixed coefficient type. For instance, a medical operation requires a specialised surgeon, one anaesthetic and one nurse along with surgical equipments. Therefore, an additional surgeon or nurse becomes redundant. Similarly, production (medical operation) becomes impossible if at least one of the required input combinations is lacking.

4.4.2 Cost Function

Following the discussion in 4.4.1, we now proceed to analyse the cost function of a hospital assuming again that only two inputs are used: human resources (H_1) and physical capital (H_2). With this, we derive the average and marginal cost functions for short-run as well long-run. In the short-run, the 'law of variable proportions' operate (where only one input is considered as variable with the other assumed to be constant), but in the long-run, both the inputs (H_1 and H_2) would vary and the 'law of returns to scale' would operate. For the short-run, we assume that the production function of healthcare can be written as $H = H(H_1)_{\overline{H}_2}$ in which input H_2 is fixed at a certain level and only first input (human capital), H_1 varies.

Law of variable proportion implies that $\frac{\partial H}{\partial H_1} > 0, \frac{\partial^2 H}{\partial H_1^2} < 0$. The total product of healthcare curve looks like inverted U-shaped which means that as H_1 increases total healthcare increases at a diminishing rate and once it reaches a maximum level, it declines. The corresponding marginal productivity also initially increases, reaches a maximum level and then sharply declines. In other words, the marginal productivity of H_1 is zero when the total healthcare product is maximum. The short run average total cost (SRATC) can therefore be written as:

$$SRATC = \frac{C}{H} = \frac{SRTVC + SRTFC}{H} = \frac{SRTVC}{H} + \frac{SRTFC}{H} = SRAVC + SRAFC \quad (4.17)$$

where C = Total Cost, H = Amount of Healthcare, $SRTVC$ = Short-Run Total Variable Cost, $SRTFC$ = Short-Run Total Fixed Cost, $SRAVC$ = Short-Run Average Variable Cost, $SRAFC$ = Short-Run Average Fixed Cost. Graphically, the short-run ATC, AVC, AFC and marginal cost (MC) would be as shown in the Fig. 4.6.

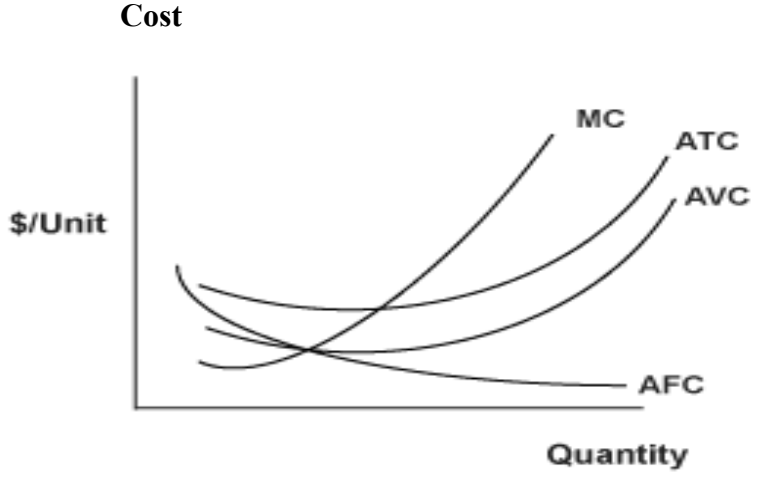

Fig. 4.6: Short Run Marginal/Average/Fixed/Total Cost Curves

As H increases, AFC declines continuously becoming asymptotic to the quantity of healthcare axis (X axis). Given H_2, due to law of variable proportions, in the short-run, the AVC of healthcare reduces because of higher productivity. Since the total cost *(TC)* function consists of total variable cost (*TVC*) and total fixed cost (*F*), in the short-run, the average cost (*AC*) is given by:

$$AC = \frac{TVC + F}{H} = \frac{W.H_1 + F}{H} = W.\frac{1}{\frac{H}{H_1}} + \frac{F}{H} = \frac{W}{AP_{H1}} + AFC \qquad (4.18)$$

where, W = unit price of H_1, H_1 is the number of units of human resources used in producing healthcare services (treated as variable input). The second input H_2, assumed as fixed, requires F amount of cost. The law of variable proportion gives us that initially average productivity of H_1 (AP_{H_1}) increases, reaches to a maximum and thereafter decreases as the unit of healthcare (H) increases. This means that the first term of right hand side of (4.18) falls, reaches to a minimum level then rises. This is the reason why in the short run, the hospital faces a U shaped AC curve. We can establish the relationship between *SRAC* and *SRMC* as:

$$\frac{d}{dH}\left[\frac{C}{H}\right] < 0 \Rightarrow \frac{H.\frac{dC}{dH} - C\frac{dH}{dH}}{H^2} < 0 \Rightarrow \frac{MC}{H} - \frac{1}{H}\frac{C}{H} < 0 \Rightarrow MC < ATC\,.$$

In the same way, we can show that when *ATC* is minimum, $MC = ATC$ and when *ATC* rises, $MC > ATC$.

The long run consists of variable inputs only. The hospital can therefore increase the size of the plant in the long run. Long run total cost refers to the minimum cost of production or the least cost of producing a given level of output. It can therefore be less than or equal to the short run average costs at different levels of output. Similarly, 'long run average cost' (LRAC) can be defined as the average of the LTC curve or the cost per unit of output in the long run. It can be calculated by the division of LTC by the quantity of output.

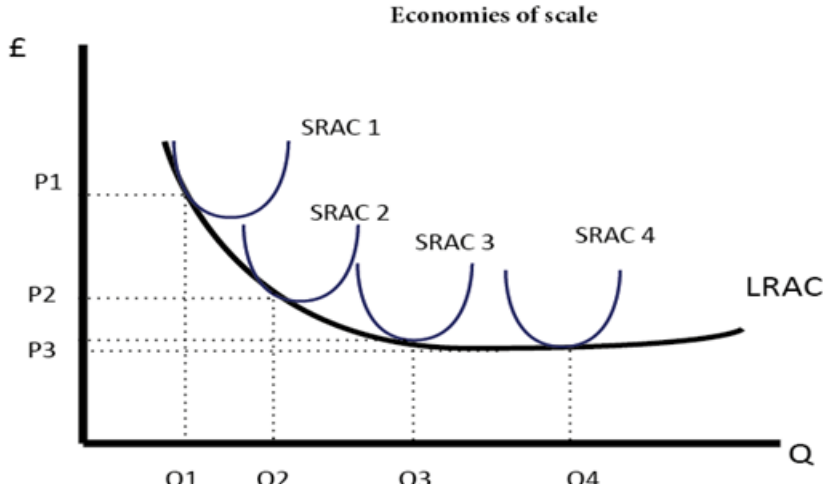

Fig. 4.7: Long Run Average Cost Curves

Graphically, LAC can be derived from the Short run Average Cost (SAC) curves. While the SAC curves correspond to a particular plant (the *plant being fixed in the short-run)*, the LAC curve depicts the scope for expansion of plant by minimising cost. The long-run cost curves are also U shaped due to 'economies of scale'. If a hospital has high fixed costs, increasing output will lead to lower average costs as in the long-run 'returns to scale' operates. This means that in the initial stage, hospital enjoys increasing returns, later constant returns but finally decreasing returns. Thus, the 'long-run average cost' (LRAC) curves can be seen as a lower 'envelope' of short run average cost curves (SRAC) i.e. the curve lies below the short run curves. Any change in inputs in the short run will carry over into the long run as well. It is therefore impossible for average costs to be driven down more in the short run than in the long run. But since the factors that are fixed in the short run can be changed in the long run, average cost (AC) reductions resulting from changing these inputs in the long run will not be felt in the short run. Hence, the long run average costs are lower. For this reason, they are also sometimes called 'envelope curves'.

4.4.3 Economies of Scale and Scope

Economies of scale for a hospital involves reduction in the **average** cost per unit arising from increasing scale of production for a **single** product type. Economies of scope involve lowering the **average** cost by producing more types of products. Economies of scope exist if a firm can produce several product lines at a given output level more cheaply than a combination of separate firms each producing a single product at the same output level. Economies of scope differ from economies of scale in that a firm receives a cost advantage by producing a complementary variety of products with a concentration on a core competency. While economies of scope and scale are often positively correlated and interdependent, strictly speaking, the benefits from scope have little to do with the size of output. Economies of scope and economies of scale are thus two different concepts used to help cut a company's costs. Economies of scope focuses on the average total cost of production of a variety of goods, whereas, economies of scale focuses on the cost advantage that arises when there is a higher level of production of one good.

Economies of scale states that the unit cost decreases as the volume of output increases. This connection can be explained in terms of the effects of volume on fixed cost, specialisation of resources and market position toward suppliers. Economies of scale arise since the initial investment of capital is spread over a larger output with increases in output. Examples are medical equipment such as CT Scan or a computer system for medical records of patients. Another source of economies of scale is that larger firms may have more market power for inputs they purchase. Economies of scope exist when the provision of one type of output makes it possible for another type of output to be produced more efficiently. For instance, consider a radiologist

practising with a general surgeon. There may be efficiency in the diagnosis and treatment of a disease if an accurate diagnosis can be followed by a timely surgical procedure. Cost complementary effect and scale effect are the explained mechanisms behind the scope effect. Cost complementary means that once the inputs have been used for producing one good they become available, at no additional cost, for use in the production of other goods. The scale effect refers to the situation where it is profitable to produce different outputs in one large unit instead of in several specialised production units.

Check Your Progress 3 [answer within the space given in about 50-100 words]

1) In practice, what does it mean when we say that the production of healthcare is of 'fixed coefficient' type? Why?

.....

.....

.....

.....

.....

2) Why are 'long run average cost (LRAC) curves' also called as 'envelope curves'?

.....

.....

.....

.....

.....

3) Distinguish between 'economy of scale' and 'economy of scope' in the context of production of healthcare services.

.....

.....

.....

.....

.....

4.5 LET US SUM UP

The unit has discussed three inter-related issues of healthcare sector viz. input market in healthcare services, physician's supply of healthcare services and cost structure of a hospital supplying healthcare services. The dual role of physician is discussed where the physician can independently supply

healthcare services and also get employed by the hospitals. The behaviour of a physician as price discriminator is explained. Exploitation of nurses (in the input market of healthcare) is analysed as a special case of monopsony power of a hospital. Similarly, the supply of healthcare service as determined by costs of inputs is discussed both for its short-run and long-run cost impacts on the sector. The concepts of 'economies of scale' and 'economies of scope' is distinguished towards the end.

4.6 KEY WORDS

Value of Marginal Product (VMP) : VMP is defined as the product of 'price' and 'marginal physical productivity' (MPP) of an input.

Marginal Revenue Product (MRP) : MMP is defined as the product of 'marginal revenue' (of the product) and MPP of an input.

Isoquant : Refers to the efficiency frontier representing all possible efficient ways to achieve a certain production level.

Monopsonistic Exploitation : Refers to the gap between MRP and the wage paid. It is given by: Gap (G) = (MRP – W_m)/W_m=1/ε where ε stands for elasticity of supply of labour with respect to wage.

Lerner's Index of Degree of Monopoly Power : Lerner's Index gives an indicator of the degree of monopoly power. It is the reciprocal of the numerical coefficient of price-elasticity of demand for the product or service. The less elastic is the demand for the product or services, the more would be the degree of monopoly power and vice versa.

Marginal Rate of Technical Substitution (MRTS) of Healthcare Production : MRTS is defined as, 'keeping the total output of health desired to be produced as constant, how much of input 1 (H_1) has to be decreased if input 2 (H_2) increases by one extra unit'? It shows the relation between inputs, and the trade-offs amongst them, without changing the level of total output of healthcare services. MRTS is the negative slope of the isoquant.

Envelope Curve : The 'long-run average cost' (LRAC) curve is a lower 'envelope' of short run average cost curves (SRAC). This means that since any change in inputs in the short run will carry over into the long run as well, it is impossible for average costs to be driven down more in the short run than the long run.

4.7 SOME USEFUL BOOKS AND REFERENCES

1) Barros Pedro. and Xavier Martinez-Giralt (2012). Health Economics: An Industrial Organization Perspectives, Routledge.

2) Sloan Frank. A and Chee-Ruey Hsieh (2012). Health Economics, The MIT Press, 55 Hayward Street, Cambridge, MA, 02142.

3) Rexford E Santerre and Stephen P Neun (2010). Health Economics, Theories, Insights and Industry Studies, Mason, Ohio: South-Western Cengage Learning.

4.8 ANSWERS/HINTS TO CHECK YOUR PROGRESS EXERCISES

Check Your Progress 1

1) Many trained persons contribute to it. The chain of services needs maintained failing which services will be sub-optimal. The sector is thus non-homogeneous with each patient's response needing to be followed individually. Lack of homogeneity in the response of patients is thus the basic characteristic of the sector making it both different and complex.

2) The services of nurses are marked for a higher degree of homogeneity whereas that of physicians is the opposite i.e. it is heterogeneous. This makes it difficult for the market to experience equilibrium in case of the latter as compared to the former.

3) Due to the gap that exists between MRP and W_m as shown in Equation (4.7).

Check Your Progress 2

1) The revenue falls due to lowered price effect. The revenue increases due to more services and fee earned. The overall or gross revenue is thus influenced by two opposing forces. The economic reasoning is the downward sloping demand curve.

2) The index shows under what conditions the monopolist physician can enjoy absolute monopoly. It depends upon the price-elasticity of demand for his services.

3) Refers to physicians charging differential prices for two segments e.g. insured, non-insured; poor, non-poor.

4) The factor of 'price elasticity of demand' in the two markets. The advantages to physicians are: profit maximisation, economies of scale and positive externalities.

5) Due to free entry and exit, entry of more physicians would keep the prices to fluctuate yielding only profits in the short run to the physicians.

In the long run, prices would become sensitive and elastic responding only to factors of search cost and geographic proximity.

Check Your Progress 3

1) This means that additional inputs will not increase efficiency just as the absence of any critical input will collapse the functioning of the healthcare sector.

2) The factors that are fixed in the short run can be changed in the long run. In view of this, the average cost (AC) reductions resulting from changing inputs (in the long run) will not be felt in the short run. Thus, the SRCCs which are all U shaped (when considered for several short stretches of time periods) provides the limits to LRAC. Fig. 4.7 illustrates this.

3) Economies of scale relates to decrease in the unit cost of production with increase in the volume of production. Economies of scope relates to lowering the average cost by producing more types of products.