Manonmaniam Sundaranar University

Directorate of Distance and Continuing Education

Tirunelveli – 627 012. Tamil Nadu.

M.A. Economics

(First Year)

Advanced Micro-Economic Theory - II

Prepared by

Dr. A. Murugapillai, Ph.D
Assistant Professor of Economics
Manonmaniam Sundaranar University
Tirunelveli – 627 012.

Advanced Micro Economic Theory – II

L	T	P	C
6	0	0	4

OBJECTIVE

To impart skills on theories of firm and distribution, welfare Economics, general Equilibrium in closed and open systems of Economic Analysis.

UNIT –I Alternative Theories of the Firm:

Recent Developments in the theory of pricing – Baumol's theory of sales revenue maximization – Cyert and March model – Maximum rate of growth hypothesis by Marris – Williamson's model of Managerial discretion – Behavioural theory of firm – Game theory.

UNIT –II Distribution:

Neo-classical approach – Marginal productivity theory – Euler's theorem- the adding up problem – Modern theories of rent, wage, interest and profit – liquidity preference theory of interest – Shackle's theory of profit – Macro theories of distribution – Ricardian, Marxian, Kalecki and Kaldor.

UNIT –III Welfare Economics:

Pigovian Welfare Economics and Externalities – the compensation criteria – Kaldor, Hicks criteria – Scitovsky criterion – Problems of existence, stability and uniqueness of general equilibrium – Maximisation of Social Welfare – marginal conditions of paretian optimum – pareto optimality and competitive equilibrium.

UNIT –IV General Equilibrium Analysis:

Walrasian General Equilibrium Analysis – 2x2x2 model – pareto optimality – tatonnement and non- tatonnement process – Social Welfare function – Arrow'S Impossibility Theorem – Second Best Theory.

UNIT –V Theories of Risk and Uncertainty:

Choice under uncertainity - formalization of uncertainity – choice under uncertainity – contingent consumption – properties of utility function and probabilities – criticisms of the expected utility theory – Risk aversion – The demand for insurance – diversification of risk – Risk spreading – Role of stock market effect of taxation on investment in risky assets.

Learning Outcome

After the completion of the course, the students must be able to

1. Have a good understanding of alternative theories of the firms.
2. Gain knowledge of Distribution Theories.
3. Get a complete knowledge about Welfare Economics and General Equilibrium Analysis.
4. Study the Theories of Risk and uncertainty.

TextBook:

Koutsoyiannis. A Modern Micro Economics, Macmillan Press, London, 1979.

Books for Reference:

1. Jhingan,M.L. Advanced Economic Theory: Brinda Publications, NewDelhi.
2. Baumol, W.J. Economic Theory and Operations Analysis: Prentice Hall of India. NewDelhi
3. Cullis and Jones; 1988, Public Economics and Public Choice: Oxford UniversityPress.
4. Gravelle and Rees;1988, Micro Economics: LongmanHouse.
5. Halvarian, Micro Economics (Advanced) Affiliated East WestPress
6. Thirlwall, A.P.; 1994, Growth Economics,ELBS.

Table of Content

Unit-I

Alternative Theories of the firm

Theories of Firm

The following points highlight the three main theories of firm.

The theories are:

1. Profit-Maximizing Theories
2. Other Optimizing Theories
3. Non-Optimizing Theories.

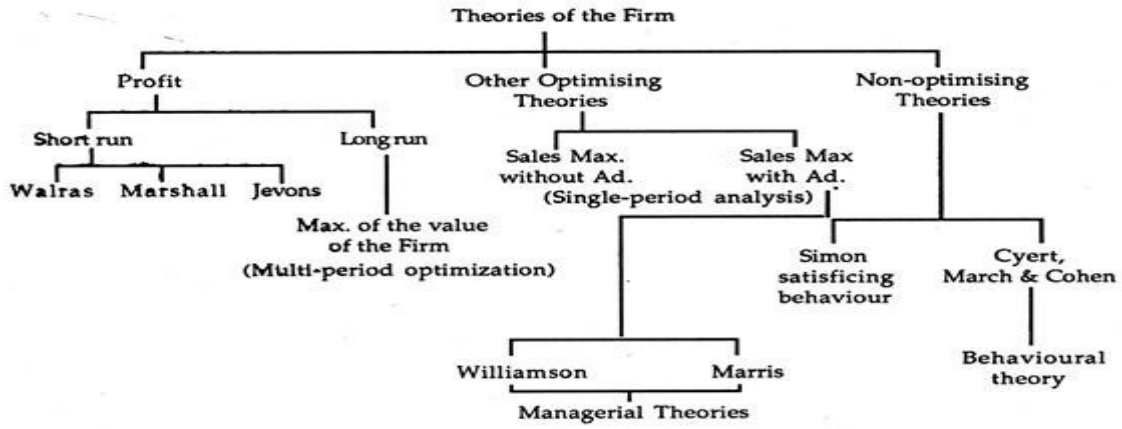


Figure 7.1
A classification of the theories of the firm

1. Profit-Maximizing Theories:

The traditional objective of the business firm is profit-maximization. The theories based on the objective of profit maximization are derived from the neo-classical marginalist theory of the firm. The common concern of such theories is to predict optimal price and output decisions which will maximize profit of the firm. We have already discussed these decisions in relation to the different forms of competitive structure from pure competition at one end of the spectrum to monopoly at the other. In essence the theories based on the profit- maximization goal suggests that firm seeks to make the difference between total revenue (or sales receipt) and total cost as large as possible. However, one pertinent question here is: does the firm attempt to maximize long term profit or short- term profit? The basic valuation model of the firm is based on the fundamental assumption that the firm seeks to maximize its long-term profit. According to this model, a firm seeks to

maximize its discounted present value. To arrive at an estimate of discounted present value of the firm we reduce future profits by a discount factor or weight, to make future profits comparable with present profits.

Let PV_f refer to the present value of the firm and $\pi_1, \pi_2, \ldots, \pi_n$ refer to profits in the next n time periods. Therefore, we can express PV_f as: $PV_f = W_1 \pi_1 + W_2 \pi_2 + \ldots + W_n \pi_n$

where $W_1, W_2, \ldots W_n$ are the weights we assign to future profits to be able to make inter-temporal comparisons of money sums. One complication that arises in this context is that the choice of weights largely depends on the firm's rate of time preference, i.e., how the firm values present profits compared to future profits. The short-run profit maximization hypothesis is based on the famous marginalist rule which we have explained. A firm maximizes profit when by producing and selling one more unit it adds as much as to revenue as to cost. The addition to revenue is called marginal revenue and the additional cost marginal cost. Thus, a firm maximizes profit when MR = MC. If this condition holds and if the MC curve intersects the MR curve from below and not from above, total profit (i.e., π = TR – TC) will be maximum.

However, if the periods are dependent short-run profit maximization will lead to incorrect decisions because of lack of provision for the future. For instance, the firm could generate higher profits now by not replacing capital goods, delaying payment on due accounts etc. all of which will surely reduce the size of future profits. By contrast, if profits are independent in different time periods, long-run profit maximization would simply amount to maximizing the series of short-term profits. But such a situation does not prevail in the real world. All firms which have made huge capital investments will observe that profits in different time periods are interdependent.

There is a trade-off between short-term and long-term profit. If more profit is derived in the long run, adequate provision has to be made for depreciation and short-term dues are to be cleared. If more profit is to be made in the short run, some long-term profit has to be sacrificed. With the above complications in mind we may now briefly discuss the traditional

theory. The essence of the traditional approach is to compare cost and revenue of a firm at different levels of output and to select the one which maximizes the absolute differences between the two. The short-run profit maximization hypothesis is illustrated in Figure 7.1. The TC and TR are shown on the vertical axis and output on the horizontal axis. The firm produces a level of output OQ* for which TR = OR* and TC = OJ and the gap between the two (R*J) is maximum. Thus Q* is indeed the profit-maximizing level of output. The slope of the TR curve measures MR and the slope of the TC curve measures MC. At points A and B, two curves have the same slope. Thus at OQ*, MR = MC. This can be verified by passing two tangents — one through A and the other through B and ensuring that they are parallel.

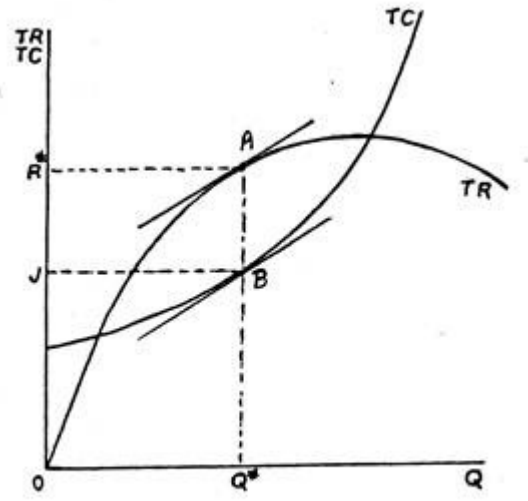


Figure 7.1
Single period profit maximization model

The total cost curve is always non-linear and has got nothing to do with the market structure. The slope of the revenue curve depends on elasticity of demand and is crucially dependent on the market structure. Since most real life markets are imperfectly competitive we assume non-linear total revenue function, too. Subtracting the TC curve from the TR curve we derive the total net profit curve π which cuts the horizontal axis where TR = TC. We reach the top of the profit hill when Q* is the level of output that is produced and sold.

In Figure 7.1 the firm produces OQ* units and makes a total revenue of OR* by charging a price of OR*/OQ*. At this stage total profit is R*J which is maximum. The hypothesis is based on a number of assumptions. Prima

facie, the decision-maker (manager or entrepreneur) is supposed to have relevant information about cost and revenue on the basis of which an optimal decision can be made. Secondly, he is assumed to have sufficient power to make a decision and implement it properly. However, the external or market forces — which are beyond the control of a firm or its management — are the major determinants of the firm's optimal decision on price and quantity. This theory is universally applicable.

Simple Mathematics of the Profit Maximization Hypothesis:

The equilibrium of the profit-maximizing firm occurs simultaneously on the input and output sides — i.e., a firm which is maximising its profit by choosing an output at which marginal cost equals marginal revenue is simultaneously minimizing the cost of producing that output, or maximizing the output subject to cost constraint. We can now prove that minimizing the cost of the prescribed level of output requires satisfaction of the same condition as doe's maximization of the output subject to cost constraint. So the latter condition is also a condition for profit maximization.

Minimizing the Cost of the Prescribed Level of Output:

Let Q = f(K,L) (1) be the production function, where Q is output and K and L are the quantities of two types of factor services. Let Q_0 = f(K,L) (2) be the prescribed output, and C = rK + wL (3) where C is total cost and r and w are the factor-service prices of K and L, respectively. Then, in order to minimize (3) subject to (2), form the function

$$S = rK + wL + \lambda[Q_0 - f(K,L)], \qquad (4)$$

$$\frac{\partial S}{\partial K} = r - \lambda\frac{\partial Q}{\partial K} = 0$$

$$\frac{\partial S}{\partial K} = w - \lambda\frac{\partial Q}{\partial L} = 0$$

$$\frac{\partial S}{\partial \lambda} = f(K,L) - Q_0 = 0$$

yield as a condition of equilibrium

$$\frac{\frac{\partial Q}{\partial K}}{\frac{\partial Q}{\partial L}} = \frac{r}{w}$$

The second-order conditions (not shown) require that the iso-quants be convex to the origin.

Maximizing the Output Subject to Cost Constraints:

Given equations (1) and (3) above, let the prescribed cost outlay be $C_0 = rK + wL$ (6) Then, in order to maximize (1) subject to (6) from the equation

Maximizing Profit:

In the case of pure competition, let the price of the goods be represented by p and profit by π. Then, from (1) and (3),

Maximum Profit

From the above hypothesis we may provide two important rationales for maximizing profit. Firstly, in a single owner firm, where the entrepreneur is both owner and manager, maximizing profit will maximize his own income. For a given amount of effort this is considered to be rational behaviour, irrespective of the structure of the market. If, however, the magnitude of profit varies with the amount of entrepreneurial effort expended, and effort has negative utility for the entrepreneur, rational behaviour would dictate something else. He must find an optimal trade-off between effort and profit to maximize entrepreneurial utility which is unlikely to lead to maximum profit. Secondly the impact of competition from rival firms forces the entrepreneur to maximize profits. Profit maximization therefore is not an aspect of discretionary behaviour but rather a compelled necessity. The entrepreneur is forced to maximize profit for his long-term survival. Thus, the justification for profit maximization depends upon the nature of competition. If competition is absent there is no such pressure, although the previous argument still holds. Under highly competitive conditions the entrepreneur has to maximize profit just for survival.

Criticisms of Marginalist Theory of the Firm:

The profit maximization hypothesis developed during 1874-1890 by Leon Walras, W. S. Jevons and Alfred Marshall has formed the basis of the neoclassical (marginalist) theory of the firm. It has not been challenged up to the 1920's. But from early 1930s it has been subject to various criticisms. Critics have argued that profit maximization is not the only objective of a firm. Modern business firms and their managers pursue certain other goals, too. Thus profit-maximization as the only goal of a firm is no longer a

tenable hypothesis. Being dissatisfied with both of the justifications, modern economists and management specialists have suggested various alternatives to profit- maximization.

The following arguments bear relevance in this context:

1. Emergence of oligopoly:

In the inter-war period it became increasingly apparent, especially in industrially advanced countries, that a modern economy was dominated by oligopoly, a market structure characterized by the existence of a few large firms. In a number of industries the structure has become gradually more and more concentrated (through merger or amalgamation) so that a few large (and dominant) firms accounted for a major portion of an industry's output. In such environments there was hardly any pressure on each firm to maximize profit independently. Instead firms arrived at joint profit maximization through such devices as collusions and cartels. Alternatively put, the pressure from rival producers was not strong enough to dictate profit maximization as an inevitable objective for each firm.

2. Separation of ownership from management:

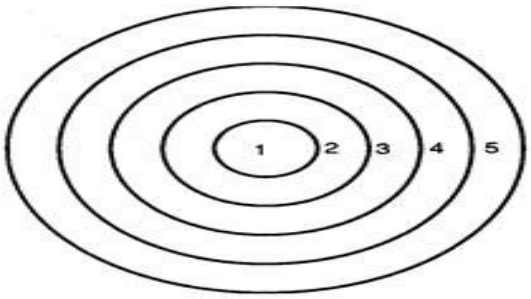
Secondly, in 1932, Berle and Means challenged, through their pioneering work, the argument that the firm would seek to maximize profits. They discarded profit-maximization as a rational behaviour because of an alleged break in the identity of purpose of the manager and his firm. They discovered that in most large U.S. companies there was separation of ownership from control. Most of such corporations were essentially in the control of the managers rather than the owners, due to fragmentation and dispersion of ownership of shares. Thus, in a handful of cases could a small group of shareholders' directly affect the decisions of the corporation? In such a situation, with managers acquiring only new shares, the identity of purpose of maximizing profits and maximizing entrepreneurial satisfaction was largely shattered.

In truth, the notion of the entrepreneur has lost relevancy with management becoming an executive function performed by a committee, rather than a simple individual taking all decisions unilaterally. The inevitable consequence of such divorce of ownership from control was that

managers may wish to pursue goals other than profit maximization, and would be forced to take into consideration the matter of profits to the extent that sufficient cash had to be generated to pay satisfactory dividends to the shareholders. J. Galbraith has noted the change in the power structure in a modern corporation. Figure 7.2 illustrates the traditional Viewpoint with shareholders holding the ultimate power and passing their decisions downwards in a chain of command — i.e., through the board of directors, management and eventually down to technicians and workers.

Figure 7.2
Traditional power structure

In Figure 7.3, we illustrate the alternative suggested by Galbraith for the modern corporation. Modern power structure consists of a series of concentric rings. Management is at the centre, controlling the firm and each ring outwards is successively less identified with the objectives of managers. Scientists and technicians are closest to managers, followed by white collar workers, blue-collar workers and finally shareholders.



1. Management
2. Scientists and technicians
3. White-collar workers
4. Blue-collar workers
5. Shareholders

Figure 7.3
Modern power structure

Criticisms of the Modern Approach:

Although this view has been accepted by many modern economists, the trend towards this type of change in power is not universal. Supporters of the traditional viewpoint would argue that the shareholders have ultimate power and, if properly motivated, can exert considerable influence. At times, at the annual general meeting of a company, shareholders are able to put a lot of pressure on managerial decisions. Secondly, it has been argued that an increase in the number of firms does not necessarily imply growing competition. There may be keen competition among 3 to 4 dominant firms in an industry. Thus the need for making maximum profit is not stronger under pure competition than under oligopoly. Those who believe that the profit- maximization is no longer a tenable hypothesis have suggested a number of alternatives.

These fall into two broad categories:

1. Those who hold that something else other than profit is maximized and
2. Those who postulate non- maximizing behaviour.

2. Other Optimizing Theories:

There are various alternative approaches to profit maximization. Here we restrict ourselves to the most important ones. Baumol's Single Period Sales Maximization subject to Profit Constraint: One alternative to profit maximization has been suggested by W.J. Baumol that firms operating in oligopoly will seek to maximize sales revenue subject to a profit constraint. His argument is largely, if not entirely, based on "public statements by businessmen and on a number of a priori arguments as to the disadvantages of declining sales, for example, fear of customers shunning a less popular product, less favourable treatment from banks, loss of distributors and a poorer ability to adopt a counter strategy against a competitor." Baumol's basic argument is summarized in Figure 7.4, which enables us to understand the difference between profit maximization and sales maximization.

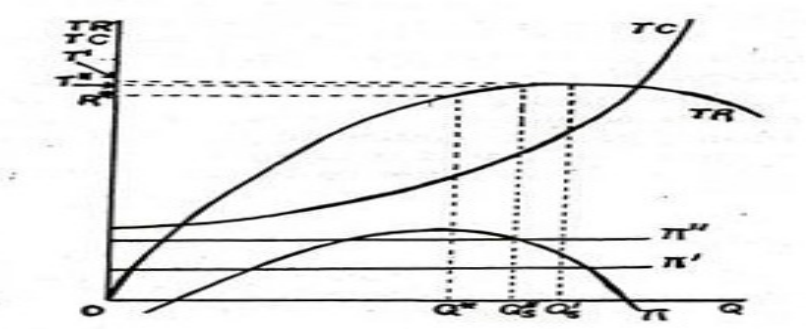


Figure 7.4
Baumol's sales-maximizing model

Total profit is maximized when the firm produces OQ* units of output (as in Figure 7.1) Sales maximization, on the other hand, refers to maximization of total revenue (= P x Q), rather than maximization of Π (It is because if a firm quotes zero price it can sell an astronomical amount but its total revenue will be zero.) Total revenue is maximum when MR = 0, and MR = 0 when the demand for a company's product is unitary elastic. In Figure 7.4 we observed that if the firm wishes to maximize total revenue (without profit constraint) it will choose output Q's, where TR is maximum (i.e., the slope of the TR curve is zero or MR = 0). However, Baumol has argued that, a constraint operates from shareholders. They require a minimum sum as dividend which would keep them content. Alternatively put, shareholder demand a level of absolute profit of some amount which is exogenous (i.e., determined outside the model). If this minimum acceptable level of profit were π', the firm could produce Q''_s and still generate profits greater than π'. Hence in this situation it will be worthwhile to produce Q'_s.

Likewise if the minimum acceptable profit is π", Q'_s will not generate sufficient profits. The firm will have to reduce output to Q''_s which is indeed the optimal output with the profit constraint specified. Baumol's model thus predicts that profits will be sacrificed for revenue. The sales-maximizing level of output will exceed the profit-maximizing level and can only be sold at a lower price under imperfectly competitive market conditions. In fact, the first main difference between the profit maximize and a constrained sales maximize is that the latter can charge a lower price to sell the extra (OQ''_s – OQ*) output. This has to be the case if both have the same demand (AR) curve. In terms of Figure 7.4, the profit maximize produces OQ* and charges

a price of OR*/OQ* (= total revenue + output). Alternatively, the sales maximize produces (in the π" constrained case) $Q"_s$ and sells at a price of $OT"/OQ"_s$.

Rationale:

Baumol's model no doubt carries enormous good sense. The motivation to maximize sales revenue is justified on the ground that the managers of large firms stand to gain more from this strategy than from profit maximization. Sales maximization implies expanding the size of the organization, enhancing the status of managers as also their promotion prospects. Again their wages and compensation are directly related to responsibility, which, in its turn, is again an increasing function of size. Conversely, as Baumol argues, it is quite irrational for managers to maximize profits for shareholders when they will get hardly anything themselves.

Implications and Limitations:

Baumol's' model is a single-period sales maximizing model. It applies at a single moment of time — i.e., it is static in nature. However, the model can be made dynamic for an in-depth study of multi-period optimization. For this it will be necessary to consider various combinations of sales and revenues over time. In that case profit would be endogenous and would form the vehicle for growth through reinvestment of funds. This would enable us to predict an optimal combination of profits and growth rate of revenue. Such a dynamic model is appended below.

With Advertising:

Secondly, advertising has been integrated into Baumol's model with consequent effect on the total revenue curve. Baumol's model has the implication that the sales-maximizing firm will spend more on advertising than the profit-maximizing firm. Here Baumol simply assumes that advertising does not affect the market price of the product. But it leads to increase in the volume of sales (with diminishing returns). Hence it is assumed that advertising will always lead to a rise in TR, i.e., MR will never be negative. Baumol's extended model is illustrated in Figure 7.5.

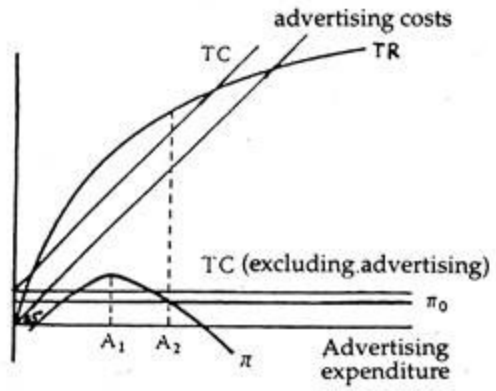


Figure 7.5
Sales maximization with advertisement

Here the TC line is derived on the basis of the assumption that advertising does not affect total non-advertising cost. Now we measure advertising expenditure on the horizontal axis, and profit, revenue and cost on the vertical axis. The TC curve is derived by superimposing the curve showing advertising cost, on the original TC. Since there is positive correlation between TR and advertising expenditure, the TR curve is upward sloping.

Since advertising will always increase TR, the businessman will go on increasing advertising expenditure until prevented by the profit constraint. In Baumol's model, therefore, A_1 will be the profit maximizing level of advertising expenditure, which, if falls short of maximum profits, will invariably be less than the constrained maximizer's expenditure A_2. Baumol's model, however, is not free from defects. It is inconsistent in one point at least. If advertising leads to greater output sold, non- advertising costs would be expected to rise. Yet, Baumol, in his simplified model, assumed that they would not. In Bamoul's scheme for a profit maximiser, this ratio is always equal to 1. As long as $\lambda > 0$, the advertising expenditure will be higher for a sales maximising firm. Thus, for a sales-revenue maximising firm, we arrive at the following

$$\left[\frac{\partial C}{\partial Q} \Big/ \frac{\partial R}{\partial Q}\right] = \left[\frac{\partial (AE)}{\partial A} \Big/ \frac{\partial R}{\partial A}\right] = \left[1 + \frac{1}{\lambda}\right]$$

The implication is that excess profit or surplus will be partly utilized for advertising and partly for enhancing production. Baumol's model can be

generalised with respect to multi-product firm. The product-mix of a revenue-maximize will not be the same as that of a profit-maximize.

The Dynamic Model:

The multi-period model of Baumol is based on the following assumptions:

The objective of the firm is to maximize the rate of growth of sales revenue over its life cycle. There is no profit constraint; profit is the main source of financing growth of sales. Profit is thus an instrumental variable whose value is endogenously determined. Demand and cost curves have traditional shape; average revenue is downward-sloping and average cost is U-shaped. Suppose sales revenue (R) grows at a rate of growth (g) per cent.

Over its whole life the firm will have the following stream of revenues:

R, R (1 + g), $R(1 + g)^2$... $R(1 + g)^n$ The present value of this stream of future revenues can be computed by applying the usual discounting procedure.

$$R, R\left(\frac{1+g}{1+r}\right), R\left(\frac{1+g}{1+r}\right)^2, \cdots, R\left(\frac{1+g}{1+r}\right)^n$$

where r is the rate of discount determined by the level of expectations and risk preferences of the firm. The total present discounted value of all future revenues is expressed as:

$$S = \sum_{i=0}^{n} R\left(\frac{1+g}{1+r}\right)^i$$

The firm seeks to maximize s by choosing an appropriate combination of current values of R and g. It is pretty obvious that

$$\frac{\partial S}{\partial R} > 0 \quad \frac{\partial S}{\partial g} > 0$$

Also note that g = g(π, R) is the growth function and π = π (R, C, g, r) is the profit function. The- growth function is derived from the profit function. Growth is mainly financed by ploughed back profits which depend on current level of revenue (R), cost(C), growth rate of sales (g) and the discount rate (r). To maximize S, the firm can choose a particular combination of R and g out of a set of alternatives. These combinations are plotted along the growth curve, shown in figure 7.6. In this diagram up to point a, which corresponds to maximum profit level, R and g increase

simultaneously. Beyond A, R increases but g tends to fall. Thus beyond $R_{\pi n}$ sales revenue level and growth rate become conflicting goals.

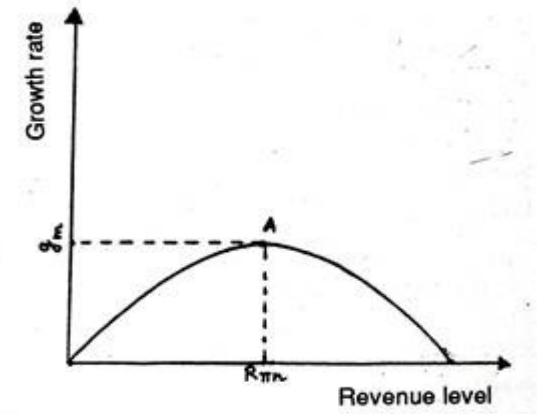


Figure 7.6
The growth curve

The optimum combination of R and g may not be a feasible one and vice-versa. Actual choice depends both on desirability and on feasibility. The desirability may be defined in terms of iso-present value curve. This curve is a locus of points showing alternative combinations of g and R which yield the same S. Here S, the aggregate discounted present value of revenue, depends on R and g, given the exogenously determined discount rate. Thus we may assume that

$$S = a \cdot g + b \cdot R \cdots \text{such that}$$
$$g = \frac{1}{a}S - \frac{b}{a}R \quad \text{and} \quad R = \frac{1}{b}S - \frac{a}{b}g$$

This is an equation of the iso-present value curve in the slope-intercept form. Thus, it is possible to think of a family of such curves, the highest one representing the maximum present value of S and the lowest one representing the minimum present value. The slope of this straight-line is given by a/b along a given curve, the level of S remains the same.

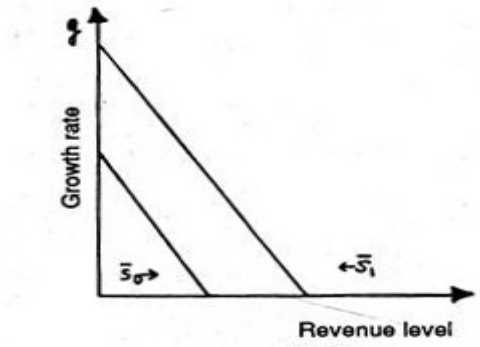


Figure 7.7
The iso-present value curve

In order the choose the optimum combination of R and g, it is necessary to put the previous two diagrams together and design it as a case of growth- constrained iso-present value of revenue maximization. In this case, the equilibrium solution is reached at point E at Figure 7.8 from which it appears that the firm will choose a combination of R* and g* to reach the highest possible level of S, subject to the growth function constraint.

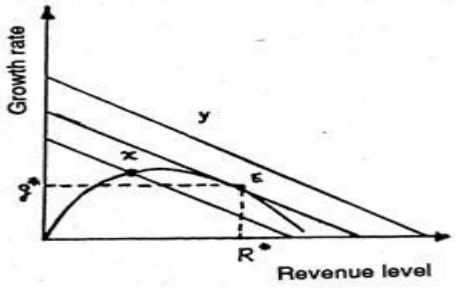


Figure 7.8
Growth-constrained revenue maximization

Empirical Evidence:

Two major studies have both verified and falsified Baumol's hypothesis. In 1962, McGvire, Chiv and Elling have discovered "that correlation between executive incomes and sales revenue is stronger than the correlation between executive income and profits". However, such correlation does not necessarily imply causation. In 1967, M. Hall in a comprehensive study has attempted to test the hypothesis, implicit in Baumol's model, that if profits above the minimum constraint are earned, ceteris paribus, firms pursue policies (for example, cut prices, increase advertising and investment) in order to increase their sales revenue. The regression results of Hall failed to justify this hypothesis.

Reconciling Short and Long-term Profitability:

There are, of course, a few practical problems in reconciling profit maximization in the short run with the long-term interest of the firm. A realistic strategy seems to be to always increase its profits in the short term by cutting down costs which ensure its long term survival, i.e., maintenance and investment. Simply put, there is danger of offending customers by excessive profit-making openly in the short run.

Williamson's Model and Maximization of Management Utility:

In his article, 'Management Discretion and Business Behaviour' in American Economic Review (1863), O.E. Williamson presents a model of managerial discretion. His model is based on the same assumption as Baumol's: a weak competitive environment, a divorce of ownership from control, and a minimum profit constraint imposed by the shareholders. He argues that managers of such large firms conduct the affairs of the firm to serve their own interests. In other words, managers are concerned with the goodwill of the firm only to the extent that it favours their own personal motives and ambitions. He argues that the most important motives of businessmen are desires for salary, security, dominance and professional excellence. All these yield additional utility or satisfaction to the manager. These can be gained by incurring additional expenditure on staff, managerial emoluments and discretionary investment. Williamson argues that managers have discretion in pursuing policies which maximize their own utility rather than seeking the maximization of profits which maximize the utility of most shareholders. In Williamson's model, each manager is supposed to have a utility function — i.e., a set of factors which provide managerial satisfaction. Such utility arises from certain aspects of the management task — e.g. responsibility, prestige, status, power, salary, etc.

These aspects can be reduced to three component terms in the utility function as follows:

$U = f(S, M, I_d)$ where U = managerial utility, S = staff, M = Managerial slack, absorbed as a cost, and I_d = discretionary power for investment. Of these, only staff is measurable. Others are non-momentary and non-operational. Still these can be measured indirectly in terms of other variables. The objective of the manager is to maximize U. An increase in the

staffing level — or an expansion of the 'span of control' (i.e., the number of people under the direct control and supervision of the manager) confers benefits to the managers in the form of higher salary. Usually other things being equal, a manager, in charge of a team of 30 people is paid more than another manager in charge of a smaller team. However, there are other positive aspects of the staff term (S) such as the higher status of managing a bigger team, and the stronger chance of promotion which comes from greater responsibility and authority.

In short, the quality and number of staff reporting to a manager enables him to gain promotion, salary and dominance as also security through greater confidence as to his departments' survival, and greater professional excellence which a large staff, by providing better services. Thus the staff term is a much wider one than simply measuring managerial salary. The second term management emolument (M) represents the type and amount of perquisites the manager usually enjoys (such as luxurious, decorated and equipped offices, personal security, allowances for the use of a car, expense account for entertainment) beyond the level necessary for efficient operation. The term M reflects the utility derived by the manager from being able to authorize expenditure of the firm to serve his own needs.

The manager is often able to undertake projects which appeal to him in particular but which may not necessarily be the best in terms of generating profits for the firm. Examples of such investment are terminals linked to a computer, mini-computers, automated equipment for data processing and record keeping. Such projects do not provide any monetary benefit to the manager but reflect his fascination for what is 'new', what can be deemed as 'scientific progress' and what may put him above other managerial staff in terms of esteem and status. Such investments permit managers to pursue their personal investment preferences and to exercise their power. Hence I_d provides utility.

In Williamson's model, the utility function is maximized subject to the constraint that satisfactory profits are earned to fulfil the shareholders' expectations. He predicts from his model that in most normal situations the firm will act in such a fashion that M and are both positive. The implication

is that 'unnecessary' expenditure is tolerated by the shareholders. The normal situation is supposed to be one in which the firm enjoys discretionary 'power' in fixing output and price if there is not much competition in the market. In such a situation S is also positive, which implies that excess staff has a positive effect on managerial utility. In times of business recession, it becomes difficult to earn satisfactory profits. Hence the components of the utility function are appropriately adjusted for cost reduction. Excess staff is laid off, expense accounts are made more stringent, and unnecessary prestige investments are cut back or postponed.

Comparison with Other Models:

Williamson's model can be compared with the traditional model presented by the marginalist school. In a highly competitive environment, M and l_d would have to fall to zero if profit is to be maximized. Moreover, excess staff has to be removed. In this limited case the predictions of Williamson's model would be similar to that of the traditional marginal approach. But this is unlikely to happen in reality. Williamson's model may also be compared with Baumol's. Whereas the profit-maximizing firm of the traditional model and the sales maximizing firm of the Baumol's model report actual profits, Williamson's firm announces only 'reported' profits.

Reported profits, i.e., the profits admitted by the firm equal actual profits less M. M is deducted because it is an expenditure and is also a deductible one for tax purposes. It is interesting to note that in Williamson's model, actual profits may not equal maximum profits if, as the model predicts, S exceeds the profit maximization level.

These are the following:

(a) The demand of the firm, P:

The firm is supposed to have a known downward sloping demand curve such that P is price per unit Q is output S is staff expenditure E is a demand shift parameter.

$$\text{Thus, } \frac{\partial P}{\partial Q} < 0; \frac{\partial P}{\partial S} > 0 \text{ and } \frac{\partial P}{\partial E} > 0$$

(b) The production cost, C:

Cost is a function of output, i.e., it depends the level of output such that

$$C = C(Q) \quad \frac{\partial C}{\partial Q} > 0.$$

(c) Actual profit, Π_a:

Actual profit is sales revenue less total costs including staff expenditure

$\pi_a = R - C - S$

(d) Reported Profit, Π_r:

This actual profit is reported to the tax authority after deducting of managerial emoluments

$\pi_r = \pi_a - M = R = C - S - M$

(e) Minimum Profit Π^*:

This is indeed the minimum amount of profit (after tax) needed to pay satisfactory dividends to the shareholders, without which the 'job security' of the manager may be at stake. Thus

$$\Pi^* \geq \pi + T \text{ , and}$$

$T = T + t\pi$, where T = lump-sum tax

t = marginal tax rate.

(f) Discretionary investment, I_d:

This is a residue of the amount left from the reported profit after setting aside the critical minimum profit and meeting the tax obligations. Thus $I_d = \pi_r - \Pi^* - T$

(g) Discretionary profit, Π_D:

This is also a residue, i.e., the amount of profit left after subtracting the minimum profits and the tax $\pi_D = \pi_a - \Pi^* - T$

Implication of the Model:

In the following table we seek to make an overall assessment of the utility-maximising firm vis- a-vis the profit maximising firm: E. Penroe, in his book The Theory of Growth of the Firm describes all types of growth to the availability of managerial resources to plan growth. At any time there will be a range of opportunities for expansion open to the firms. Since all these cannot be exploited choice must be restricted to those which can be effectively planned. The implication is that there must be available spare managerial capacity at the right level over and above that needed for the smooth conduct of current operations. In her scheme the ultimate limit to

growth, under static conditions, is set by managerial discretion. If, however, improvements in these occur under more realistic conditions, the limits recede to the background for extended periods of time.

Marris's Model of Managerial Enterprise:

An alternative managerial theory of the firm has been developed by Robin Marris. It also stems from the so-called dichotomy between ownership and control. He suggests that a possible goal which has connections with both sales and profits is that of growth of the firm. So managers will have varying objectives apart from profit. These non-profit objectives are strongly correlated with the size of the firm, examples being salary, power and status. An important exception is that of security, since in recent years managers, even in larger firms, have found themselves declared redundant.

In fact, Marris, like Williamson, hypothesizes that managers have a utility function in which salary, prestige, status, power, security, etc. all assume significance. On the contrary, the owners (shareholders) are usually more concerned with profits, market share and output.

In contrast to Williamson, Marris suggests that on one aspect at least, there is no conflict between the two groups — the management team and the shareholders. Rather there is harmony of interest. They have a common interest in the size of the firm. Thus he postulates that members of the management team will be primarily concerned with maximization of the rate of growth of size. By size he means: 'corporate capital, that is, the book value of fixed assets, plus inventory, plus net short- term assets, including cash revenue.

Managers feel interested in growth rate of size because positive growth is supposed to enhance the promotion prospects of managers. In Marris' model the stress is on an alleged preference of managers for internal promotion (rather than through changing firms). This is possible if and only if the firm expands rapidly over time. However, Marris suggests that there are certain factors which operate within the firm to limit the growth process such as:

- The ability of managers to cope with and administer a rapidly growing organization without any loss of control,

- The ability of managers to develop and introduce new products to neutralize the losses inflicted by products experiencing falling market shares and
- The ability of the research and development expenditure to generate an expanding flow of potential new products.

However, the major constraint on growth seems to stem from the managers' desire for security, which largely, if not entirely, depends on the financial side of the enterprise. Managers of big companies do not want to lose their jobs. Thus they never pursue the growth objective beyond limit so that the company suffers from financial stringency and its very existence is at stake. In other words, the desire of the management for job security implies a deliberate brake on the growth process. If job security is accorded the highest priority among managerial objectives the firm has to grow in such a fashion that its financial side is not damaged.

Since excessive dependence on external finance implies loss of control or too much borrowing may enhance the chance of take-over by another firm and pose a threat to the job-security motivation of the managers. Hence there is desire for growth financed mainly from the profit levels being generated by the existing products. Hence Marris postulated a theory of balanced growth, i.e., growth in demand for the firm's products balanced by growth in supply. The need for balanced growth is felt for two reasons. Prima facie, there are risks in expanding too fast by undertaking very risky projects, by putting undue pressure on the managerial input, and/or by incurring huge debt to finance the expansion. By contrast, there are dangers associated with slow growth such as lack of initiative in identifying new products or markets, excessive revenues not being invested into new projects, and, above all, allegations of slack or uninventive management. The failure on the part of a firm to expand rapidly enough could lead to take-over bids by other firms with more active, energetic and dynamic managers who are aware of the potential which is not being utilized in the slow-growing firm.

Formal Presentation of the Model:

Thus, Marris has presented a dynamic model of the firm, by stating clearly the objectives and constraints. However, he presents his model formally in mathematical and/or graphical form. We present below the model in a simple mathematical form.

Objectives and Constraints:

In Marris's model the optimization goal of the firm is maximization of the balanced rate of growth (G) which internally depends on two factors: the rate of growth of demand for the firm's product (G^d), and the rate of growth of capital supply (G^s). Thus $G = G^d = G^S$ The firm seeks to pursue this balanced growth objective, subject to two major constraints: managerial and financial. The managerial constraint is set by the skill and efficiency of available manager's team. The financial constraint is set by the desire of managers to attain the maximization of their own utility function and their owner's utility function. In a modern organization, there is separation of ownership from management. This is why owners and managers are supposed to have conflicting interests. But at times there may be harmony of interests. One such common area of interest not only ensures fair returns on owner's capital but also continued trust and faith in managers who have succeeded in achieving it.

If the firm is unable to achieve balanced growth, managers run the risk of losing their jobs as owners' capital is at stake. Thus when the goals of managers and that of owners coincide, they may collaborate in their endeavour to achieve a common goal, viz., balanced growth of the firm. It is against this backdrop that Marris specifies two different utility functions one of the manager and the other for the owner. The utility functions of the manager (U_m) include such variables like salaries, power, status, job security etc., while treat of the owner (U_0) includes variables like profits, capital, output, market-share, public esteem, etc. But in the ultimate analysis most of these explanatory variables are related to the size and steady growth of the firm.

Thus in Marris model we have: $U_m = m(G^d,S)$ $U_0 = O(G^s)$ At the outset, Marris treats S as an exogenously determined constraint by assuming that

there is a saturation level of job security. Above that critical level, $(\partial U_m/\partial s) = 0$, while below that level, $(\partial U_m/\partial s) = \alpha$. If this assumption is made the managerial utility function may be expressed as $U_m = m (G^d)^{\bar{s}}$ where

$s = \bar{s}$ is the job security constraint. We may now have a fresh look at the above constraints. We may first focus on the managerial constraint. Marris adopts Penrose's thesis that there exists a definite limit on the rate of managerial expansion such that 'managerial ceiling' sets an upper limit to the growth of a firm. Secondly the financial constraint can also set a limit to growth and this constraint originates in the job- security considerations. In view of job-security, the managers become risk-averters by choosing a prudent financial policy which consists of determining optimum levels of the following critical financial ratios: These three ratios may now be combined into a single parameter, r, to represent the financials security constraint.

To affect the balanced growth of the firm, Marris make use of instrumental variables.

r, the financial security co-efficient

d, the rate of product diversification

p, the average profit margin.

Now by combining the set of objectives, constraints and instruments, we may present the complete model thus:

Structure of the Model:

(1) Demand growth equation	:	$G^d = D(P,d)$
(2) Profit equation	:	$\pi = \pi(p,d)$
(3) Supply of capital equation	:	$G = r[\pi(p,d,)]$
(4) Security constraint	:	$\bar{r} \leq r$
(5) Balanced growth equilibrium condition	:	$G^d = G^s$

In this model the level of profit, n, is endogenously determined, i.e., determined from within the system whereas the security constraint, r, is exogenously determined by the attitude of the managers toward risk. Given this, the balanced growth can be ensured through the operation of two instrumental variables, p and d. In balanced growth formulation we have, in fact, one equation in two unknowns:

$D(p, d) = \bar{r}[\prod(p,d)]$,

Thus from Marris' dynamic model emerges the possibility of an optimum growth rate of the firm through time in a more or less unchanged environment. In finding this optimum Marris has referred to two major constraints — management capacity to successfully generate greater demand (i.e., managerial constraint) and ability of existing products to generate sufficient after-tax profit for reinvestment (i.e, financial constraint). There is also a third constraint, namely, profit. On this point Marris' model resembles that the Baumol and Williamson very closely. Likely them, Marris also includes a profit constraint so that if the growth-maximizing solution fails to generate sufficient profits, growth will have to be sacrificed somewhat to increase current dividend payments so as to fulfil the expectations of shareholders. From the point of management Marris's model throws light on two important factors: the attitude to risk and unvertainty and the desire for utility (subject to security maximization) which may not be maximized by the pursuit of maximum profits.

We have noted that the most celebrated managerial models are those of Baumol, Marris and Williamson. They are distinguished primarily by the assumed objectives of the managers. Baumol suggested that managers maximise sales revenue, Marris that they maximise growth, and Williamson that they maximise a utility function including 'staff or 'emoluments'. In each case the existence of monitoring from outside and limits to managerial discretion were explicitly recognised. Baumol included a minimum profit constraint in his model, and Marris similarly incorporated a valuation ratio constraint to reflect external pressure, i.e., from shareholders. The value of the assets of a firm. According to Marris too low a ratio will involve a risk of takeover 'unacceptable' to the management.

In many ways figure 7.9 is absolutely typical of diagrammatic representations of managerial models of the firm. In Williamson's managerial firms the constraint OW derived as the summation of marginal revenue minus marginal cost. In other words Williamson's firm is a monopo-list. For Marris the diagram is again basically the same with the horizontal

axis now measuring the rate of growth and the vertical axis the valuation ration. The constraint is not supposed to emanate from the origin but is likely to have the same concave shape. If growth is pushed past a certain point the value of shares on the market will fall as diseconomies associated with staff training and encountered (Penrose effects) and as a greater proportion of earnings are retained in the firm to finance expansion instead of being distributed as dividends to shareholders.

3. Non-Optimizing Theories:

By criticizing the profit-maximization hypothesis modern economists have developed certain theories of the firm which do not hypothesize any optimizing behaviour. We have noted that the most celebrated managerial models are those of Baumol, Marris and Williamson. They are distinguished primarily by the assumed objectives of the managers. Baumol suggested that managers maximise sales revenue, Marris, that they maximise growth, and Williamson, that they maximise a utility function including.'staff or 'emoluments'. In each case the existence of monitoring from outside and limits to managerial discretion were explicitly recognised. Baumol included a minimum profit constraint in his model, and Marris similarly incorporated a valuation ratio constraint to reflect external pressure, i.e., from shareholders. The valuation ratio is the market value of outstanding equity shares divided by the book value of the assets of a firm. According to Marris too low a ratio will involve a risk of takeover 'unacceptable' to the management.

In many ways absolutely typical of diagrammatic representations of managerial models of the firm. In Williamson's managerial firms the constraint OW is derived as the summation of marginal revenue minus marginal cost. In other words Williamson's firm is a monopolist. For Marris the diagram is again basically the same with the horizontal axis now measuring the rate of growth and the vertical axis the valuation ratio. The constraint is not supposed to emanate from the origin but is likely to have the same concave shape. If growth is pushed past a certain point the value of shares on the market will fall as diseconomies associated with staff training are encountered (Penrose effects) and as a greater .proportion of

earnings are retained in the firm to finance expansion instead of being distributed as dividends to shareholders.

Satisfying:

Being dissatisfied with the profit- maximization models of economists in 1955, H. A. Simon (the 1978 Nobel Laureate in Economics) has put forward the hypothesis that firms run by single enterprisers (who are also the owners) are likely to have different objectives from firms operated by modern executives in large corporations. Simon argues that managers in most cases have imperfect knowledge and inadequate information on the basis of which to take decisions. In fact, if perfect knowledge and complete information were not available, the calculations involved in the decision-making process would be too complex to be practicable; and that, given this and the other inevitable uncertainties surrounding the decision making process in reality, business people can never be confident whether they are maximizing profits or not. Instead, business people "satisfy" rather than maximize, i.e., their aim is to earn just satisfactory profits.

Simon basically puts forward the proposition that firms have an 'aspiration level' which they seek to reach. In fact, what the satisfactory aspiration level of profits will be depends on past experience and will take account of future uncertainties, too. This level may be in terms of sales, market share, profits, etc. For any -fixed period of time actual results are compared with the aspiration level. If actual performance exceeds the aspiration level, no corrective action is called for. Instead the aspiration level for the next period is revised upward. On the contrary, if actual performance falls short of the aspiration level, the firm attempts to identify or search out the causes of discrepancy by spending sufficient time, effort and money.

Alternatively, if no apparent inefficiency is found (and the shortfall is believed to be due to external factors — factors beyond the control of the firm or its management) the firm will be constrained to revise its aspiration level for the next period downward. The aspiration level is, of course, the consensus of what can reasonably be expected in near future in the light of past performance.

However, since the cost of gathering information is high, all the alternatives will not be explored. A satisfactory alternative course of action if likely to be selected. This will probably not be the profit- maximising alternative. Simon also argues that if neither searches behaviour nor the lowering of aspiration levels quickly results in the achievement of a 'satisfactory' situation then the manager's behaviour pattern will become one of apathy or of aggression. In this sense this model does not have managerial usefulness. Simon has also argued that the effort of trying to squeeze the last rupee of profitability out of the operation of the firm is likely to put extra strains and stresses on the business manager which is most cases, may not be liked by him. He therefore seeks to reach a level of profit which yields an income which he regards as satisfactory and does not put any special effort to extract any extra rupee of benefit. He satisfies rather than maximises.

The validity of Simon's hypothesis (i.e., the desire of businessmen for quiet life) largely depends on the business environment. In a highly competitive environment, a businessman has to work hard in order to safeguard his position (and thus protect his market share), whether he likes it or not. On the contrary, if there is not much competitive rivalry in the area of business in which he is operating, he can afford the luxury of 'quiet life' and Simon's hypothesis may carry enormous good sense. However, a related point may be noted in this context. In a single-owner firm (i.e., sole proprietorship concern) it is possible for the owner-manager to 'satisfice' rather than maximize. But it is not possible for the head of a managerial team in a joint stock company to behave like this.

He may well be subject to various pressures from below to pursue a more expansionist policy. The pressure may come from those who are ambitious but are placed less comfortably than he is (i.e., at a lower point in the organization chart). Shareholders may also demonstrate this type of 'satisfying' behaviour. A private shareholder is always at liberty to sell the share of a company if he is not satisfied with its performance and feels that he can secure a better return on his investment elsewhere. But he is usually constrained by a lack of information. Thus he tends to act as a 'satisfier' so

that if, for instance, dividends are held at a customary level, shareholders do not usually inquire whether they should be higher if management were better.

Comments:

Simon's hypothesis is not altogether wrong. There are firms or business people in reality which pursue 'satisfying' behaviour. As noted by W. D. Reekie and J. N. Clook, "It does help explain why some firms, faced with a falling market share, act more vigorously than competitors, in an attempt to halt the decline, while others, conversely, in the same situation, act as though they were commercially moribund." However, Simon's theory of satisfying rather than optimising behaviour forms the basis for a more detailed analysis of the objectives of firms as spelled out by Cyert and March in their behavioural theory of the firm. We may now turn to the behavioural theory.

The Behavioural Theory of the Firm:

In their book A Behavioural Theory of the Firm (1963), Cyert and March go a step ahead of Simon in making an in-depth study of the way in which decisions are made in the large modern (multi- product) firm (characterized by divorce of ownership from management) under uncertainty in an imperfect, market. They have expressed more concern in the decision making process than in the objectives or motivations of such firms (e.g., profit/sales maximization and satisfying). They look at the bureaucratic structure of the firm and study the nature of interrelationships of its various parts. At the outset Cyert and March declare that if we are to develop a theory that predict and explain business decision making behaviour, the following two points have to take note of:

I. People (i.e., individuals) like organizations have goals,
II. In order to define a theory of organizational decision making, we need something analogous — at the organizational level — to individual goals at the individual level.

Cyert and March set the formation of organizational goals through the notion of a coalition. The firm itself is visualized as a coalition of individuals who are organized in sub-coalitions. So they differ from Baumol and Simon

who have assumed that the firm is dominated by a single person who makes the decisions and whose authority is unquestioned. Instead Cyert and March assume that the firm is a decision-making entity in its own right. They have recognized that management must achieve an objective, or possibly a set of objectives, through the efforts of a group of persons or through a coalition. The coalition consists of the various units or parties associated with the firm such as managers, workers, shareholders, customers, suppliers as also professional people like accountants, auditors, lawyers, etc.

As with most others, such coalitions are not necessarily stable. Membership may change over time and also when particular decisions are involved. Within any group there is unlikely to be any permanent unanimity of purpose, although it may be worthwhile or expedient to act for a time as though there were. There is still less chance of acceptance of the goals of the firm by all the members of the coalition. Thus the overriding problem of the leader of the coalition, who may be designated as the entrepreneur, is to attempt to resolve the conflict of goals and to keep all members pulling, more or less, in the same direction as long as possible. However, he must always be prepared for an unforeseen situation or sudden emergency.

The starting point of the behavioural theory is "where the entrepreneur makes a contract with the individual whereby the latter agrees to carry out instructions and to accept the organizational goal, or goals, as interpreted by the entrepreneur." In order to get full support from the subordinates, the entrepreneur has to make 'side payments'. Alternatively put, the goals keep on changing through a process of bargaining, in which side payments are involved. Side payments not only involve money but non-pecuniary benefits also like authority, personal treatment, etc. At the management level these involve matters outside the normal contract of employment (salaries, paid holidays, hours of work, etc.). The most important one seems to be policy commitments of one kind or another. This is known as policy side payment. Finally, a winning coalition forms and the goals are set. However, the position is not static. Due to continuous changes in circumstances the

bargaining is going on most of the time so that the coalition and its goals are liable to alter frequently.

In other words, a process of unrestricted bargaining would be inconsistent with stability in the organization. However, stability can be secured by working outside payments for those situations that are thought likely to occur. There is, of course, likely to be conflict within such a coalition. Thus it is quite likely that some of the goals may be incompatible. However, such conflict resolution is possible in two ways. Firstly, decisions may be decentralised into divisions and departments. Therefore conflict may be isolated geographically to ensure that all conflicts do not arise within the same unit. Secondly, crises and conflicts may be dealt with sequentially, i.e., they can be spaced out intertemporally and can be tackled as and when they arise.

Five Goals:

(a) Production goal:

This goal will be set as a target for the period and will have two aspects: level and smoothness. For example, a division may be set up to reach a specified goal (say, producing 100 units of a commodity per day) with the restriction that output should not deviate by more than 10% from this figure.

(b) Inventory goal:

Business firms have to hold inventories because production and sales do not always coincide. It is absolutely essential to hold sufficient stocks of finished goods to meet consumer demand (as and when it arises). At the same time, it is to be ensured that there is no excessive stock holding at high cost. This goal may be specified in terms of a target level and upper and lower limits may be set.

(c) Sales goal:

This goal may be specified for the future either in volume or in value terms. Moreover it may again be expressed in terms of a level and/or range.

(d) Market share goal:

The firm may set a target related to its share of the market (i.e., the industry of which it is a part for the product concerned). In some cases this

may be a substitute for the sales goal, but in other cases it may be a supplementary goal.

(e) Profit goal:

The purpose of setting this goal is twofold: to measure the effectiveness of management and to act as a source of payment of dividends to shareholders.

Search Activity:

The behavioural theory does not postulate goal maximization but seeks sub-optimization or attainable goals. Like Simon, Cyert and March state that firms compare performance with goals. What will be sought at any time largely depends on the level of aspirations. If the goal is met no action will be taken. But in practice the level of aspirations, in most cases, outstrip achievements. In contrast, if achievement improves rapidly, then it may out-strip aspirations, which may then be revised upwards. In a like manner, where achievement worsens there may be a tendency for a downward revision of aspirations to occur. There is thus likely to be a certain adjustment of goals in the light of experience.

If, however, performance falls short of aspirations (i.e., the goal is not met) a search activity is initiated to identify the causes of non- attainment. If the reason is within the firm's compass, steps are taken to rectify the non-attainment (i.e., alternative courses of action will be stimulated). This imposes extra costs on the firm and will not be carried beyond the point where a satisfactory solution is found. If a number of alternatives are found, the best one will be selected and no additional search will be carried out to see whether any further improvement is possible. If the reason is outside the control of the firm (e.g., depressed market conditions due to recession in the economy) the goal for the next period may be revised downward.

Organizational Slack:

Cyert and March argue that the coalition will remain viable so long as the payments are sufficient to keep the members within the organization. So it is absolutely essential to develop a satisfactory 'package' of money together with other benefits which will prevent the individual manager from looking for openings elsewhere. In practice, however, there is likely to be dis-

parity between the actual payment which is made and that which is necessary to keep the individual in the organization.

However, it is not that easy to calculate side payments accurately. Usually payments made, tend to exceed what is really necessary. Such excess payment is termed organizational slack. The concept is of considerable importance in rectifying the non-attainment of goals.

The following three examples bear relevance in this context:

I. Shareholders are often paid more than what is required to keep them holding shares.
II. Wages are often in excess of those required to keep workers within the organization.
III. Executives in most cases are provided with luxuries and services in excess of what they really need.

Cyert and March argue that organizational slack (OS) grows naturally as the firm itself grows and prospers over time; it is not a deliberate objective. However, when circumstances become more and more adverse, OS provides the first means of making economies on costs. Under difficult conditions there will be real pressure to reduce those side payments which can no longer be afforded at their original level. This slimming operation will, in all likelihood, reduce the organizational slack, while, at the same time, still leave the members of the organization sufficiently satisfied to stay within it.

Conclusion:

The behavioural approach of Cyert and March is a dynamic one.

Three major points that emerge from the approach are as follows:

1. The goals and objectives of a firm will emerge from the coalition in existence, at any given point of time.
2. However, there is likely to be a change in coalition, and with it, the objectives pursued by the organization as a whole.
3. Hence, not only different firms will have different objectives at the same point of time, but the same firm may have different aims and objectives at various time periods.

Growth Maximisation Theory of Marris: Assumptions, Explanation and Criticisms

Robin Marris in his book The Economic Theory of 'Managerial' Capitalism (1964) has developed a dynamic balanced growth maximising model of the firm. He concentrates on the proposition that modem big firms are managed by managers and the shareholders are the owners who decide about the management of the firms. The managers aim at the maximisation of the growth rate of the firm and the shareholders aim at the maximisation of their dividends and share prices. To establish a link between such a growth rate and the share prices of the firm, Marris develops a balanced growth model in which the manager chooses a constant growth rate at which the firm's sales, profits, assets, etc. grow.

If he chooses a higher growth rate, he will have to spend more on advertisement and on R & D in order to create more demand and new products. He will, therefore, retain a higher proportion of total profits for the expansion of the firm. Consequently, profits to be distributed to shareholders in the form of dividends will be reduced and the share prices will fall. The threat of take-over of the firm will loom large among the managers. As the managers are concerned more about their job security and growth of the firm, they will choose that growth rate which maximises the market value of shares, give satisfactory dividends to shareholders, and avoid the take-over of the firm. On the other hand, the owners (share-holders) also want balanced growth of the firm because it ensures fair return on their capital. Thus the goals of the managers may coincide with that of owners of the firm and both try to achieve balanced growth of the firm.

Assumptions:

The Marris model is based on the following assumptions:

1. It assumes a given price structure.
2. Production costs are given.
3. There is no oligopolistic interdependence.
4. Factor prices are constant.
5. Finns are assumed to grow through diversification.

6. All major variables such as profits, sales and costs are assumed to increase at the same rate.

Explanation:

Given these assumptions, the objective of the firm is to maximise its balanced growth rate, G. The G itself depends on two factors: First, the rate of growth of demand for the firm's product, GD; and second, the rate of growth of capital supply, GS. Thus G = GD == GS. Despite the fact that in modem big firms ownership is divorced from management, owners and managers have a common goal of balanced growth of the firm. According to Marris, there are two different utility functions for the manager and the owner of the firm. The utility function of the manager consists of his emoluments, status, power, job security, etc. On the other hand, the utility function of the owner includes profits, capital, output, market share, etc.

Thus the manager of a firm aims at maximising his utility, and his utility depends upon the rate of growth of the firm. Though promoting the growth of the firm is the main aim of the manager, yet he is also motivated by his job security. The manager's job security depends upon the satisfaction of shareholders who are concerned to keep the firm's share prices and dividends as high as possible. Thus the manager aims at maximising the rate of growth of the firm and the shareholders aim at maximising their profits in the form of dividends and share prices. Marris analyses the means by which the firm tries to achieve its growth-maximisation goal.

The firm may grow in size through the creation of new products which create new demands. Marris calls it differentiated diversification. The introduction of new products depends upon the rate of diversification, advertising expenses, R&D expenditures, etc. Marris establishes the relationship between growth and profits on the demand side through diversification into new products. The links between growth and profits are different at different levels of growth. In this growth-profits relationship, growth determines profits. When the rate of growth of the firm is low, the relationship is a positive one. As new products are introduced, the firm expands and profits increase. With the further increase in the growth rate

due to greater diversification into new products, the growth-profits relationship becomes negative. This is because there is the managerial constraint which sets a limit on the rate of managerial growth that restricts the growth of the firm. The firms' managerial ability to cope with a great number of changes at once is limited. It is not possible to develop a larger management team for the development and marketing of new products. The higher rate of diversification requires higher expenditures on advertising and R &D. As a result, beyond a certain growth rate, the higher growth rate leads to a lower rate of profit.

The other aspect of the growth-profits relationship is the rate of growth of capital supply. The aim of the shareholders is to maximise the growth rate of capital stock. The main source of finance for its growth is profits. Thus profits determine growth on the supply side. A higher level of profits provides more funds directly for reinvestment. It also allows more funds to be raised on the capital markets. It, therefore, allows a higher rate of growth to be funded. This gives a direct and positive relationship between profits and growth. This is shown in Figure 4 as a straight line GS from the origin.

For the equilibrium of the firm, the growth-demand and growth-supply relationship must be satisfied. This is achieved when the two curves GD and GS intersect at a point where the growth-profits combination gives the optimum solution. Suppose in the figure the GS_2 curve intersects the GD curve at point M where profits are maximised. This point does not provide an optimum solution because the managers desire more growth than is consistent with long-run profit maximisation. The extent to which they can increase the growth rate beyond point M depends upon their desire for job security. Their job security is threatened if the shareholders feel that the share prices and dividends are falling and there is the threat of take-over by other firms. This will affect the growth rate of capital supply (GS). Thus it is the financial constraint which sets a limit to the growth of the firm on the supply side.

According to Marris, it is the retention ratio which determines the growth rate of capital supply. The retention ratio is the ratio of retained

profits to total profits. If the retention ratio is very low, it means that almost all profits have been distributed to the shareholders. As a result, there are limited funds available with the managers for the growth of the firm and the growth rate will be very low. The growth-supply curve will be very steep as shown by GS_1 curve. The firm's equilibrium will be at point L where the GS_1 curve intersects the GD curve. This is again not the optimal equilibrium point of the firm because here the growth rate is low and profits are below the maximum level.

Larger retained profits are required by managers to invest larger funds for the growth of the firm. These raise the retention ratio which, in turn, leads to higher profits and higher growth rates until point M of maximum profits is reached. This is again not the optimum equilibrium point of the firm because the managers feel that this combination of higher growth rate and higher profits is approved by the shareholders and there is no threat to their job security. They will, therefore, be encouraged to raise the retention ratio further, invest more funds, expand and increase the growth rate of the firm.

As a result, the growth-supply curve will become flatter and take the shape of GS_3 curve as in the figure where it intersects the DS curve at point E. At this point, distributed profits to shareholders fall. But they are adequate to satisfy the shareholders so that there is no fear of fall in the prices of shares and of the threat of take-overs. There is also job security for managers. Thus point E is the optimal equilibrium point of the firm. If the managers adopt a higher retention ratio than this, the distributed profits will fall further and the shareholders will not be satisfied which will endanger the job security of managers. The existing shareholders may decide to replace the managers. If the distribution of low profits to shareholders brings a fall in the market prices of shares, it may lead to take-over of the firm.

Criticisms:

Marris's growth-maximisation model has been severely criticised for its over-simplified assumptions by Koutsoyiannis and Hawkins.

1. Marris assumes a given price structure for the firms. He, therefore, does not explain how prices of products are determined in the market. This is a serious weakness of his model.
2. Another defect of this model is that it ignores the problem of oligopolistic interdependence of firms in non-collusive market.
3. This model also does not analyse interdependence created by non-price competition.
4. The model assumes that firms can grow continuously by creating new products. This is unrealistic because no firm can sell anything to the consumers. After all, consumers have their preferences for certain brands which also change when new products enter the market.
5. According to Koutsoyiannis, "Marris's model is applicable basically to those firms which produce consumers' goods. The model is not appropriate for analysing the behaviour of manufacturing businesses or traders."
6. Marris lumps together advertising and R&D expenses in his model. This is a serious shortcoming of the model because the effectiveness of these two variables is not the same in any given period.
7. Marris assumes that firms have their own R&D department on which they spend much for creating new products. But, in reality, most firms do not have such departments. For product diversification, they imitate the inventions of other firms and in case of patented inventions they pay royalties for using them.
8. The assumption that all major variables such as profits, sales and costs increase at the same rate is highly unrealistic.
9. It is also doubtful that a firm would continue to grow at a constant rate, as assumed by Marris. The firm might grow faster now and slowly later on.
10. It is difficult to arrive at the growth rate which maximises the market value of the firm's shares and the rate at which the take-over is likely to take place.

11. Despite these criticisms, Marris's theory is an important contribution to the theory of the firm in explaining how a firm maximises its growth rate.

Williamson's Model of Managerial Discretion

Williamson argues that managers have discretion in pursuing policies which maximise their own utility rather than attempting the maximisation of profits which maximises the utility of owner-shareholders. Profit acts as a constraint to this managerial behaviour, in that the financial market and the shareholders require a minimum profit to be paid out in the form of dividends, otherwise the job security of managers is endangered.

The managerial utility function includes such variables as salary, security, power, status, prestige, professional excellence. Of these variables only the first is measurable. The others are non-pecuniary and if they are to be operational they must be expressed in terms of other variables with which they are connected and which are measurable. This is attained by the concept of expense preference, which is defined as the satisfaction which managers derive from certain types of expenditures.

In particular, staff expenditures on emoluments (slack payments), and funds available for discretionary investment give to managers a positive satisfaction (utility), because these expenditures are a source of security and reflect the power, status, prestige and professional achievement of managers. Staff increases are to a certain extent equivalent to promotion, since they increase the range of activity and control of managers over resources. Being the head of a large staff is a symbol of power, status and prestige, as well as a measure of professional success, because progressive and increasing staffs implies successful expansion of the particular activity for which a manager is responsible within a firm.

Managers' prestige, power and status are to a large extent reflected in the amount of emoluments or slack they receive in the form of expense accounts, luxurious offices, company cars, etc. Emoluments are economic rents accruing to the managers; they have zero productivity in that, if removed, they would not cause the managers to leave the firm and seek employment elsewhere. They are discretionary expenditures which are made

possible because of the strategic position that managers have in the running of the business. Emoluments are probably less attractive than salary payments since there are certain restrictions in the way in which they may be spent. However, they may have tax advantages (since they are tax deductible) and furthermore they are less visible remunerations to the managers than salary, and hence are less likely to attract the attentions and cause dissatisfaction of the shareholders or the labour force of the firm.

Finally the status and power of managers is associated with the discretion they have in undertaking investments beyond those required for the normal operation of the firm. These minimum investment requirements are included in the minimum profit constraint together with the amount of profits required for a satisfactory dividend policy. Discretionary investment expenditure gives satisfaction to the managers because it allows them to materialise their personal favourite projects. This is an obvious measure of self-fulfilment for managers and top executives.

Staff expenditures, emoluments and discretionary investment expenses are measurable in money terms and will be used as proxy-variables to replace the non-operational concepts (power, status, prestige, professional excellence) appearing in the managerial utility function. Thus the utility function of the managers may be written in the form $U = f_1 (S, M, I_D)$ where S = staff expenditure, including managerial salaries M = managerial emoluments, I_D = discretionary investment

The Demand of the Firm:

It is assumed that the firm has a known downward-sloping demand curve, defined by the function, $X = f^*(P, S, \varepsilon)$, or $P = MX, S, \varepsilon)$, where X = output, P = price, S = staff expenditure, ε = the condition of the environment. It is assumed that the demand is negatively related to price, but positively related to staff expenditure and to the shift factor. An increase in staff expenditure is assumed to cause a shift in the demand curve upwards and thus allow the charging of a higher price. The same holds for any other change in the environment which shifts upwards the demand curve of the firm.

The Production Cost:

The total production cost (C) is assumed to be an increasing function of output

Actual profit II:

The actual profit is revenue from sales (R), less the production costs (C), and less the staff expenditure (S)

$\Pi = R - C - S$

Reported profit Π_R:

This is the profit reported to the tax authorities. It is the actual profit less the managerial emoluments (M) which are tax deductible

$\Pi_R = \Pi - M = R - C - S - M$

Minimum profit Π_0:

This is the amount of profits (after tax) which is required for an acceptable dividend policy by the shareholders. If shareholders do not receive some profit they will be inclined to sell their shares or to vote for a change in the top management. Both actions obviously reduce the job security of managers. Hence they will make sure to have a minimum profit Π_0 adequate to keep shareholders satisfied. For this the reported profits must be at least as high as the minimum profit requirement plus the tax that must be paid to the government. In summary, staff expenditure, managerial slack and discretionary investment spending will be larger for a firm that maximises utility than for a firm that maximises profits. However, no general conclusion can be drawn regarding the level of output in the two models.

We may compare further the two models by examining their predictions about the changes of the policy variables when some change in the environment takes place. We will examine the effects on the levels of the policy variables (X, S, ρ) of a shift in demand, a change in the profit tax rate, t, and of the imposition of a lump-tax T to the firm. We will examine only the directions of these changes.

A Shift in the Market Demand:

This may be denoted by a change in the shift factor ε, appearing in the demand function. A shift in demand will increase output X, and staff expenditure S, in both models. The two models give identical predictions of

the direction of changes in X and S, so that by observing only changes in these variables in response to a shift in demand we cannot tell whether the firm is a utility-maximiser or a profit-maximiser. The two models cannot be verified by simply looking at what happens to X and S as demand changes. However, an upward shift in demand, while not affecting ρ in a classical model, will lead to a reduction in ρ if the firm is utility-maximising. A shift in demand will increase slack payments faster than the increase in the actual profits. Thus an increase in slack payments in booms, and a decrease of slack in recessions, suggests that the firm is of the Williamson type rather than a profit maximiser.

An increase in the profit tax rate t:

The effects of this change on X, S and p are summarised below: An increase in the profit tax rate will not change the equilibrium X and S of a profit- maximising firm. A profit-maximising firm cannot avoid the burden of an increase in the profit tax rate by changing its output (or its price) or its staff expenditure, unless the burden is so high as to lead the firm to close down. A utility-maximising firm, on the other hand, will be able to avoid part of the tax burden by increasing its staff expenditure and its slack payments, and reporting a lower level of profit for taxation.

Effects of the imposition of a lump-tax T:

The effects on output, staff expenditure and on slack payments are summarised below: The imposition of a lump-tax will not change the short-run equilibrium X and S of a profit-maximising firm, which cannot avoid the burden of T. On the other hand, the imposition of T leads a utility maximiser to a reduction of his output, reduction of his staff expenditure and reduction of the slack payments.

Change in fixed costs:

Since a lump-sum tax is similar in its impact on the firm's activities to an increase in its fixed costs, we may infer from the above analysis that an increase in the fixed costs will not affect the short-run equilibrium X and S of a profit maximiser (unless he is driven out of business completely), while it will lead to a change in the level of output, the staff expenditure and the slack payments of a utility maximiser.

Empirical Evidence:

Williamson conducted several tests of his model. We will summarise the evidence he presented from several case studies.

Principal-firm analysis:

Williamson attempted to test the hypothesis that managerial discretion influences the expenditures for which managers have a strong expense preference. He fitted the above model to cross-section samples for the years 1953, 1957 and 1961. Each sample included the two largest firms from twenty-six industries. The samples were not random, but included the large firms for which the managerial model is thought to be more appropriate. Surprisingly enough the dependent variable is only a small fraction of staff expenditure, not any one or all of the expenses for which managers have a definite preference. Williamson argues that the remuneration of the top executive is determined within a carefully designed scale for the salaries of the rest of the managerial group. 'Payments between executive levels are carefully scaled so that the factors which influence compensation of the top executive can be presumed to affect the level of staff compensation generally.'

Even more questionable is the use of S, staff expenditure, as a determinant of the remuneration of the top executive. Williamson uses S as a proxy for the 'compensation which the top executive would receive strictly on a profit-maximising basis', on the grounds that staff expenditure reflects the size of personnel over which the top executive has responsibility. This may be so, but the logic for including the profit-maximising remuneration as a determinant of the actual compensation of the top executive does not seem clear to us, given the hypothesis being tested. If anything, staff expenditure, S, is probably the most important element in the discretionary expenses of managers, and, hence, should be the dependent rather than an explanatory variable.

The concentration ratio, C_i, and the height of the barriers to entry, H_i, are used as measures of opportunities for managerial discretion. Williamson argues that the higher the concentration (the fewer the firms in an industry) and the stronger the barriers to entry, the greater the power of managers for

discretionary spending. Williamson, anticipating the criticism that C_i, and H_i, are another measure of 'size', estimated the correlation coefficient between sales and C_i, and sales and height of barriers to entry.

He found the values of these correlations sufficiently low (-013 and — 014) for the firms included in his samples, and he thus concluded that C_i and H_i are not proxies for size. However, he found that the correlation between sales and S was considerable (0-75). Yet, he does not comment on the implications of this correlation for his estimated regression. Another criticism might be the simultaneous use of C_i and H_i as measures of the opportunities of discretion for managers. Why is neither of these measures adequate for capturing the effect of 'opportunity for managerial discretion'? Williamson does not answer this question; hence one might suspect that the combined use of C_i and H_i was chosen in order to improve the statistical fit, rather than on grounds of theoretical importance of these variables in explaining the dependent variable.

Williamson, anticipating the criticism that C_i and H_i are a measure of profitability, estimated several equations in which profits were included either in combination with C_i and H_i or replacing them. From his statistical findings he concluded that profits give a worse fit to his data. We should think that the evaluation of the regression findings on the basis of statistical results alone is not adequate for concluding that C_i and H_i are not in fact a proxy for profits. The variable B_i, proportionate representation of management on the board, is used by Williamson as a measure of the 'desire of managers to act free from outside interference', that is, free from the interference of owner-shareholders. The greater the number of managers on the board, the greater the 'desire' of management for discretionary action.

It seems to us that the distinction between the 'desire of managers' and the 'opportunity of managers' for discretionary behaviour cannot be disentangled and measured independently by the three variables C_i, H_i, and B_i. Obviously one can argue that the greater the representation of management on the board, the greater (not only their desire but also) their opportunity for discretionary action. Thus, we should think that C_i H_i and B_i are largely overlapping measures of the same factor the degree of

discretion that managers have in the allocation of resources in the firm. In summary, the causality implied by staff, S, plus the fact that S and sales are highly correlated, that H_i and C_i may reflect profitability, that C_i H_i and B_i may measure (at least partly) the same factor, cast serious doubt on the suitability of the fitted model for testing the hypothesis of managerial discretion.

Evidence from field studies:

Williamson conducted several case studies from which he infers that his model is better suited for the explanation of some real-world phenomena, such as:

1. Increase in S and M in booms and drastic cut of these expenditures in recessions.
2. Reaction of firms to taxation changes.
3. Changes of the level of X, S and M in response to changes in the fixed costs of the firm. O. Williamson's Model of Managerial Discretion
4. Drastic cuts in staff expenditure by newly appointed top management, without affecting the productivity of the firm.
5. Allocation of 'fixed overheads' of multiplant-multiactivity corporations to their different plants and activities so as to obtain the effects of a lump-tax.

Such phenomena, Williamson argues, while incompatible with a profit-maximising behaviour, can be explained by his model of 'rational managerial behaviour'. We think that the available evidence is not enough for the verification of the theory. The above arguments of Williamson rest on an implicit 'ceteris paribus' clause, which is not at all sure to be fulfilled in dynamic situations, such as shifts in demand and costs in booms and recessions. Furthermore, Williamson's model fails to deal with the core problem of oligopolistic interdependence and of strong oligopolistic rivalry. Williamson's model is applicable in markets where rivalry is not strong, or for firms who have some advantage over their rivals. However, in the long run such advantages which shelter a firm from competition are usually weakened, and competition is enhanced. When rivalry is strong a profit-maximising model may be more appropriate, unless some form of collusive agreement is achieved and firms adhere to it.

The **behavioral theory of the firm** first appeared in the 1963 book A Behavioral Theory of the Firm by Richard M. Cyert and James G. March. The work on the behavioral theory started in 1952 when March, a political scientist, joined Carnegie Mellon University, where Cyert was an economist. Before this model was formed, the existing theory of the firm had two main assumptions: profit maximization and perfect knowledge. Cyert and March questioned these two critical assumptions. A behavioural model of rational choice by Herbert A. Simon paved the way for the behavioural model.[4][5] Neo-classical economists assumed that firms enjoyed perfect information. In addition the firm maximized profits and did not suffer from internal resource allocation problems. Advocates of the behavioural approach also challenged the omission of the element of uncertainty from the conventional theory. The behavioural model, like the managerial models of Oliver E. Williamson and Robin Marris, considers a large corporate business firm in which the ownership is separate from the management.

Cyert and March

These researchers offered four major research themes:

- A small number of key economic decisions
- Development of a general theory, generalizing the results from studies of specific firms
- Linkage of empirical data to models
- Orientation towards process rather than outcomes

Theory construction

The behavioural approach takes the firm as the basic unit of analysis. It attempts to predict behaviour with respect to price, output and resource allocation decisions. It emphasizes the decision-making process.[8]

The firm as a coalition of groups

The theory argues that while small firms may operate under the guidance of the entrepreneur, such a simple model does not describe larger corporations. These larger firms are coalitions of individuals or groups, which may include managers, stockholders, workers, suppliers and so on. According to Cyert and March, these groups participate in setting goals and

making decisions. Priorities and information may vary by group, potentially creating conflicts. Cyert and March mentioned five goals which real world firms generally possess: production; inventory; market share; sales and profits. According to the behavioural theory, all the goals must be satisfied, following an implicit order of priority among them.

Satisfying behaviour

Cyert and March proposed that real firms aim at satisfying rather than maximizing their results. I.e., some groups may settle for "good enough" achievements rather than striving for the best possible outcome. This came from a concept known as bounded rationality, which was developed by Herbert Simon. Bounded rationality means prudent behaviour under a given set of circumstances. In this model goals are not set to maximize relevant magnitudes such as profits, sales and market share. Instead, goals are compromises negotiated by the groups.

The process of decision making

In the model, top management sets the goals of the organization. But these goals are implemented through decision making at two levels, one at the top and the second at lower management levels. During approval of proposals of various departments, two criteria are generally employed. A financial measure assesses the availability of the required funds given resources. An improvement measure assesses whether the proposal improves the health of the organization. According to Cyert and March, information is required to take the most appropriate decisions. However, information gathering itself is not Costless and requires resources.

Organizational slack

To keep the various groups in the organization, payments had to be in excess of what was required for the efficient working of the firm. The difference between the total resources and the necessary payments is called the organizational slack. In conventional economic theory organizational slack is zero, at least at equilibrium. Cyert and March claim that organizational slack plays a stabilizing and adaptive role. Cyert and March gave many examples of organizational slack such as high dividends paid to

shareholders, prices set lower than necessary and wages paid in excess of those required.

Critical evaluation

The behavioural model made a great impact on the theory of the firm. It gave insights in the process of goal formation and fixation of aspiration levels and resource allocation. Its critics claim that the theory is unnecessarily complicated. The virtual assembly of the firm, with the decision-making process as the unit, for the purpose of predicting their behaviour is highly questioned by critics. There has also been staunch support for profit maximization rather than satisfying behaviour, which is one of the core elements of the model. The behavioural theory of the firm has become important for much later research in organization theory and management, and has led to empirical studies and simulation modelling in organizational learning, as well as work on the cognitive foundations of firm strategy.

Game theory is a Mathematical subject that is commonly used in practical life. It is applied to various other non-mathematical fields too. Game theory explains how a strategic game is played. It determines the way or order in which the players should make moves. It considers the information for the players at each decision point. In-game theory, the interdependence of actions of players is the essence of the game. The game has two kinds of strategic interdependence – one is sequential, and the other is simultaneous. In sequential interdependence, players act in a sequence, aware of other players' actions. While, in simultaneous interdependence, players act at the same time, ignoring other players' actions. The game theory is all about such strategies. Let us go ahead and learn more about game theory.

Game Theory

The game theory is said to be the science of strategies which comes under the probability distribution. It determines logical as well as mathematical actions that should be taken by the players in order to obtain the best possible outcomes for themselves in the games. The games studied in game theory may range from chess to tennis and from child-rearing to

takeovers. But there is one thing common that such an array of games is interdependent, i.e. outcome for each player depends upon the strategies of all. In other words, game theory deals with mathematical models of cooperation and conflicts between rational decision-makers. Game theory can be defined as the study of decision-making in which the players must make strategies affecting the interests of other players.

Zero-Sum Game Theory

There is a special kind of game studied in game theory, called zero-sum games. They are constant-sum games. In such games, the available resources can neither be increased nor decreased. Also, the total benefit in zero-sum games for all combination of strategies, always adds to zero. We can say that in zero-sum games, one wins and exactly one opponent loses. The sum of benefits of all the players for any outcome is equal to zero is called a zero-sum game. Thus, the interest of the two players is opposed. Several games, game theory are non-zero-sum games, since net result of outcome is less than or greater than zero. So, when one player's gain does not correspond to other's loss, it is called a non-zero sum game.

Game Theory Applications

The game theory is widely applied to study human as well as animal behaviours. It is utilized in economics to understand the economic behaviours, such as behaviours of consumers, markets and firms. Game theory has been commonly used in social sciences as well. It is applied in the study of sociological, political and psychological behaviours. The use of analysis based on game theory is seen in biology too. In addition to behavioural prediction, game theory utilized in the development of theories of normative or ethical behaviour.

Game Theory Example

The best example of game theory is a classical hypothesis called "Prisoners Dilemma". According to this situation, two people are supposed to be arrested for stealing a car. They have to serve 2-year imprisonment for this. But, the police also suspect that these two people have also committed a bank robbery. The police placed each prisoner in a separate cell. Both of

them are told that they are suspects of being bank robbers. They are inquired separately and are not able to communicate with each other.

The prisoners are given two situations:

- If they both confess to being bank robbers, then each will serve 3-year imprisonment for both car theft and robbery.
- If only one of them confesses to being a bank robber and the other does not, then the person who confesses will serve 1-year and others will serve 10-year imprisonment.

According to game theory, the prisoners will either confess or deny the bank robbery. So, there are four possible outcomes:

	2-Confess	**2-Deny**
1-Confess	Both punished 3 years	Prisoner 1 punished 1 year Prisoner 2 punished 10 years
1-Deny	Prisoner 1 punished 10 year Prisoner 2 punished 1 year	Both punished 2 years

Here, the best option is to deny. In this case, both will have to serve 2 years sentence. But it cannot be guaranteed that others would not confess, therefore most likely both of them would confess and serve the 3-year sentence.

Unit-II

DISTRIBUTION

An Introduction to Neoclassical Theory

Neoclassical theory was introduced based on classical theory. It has added, modified, and in some way extended classical theory. The basic assumption of this theory is that the physiological and social aspects of a worker as an individual and his workgroup ought to be focused on. In classical theory, the organization emphasized order, structure, economic factors, formal organization, and objective rationality. Whereas neo-classical theory emphasized social factors and emotions at work. Human relation is a general term that is frequently used to describe how a manager interacts with their employees. The importance of human relations is included in two aspects: The organization situation should be observed in social terms as well as economical and technical terms, and in terms of clinical method, it is similar to the doctor's diagnosis of the human organisms.

Neo Classical Theory

The Neoclassical approach was developed many years ago because it was believed that the classical theory did not achieve complete production efficiency and workplace harmony. Managers still observe the frustrations and difficulties because people always do not prefer to follow estimated or rational patterns of behaviour. Hence, there was a greater interest in assisting managers to deal more effectively with the "people side" of the organization. The neo-classical approach reflects a slight modification over the classical approach. The neoclassical approach identifies the importance of physiological and social aspects of workers as an individual and their relationships within and among the group of the organization. The Neoclassical Theory gained importance specifically in the rise of the "Hawthorne Experiment" at Western Electric Company by the father of human relation management named Elton Mayo from 1924 to 1932.

The Inception of Neo-Classical Theory

The neoclassical theory is the extensive version of the classical theory that includes behavioural science in business management. In this theory, the organization is the social system, and its performance is affected by

human efforts. The classical theory emphasized the mechanical and physiological variables and deemed them the primary factor in inferring its efficiency. However, when the efficiency was checked, it was found that the positive aspect of these factors did not evoke a positive response in work behaviour.

Hence, the researchers tried to specify the reasons for human behaviour at a job. This eventually led to the arrangement of the neoclassical theory of economists. This mainly emphasizes human behaviour in the organization. This approach is often referred to as human relations or behavioural theory of organization. The neoclassical theory states that an organization is a mix of both informal and formal aspects of the organization. This aspect was ignored in classical theory. The organization's informal structure is majorly formed due to the social interaction with workers; this affects and gets affected by an organization's formal structure. Generally, disputes between the organization and workers often exist but this needs to be resolved immediately as the problems persist.

Features of Neo-Classical Theory

- Business Organizations are identified as a social system.
- Human factors are regarded as the most important elements in the organization.
- The theory revealed the importance of social and psychological factors in determining the worker's productivity and satisfaction.
- The management aims to develop social and leadership skills along with technical skills. It must be done for the welfare of the workers and the organization.
- Morale and productivity work together in an organization.

Organization Structure in Neo-Classical Theory

The neoclassical writers offered the following organizational structure:

Flat Structure: In the case of flat structure, the wide span of control in an organization helps in motivating the employees more effectively, a shorter chain of communication and it is free from hierarchical control.

Decentralization: Neo-classical theory adopted a decentralized organizational structure which is close to the flat structure due to the wider span of control. It permits autonomy and initiative at lower levels. It also encouraged people to occupy higher positions in the organization.

Informal Organizations: The neoclassical theorists emphasized the need for both formal and informal organizations. The formal organization indicates the motive of top management for interactions among the people. Informal organization is significant to promote the inadequacy of formal organization and to satisfy the social and psychological needs of people. The management uses informal organization for overcoming resistance to change on the part of workers and also for a fast communication process. Both formal and informal organizations are interdependent upon each other.

The Criticism of Neoclassical Theory

The assumption on which neoclassical theory is formulated is sometimes not true. Thinking that there is always the possibility to find a solution that is acceptable to all is not always true. There are several conflicting interests among distinct groups that are merely structural and not physiological. This aspect is not covered in neo-classical theory. No specific organizational structure is suitable for all organizations. Various organizational formats introduced by the neo-classicists are not acceptable in all situations. Neoclassical theory is only a modification of classical organization theory. It also suffered from similar drawbacks from which classical organization theory suffered. It lacked a unified approach to organization. This theory has been criticized by the fact it is nothing more than “a common place of descriptive and empirical information as it has mainly relied on the Hart word Experiment”.

Fun Facts

- Neoclassical theory is based on the assumption that makes its structure irrelevant for different organisation situations.
- This theory is a modified version of the classical theory that includes behavioural sciences in business management.

Conclusion

The neoclassical theory tried to overcome the drawbacks of classical organization theory. It introduced the concept of informal organization and the human behaviour approach in the study of organizational functioning. However, the neoclassical theory is also not free from several drawbacks. Like the classical theory of organization, the neo-classical theory is also suffered from incompetency, a short-sighted approach, and lack of integration among many facts of human behaviour studied by it.

Marginal Productivity Theory of Distribution

The oldest and most significant theory of factor pricing is the marginal productivity theory. It is also known as Micro Theory of Factor Pricing. It was propounded by the German economist T.H. Von Thunen. But later on many economists like Karl Mcnger, Walras, Wickstcad, Edgeworth and Clark etc. contributed for the development of this theory. According to this theory, remuneration of cache factor of production tends to be equal to its marginal productivity. Marginal productivity is the addition that the use of one extra unit of the factor makes to the total production. So long as the marginal cost of a factor is less than the marginal productivity, the entrepreneur will go on employing more and more units of the factors. He will stop giving further employment as soon as the marginal productivity of the factor is equal to the marginal cost of the factors.

Definitions:

"The distribution of income of society is controlled by a natural law, if it worked without friction, would give to every agent of production the amount of wealth which that agent creates." -J.B. Clark. "The marginal productivity theory contends that in equilibrium each productive agent will be rewarded in accordance with its marginal productivity." -Mark Blaug. "The marginal productivity theory of income distribution states that in the long run under perfect competition, factors of production would tend to receive a real rate of return which was exactly equal to their marginal productivity." -Liebhafasky

Assumptions of the Theory:

1. Perfect Competition:

The marginal productivity theory rests upon the fundamental assumption of perfect competition. This is because it cannot take into account unequal bargaining power between the buyers and the sellers.

2. Homogeneous Factors:

This theory assumes that units of a factor of production are homogeneous. This implies that different units of factor of production have the same efficiency. Thus, the productivity of all workers offering the particular type of labour is the same.

3. Rational Behaviour:

The theory assumes that every producer desires to reap maximum profits. This is because the organizer is a rational person and he so combines the different factors of production in such a way that marginal productivity from a unit of money is the same in the case of every factor of production.

4. Perfect Substitutability:

The theory is also based upon the assumption of perfect substitution not only between the different units of the same factor but also between the different units of various factors of production.

5. Perfect Mobility:

The theory assumes that both labour and capital are perfectly mobile between industries and localities. In the absence of this assumption the factor rewards could never tend to be equal as between different regions or employments.

6. Interchange ability:

It implies that all units of a factor are equally efficient and interchangeable. This is because different units of a factor of production are homogeneous, since they are of the same efficiency, they can be employed inter-changeable, and e.g., whether we employ the fourth man or the fifth man, his productivity shall be the same.

7. Perfect Adaptability:

The theory takes for granted that various factors of production are perfectly adaptable as between different occupations.

8. Knowledge about Marginal Productivity:

Both producers and owners of factors of production have means of knowing the value of factor's marginal product.

9. Full Employment:

It is assumed that various factors of production are fully employed with the exception of those who seek a wage above the value of their marginal product.

10. Law of Variable Proportions:

The law of variable proportions is applicable in the economy.

11. The Amount of Factors of Production should be Capable of being varied:

It is assumed that the quantity of factors of production can be varied i.e. their units can either be increased or decreased. Then the remuneration of a factor becomes equal to its marginal productivity.

12. The Law of Diminishing Marginal Returns:

It means that as units of a factor of production are increased the marginal productivity goes on diminishing.

13. Long-Run Analysis:

`Marginal productivity theory of distribution seeks to explain determination of a factor's remuneration only in the long period.

Explanation of the Theory:

The marginal productivity theory states that under perfect competition, price of each factor of production will be equal to its marginal productivity. The price of the factor is determined by the industry. The firm will employ that number of a given factor at which price is equal to its marginal productivity. Thus, for industry, it is a theory of factor pricing while for a firm it is a factor demand theory.

Analysis of Marginal Productivity Theory from the Point of View of an Industry:

Under the conditions of perfect competition, price of each factor of production is determined by the equality of demand and supply. As the theory assumes that there exists full employment in the economy, therefore, supply of the factor is assumed to be constant. So, factor price is determined by its demand which itself is determined by the marginal productivity. Thus, under such conditions, it becomes essential to throw light on the demand curve or marginal productivity curve of an industry. As the industry consists of a group of many firms, accordingly, its demand curve can be drawn with the demand curves of all the firms in the industry. Moreover, marginal revenue productivity of a factor constitutes its demand curve. It is only due to this reason that a firm's demand or labour depends on its marginal revenue productivity. A firm will employ that number of labourers at which their marginal revenue productivity is equal to the prevailing wage rate.

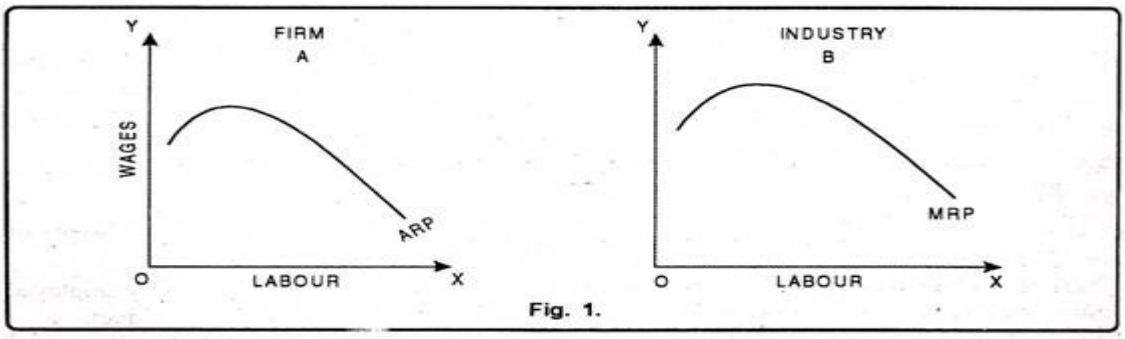


Fig. 1.

Fig. 2 shows that at wage rate OP_1, the demand for labour is ON_1 and marginal revenue productivity curve is MRP_1. If wage rate falls to OP, firms will increase production by demanding more labour. In such a situation the price of the commodity will fall and marginal revenue productivity curve will also shift to MRP_2. At OP wages, the demand for labour will increase to ON. DD_1 is the firm's demand curve for labour. The summation of demand of all the firms shows demand curve of an industry. Since the number of firms is not constant under perfectly competitive market, it is not possible to estimate the summation of demand curves of all firms. However, one thing is certain that is the demand curve of industry also slopes downward from left to right. The point where demand for and supply of a factor are equal will determine the factor price for the industry. This theory assumes the supply of a factor to be fixed.

Thus factor price is determined by the demand for factor i.e. factor price will be equal to the marginal revenue productivity. Labour has been taken on OX axis whereas wages and MRP have been taken on OY axis. DD_1 is the industry's demand curve for labour. This is also the Marginal Revenue Productivity curve. Factor Price (OW) = Marginal Revenue Productivity MRP. Thus under perfect competition, factor price is determined by the industry and firm demands units of a factor at this price.

Analysis of Marginal Productivity Theory from the Point of View of Firm:

Under perfect competition, number of firms is very large. No single firm can influence the market price of a factor of production. Every firm acts as a price taker and not a price maker. Therefore, it has to accept the prevailing price. No employer would like to pay more than what others are paying. In other words, a firm will employ that number of a factor at which

its price is equal to the value of marginal productivity. Therefore, from the point of view of a firm, the theory indicates how many units of a factor it should demand.

It is due to this reason that it is also called Theory of Factor Demand. Other things remaining the same, as more and more labourers are employed by a firm, its marginal physical productivity goes or- diminishing. As price under perfect competition remains constant, so when marginal physical productivity of labour goes on diminishing, marginal revenue productivity will also go on diminishing. Therefore, in order to get the equilibrium position, a firm will employ labourers up to a point where their respective marginal revenue productivity is equal to their wage rate.

Table 2. Factor Demand by the Firm

Labourer	MPP	Price of Product	MRP (MPP × MR)	Wage Rate
1.	20	5	20 × 5 = 100	55
2.	17	5	17 × 5 = 85	55
3.	14	5	14 × 5 = 70	55
4.	11	5	11 × 5 = 55	55
5.	8	5	8 × 5 = 40	55

Table 2 indicates that wage rate of labour is Rs. 55 per labourers. Price of the product produced by the labourer is Rs. 5 per unit. Now, when a firm employs one labourer, his marginal physical productivity is 20 units. By multiplying the MPP with price of the product we get marginal revenue productivity. Here, it is Rs. 100 for the first labour. The marginal revenue productivity of second labourer is Rs. 85 and of third labourer it is Rs. 70. The marginal revenue productivity of fourth labourer is Rs. 55 which is equal to wage rate. The firm will earn maximum profits if it employs up to the fourth labourer. If the firm employs fifth labourer, it will have to suffer losses of Rs. 15. Therefore, to get maximum profits, a firm will employ a factor upto a point where MRP is equal to price.

Under perfect competition, in long period in the equilibrium position, not only the marginal wages of a firm are equal to marginal revenue productivity, even the average wages of the firm are equal to average net revenue productivity as has been shown in Fig. 5. The fig. 5 shows that at point 'E' marginal wages of labour are equal to marginal revenue

productivity and the firm employs OM number of workers. At this point, even the average net revenue productivity is equal to average wages. Thus firm earns only normal profit. If wage line shifts from NN to N[N] then the demand for labour increases from OM to OM_1.

Determination of Factor Pricing under Imperfect Competition:

Marginal productivity theory applies to the condition of perfect competition. But in real life we face imperfect competition. Therefore, economists like Robinson, Chamberlin have analyzed factor pricing under imperfect competition. There are various firms under imperfect competition. But here we shall analyze only Monopsony. Under monopsony, there is perfect competition in product market. Consequently MRP is equal to VMP. There is imperfect competition in factor market. It indicates that there is only one buyer of the factors. Therefore, monopsony refers to a situation of market where only a single firm provides employment to the factors. If the firm demands more factors, factor price will go up and vice-versa. However, the determination of factor price under monopsony can be explained with the help of Fig. 6.

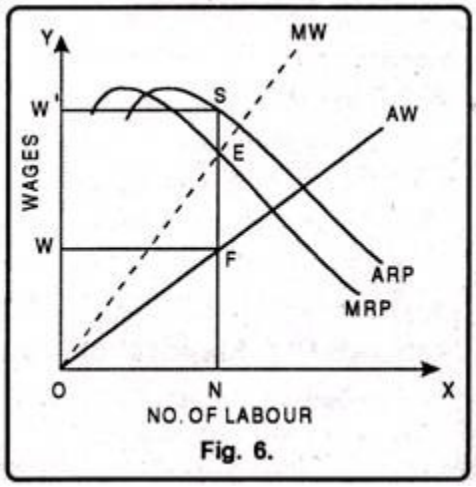


Fig. 6.

In Fig. 6 number of labourers has been shown on X-axis and wages on Y-axis. MW is marginal wage curve and ARP is the average wage curve. MRP is the marginal revenue productivity curve and AW is the average revenue productivity curve. In the fig. 6 a monopsony will employ that number of labourers at which their marginal wage is equal to MRP. In the fig. 6 firm is in equilibrium at point E. Here, firm will employ ON labourers and they will be paid wages equal to NF. In this way, ON labourers will get less wages

than their MRP i.e. EN. Monopsony firm will have EF profit per labourer which arises due to exploitation of labourers. Total profit SFWW' is due to exploitation of labour.

Euler's Product Exhaustion Theorem

Euler's theorem can be solved as under. Let C and L be the quantities of two factors of production, capital and labour respectively and P the total product of these factors. Then P = f(C, L).

In other words, if P is a linear homogeneous function (f) of C and L, the following equation will hold:

P = (∂f/∂C) C + (∂f/∂L) L ..... (1)

If the quantities of all inputs C and L are increased k-fold, the output P will also increase k-fold. Then the production function becomes kP=f(kC, kL)

By taking the total derivate of kP with respect to k, we have

(dk/dk)P = ∂f/∂kC. dkC/dk + ∂f/dkL. dkL/dk

Or P= (∂f/∂kC)C + (∂f/dkL)L [By eliminating dk/dk]

P=(∂f/∂C)C + (∂f/dL)L [k=1], where ∂f/∂C is the marginal product of capital and ∂f/∂L is the marginal product of labour. And ∂f/∂C. C is the share of capital in the product P, and ∂f/∂L. L is the share of labour in the total product. The above equation states that the marginal product of capital (∂f/∂C) multiplied by units of capital employed (C) plus the marginal product of labour (∂f/∂L) multiplied by the number of labourers (L) exactly equals the total product, P. Thus total factor payments exhaust the total value of the product.

Assumptions:

First, it assumes a linear homogeneous production function of first degree which implies constant returns to scale, Second, it assumes that the factors are complementary, i.e., if a variable factor increases, it increases the marginal productivity of the fixed factor, Third, it assumes that factors of production are perfectly divisible, Fourth, the relative shares of the factors are constant and independent of the level of the product, Fifth, there is a stationary, riskless economy where there are no profits, Sixth, there is perfect competition, Last, it is applicable only in the long-run.

Explanation:

Given these assumptions, Wick-steed proved with the help of Euler's theorem that when each factor was paid according to its marginal product, the total product would be exactly exhausted. This is based on the assumption of a linear homogeneous function. Wick-steed did not differentiate between the laws of increasing, constant and diminishing returns. He held that under perfect competition and constant returns to scale, the product exhaustion theorem was universally valid. Wick-steed's solution was treated by Edge worth with mockery and Pareto objected to the assumption of constant returns to scale. Wicksell, Walras and Barone also criticised him. They pointed out that the production function does not yield a horizontal long-run average cost curve (LRAC) but a U-shaped LRAC curve. The U- shaped LRAC curve first shows decreasing returns to scale, then constant and in the end increasing returns to scale, "Where Wick-steed went wrong," writes Hicks, **"was his assumption that he could argue from the shape of the curve at one particular point to the general shape of the curve."** Wick-sell proved that the product exhaustion problem held under perfectly competitive conditions in the long-run when profits were zero. He regarded it as a condition of equilibrium at the minimum point of firm's long-run average cost curve (LRAC) where the linear homogeneous production function was satisfied.

Suppose an entrepreneur is left with more than the marginal product of the resource he owns after paying all other resources their marginal products. Then all owners of resources are induced to become hiring agents and in the process the difference between the total product and the rewards to factors is eliminated. Conversely, if the residual left with the entrepreneur is less than his marginal product, after paying the other resources their marginal products, he will cease to be a producer and lend his services for its marginal product. Thus a firm under competitive conditions will produce at a level where the total product is exactly distributed according to the marginal product of the factor.

This solution of the product exhaustion theorem is based on a profitless long-run, perfectly competitive equilibrium position of a firm which

operates at the minimum point, E of its LRAC curve, as shown in panel (A) of Figure 2. At this point the firm is in full equilibrium, the marginal revenue productivity (MRP) of the factors being equal to the combined marginal cost of the factors (MFC). This is shown in panel (B) of Figure 2 where MRP = MFC at point A. It is at point A that the total product OQ is exactly distributed to OM factors and nothing is left over.

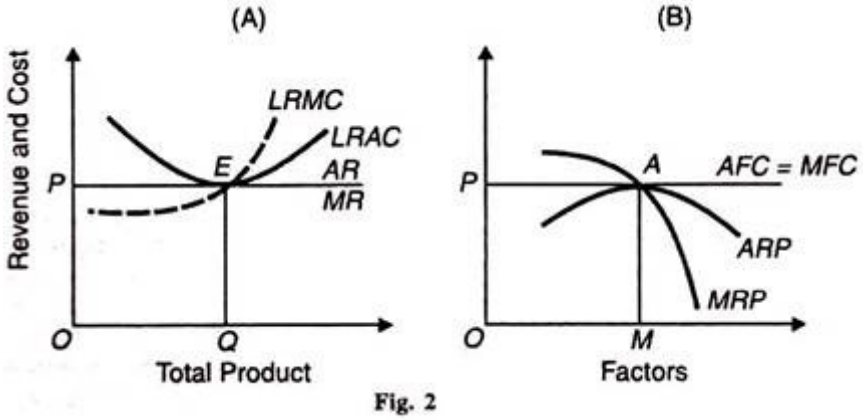


Fig. 2

As studied above, the product exhaustion problem is solved with a linear homogeneous production function: P = (∂P/∂C) +C (∂P/∂L) L. If, however, there are diminishing returns to scale, less than the total product will be paid to the factors: P> (∂P/∂C) +C (∂P/∂L)L. In such a situation, there will be super-normal profits in the industry. They will attract new firms into the industry. As a result, output will increase, price will fall and profits will be eliminated in the long-run. In this way, the distributive shares of the factors as determined by their marginal productivities will completely exhaust the total product.

Criticism:

In reality, constant returns to scale are incompatible with competitive equilibrium. For if long-run cost curve of the firm is horizontal and coincides with the price line the size of the firm is indeterminate; if it is below the price line the firm will become a monopoly concern; and if it is above the price line, the firm will cease to exist. While in the case of increasing returns to scale more than the total product will be distributed, because doubling the factors wills more than double the total product. But increasing returns are incompatible with perfect competition, since the economies of

production lead to the lowering of the cost of production and in the long-run there is a tendency towards the establishment of a monopoly

The whole analysis is based on the assumption that factors are fully divisible. Since the entrepreneur cannot be varied, we have not taken him as a separate factor. In fact, entrepreneurship disappears in the stationary economy. When there is full equilibrium at the minimum point of the LH4C curve, there is no uncertainty and profits disappear altogether. So the assumption of an entrepreneur less economy is justified for the solution of the adding-up problem. But once uncertainty appears, the entrepreneur becomes a residual claimant and the exhaustion of the production problem disappears,

Under imperfect or monopolistic competition the total product adds up to more than the share paid to each factor, that is, P is greater than C and L. Taking an imperfect labour market, the average and marginal wage curve (AW and MW) slope upward and the average and marginal revenue product curves (ARP and MRP) are inverted U-shaped, as shown in Figure 3.

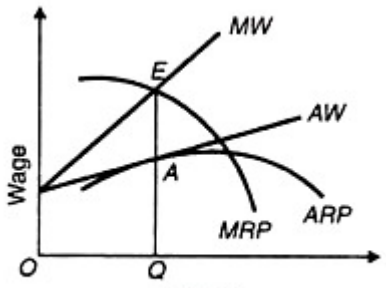


Fig. 3

Equilibrium is established at point E where the MRP curve cuts the MW curve from above. The firm employs OQ units of labour by paying QA wage which is less than the marginal revenue product of labour QE. Thus workers are paid less than their marginal productivity when there is imperfect competition. This argument applies not only to labour but to all shares even under constant returns to scale in the industry.

The product exhaustion theorem, however, holds true under monopolistic competition when the firm is in equilibrium. At equilibrium, the marginal cost curve cuts the marginal revenue curve and the average revenue curve is tangent to the average cost curve. It follows that the total

outlay for factors and the total revenue product will be equal. If now a small change in factors is made, keeping their prices constant, the increase in the total revenue product is approximately proportional to the increase in the outlay for factors. Thus if each factor included in the cost curve is paid according to its marginal revenue product at equilibrium, the total product of the firm will be exactly exhausted among them. But if there is monopoly, payment in accordance with marginal product will not exhaust the total product.

Modern Theory of Rent

Modern theory of rent does not confine itself to the reward of only land as a factor of production. Rent in modern sense can arise in respect of any factor of production, and not merely land. Rent is a surplus. In the sense of surplus, rent is a payment in excess of transfer earnings. Transfer earnings mean the amount of money which any particular unit could earn in its next best alternative use. Suppose a piece of land under cotton is yielding Rs. 150 and its next best use wheat fetches Rs. 100. The transfer earnings are Rs. 100 and, therefore, in its present use it is giving a surplus of Rs. 50.

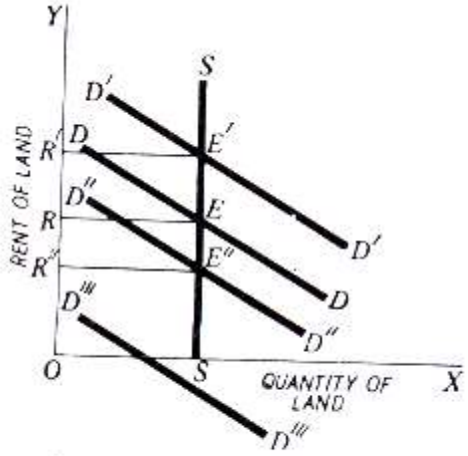
We can also define transfer earnings as the minimum sum which must be paid for a unit of a factor of production in order to induce it to stay in its present use or employment. In the above example, a sum of Rs. 100 at least must be paid for the land under cotton in order to retain it under cotton; otherwise it will shift to wheat, which is its next best alternative use where it can fetch Rs. 100. Actually, this piece of land is earning Rs. 150, i.e., Rs. 50 extra or in excess of its transfer earnings. This is economic rent. Economic rent in this sense is thus the difference between the present earnings and the transfer earnings. This concept of rent is applicable not merely to land but also to all factors of production i.e. labour, capital and entrepreneur's earnings too. They can all earn economic rent in the sense that the modern economists use the term 'rent'.

How Rent arises:

Rent in the sense of surplus arises when the supply of land, or for that matter that any other factor service, is less than perfectly elastic.

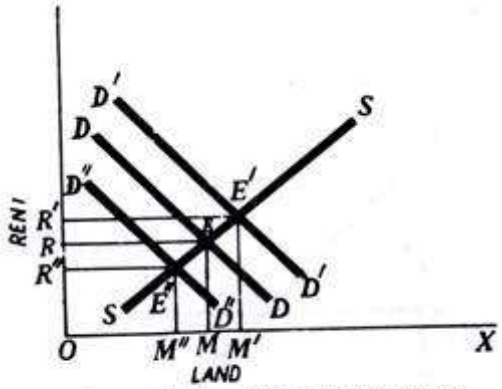
From the point of elasticity of supply, there are three possibilities:

(a) The supply may be perfectly elastic, which can be shown as a horizontal straight line, as in Fig. 33.5 above.



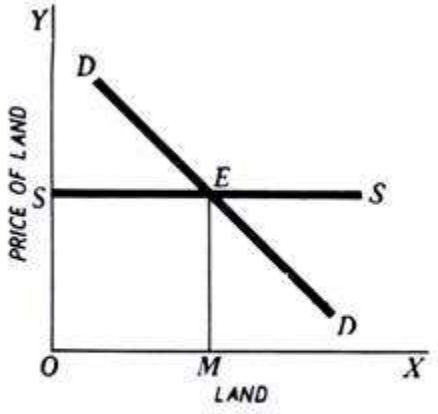
Determination of Rent Under Inelastic Supply of Land

Fig. 33.3



Determination of Rent Under Elastic Supply of Land

Fig. 33.4



Determination of Rent Under Perfectly Elastic Supply of Land

Fig. 33.5

(b) The supply of land may be absolutely inelastic. This is shown in Fig. 33.3 by a vertical straight line.

(c) There is the situation in between these two extremes, i.e., it is elastic, but not perfectly elastic. This is shown in Fig. 33.4.

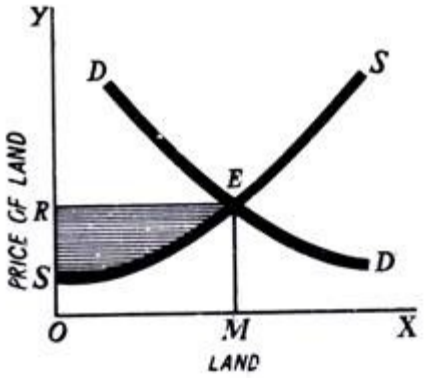
In these three conditions, rent as a surplus over transfer earnings will be different. If the supply is absolutely inelastic (see Fig. 33.3), the transfer earning is zero, because land cannot be transferred to any use; the supply of land is fixed, and it has only one use, whether it is used or not. In this case, the entire income from land is surplus, and hence rent. When the supply of land is perfectly elastic, there will be no surplus and the actual earnings and transfer earnings will be equal. For example, for an individual firm or farmer, the supply of land is perfectly elastic. Suppose the supply is elastic but not perfectly elastic, then a part of income from land is rent (in the sense of surplus over transfer earnings), and a part is not rent.

These three conditions are represented in the diagrams as mentioned below:

In Fig. 33.3, DD is the demand curve and SS a vertical straight line fixed supply curve. They intersect at E. Here OS is the quantity of land used. OR (=SE) is the rent per unit and total earnings are OSER. Since land is fixed in supply and cannot be transferred to any other use, its transfer earnings are zero. Hence its entire earnings OSER are rent as surplus over transfer earnings. For the economy as a whole, land has no alternative use at all. Hence the transfer earnings of land, from the point of view of economy as a whole, are zero and all the earnings are rent.

In Fig. 33.5, the supply curve SS of land is a horizontal straight line which is perfectly elastic. DD is the demand curve. The two intersect at E. In this case, OM land is put to use. The rent per unit is OS (=EM) and the total earning is OMES. The transfer earning is also OMES. If this firm does not pay OS rent, the land will be transferred to some other use or firm. Since transfer earnings and actual earnings are equal, there is no surplus or rent. In Fig. 33.6, SS supply curve is somewhat elastic. It cuts DD demand curve at E. In this case, OM land is used and the rent per unit is OR (= ME). The total earnings are OMER and the transfer earnings are OMES. If we deduct

transfer earnings OMES from the actual earnings OMER, we get RES (shaded area). This is surplus or rent.



Rent As Surplus Over Transfer Earnings
Fig. 33.6

Surplus in Other Factors Too:

It should be borne in mind that the above analysis with regard to rent as surplus over transfer earnings is applicable not only to the share of land, but also to the shares of other factors, viz., wages, interest and profits. This applies to all cases where the supply of a factor is less than perfectly elastic. In such cases, a part of present earnings is the transfer earnings and the remainder is economic rent.

Comparison between the Ricardian Theory and the Modern Theory of Rent:

Now that we have studied the two main theories of rent, viz., the Ricardian theory and the modern theory of rent, we should be in a position to distinguish between the two. We can see that both theories regard rent as a surplus. In Ricardo's theory, the surplus is due to superiority of the land in question over the marginal one. The superiority may be due to either quality of the land or better situation. Also both theories of rent have the same concept of land, i.e. a natural factor rather than a man-made factor like capital, but then where is the difference between the two theories. The difference between two is basic and it lies in this that while Ricardo takes agricultural land (the cultivation of which is subject to the law of diminishing returns sooner or later), the modern economists, on the other hand, do not confine the concept of rent to agricultural land only.

As we have said earlier, rent can arise in the sense of surplus in the case of other factors of production also and even in a situation of increasing returns. Rent represents the opportunity cost or transfer earnings. In this sense, rent is of a more general nature applicable to all factors. That is why it is said. “It (land rent) is leading specie of large genus”. That is, land rent is not a separate class by itself. It is only a prominent example of its type.

Modern Theory of Wages: Demand & Supply of Labour

According to the modern theory of wages, wages are the price of services rendered by a labour to the employer. As products the prices are determined with the help of demand and supply curve. Similarly, the wages (prices of services rendered by labour) is also obtained with the help of demand and supply of labour. Therefore, for the determination of wage level, it is necessary to study the demand for labour, supply of labour, and the interaction between them.

Demand for Labour:

The demand for labour is dependent on various factors.

Some of these factors are as follows:

i. Demand for a product:

The demand for labour is derived from the demand of the product it produces. In case, the demand for the product increases, the demand for labour would also increase However, this is the expected demand of the product and not the current demand. Therefore, the expected demand of the product determines the demand for labour. Moreover, along with the magnitude of demand, the elasticity of demand for labour is also need to be determined. The elasticity of output helps in determining the elasticity of labour.

The following are the conditions for determining the elasticity of demand of labour:

a. Condition

Labour would be inelastic if their wages contribute only a small amount to the total wages of industry

b. Condition

Labour would be elastic if the product produced by him is elastic

c. Condition

Labour would be elastic if cheaper substitutes of products are available

Elasticity of demand of labour depends on two factors, which are technical aspects of production and elasticity of demand for the product. The long-term demand for labour is more elastic than the short-term demand of labour.

ii. Other factors of production:

The price and amount of other factors of production employed affects the demand for labour. For example, if other factors of production are expensive then the demand for labour would be high. However, if other factors are available at cheaper quantity, then the demand for labour would reduce. Similarly, an increase in the demand of technology would reduce the demand for labour.

iii. Marginal productivity:

Refers to one of the most important factor that helps in the determination of demand for labour. An employer hires labour to increase his/her profit. For this, the employer needs to provide wages to avail the services of labour' He/she would employ labour until the increase in number of labour would increase the net output but at the diminishing rate. The employer would not hire any more labour when the output produced by an additional labour is equal to the additional cost incurred to hire that labour. Therefore, the wages paid to the labour is equal to the additional output/marginal output produced by that labour.

However, labour is considered as the homogeneous commodity; therefore, the amount of wage paid to one additional labour is similar to the amount of wage paid to the rest of the labour. The demand schedule of labour shows that the decrease in wage would increase the demand for labour. It is similar to the demand schedule of a product. Increase in number of labour would increase the output of product that would result in lowering down the product's price. This results in the decrease of marginal productivity of the industry. The change produced in the demand of labour can be determined with the help of change produced in wage rate of labour. However, the degree of this change is obtained with the help of elasticity of demand of labour. If smaller change in the wage rate of a labour produces a larger change in the demand of labour, then the demand of labour is elastic and vice-versa.

Supply of Labour:

Supply of labour refers to the number of hours spent by labour in the factor market. In an economy, there are several factors that influence the supply of labour. Some of the factors are wage rate, population size, age structure, availability of education and training employment opportunities for women, and social security programs. On the other hand, in an industry, the supply of labour is less elastic in the short-run. In this case, the supply of labour is dependent on the accessibility of workers in the nearby areas and their willingness for overtime work. However, the supply of labour becomes more elastic in the long-run. Industries attract labour by providing higher wages, training facilities, and good working conditions. Therefore, the supply curve of labour for an industry is upward sloping.

The Modern Theory of Interest

In this article we will discuss about the modern theory of interest with its criticisms.

An adequate theory to be determinate must take into consideration both the real and monetary factors that influence the interest rate. Hicks has utilized the Keynesian tools in a method of presentation which shows that productivity, thrift, liquidity preference and money supply are all

necessary elements in a comprehensive and determinate interest theory. According to Hansen, "An equilibrium condition is reached when the desired volume of cash balances equals the quantity of money, when the marginal efficiency of capital is equal to the rate of interest and finally, when the volume of investment is equal to the normal or desired volume of saving. And these factors are inter-related." Thus in the modern theory of interest rate, saving, investment, liquidity preference and the quantity of money are integrated at various levels of income for a synthesis of the loanable funds theory with the liquidity preference theory. The four variables of the two formulations have been combined to construct two new curves, the IS curve representing flow variable of the loanable funds formulation (or the real factors of the classical theory) and the LM curve representing the stock variables of liquidity preference formulation. The equilibrium between IS and LM curves provides a determinate solution.

The IS Curve:

The IS curve has been derived from the loanable funds formulation. It is a curve which explains the relationship between a family of saving schedules and investment schedules. In other words, this curve shows the equality of saving and investment at various combinations of the levels of income and the rates of interest. In Figure 8 (A), the saving curve S in relation to income is drawn in a fixed position, since the influence of interest on saving is assumed to be negligible. The saving curve shows that saving increases as income increases, viz., saving is an increasing function of income. Investment, on the other hand, depends on the rate of interest and the level of income. Given a level of interest rates, the level of investment rises with the level of income. At a 5 per cent rate of interest, the investment curve is I_2. If the rate of interest is reduced to 4 per cent, the investment curve will shift upward to I_3.

The rate of investment will have to be raised to reduce the marginal efficiency of capital to equality with the lower rate of interest. Thus the investment curve I_3 shows more investment at every level of income. Similarly when the interest rate is raised to 6 per cent, the investment curve

will shift downward to l_1The reduction in the rate of investment is essential to raise the marginal efficiency of capital to equality with the higher interest rate. In Figure 8 (B), just below Figure 8 (A), we derive the IS curve by marking the level of income at various interest rates. Each point on this IS curve represents a level of income at which saving equals investment at various interest rates. The rate of interest is represented on the vertical axis and the level of income on the horizontal axis. If the rate of interest is 6 per cent, the S curve intersects the 7, curve at E which determines $OY_;$ income. From this income level which equals Rs100crores we draw a dashed line downward to intersect the extended line from 6 per cent at point A. At interest rate 5 per cent, the S curve intersects the I_2 curve at E_2 so as to determine OY_2 income (Rs200 crores).

In the lower Figure 8 (B), the point B corresponds to 5 per cent interest rate and Rs200crores income level. Similarly, the point C corresponds to the equilibrium of S and I_3 at 4 per cent interest rate. By connecting these points A, B and C with a line, we get the IS curve. The IS curve slopes downward from left to right because as the interest rate falls, investment increases and so does income.

The LM Curve:

The LM curve shows all combinations of interest rates and levels of income at which the demand for and supply of money are equal. The LM curve is derived from the Keynesian formulation of liquidity preference schedules and the schedule of supply of money. A family of liquidity preference curves L_tY_1, L_2Y_2 and L_3Y_3 is drawn at income levels of Rs100crores, Rs200crores and Rs300crores respectively in Figure 9 (A). These curves together with the perfectly inelastic money supply curve MQ give us the LM curve. The LM curve consists of a series of points, each point representing an interest-income level at which the demand for money (L) equals the supply of money (M). If the income Level is Y (Rs. 100 crores), the demand for money (L_1Y_1) equals the money supply (QM) at interest rate OR_r At the Y_2 (Rs. 200 crores.) income level, the L_2Y_2 and the QM curves

equal at OR^ interest rate. Similarly at the Y_3 (Rs. 300 crores) income level, the L_3Y_3 and QM curves equal at OR_3 interest rate.

The supply of money, the liquidity preference and the level of income and the rate of interest provide data for the LM curve shown in Figure 9 (B). Suppose the level of income is Y_t (Rs100crores), as marked out on the income axis in Figure 9 (B). The income of Rs.100crores generates a demand for money represented by the liquidity preference curve L_1Y_1. From the point Â£, where the L_1Y_1 curve intersects the MQ curve, extends a dashed line horizontally to the right so as to meet the line drawn upward from Y_1 at K in Figure 9 (B). Points S and T can also be determined in a similar manner. By connecting these points K, S and T with a line, we get the LM curve. This curve relates different income levels to various interest rates, but it does not show what the rate of interest will be.

The LM curve slopes upward from left to right because given the quantity of money, an increasing preference for liquidity manifests itself in a higher rate of interest. It also becomes gradually perfectly inelastic shown as the vertical portion from T above on the LM curve in Panel (B) of Figure 9. This is because at higher income levels the demand for transaction and precautionary motives increases so that little is left to satisfy the demand for speculative motive out of a given supply of money. We may also note that at the extreme left the LM curve is perfectly elastic in relation to the rate of interest. This is shown as the horizontal portion of the LM curve which starts from the vertical axis in Panel (B) of Figure 9. With the decline in the level of income, the demand for transactions and precautionary motives also declines. Thus a larger amount is available in the form of idle balances but it does not lead to the lowering of the interest rate because we have reached the limit to which the rate of interest will fall. This lower limit to which the rate of interest will fall is the Keynesian liquidity trap already explained above in Keynes's theory of interest.

Determination of the Rate of Interest:

The IS and LM curves relate to income levels and interest rates. Taken by themselves they cannot tell us either about the level of income or the rate

of interest. It is only their intersection that determines the rate of interest. This is illustrated in Figure 10 where the LM and IS curves intersect at point E and OR rate of interest is determined corresponding to the income level OY. The income level and the interest rate lead to simultaneous equilibrium in the real (saving-investment) market and the money (demand and supply of money) market. This general equilibrium position persists at a point of time. If there is any deviation from this equilibrium position, certain forces will act and react in such a. manner that the equilibrium will be restored. At the income level OY_t the rate of interest in the real market is Y_1B and it is Y A in the money market. When the former rate is higher than the latter rate ($Y_1B > Y_1A$), the businessmen will borrow at a lower rate from the money market and invest the borrowed funds at a higher rate in the capital market.

This will tend to raise the level of income to OY via the investment multiplier and the equilibrium level of OR interest rate will be reached. On the other hand, at the income level OY_2 the rate of interest in the real market is less than the interest rate in the money market ($Y_2C < Y_2D$). In this situation, the businessmen will try to discharge debts in the money market rather than invest in the capital market. As a result, investment will fall and reduce income by the multiplier to OY and the equilibrium rate of interest OR will be established, Shifts or changes in the IS curve or the LM curve or in both change the equilibrium position and the rate of interest is determined accordingly. These are illustrated in Figure 11. Let IS and LM be the original curves. They intersect at E where OR interest rate is determined at OY income level. If the investment demand schedule shifts upward, or the saving schedule shifts downward, the curve IS would shift to the right as IS_1 curve. Given the LM curve, equilibrium will take place at E_1. The rate of interest would be OR_1 and the income level OY_1. If the quantity of money is increased or the liquidity preference curve is lowered, the LM curve would shift to the rights as LM_1 .It intersects IS_1curve at point E_2.

The new equilibrium rate of interest is OR and the income level is OY_2 Thus with a given LM curve, when the IS curve shifts to the right income increases and along with it the rate of interest also rises. Given the IS curve, when the LM curve shifts to the right, income increases but the

rate of interest falls. The Hicks-Hansen analysis is thus an integrated and determinate theory of interest in which the two determinates, the IS and LM curves, based on productivity, thrift, liquidity preference and the supply of money, all play their parts in the determination of the rate of interest.

Criticisms of the Modern Theory of Interest:

1. Static Theory. It is a static theory that explains the short-run behaviour of the economy. Thus it fails to explain how the economy behaves in the long run.

2. Interest Rate not Flexible. The theory is based on the assumption that the interest rate is flexible and varies with changes in LM or/and IS curves. But it may not always happen if the interest rate happens to be rigid because the adjustment mechanism will not take place.

3. Investment not Interest Elastic. The theory assumes that investment is interest elastic. But if investment is interest inelastic, as is generally the case in practice, then the Hicks-Hansen theory does not hold good.

4. Highly Artificial. According to Don Patinkin, the Hicks-Hansen theory is highly artificial and oversimplified because it divides the economy into real and monetary sectors. In reality, the real and monetary sectors of the economy are so interrelated and interdependent that they act and react on each other.

5. Closed Model. According to Prof. Rowan, the Hicks-Hansen theory is a closed model which does not take into consideration the effect of international trade. This restricts its usefulness for the study of policy.

6. Price Level Exogenous Variable. The price level is treated as an exogenous variable in this model. This is unrealistic because price changes play an important role in the determination of income and interest rates in an economy. Despite these weaknesses, this theory does not undermine the utility of the IS-LM technique in explaining the determination of interest rate in an economy.

Theories of Profit

The following points highlight the eight theories of profit in economics. The theories are: 1. The Rent Theory of Profit 2. The Wage Theory of Profit 3.

The Marginal Productivity Theory of Profit 4. The Dynamic Theory of Profit 5. F.W. Hawley's the Risk Theory of Profit 6. Knight's Theory or the Uncertainty-Bearing Theory 7. Modern Theory or Perfect Competition or Demand and Supply Theory of Profit 8. Prof. Schumpeter's Innovation Theory of Profit or "Profit is the Reward for Successful Innovation".

1. The Rent Theory of Profit:

This theory was developed by an American Economist Francis L. Walker. Walker has said that Profit is the rent of ability. He has made a comparative study between different grades of land and entrepreneur's different abilities. Entrepreneurs of superior ability earn Profits just as superior land earns rent.

According to Walker:

"Just as there is the marginal or no rent land, similarly there exists a marginal or no Profit entrepreneur who earns only wages of management. The marginal or no-profit entrepreneur is the least efficient one earning Profit not beyond an amount just sufficient to keep him or to carry on in his present industry. The industry managed and run by the marginal entrepreneur is similar to marginal land. Just as the land which is at margin is no rent, land, similarly, the marginal entrepreneur earns no profit." But there are other industries under the control of entrepreneurs possessing super abilities which yield Profits. The entrepreneur with superior ability earns Profit as the reward over the ability of the marginal or no-profit entrepreneur. Thus it can be said that the essential nature of Profit does not differ from that of rent because we are aware that rent is a differential surplus accruing to the superior land over the marginal or no rent land, similarly profit is a differential surplus which accrues to the superior ability entrepreneur over the marginal or no-profit entrepreneur.

Criticisms:

The important criticisms of this theory are as follows:

a. This theory is unrealistic:

Walker's view of Profit as a surplus like rent is unrealistic and it cannot be accepted as true approach of Profit.

b. It is not a true surplus as Marshall has said:

In this connection Marshall has said that land can earn positive or zero rent. But in the case of firm's entrepreneurs may have negative profits or losses.

c. Profits only in a dynamic state:

Rent can emerge in both static and dynamic conditions whereas profits we can find only in a dynamic state.

d. Profit is not gift of ability:

Profit does not arise always due to the superior ability of the entrepreneur. It may arise due to monopoly, innovation, risk, uncertainty etc.

e. This theory overlooks the important function of the entrepreneur as a risk-bearer:

From the profits of entrepreneur we must deduct the losses sustained by some others, who have been driven to bankruptcy. When this is done, there may be no surplus element in Profit and the analogy to rent vanishes. Moreover, it fails to explain the Profit of the ordinary shareholder of a joint-stock company.

f. This theory fails to explain the main causes of the size of Profits:

The differential gain arises because of the scarcity of superior units, either of land or of entrepreneurs. But the real thing is the explanation of the causes of the scarcity of the superior units. In the case of the rent of land, the point is not of great importance because the limitation is due to nature. Here the rent theory can throw no light on the fundamental questions.

g. Profits do not enter into price this cannot be said here:

The reward for risk-bearing must enter into long-period cost of production. In the short-period, Profits may not enter into price. But in the long-run, supply of entrepreneurs not being fixed by nature, normal Profits must form a part of cost of production.

2. The Wage Theory of Profit:

This theory was popularized and put forward by Prof. Taussig and Davenport the two most prominent economists. According to them**—"Profits are best regarded as simply a form of wages. They accrue to the**

entrepreneur on account of his special ability." They have argued that there is very close similarity between a labourer and entrepreneur. Just as labourers receive wages for his services, similarly entrepreneurs receive profit for his service.

The entrepreneur performs mental labour like—teachers, doctors, lawyers etc. But the only difference between entrepreneur and other mental workers is that the entrepreneur receives profit for his special ability and hard work. This is a surplus amount which the entrepreneurs receive after meeting all expenses of production where as the wage forms a part of the cost of production.

Criticisms:

a. Element of risk and uncertainty:

The entrepreneur's work is full of risk and uncertainty and profit is given to face this risk. But the workers receive wages simply for his labour. Risk and uncertainty part do not incorporate anywhere in his activities. For labourer risk is of losing the job which is an extreme step.

b. Profit is flexible, it may vary:

Profits may rise or fall. It depends upon the business conditions and situations. But wage may remain stable and cannot fluctuate more in the short- period.

c. This theory is silent over the payment to shareholders:

The shareholders of any organisation or company do not perform any function but they receive the share of profits in the form of dividend for undertaking risk of money invested. This theory fails to explain this contention as to why they are paid.

d. Entrepreneurs windfall or chance profits:

The entrepreneur may receive windfall or chance profits but a worker cannot have opportunity to get wages of chance or windfalls.

3. The Marginal Productivity Theory of Profit:

This theory was propounded by Prof. Marshall. According to him, **"Profit is equal to the marginal productivity of the entrepreneur. He has said that the amount which the community is liable to produce**

with the help of entrepreneur over and above what it could produce with his help." Recently Stigler and Stonier and Hague have said that "Profit is the reward of an entrepreneur which is determined by its marginal revenue productivity, the higher are the profits and lower the marginal revenue productivity, the lower are the profits of an entrepreneur."

Criticisms:

Important criticisms given by various economists are as follows:

a. This theory is based on unrealistic assumptions:

These unrealistic assumptions are homogeneity of entrepreneurs in an industry. As entrepreneurs' efficiency differ, therefore it is not possible that there will be one marginal revenue productivity curve for all entrepreneurs. So Profit cannot be same.

b. This theory fails to determine profit accurately:

Because efficiency of entrepreneurs differs, systems and methods of doing work differ, therefore. Profit cannot be calculated accurately.

c. The concept of marginal revenue productivity of entrepreneurship is a meaningless concept:

Because unlike other factors, there can be only one entrepreneur in a firm.

d. It is one sided theory:

This theory takes into account only the demand for entrepreneurs and does not take into account the supply or availability of entrepreneurs.

e. This is a static theory:

Where all entrepreneurs earn only normal profits, they have not considered that the world is dynamic also where some entrepreneurs can earn more than normal profits. This theory has not taken into account the windfall or chance or gain or even monopoly profits.

4. The Dynamic Theory of Profit:

Prof. J. B Clark propounded this theory in the year 1900. According to him**—"Profit is the difference between the price and the cost of the production of the commodity".** But Profit is the result of dynamic change. Further, Prof. Clark was of this opinion that in a stationary state having static economic conditions of demand and supply, there can be no real or

pure profit as a surplus. In a stationary economy, the quantum of capital invested, methods of production, managerial organisation, technology, demand pattern etc. remain constant. Under competitive conditions, price tends to equal average costs; hence, the surplus is zero. So, no pure profit but there may be some frictional profits emerging due to frictions in the system. But, this cannot be regarded as real Profits.

Profit is the result exclusively of six dynamic changes i.e.:

- Changes or increase in population,
- Changes in tastes and preferences,
- Multiplication of wants,
- Capital formation,
- Technological advancement and
- Changes in the form of business organisation.

On account of these changes the economy tends to be dynamic. Demand and supply conditions are altered. Some entrepreneurs may get advantageous business positions against others and may reap surplus over costs, as a real profit. In short, those who take advantage of changing situation can earn real profits according to their efficiency. Inefficient and careless producers who fail to move with dynamic changes may not get any real profit and may even incur losses. Thus, Clark's dynamic theory of Profit has an element of truth as it emphasis the dynamic aspect of Profit.

Criticisms:

a. All changes are not foreseen:

Clark's theory fails to make any difference between a change that is foreseen and one that is unforeseen in advance. If the six generic changes as assumed by Prof. Clark are to be foreknown in advance then the effects of changes will not hold at all. In reality, all changes are not foreseen. Some are foreseen and some are not. So, to have a clear understanding of the problem, it is essential to separate its effects from those of change as such.

b. This theory gives artificial dichotomy:

In this connection Taussig has said that Clark's theory gives an artificial dichotomy of 'Profit' and 'Wages of management'.

c. All changes do not lead to Profit:

Clark's theory suggests that all dynamic changes lead to Profit. But critics are of this opinion that only unpredictable changes would give rise to profits. Predictable changes will not cause surplus to emerge on account of precise adjustments.

d. Here, the concept of frictional Profit is vague:

Clark's theory indicates that in a stationary state, there is only a frictional profit. But the concept of frictional profit is vague. But it is the normal profit which is earned in a stationary state.

e. Element of risk involved in business:

Clark's theory of Profit do not stress the element of risk involved in business due to dynamic changes. The best course is to combine elements of risk dynamic changes to understand the true nature of profit in a modern economy.

5. F.W. Hawley's the Risk Theory of Profit:

This theory of Profit is associated with F. B. Hawley who has considered the risk-taking as the important function of an entrepreneur. The entrepreneur exposes his business to risk, and in turn he receives a reward in the form of Profit because the task of risk-taking is irksome. It is definite that no entrepreneur will like to undertake risks if he gets only the normal return. Therefore, the reward for risk-taking must be higher than the actual value of the risk. Further, it has been said that the actual value of the risk. Further, it has been said that more risky the business, the higher is the expected Profit rate. As Professor D. M. Holland has said that "riskier the industry or firm, the higher is its Profit rate." But he was warned that this tentative view must be tested in depth.

Criticisms:

a. There cannot be functional relationship between Risk and Profit:

Those persons who dare to take high risks in certain businesses may not necessarily earn high profits.

b. Profit is not based on entrepreneur's ability:

In this connection Prof. Carve has said that "Profit is not based on entrepreneur's ability to undertake the risks of the business, but rather as his capability of risk avoidance."

c. It is an incomplete theory:

From business point of view, all enterprises are risky and an element of uncertainty is present there. But every entrepreneur aims at making large profits which is also uncertain. Therefore, Hawley's Risk Theory can also be called as an incomplete theory of Profit.

d. Amount of Profit not related to size of risk involved:

The amount of Profit is not in any way related to the size of the risk undertaken. If it were so related then every entrepreneur would involve himself into huge risks in order to earn larger profits.

e. Concentrates mostly on risk and not on anything else:

This theory mostly disregards many other factors attributable to Profit and just concentrate on risks and risks alone.

6. Knight's Theory or the Uncertainty-Bearing Theory:

Prof. Knight's theory of uncertainty bearing theory of Profit is an improvement and refinement theory of Profit over Hawley's risk-bearing theory of Profit. Here, Profit according to Knight, is the reward of bearing non-insurable risks and uncertainties. It is a deviation arising from uncertainty. Uncertainty prevails in the entire society and profits, positive or negative, in a way accrues to all factor services. In other words, there is profits element in all types of income. But the division of social income between Profit and contractual income depends on the supply of entrepreneurial ability. Uncertainty bearing is the most important function in a dynamic state. It is the entrepreneur who either delegates this function among different personnel or assumes it himself. The expectation of Profit is, in a way, the supply price of entrepreneurial uncertainty-bearing. In a competitive economy where there is no risk, every entrepreneur will have a minimum supply price.

In short Knight's theory implies that:

(i) Profit is reward for uncertainty-bearing.

(ii) The un-measurable risks are termed as uncertainty. These un-measurable risks are true hazards of business.

(iii) Pure Profit is, however, a temporal and unfixed reward. It is turned with uncertainty. Once the unforeseen circumstances become known, necessary adjustment would be possible. Then pure Profit disappears.

Criticisms:

a. This theory does not give clear notion of entrepreneurship therefore it has been called unrealistic:

In this theory there is no indication as to who are the real owners because owners are shareholders and policy decision-makers are salaried people.

b. Difficulty in the distribution of profit:

This theory does not solve the problem of allocation or distribution of profit among the controlling and ownership group, therefore, this theory keeps the problem of the determination of Profit unsolved.

c. This theory fails to expose the phenomenon of monopoly profit:

The theory does not suit well to expose the phenomenon of monopoly profit. When there is least uncertainty involved in a monopoly business.

d. Profit is not a residual income:

Knight has mentioned in his theory that Profit is a residual income but J. F. Weston has said that "the exercise of judgment of Profit may be sold on a fixed-price basis or on a variable price-basis." This is how the expert manager sell their services to earn Profit.

e. This theory has not said anything on monopoly profit:

This theory does not throw any light on the monopoly profit. As we have studied that monopoly firms earn much larger profits than competitive firms and they are not due to the presence of uncertainty. Above all, the uncertainty element cannot be qualified to improve profits. In-spite of the weaknesses as mentioned above, this theory of Knight is regarded as the only satisfactory explanation of the nature of profit.

7. Modern Theory or Perfect Competition or Demand and Supply Theory of Profit:

This modern theory of Profit defines the entrepreneur as a business enterprise itself and 'Profits' as his net income. In this theory profits have been regarded as the reward of an entrepreneur and are governed by the demand for and supply of entrepreneur.

Demand for Entrepreneurs:

The demand for entrepreneurs mostly depends upon the level of industrial development, the elements of uncertainty in the industry, the scale of production and the marginal revenue productivity of entrepreneurship. If the level of industrial progress is high, the scale of production is large and efficiency and productivity increase, the profits will be high. The marginal revenue productivity of entrepreneurship is the most important factor in influencing the demand for entrepreneurs.

Supply of Entrepreneurs:

Similarly, the supply of entrepreneurs depends upon various factors like the availability of capital, the existence of managerial and technical personal, the number of entrepreneurs and the condition of society etc. The larger the availability of capital, the larger is the supply of entrepreneur's capital may be available in sufficient amount, but an entrepreneur has to depend largely on the managers and other technical personal for organising and running the business successfully. If trained managerial and other personal are available in the market, the supply of entrepreneurs is bound to increase. Further, the economists are also of this opinion that the size of population is another factor that influences entrepreneurship. The larger the size of population, the higher will be the demand for various products which will attract more people to entrepreneurship and the supply of entrepreneurs will increase.

Criticisms:

While criticising this theory Knight has said that Profit has been regarded as the reward for bearing non-insurable risks and uncertainties, then under perfect competition there can be no profit in the long-run. It is a

static state where population, capital, technology, tastes, business organisation and income do not change. If they change they can be predicted. Thus, there is no risk and uncertainty. The marginal revenue productivity curve of entrepreneurship would be zero. Therefore, Profit will also be zero. In a static state, profits exist because Profits are not competed away due to the presence of imperfect competition. So what entrepreneurs earn are monopoly profits rather than pure profits. It should be remembered that Manager-entrepreneurs earn wages of management and capitalist—entrepreneurs earn interest.

8. Schumpeter's Innovation Theory of Profit

Schumpeter deemed Profit as the reward to enterprise and innovation. In his opinion, the entrepreneur initiates innovation in the business and when he succeeds, he earns Profit as his reward. Now, the question is what is innovation? "Innovation means commercial application of new scientific inventions and discoveries." An innovator is, therefore a businessman with vision, foresight, and originality and is bold enough to bear high risks involved in undertaking new activities on a new basis. The innovator is not a scientist, but he successfully introduces new inventions on a commercial basis. In giving opinion over this Samuelson has written as an example— **"The scientific theory of radio wave was the brain-work of Maxwell. It was experienced upon by Hertz and its commercially profitable use was carried out by Marconi and Sarnoff, who are the innovators in radio manufacturing."**

Innovation is of two types:

(i) Product innovations, and

(ii) Market innovations.

Product innovations affect the cost and quality of the product while market innovations include discovery and exploitation of new market, introducing new variety of products and product improvement, modes of advertising and sales propaganda etc. It has been said that any form of

innovation leads to a Profit. It is called as innovation profit. This Profit is uncertain and unpredictable. It is temporary in nature.

Criticisms:

a. Schumpeter has never considered Profit as the reward for risk-taking:

He is of this opinion that risk-taking is the function of the capitalist and not of the entrepreneur. It is the shareholders who undertake risks and thus earn profits.

b. There is no place of uncertainty in Schumpeter's innovation theory:

Profit is not the reward of uncertainty it is simply the wages of management.

c. This theory is incomplete:

Profit accrues to the entrepreneur for his organisational ability and nothing else. Therefore, this theory has been called as an incomplete explanation of the emergence of profits.

The Liquidity Preference Theory of Interest

The Liquidity Preference Theory presented by J. M. Keynes in 1936 is the most celebrated of all. According to Keynes, the rate of interest is a purely monetary phenomenon. It is the reward for parting with liquidity for a specific period of time. Thus, like the price of a commodity, the rate of interest is determined by the demand for and the supply of money. It is, therefore, necessary to introduce the concepts of demand for money and supply of money.

The supply of money refers to the stock of money in circulation and is a fixed quantity at a particular point of time. It is the sum of currency (notes and coins) and commercial bank deposits. It remains fixed in the short rim because it is determined and controlled by the central bank of a country. So it plays a passive role in interest rate determination. By contrast, the demand for money plays an active role in determining the equilibrium rate of interest. Therefore, background knowledge of demand for money is essential in order to understand Keynes' theory.

The Demand for Money:

Wealth can be held in various forms— money, fixed interest securities (bonds), shares, property, jewellery, valuable paintings etc. Keynes first analysed, in detail, the reasons why people will hold wealth in the form of money. At a fixed point of time, a certain stock of money is held, i. e., people wish to hold a certain amount of wealth in 'liquid' form. 'Liquidity' refers to the ease with which assets can be changed into cash without loss or delay. It is property which is enjoyed by all assets to some extent. Obviously money is the most liquid of all assets. The demand for money was, therefore, termed by Keynes 'liquidity preference'. J. M. Keynes gave three reasons for holding money the transactions motive, the precautionary motive and the speculative motive.

1. Transactions Motive:

Individuals and business firms hold money in order to carry out day-to-day transactions. Each individual or firm has a time gap between receipts and payment and will need to hold money to cover this. The average amount held will depend primarily on the system of payments, i.e., on the frequency of the receipts. For example, if a weekly paid person receives Rs. 300 a week and he has spent it all by the next pay-day, his average cash holding is Rs. 150, i.e., the amount he had at the beginning (Rs. 300) and the amount he has at the end (zero), divided by 2. If he receives monthly salary of Rs. 1,200 then, assuming that his spending habits do not alter, his average cash holding will rise to Rs. 600, i.e., (Rs. 1200 + 0) ÷ 2. The amount of cash held for transactions and precautionary purposes also depends on incomes and prices. If income increases, then more money will be held. Similarly, if prices rise, more money will be required to purchase the same amount of goods and services.

2. Precautionary Motive:

People and business firms hold some money as a reserve to meet unforeseen contingencies, such as sickness or accidents or the need to take advantage of an opportune to buy something which is being offered at a specially reduced price for only a limited period, e.g., during a sale.

3. Speculative Motive:

The classical economists considered it irrational for people to hold wealth in the form of money other than that held for transactions and/or precautionary purposes. It is because any money left over could be invested in interest-earning assets like bonds. Keynes, however, argued that it was not necessarily irrational to hold idle money balances.

He pointed out that at times it might be preferable to hold idle money (cash) than to buy government securities (bonds). If a person holds money, he loses interest but he does not suffer capital loss (due to fall in the value of his assets) either. In fact, it costs money to hold money. Therefore, the rate of interest is called the opportunity cost of money holding. By holding money an individual loses the opportunity to earn interest. (Here we ignore the effect of inflation and leave aside any reduction there-from). By holding securities, however, he earns a fixed sum as interest, but its market value can vary. Therefore, in certain situations, money is preferable to securities. For example, if a person pays Rs. 100 for a Rs. 100 bond whose rate of interest is 10%, then at the end of the year he receives Rs. 10. But if in the meantime the value of the bond has fallen to below Rs. 90, the loss on this amount more than offsets the interest. The market value of a bond is inversely related to the market rate of interest. Thus, if the rate of interest goes up, the market value of a bond will fall.

The market value is shown in the following formula:

Market value = Original value x Original rate of interest Market rate of interest In the case of a Rs. 100 bonds whose original rate of interest is 10%, Rs. 10 interest will be paid at the end of the year. If the market rate of interest raises to 20% and the bondholder wishes to sell it for some reason, he will not find a buyer ready to give him Rs. 100 for it. The reason is very simple. If the buyer pays Rs. 100 and at the end of the year receives Rs. 10 as interest, then his investment has yielded interest of only 10%, whereas elsewhere he could have gained 20%. The buyer can at best offer Rs. 50 for the bond, so that when he receives the Rs. 10 interest, his investment has earned 20%. Thus, because the market rate of interest has risen, the market value of the bond has fallen. The converse is also true if the market rate of

interest falls, the market value of a bond will rise. If, in an example, the rate of interest falls to 5%, the value of the bond will rise to Rs. 400. It is because the return from this bond at 5% interest will now be Rs. 20.

According to Keynes, the speculative demand for money will be determined by people's expectations regarding the market rate of interest. If the rate of interest is very low and people expect it to rise, then they will consider it more judicious to hold money rather than bonds. If, on the other hand, the rate of interest is very high and people expect it to fall, then they will prefer to hold bonds instead of money. Thus, there is an inverse relation between the rate of interest and the demand for money. At high rates of interest people hold less money and vice versa. Another reason is that if the rate of interest is high, it is 'more expensive' to hold money, i.e., the interest which is foregone by not investing the money is at a high level. For these two reasons, the demand for idle money balances is inversely related to the rate of interest. Keynes assumed that the demand for money for the other two motives is not affected by changes in the rate of interest, i.e., is perfectly inelastic with regard to the rate of interest. Therefore, if all three elements in the demand for money are added together to derive the total demand curve for money, the result would be a curve of the type shown in Fig. 15.2.

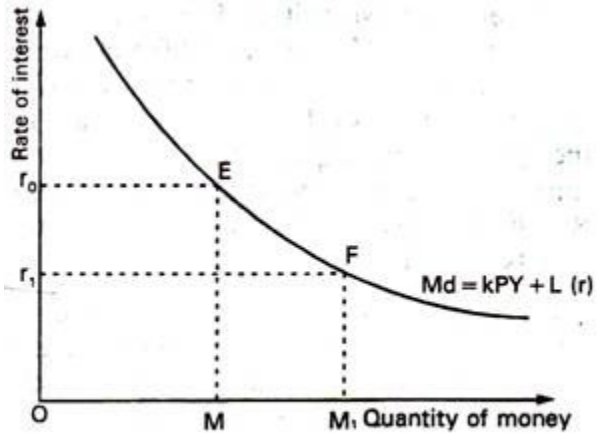


Fig. 15.2. The Demand Curve for Money

The curve shows that if the rate of interest falls, e.g., from O_0 to Or_1, the demand for money increases, from OM to OM_1. According to Keynes, at some low rate of interest the demand for money becomes perfectly elastic because if the rate falls below this level, no one would be prepared to buy bonds.

Determination of the Rate of Interest:

The rate of interest, which is the 'price' of money, is determined by demand and supply in a competitive situation. We have seen that the demand curve for money is downward sloping. We assume that the supply curve will be perfectly inelastic with regard to the rate of interest, i.e., that the supply is determined by the monetary authorities and does not vary with the rate of interest in the short run. Thus, in Fig. 15.3(a) the supply of money is represented by the perfectly inelastic supply curve, M.

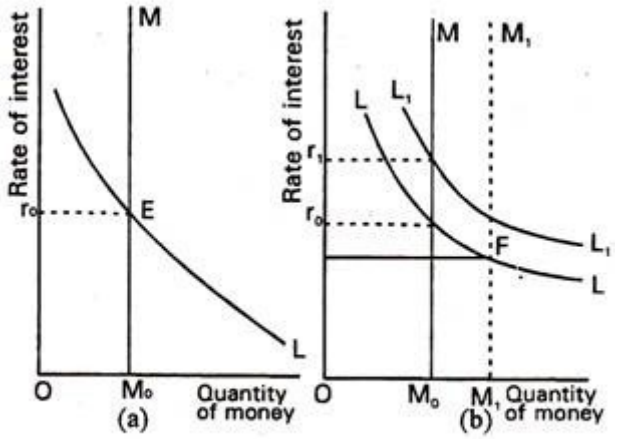


Fig. 15.3. The Determination of the Rate of Interest

The equilibrium rate of interest is Or_0 because it is the only rate of interest at which the money market is in equilibrium, e.g., the demand for money is equal to its supply. What is the logic of this equilibrium? If the rate of interest goes above the equilibrium level there will be excess supply of money or excess demand for bonds. The price of bonds will rise or the rate of interest will fall. On the other end, if the rate of interest goes below the equilibrium level, there will be excess demand for money, i.e., people will need more money to hold that is currently being supplied by the central bank. To meet this demand people will sell bonds. There will be excess supply of bonds. Bond price will fall or, what comes to the same thing, the rate of interest will rise.

Shifts of Supply and Demand Curves:

Both the demand and supply curves may shift to left or right if circumstances change. For instance, if incomes rise the demand for money will shift to the right, because people will need more money for transaction purposes. Consequently, the rate of interest will rise. This is indicated by the curve L_1 which intersects the supply curve of money at point E so as to cause the rate of interest to rise to Or_1. By contrast, the supply curve will shift if the monetary authority (i.e., the central bank) increases or reduces the supply of money. The effect of an increase in the supply of money is illustrated in Fig. 15.3(b). If supply of money is increased from OM_0 to OM_1 the rate of interest will fall from Or_0 to Or_1. The reason is that at the old rate of interest the supply of money has become greater than demand and people use the surplus money to buy bonds. The increased demand for bonds causes the price of bonds to rise. The rate of interest will, therefore, fall.

Criticisms of the Theory:

Keynes' liquidity preference theory has been severely criticised.

These are discussed here:

1. In the construction of the figure, speculative demand for money is included and the other two sources of demand are ignored. It implies that they are known and subtracted from total money supply. But they can be known only when income is in equilibrium, i.e., Y – C + I or S = I. Hence, liquidity preference theory requires as pre-condition saving-investment equality, already postulated by classical scholars. Hence, the rate of interest is neither a purely monetary phenomenon nor a purely real phenomenon.
2. So far as the main content of the Keynesian interest theory is concerned, it is the determination of the rate of interest through equality between demand for, and supply of, money. But one of the components of total money demand—known as speculative demand—is assumed to depend on rate of interest. Hence, the logical circularity in the model can be mentioned as one of its principal sources of weakness.

3. Keynes ignored real factors like productivity of capital and thriftiness in the determination of interest rate.

4. As Jacob Viner has remarked: **"Without saving there can be no liquidity to surrender."** According to Keynes, interest is a reward for parting with liquidity and in no way an inducement for saving, but it is ridiculous to think of surrendering liquidity if one has not already saved money.

Liquidity Trap:

Liquidity trap refers to a situation where the rate of interest is so low that people prefer to hold money rather than invest it in bonds. Keynes pointed out that at low rates of interest the demand curve for money becomes completely elastic. So the liquidity preference curve is not downward sloping throughout. This usually happens during depression. During depression any attempt by the central bank to reduce the rate of interest by increasing the stock of money will be futile. In such a situation, no change in money supply is sufficient to alter the rate of interest. So it is not possible to stimulate more investment. In fact, any increase in the stock of money by the central bank will be held by the people in the form of liquid balance. This will prevent the rate of interest from falling further. See Fig. 15.4 where the completely elastic position (EFMd) of the liquidity preference curve is called liquidity trap.

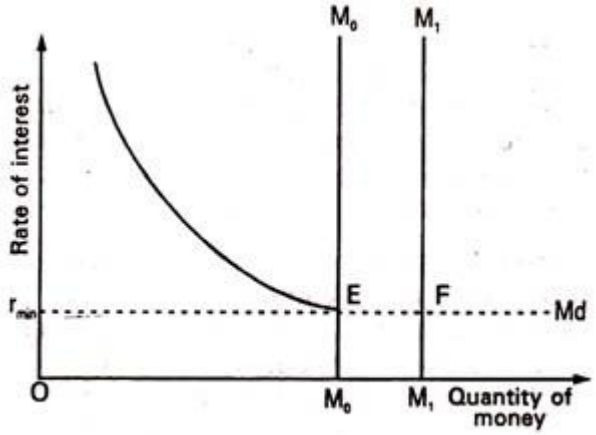


Fig. 15.4. Liquidity Trap

The implication is that monetary policy loses its effectiveness if there is a liquidity trap in the demand curve of money. Keynes argued that the only way to stimulate investment in a depressed economy is to use a positive

fiscal policy. Such a policy works through an increase in government expenditure or reduction in taxes in order to increase aggregate effective demand. The reason is simple. People feel that the rate of interest has fallen enough. It cannot fall further. Thus, if it rises in near future the price of bonds will fall. So purchase of bonds is risky. Money-holding is not that costly because the rate of interest is low. Thus, people prefer to hold as much money as possible, with the expectation that the rate of interest will rise in future. As soon as it rises they will buy bonds. In such a situation any additional money supplied by the central bank will be absorbed by the people and this will prevent the rate of interest from falling further.

The Possibility of Zero Rate of Interest:

Interest is treated as a price paid by borrowers to lenders and will depend on the supply for and demand of loanable money for various purposes. Generally, a large supply of capital relating to the demand means low rates of interest and a large demand relative to supply means high interest rates. According to some writers, the rate of interest would fall to zero in a static economy where the demand for loanable funds is nil. In a static economy, there is no fresh investment, the demand for loanable funds is nil and so the rate of interest would be zero. From the point of view of the demand for loans, zero rate of interest means that the marginal net product of capital is nil. This means that we cannot increase society's total product further by employing more capital. We have reached a state in which our productivity is maximum. It means that all our wants have been satisfied. So the demand for capital will be zero. But, it is not likely that the demand for loan-capital will be ever zero. But, in reality, we cannot think of a state of society in which people will have no wants, and no desires. So long as they remain, there will always be endless possibilities for employing capital. The rate of interest, therefore, cannot be zero.

There are certain dynamic forces like inventions and discoveries, growth of population, etc. which keep always up the demand price of loan-capital. Similarly, from the supply side, a zero rate of interest means that people will go on lending without expecting any return in exchange. But the liquidity preference will not drop to zero for a number of reasons. As the rate

of interest falls, more money will be absorbed by people to satisfy transactions demand for money. Moreover, the zero interest rate means that the liquidity-preference also becomes zero; people lend money without any interest. But such a situation is most unlikely to appear in reality. There are always reasons why the liquidity-preference would never drop to zero. As the holding of cash-money has the distinct advantages over the holding of other assets, people will always prefer cash money to other assets. It means that the liquidity- preference cannot drop down to zero, and from this it follows that the rate of interest will never fall to zero. All these set a limit much above zero to the practical decline in the rate of interest. In the Keynesian theory it is also seen that the rate of interest cannot fall below a certain level where the demand for liquidity becomes infinitely elastic and that situation has already been described as the liquidity trap (Fig. 15.4), a term first used by D. Robertson.

Shackle's Theory:

Shackle has extended Knight's theory of profit by introducing expectations under conditions of uncertainty. According to Shackle, expectations are of two types: general and particular. General expectations relate to variables general to the economy as a whole. They are associated with future macro-variables such as the general price level, GNP. Balance of payments, etc. On the other hand, particular expectations relate to variables particular to a firm or industry. They are associated with such micro-variables as the future reaction of a particular marketing strategy adopted by a firm, the future pricing policy of a competitive firm, etc. The decisions of the business community are generally based on general expectations. If it regards them favourable, investments are made. But there is 'subjective certainty' in the case of general expectations. Their time horizon is about 12 months.

As the general expectations have subjective certainty and their time horizon is also of reasonable duration, the business community is able to anticipate price and income increases correctly for the economy as a whole, and by adopting appropriate inventory policies, it earns windfall profit. But

in the case of particular expectations, there is "subjective uncertainty" and the time horizon is also quite long ranging between 100 to 150 months. Under particular expectations, a firm or an industry may earn either innovative profit or monopoly profit depending upon its policies and competitors. Under perfect competition, the number of buyers and sellers of a similar product is very large.

A firm which innovates in introducing new techniques of production or new products or new techniques of management earns innovator's profit. On the other hand, when there is monopolistic competition and the product is differentiated, it is the marketing policy that leads to profit. As there is subjective uncertainty and the time horizon is quite long, it is the taking of correct decisions by a firm about marketing, advertising, etc. of its products in relation to the products of its competitors that lead to monopoly profit. Thus profit whether monopoly or innovative arises under subjective uncertainty depending upon correct decision-making by a firm. Who takes such decisions in a firm and what is the basis? According to Shackle, decision-making under uncertainty is done by the entrepreneur of a firm. The routine types of decisions which often require **"weighing the evidence"** are made by the respective heads of departments in the firm.

So far as the basis of decision-making is concerned, Shackle adopts a psychological approach. According to him, the entrepreneur formulates hypotheses about the future consequences of his decision. He imagines a neutral point to the right of which he places those hypotheses that are pleasing and to the left of it, those that are displeasing. All pleasing or displeasing consequences that are close to the neutral point appear "very plausible" and have a low degree of "potential surprise". But more pleasing and more displeasing hypotheses that are moving away from the neutral point on both directions have a growing degree of potential surprise.

To take one hypothesis, it is a combination of its plausibility and its relative pleasantness and unpleasantness. When the entrepreneur moves towards the right of the neutral point, the hypothesis grows in pleasantness faster than it grows in implausibility. But after a point, the increasing pleasantness offsets the increasing implausibility of the hypothesis.

Ultimately, there is "peak" hypothesis on the pleasant side. On the other hand, when the entrepreneur moves towards the left of the neutral point, the hypothesis grows in pleasantness faster than it grows in plausibility. But after a certain point, the increasing unpleasantness offsets the increasing plausibility of the hypothesis.

Ultimately, there will also be a peak hypothesis on the unpleasant side. Shackle calls the pleasant side peak the "focus gain" and unpleasant side peak the "focus loss". If the focus gain exceeds the focus loss, the entrepreneur will make a positive decision. He will make investments and earn profit. On the contrary, if the focus loss exceeds the focus gain, the entrepreneur will make a negative decision and refrain from making investments because his particular expectations are likely to be unfavourable. Thus in Shackle's theory, the entrepreneurial decision-making is neither irrational nor whimsical. Rather, it is based on his intuitive perception.

Criticisms:

Prof. Shackle has formulated a psychological theory of profit which is highly abstract. But it contains within it the elements of Knight's uncertainty theory of profit, Schumpeter's innovations theory of profit and monopoly theory of profit. However, it is essentially a decision theory which is based on the psychology of the entrepreneur. As pointed out by Prof. Kierstead, "Professor Shackle himself uses the device of introspection effectively, but introspection can allow him to discover how he makes a decision; it cannot tell with any certainty how an entrepreneur or a Board of Directors makes a decision."

Alternative Theories of Distribution

The theory of distribution deals essentially with the determination of the levels of payment to the various factors of production, i.e., the prices of the economy's productive resources. The theory of income distribution is related to factor pricing. It is a segment of general equilibrium theory, inasmuch as a change in the level of wages, interest rates, or rents has significant effects on the whole economy. As W.J. Baumol has put it- **"Since**

general equilibrium analysis seeks to account for the determination of every price in the economy, it includes the pricing of inputs within its scope."

Two Sets of Theories:

The theories of distribution can be broadly divided into two categories, viz., microeconomic theories and macroeconomic theories. The most celebrated microeconomic theory is the marginal productivity theory of distribution. It was developed by J.B. Clark in 1899 and then modified by Philip Wick steed. The two macroeconomic theories are the classical (Ricardian) theory and the Cambridge (Kaldor) theory. Although Karl Marx was very much concerned about the ethical aspects of distribution theory, he never formulated any model (theory) of distribution Marxian economic analysis is related primarily to production.

1. Marginal Productivity Theory:

The marginal productivity theory is an approach to explaining the rewards received by the various factors of production that jointly produce output. It holds that the wage rate or payment for the services of a unit of a factor is equal to the decrease in the value of commodities produced that would result if any unit of that factor were withdrawn from the productive process, the amounts of all other factors remaining the same. The basic justification of this assertion is simple enough. It rests on three assumptions: that the products sold are produced by technologies that satisfy the 'law of variable proportions' which holds that successive equal increments of one factor of production, the amounts of all other factors remaining unchanged, will yield successively smaller increments of physical output.

It follows immediately from these assumptions that if the wage of any factor exceeds the value of the output that would be lost if a unit less of that factor were employed, then a unit less of that factor will be employed, and successive units will be released until the inequality is annihilated. Similarly, if the wage of any factor is less than the value of the output that an additional unit could produce, successive units of that factor will be employed until the inequality disappears. The essence of the marginal

productivity (MP) theory is very intuitive: under pure competition the profit-maximising firm will hire any factor (such as labour) up to the point where its price (wage) equals the value of its marginal product, i.e., MPP_L x P. The reason is that if the VMP exceeds the price of the factor, the firm can increase its profits by acquiring additional units of the input since more units bring in more revenue to the firm than they cost. The reverse is true if the price of the factor exceeds its VMP. If there is imperfection in the commodity market and price of the product of a firm varies with output, the return to the input will be equated to its MRP (= P x MPP – the loss in revenue of the firm that results due to the fact that increased production forces it to reduce its price of all the previous units).

In this case the profit-maximisation requirement becomes:

Suppose an additional unit of input increases output from q_a to q_b (MPP = q_b– q_a). Thus, the VMP is equal to the price multiplied by (q_b– q_a). This is shown by the area Cq_aq_bB. However, price fall leads to a loss on the initial units, shown by the area P_aP_bCA. The difference between the two shaded areas is the MRP of the output. The firm will hire the input until its price is equal to that MRP.

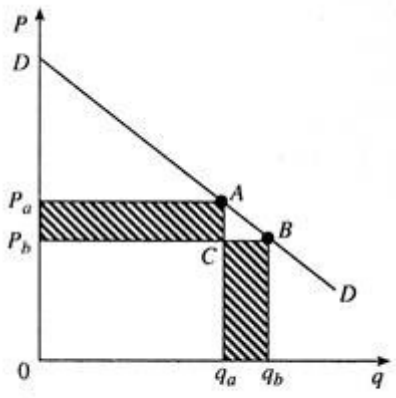


Fig. 1: Demand Curve of a Variable Factor

Evaluation:

At present the marginal productivity principle is used to explain the demand for factors of production in both a two-factor version using aggregate capital and aggregate labour as the factors, and an n-factor version, where n is the number of distinguishable factors used in the

production process. To use the two-factor version it is necessary to establish quantitative measures of the aggregates of dissimilar objects that are given the names 'capital' and 'labour', a task that has never been performed to anyone's satisfaction. According to Milton Friedman and W.J. Baumol, the marginal productivity theory is essentially a theory that helps us to determine the firm's derived demand for any given input. It shows how the quantity of the input demanded by a profit-maximising firm will vary with the input's price and makes it abundantly clear that, for such a firm, this demand relationship depends directly on the demand for the final product as well as the input's marginal physical product.

In truth, the marginal productivity theory is not a theory of input price determination. It analyses how a firm takes any decision regarding the optimal usage of an input. But it fails to explain how usage and prices of other inputs are determined. The reason is that it ignores the supply side of the input market completely. Very frequently, if the problem of finding the combination of factor inputs that maximises profits is solved in a straightforward way, some of the input levels in the solution turn out to be negative—which is nonsense. The essential perceptions of marginal productivity theory still apply, but they can no longer be expressed by equalities between price ratios and ratios of marginal changes.

Moreover, the marginal productivity theory has to be cast in general equilibrium framework of the Walrasian type by collecting information on each input (whether purchased from another firm or a private individual like a worker selling his labour power) as also on demand for and supply of every good produced in the economy. Otherwise it is not possible to find out a set of equilibrium prices and quantities for every item in the economy, including wages for different types of labour (skilled, semi-skilled and unskilled), rents for different qualities of land, etc. If this is done, then only the marginal productivity theory will turn out to be a generalized theory of factor price determination.

No partial model is adequate for the purpose of building such a theory because a rise in wages in industry 'A' will sooner or later raise labour costs in industry 'B'. A rise in the price of fuels will affect the relative demands of

other inputs. The relevant question here is not whether marginal productivity theory (with necessary modifications) is valid or logically defective. Instead, the issue is the degree to which it is useful. The truth is that the theory, with all its assumptions, is fundamentally valid but perhaps not as illuminating as one might expect.

Theory and Evidence:

The real test of a theory lies in its empirical verification. By using the Cobb-Douglas production function (CDPF) at the aggregate level, economists have attempted to test the empirically observed fact that the share of wages in the national income of the USA has remained relatively constant for a fairly long period of time.

At the macro-level the CDPF takes the following form:

$Y = mL^{\alpha}K^{1-\alpha,}$ where, 'm' and a are positive constants (and $\alpha < 1$). Here 'F is national income, 'L' is the quantity of labour input and 'K' is the quantity of capital employed. If labour is paid a wage equal to its marginal product this production function will yield a share of wage relative to total output which has a fixed value and is independent of the values of the variables Y, L and K, as the empirical evidence suggests. In fact, the ratio between total wage income and total output, K, must, in this case, be exactly equal to a, the exponent of labour (L) in the CDPF. The same result is obtained in case of the income of capital.

Proof:

Here $MP_L = \alpha\, mL^{\alpha-1}\, K^{1-\alpha}$.

Since this is the wage per worker, total wage payments must equal its amount multiplied by the number of workers, L; i.e., the total wage bill must be:

$L\ \alpha mL^{(\alpha-1)}K^{(1-\alpha)} = \alpha\, mL^{\alpha}K^{(1-\alpha)} = \alpha Y.$, Thus, total wage bill equals α times total output.

One question which remains unanswered is why, in spite of rapid technological progress in the USA, the exponent of the production function (α) has not changed. Thus, the explanation of a constant wage share goes in terms of a constant α, for which no explanation is offered. Critics also

comment that there is no valid reason for accepting the basic proposition that the CDPF gives an accurate depiction of technology at the macro-level. It is just an empirical thesis which has been proposed to explain an empirical observation.

Euler's Theorem and the Adding-up Controversy:

The second application of the marginal productivity theory was in the area of distributive justice. J.B. Clark believed that distribution of factor incomes according to the marginal product of each factor gives every factor an amount of social output the factor (or the agent production) creates. Thus, the distribution of income on the basis of the marginal productivity theory seems to be equitable in nature. For this reason some economists attempted to use the theory as a basis for showing that the distribution of income under free competitive capitalism must be morally just. As a result, it became important to the proponents of the theory to show that the sum of the marginal products added up to exactly the total product, leaving neither a deficit nor a surplus for the entrepreneur to extract. In this context, the Euler's theorem comes to our aid. The theorem tells us that if the production functions are linearly homogeneous, i.e., it shows CRS, the sum of the marginal products, will actually add up to the total product. **This means that if each input 'i' is paid $r_i = p\ \partial q/\partial f_i$, the value of its marginal product, we must have:**

$$pq = \sum f_i \frac{\partial q}{\partial f_i} = \sum r_i f_i$$

Philip Wicksteed's injection of linear homogeneous production function into the discussion of distribution theory opened a heated and prolonged controversy over the plausibility of the hypothesis that the production function will really take this form. According to Samuelson, whether there are any profits of exploitation left over for the capitalist to realise is really a matter of market conditions. Only in monopoly there will be profit in excess of other factor incomes. But, in perfect competition, long-run profit will be zero—since each factor is paid on the basis of its marginal product.

It was initially proposed by Leon Walras and then rediscovered by J.R. Hicks that whether or not the production function is linearly homogeneous in the vicinity of a competitive equilibrium point it must be locally linearly homogeneous, that is, all of its values and derivatives must be the same as those of a linearly homogeneous function. Thus, at that point all of the marginal products (the partial derivatives $\partial qi/\partial fi$) must coincide with those of a linearly homogeneous function and so they too must satisfy the Euler's theorem condition which says that marginal products add up to the total product. In this context, W.J. Baumol has suggested an explanation of why the production function must be locally linearly homogeneous in competitive equilibrium.

It may be noted that the simple function, $C = r_1 f_1 + r_2 f_2 + \ldots + r_n f_n \ldots$ (2), must be linearly homogeneous in the input quantities f_1 , f_2,... ,f_n , since if each f_i is multiplied by λ then C (total factor cost) will also be multiplied by λ and that is the implication of linear homogeneity. Now equation (2) is the total factor cost of the firm and its graph is the (hyper) plane through the origin, $C_aC_bC_cC_d$, in the two-input case shown in fig. 2. The shaded area (surface) of the diagram represents the production function (or, in this context, the value of output) PI = p*f*(L, K), in case of two variable factors. If the second order conditions hold at the point of equilibrium, T, the two surfaces must be tangent there, since the requirement of zero profit ensures that at no combination of inputs and outputs will the value of output exceed the cost of the corresponding inputs and at the equilibrium point the two will be the same. In fact, the tangency of the two surfaces at T means that they will have the same derivatives.

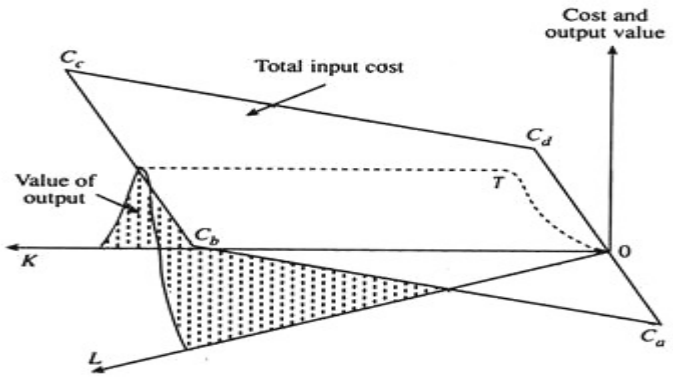


Fig. 2: Linearly Homogeneous Production Function and Product Exhausion

Alternatively stated, pf (L, K) must, indeed, be linearly homogeneous locally at T. Euler's theorem must, therefore, apply, and the payment to each factor on the basis of its marginal product must exhaust total product.

Criticisms:

No doubt, the marginal productivity theory is the basis of most theoretical discussions on the issue of distribution. However, with all its restrictive assumptions, most notably those of universal perfect competition and stationary equilibrium, it is not a very accurate representation of reality.

As Baumol has put it:

"What is claimed is that it describes a consistent mechanism which bears at least some resemblance to the workings of our economic institutions and that embodied within its general equilibrium relationship, there are forces which determine the payments going to labourers, capitalists, landlords, etc. It used to be thought that these complex relationship in fact followed certain simple patterns, at least, roughly, and that from these patterns one could safely formulate intuitive generations and draw conclusions relevant for policy." However, the members of the Cambridge School, such as N. Kaldor and P. Sraffa, contended that no such generalisations are possible. This means that any simple conclusions drawn from the general equilibrium models will encounter so many exceptions of such significance that they become untenable.

Thus, economists are left with the suspicion that the marginal productivity theory, with all its assumptions, is fundamentally valid but perhaps not so illuminating as one might wish. To tackle this problem, neoclassical economists have sought to aggregate large sectors of the marginal productivity model, permitting it to maintain its general equilibrium character but reducing its scope by restricting their analysis to two or three homogeneous inputs. To be more specific, models have been constructed containing only labour and capital, and certain qualitative conclusions have been derived from them. We may now discuss some macroeconomic models of distribution against this backdrop.

Macroeconomic Models of Distribution:

The macroeconomic models of distribution lump together large numbers of moderately diverse economic variables and relationship and treat the resulting aggregates as homogeneous economic elements. This is how manageable models involving small numbers of variables and relationships are derived. However, in the process of such aggregation there is wide abstraction from reality.

As Baumol has put it:

"The statement that the labour market is in equilibrium, when the total effective demand for labour equals the total supply, can conceal serious difficulties of oversupply in some industries and shortages in others. One must, therefore, seek fruit-lessness rather than vigour in a macroeconomic model. A completely formalistic macro-model is likely to be the worst of both the worlds because it is apt to offer neither empirical insights nor an accurate analytic mechanism.", Two macroeconomic models of distribution are the classical theory of David Ricardo and the Cambridge version of Nicholas Kaldor. These two theories differ from the marginal productivity theory on the ground that they address themselves to the burning issues of distribution theory, such as the magnitude of the income gap between the rich and the poor and its relationship to their role in the production process.

Ricardo's theory of distribution has four central components:

(i) Diminishing returns to labour working on a fixed supply of land

(ii) The theory of rent

(iii) The tendency of universal competition to equalize returns to investment

(iv) The Malthusian theory of population from which emerges the iron law of wages (i.e., actual wages will always tend to come back to the subsistence level due to population growth).

In Ricardo's model, society's output is distributed among, three main classes—landlords, workers and capitalists—in the form of rent, wages and profits. Rise in rent leads to a fall in wages and profits. That is why there is a clash of interests of the landlords on the one hand and that of workers and capitalists on the other. Ricardo's theory of distribution is illustrated in Fig. 3, which shows the behaviour of population, wages, rent and output in

the context of growth. The basic assumption of the model is that the ratio between the size of the population and that of labour force remains constant.

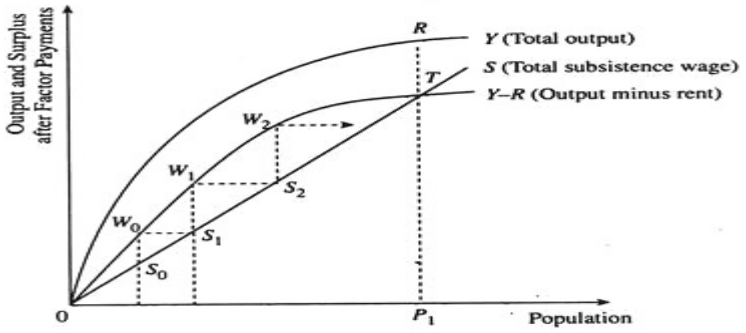


Fig. 3: Distribution of Income in Ricardo's Model

Total rent payments increase steadily with population growth and the consequent increase in the use of land. Thus, curve Y-R, that is, total output minus rent, also levels off as we move to the right, i.e., as population grows. Here Y-R is the amount of output left for distribution between wages and profits. Finally, the line OS shows how much output is required to pay every worker a fixed subsistence wage (due to the assumption of full employment in the classical model, the number of workers = the size of the population). Since the equation of this curve is S = sP, where P is the size of the population, this is a straight line through the origin.

Pattern of Income Distribution in the Process of Economic Growth:

Ricardo discussed the process of income distribution in the context of economic growth. Let us suppose that population is initially P_0 and that the rate of capital formation is initially so high that the level of wages is pushed up to a point where the whole of output after rent payment (W_0) is almost exhausted through wage payment. In such a situation the wage rate goes above the subsistence level, P_1S_1. This will encourage population to grow to P_1 at which the wage payment covers no more than subsistence P_1S_1. At this point profits will be high (S_1 W_1). This will induce increased accumulation which will raise the demand for labour and thus push total wage upwards once again, this time towards W_1. The process repeats itself, the economy moves towards point T, through the sequence of steps W_0S_1, W_1S_2, W_2,.

At point T, output after rent payment is just sufficient to pay subsistence wages. As the population approaches P_t, the level corresponding to point T, the economy approaches the stationary state. In this state, profits, capital accumulation and population growth remain zero forever, wage payments remain at the subsistence level and rent payments at maximum attainable level, TR. Thus, in the Ricardian model, workers gain very little, while capitalists lose during the process of economic growth. Only landlords' gain due to ever-increasing rents caused by the rise in the demand for land as population grows and inferior grades of land are brought under cultivation. Thus, the interests of landlords are diametrically opposed to those of workers and capitalists.

Kaldor Model:

The primary aim of Kaldor's macroeconomic model of distribution (which is based on the Keynesian income and employment model) is to analyse the share of wages in the total output of the society (national product). The model appeared in 1955. The Kaldor model is based on the crucial assumption that workers and capitalists have different propensities to save. This implies one thing, at least. Given the level of investment (at full employment) and total income, there will be only one proportion between workers' and capitalists' shares of national income, at which total saving will equal total investment, i.e., at which the total demand for output will equal its total supply.

Employment is a function of national output, Y. The level of employment $f(Y)$ times the wage rate, w, is the total wage bill $wf(Y)$. The residue, $Y- wf(Y)$ is the income that goes to other factor owners. Let us suppose, for the sake of simplicity, that there are two classes in the society—workers and capitalists (who represent the non-workers). Kaldor assumes that workers save a smaller proportion of their incomes (say s_1) than do capitalists (s_2). By assumption $s_1 < s_2$. Total desired saving will, thus, be equal to that of the workers, $s_1.wf(Y)$ plus that of the capitalists $s_2[Y- wf(Y)]$. If I is the fixed level of investment, equilibrium is attained when desired saving equals the level of investment, i.e., $S_1Wf(Y) + s_2[Y-wf(Y)] = I$ … (3), where I, s_1 and s_2 are assumed to be known constants. If we substitute

the full employment level of output Yf for Y, then the above equations becomes a single equation with one unknown, w, which can be solved for the equilibrium level of wages, w_e.

Policy Implication:

Kaldor's analysis has an interesting policy implication. Let us suppose, at some other wage rate, equilibrium national income is below the full employment level and that the employment function $f(Y)$, is independent of the level of wages. In such a situation a rise in wage level will not depress the demand for labour. On the contrary, it will transfer income from a group with low propensity to consume to a group with a high propensity to consume so that total effective demand and, hence, employment and the level of national income will tend to rise. This point may easily be proved.

Since consumption equals income minus saving, a rise in the wage rate by Δw will raise workers' consumption spending to $(1-s_1)(w + \Delta w)f(Y)$, and capitalists' consumption spending will now be $(1 - s_2)[Y - (w + \Delta w) f(Y)]$ so that total consumption demand will have changed from $(1 - s_1) wf(Y) + (1 - s_2)[Y - wf(Y)]$ to $(1 - s_1)(w + \Delta w) f(Y) + (1 - s_2)[Y-(w + \Delta w) f(Y)]$. By subtraction, we find that demand will have changed by $(s_2 - s_1) \Delta wf(Y)$, i.e., effective demand must have risen, since $s_1 < s_2$. Thus, the implication of the model is that during depression a wage rise is likely to be a good thing and may produce at least part of the income necessary to pay for it. It is interesting and at the same time a bit surprising that the payments of higher wages out of national income helps to produce the wherewithal to pay workers by increasing demand and thus raising revenues of business firms (producing units).

Another surprising implication of Kaldor's model is that capitalists can always increase their shares of income by increasing their consumption, i.e., by gradually reducing their saving rate, s_2, until it is less than that of s_1.

Let us suppose total desired saving was initially equal to investment. So after the fall in s_2 desired saving is less than investment. If s_2 is not much less than s_1 for a given transfer of income from workers to capitalists, then a given transfer will leave total saving virtually unchanged. So it will require a large transfer to the capitalists to enable the economy to reach full

employment equilibrium in which desired saving is again equated to investment. This point may now be proved. From equation (3), $s_1 wf(Y) + s_2[Y - w'(Y)J = I$, we get $(s_1 - s_2)\, wf(Y) + s_2 Y = I$, so that total wage earnings equal $wf(Y) = [I - s_2 Y)/(s_1 - s_2)$ and total profits $\pi = y - wf(y) = [Y(s_1 - s_2) - I + s_2 Y]/(s_1 - s_2) = (s_1 Y - I)/(s_1 - s_2)$. Since $s_1 < s_2$, by assumption, this will be positive if $I > s_1 Y$. Then as s_2 moves towards s_1, so that the value of the denominator falls, total profits must rise. Thus, capitalists will find that the more they spend the more they earn in the form of profits. So capitalists have access to their own resources.

Criticisms:

There are two main criticisms of the Kaldor model:

1. It is not at all clear why the economy in this model has an automatic tendency to approach the level of full employment
2. The promise that employment depends only on output and not on wage level denies that higher wages will induce the adoption of labour-saving inventions.

Conclusion:

Kaldor's model is not sufficiently comprehensive to show clearly how labour's share in national income is determined. Yet the model can be readily interpreted to suggest policy measures for changing the pattern of income distribution in a capitalist economy.

Unit-III

WELFARE ECONOMICS

Introduction

Market structure, in economics, refers to how different industries are classified and differentiated based on their degree and nature of competition for goods and services. It is based on the characteristics that influence the behaviour and outcomes of companies working in a specific market.

Market Structure

Some of the factors that determine a market structure include the number of buyers and sellers, ability to negotiate, degree of concentration, degree of differentiation of products, and the ease or difficulty of entering and exiting the market.

Understanding Market Structures

In economics, market structures can be understood well by closely examining an array of factors or features exhibited by different players. It is common to differentiate these markets across the following seven distinct features.

1. The industry's buyer structure
2. The turnover of customers
3. The extent of product differentiation
4. The nature of costs of inputs
5. The number of players in the market
6. Vertical integration extent in the same industry
7. The largest player's market share

By cross-examining the above features against each other, similar traits can be established. Therefore, it becomes easier to categorize and differentiate companies across related industries. Based on the above features, economists have used this information to describe four distinct types of market structures. They include perfect competition, oligopoly market, monopoly market, and monopolistic competition.

Types of Market Structures

1. Perfect Competition

The perfect competition occurs when there are a large number of small companies competing against each other. They sell similar products (homogeneous), lack price influence over the commodities, and are free to enter or exit the market. Consumers in this type of market have full knowledge of the goods being sold. They are aware of the prices charged on them and the product branding. In the real world, the pure form of this type of market structure rarely exists. However, it is useful when comparing companies with similar features. This market is unrealistic as it faces some significant criticisms described below.

A. No incentive for innovation: In the real world, if competition exists and a company holds a dominant market share, there is a tendency to increase innovation to beat the competitors and maintain the status quo. However, in a perfectly competitive market, the profit margin is fixed, and sellers cannot increase prices, or they will lose their customers.

B. There are very few barriers to entry: Any company can enter the market and start selling the product. Therefore, incumbents must stay proactive to maintain market share.

2. Monopolistic Competition

The Monopolistic competition refers to an imperfectly competitive market with the traits of both the monopoly and competitive market. Sellers compete among themselves and can differentiate their goods in terms of quality and branding to look different. In this type of competition, sellers consider the price charged by their competitors and ignore the impact of their own prices on their competition. When comparing monopolistic competition in the short term and long term, there are two distinct aspects that are observed. In the short term, the monopolistic company maximizes its profits and enjoys all the benefits as a monopoly. The company initially produces many products as the demand is high. Therefore, its Marginal

Revenue (MR) corresponds to its Marginal Cost (MC). However, MR diminishes over time as new companies enter the market with differentiated products affecting demand, leading to less profit.

3. Oligopoly

An oligopoly market consists of a small number of large companies that sell differentiated or identical products. Since there are few players in the market, their competitive strategies are dependent on each other. For example, if one of the actors decides to reduce the price of its products, the action will trigger other actors to lower their prices, too. On the other hand, a price increase may influence others not to take any action in the anticipation consumers will opt for their products. Therefore, strategic planning by these types of players is a must. In a situation where companies mutually compete, they may create agreements to share the market by restricting production, leading to supernormal profits. This holds if either party honors the Nash equilibrium state or neither is tempted to engage in the prisoner's dilemma. In such an agreement, they work like monopolies. The collusion is referred to as cartels.

4. Monopoly

In a monopoly market, a single company represents the whole industry. It has no competitor, and it is the sole seller of products in the entire market. This type of market is characterized by factors such as the sole claim to ownership of resources, patent and copyright, licenses issued by the government, or high initial setup costs. All the above characteristics associated with monopoly restrict other companies from entering the market. The company, therefore, remains a single seller because it has the power to control the market and set prices for its goods.

Meaning of pricing

The price is the value that is put to a product or service and is the result of a complex set of calculations, research and understanding and risk taking ability. A pricing strategy takes into account segments, ability to pay,

market conditions, competitor actions, trade margins and input costs, amongst others.

Price and Output Determination

The equilibrium conditions of an individual firm under perfect competition are to be studied under two heads. They are: 1. Short-Run Equilibrium 2. Long-Run Equilibrium.

1. Short-Run Equilibrium:

A short period has two major characteristics:

(a) Fixed production scale:

In the short period a firm cannot change its scale of production. It has to increase its supply by utilising fully the existing machinery and capital equipment. So, it has to increase fixed and variable costs in such a period.

(b) No free entry or free exit of firms:

In the short period no new firms can enter into the industry, nor can an old firm go out of the industry. So, the number of firms in the industry remains the same.

Conditions of a short-run competitive equilibrium:

In such a short period there are two major conditions of equilibrium of a firm:

(a) P = SMC:

The total profits of a firm become maximum at the output where marginal cost is equal to marginal revenue. Under perfect competition a firm is to sell all the units of its output at the same market price. For this reason market price becomes equal to a firm's marginal revenue in this type of market.

So, it follows that the total profits of a competitive firm become maximum at the output level at which P=MC. So a competitive firm produces an output up to this level. It may be noted that in equilibrium position the marginal cost must be a rising one; otherwise there cannot be any stable equilibrium of a competitive firm.

(b) The Short-Run P and SAC—(Break-Even and Shut-Down Points):

In the short run, the competitive price may be greater than SAC, creating excess profits as no new firms can enter the industry. Sometimes it may be

equal to AC and then the firm earns just normal profits, and this point (P=SMC=SAC) is called the break-even point as it equals TR with TC. Again, sometimes the competitive price may even be less than average cost yielding loss to the firm, provided it is greater than, or equal to, average variable cost.

If the market price is just equal to the average variable cost, the firm reaches what is called the shutdown point. Unless the position improves; or if the price goes below it, the firm will have to close down at this point. The short-run equilibrium output and price of a competitive firm is illustrated in Fig. 4.

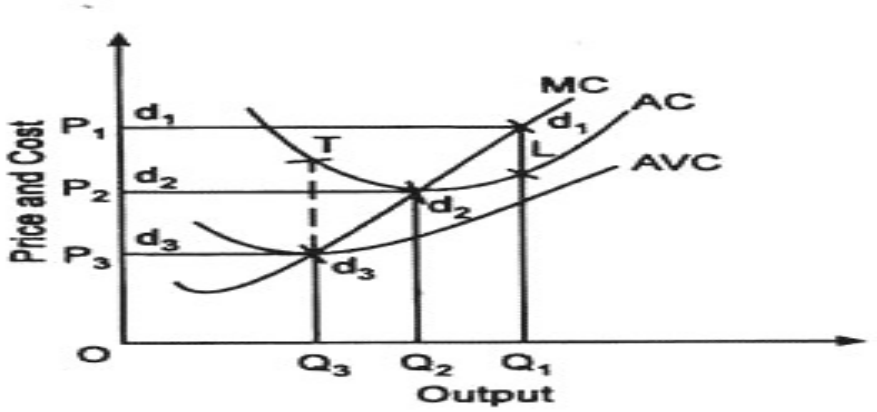


Fig. 4 : Price Cost and Output

It shows that at price OP_1 (the demand curve being d_1d_1) the competitive firm produces OQ_1 units of output because at this output level the price (OP_1 or Q_1d_1) is equal to the marginal cost (Q_1d_1). Here the price is greater than the average cost (Q_1d_1), creating an excess profit (Ld_1) is possible in the short run as no new firms can enter into the industry. If the price is lower, i.e., OP_2 (the demand curve being d_2d_2), the firm produces OQ_2 units of output, because at this output now the price (OP_2 or Q_2d_2) is equal to marginal cost (Q_2d_2).

Here the price is also equal to the average cost (Q_2d_2), giving the firm only normal profits. Such a point (P=SRMC=SRAC) is known as the break-even point (the point without either loss or profit, the normal profit being ignored). Again, if the price is still lower at OP_3 (the price line being d_3d_3), the firm would produce OQ_3 because at this output the present price (OP_3 or Q_3d_3) is equal to the marginal cost (Q_3d_3).

But, here the price is less than the average cost (Q_3T) creating a loss for the firm. Such a price is possible in the short run as it fully covers the average variable cost (Q_3d_3). The point (P=AVC) is called the shut-down point; the firm will shut down its business, unless the position improves or if the price falls below this short-run minimum limit of the average variable cost.

2. Long-Run Equilibrium:

A long period has also two characteristics:

(a) Variable Scale of Production:

In the long run a firm can change it scale of production. So, it has to increase only variable costs in the long run.

(b) Free Entry or Free Exit of Firms:

In the long run new firms may enter the industry or the existing firms can go out of it, depending on profits and losses.

Conditions of the long-run equilibrium of a competitive firm:

In such a long period there are also two important conditions of equilibrium of a competitive firm.

(a) P = LMC:

Even in the long run a competitive firm would produce an output up to that level at which price is equal to marginal cost which must be raising one. Although the equality of P and MC is an essential condition of the long-run equilibrium, it is not a sufficient one; another important condition is to be fulfilled for a stable long-run equilibrium.

(b) P = LAC (minimum):

The second condition is that the long-run competitive price must also be equal to average cost. It is obvious due to the fact that in the long run the total revenue of a competitive firm must cover its total cost. If the long-run price is greater than the average cost, the firm will make excess profits, which will attract new firms into the industry.

For this reason, the number of firms will increase, the total supply will be larger and the market price will fall to the level of average cost. So, the firm under consideration will incur a loss; as a result some firms will leave the industry, the number of firms will be smaller, the market supply will fall

and the market price will rise again to the level of average cost. Thus, it is clear that the long-run competitive price must equal both marginal cost and average cost, and the firm is then able to make only normal profits. At this stage the average cost is the minimum (at the point of equality of P and LAC). So, in long-run equilibrium the competitive price becomes equal to the minimum average cost.

Hence, the condition of the long- run equilibrium of a competitive firm becomes:

P=LMC=LAC (minimum).

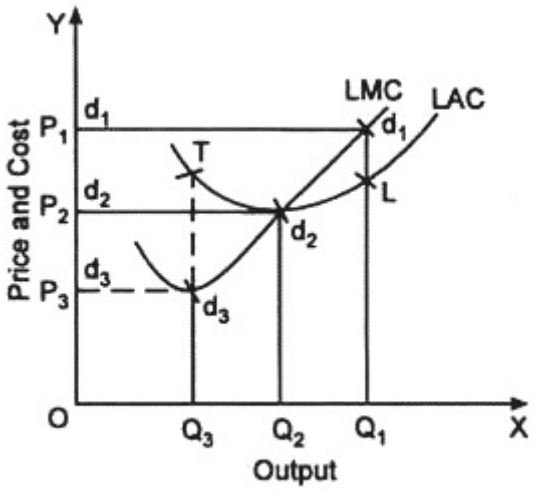


Fig. 5 : Long-run equilibrium

The long-run equilibrium of a competitive firm is shown in Fig. 5. Here at the output level OQ_1 the price OP_1 is equal to the marginal cost (Q_1d_1). This price cannot be stable as it is higher than the average cost (Q_1L). This price will attract new firms into the industry, causing an increase in market supply and a consequent fall in market price. If, on the other hand, market price is lower at OP_3, the firm produces OQ_3 units of output where the price is equal to the marginal cost (Q_3d_3).

This price also cannot be stable as it is lower than the average cost (Q_3T). But when the price is OP_2, the firm produces OQ_2. This price becomes stable as it is equal to both marginal and average costs (Q_2d_2). Here the average cost is the minimum. So, a competitive firm attains the long-run stable equilibrium position at the point of intersection of the three curves

(d_2d_2, LMC and LAC), i.e., at the output level at which P=LMC=LAC (minimum).

What is 'Perfect Competition'

Definition: Perfect competition describes a market structure where competition is at its greatest possible level. To make it more clear, a market which exhibits the following characteristics in its structure is said to show perfect competition:

1. Large number of buyers and sellers
2. Homogenous product is produced by every firm
3. Free entry and exit of firms
4. Zero advertising cost
5. Consumers have perfect knowledge about the market and are well aware of any changes in the market. Consumers indulge in rational decision making.
6. All the factors of production, viz. labour, capital, etc, have perfect mobility in the market and are not hindered by any market factors or market forces.
7. No government intervention
8. No transportation costs
9. Each firm earns normal profits and no firms can earn super-normal profits.
10. Every firm is a price taker. It takes the price as decided by the forces of demand and supply. No firm can influence the price of the product.

Description: Ideally, perfect competition is a hypothetical situation which cannot possibly exist in a market. However, perfect competition is used as a base to compare with other forms of market structure. No industry exhibits perfect competition in India.

Summary

- **A perfectly competitive market is defined by both producers and consumers being price-takers. Price-takers are unable to affect the market price because they lack substantial market share.**
- **The three primary characteristics of perfect competition are (1) no company holds a substantial market share, (2) the industry output is standardized, and (3) there is freedom of entry and exit.**
- The efficient market equilibrium in a perfect competition is where marginal revenue equals marginal cost.

Price-Takers

Price-takers are market participants that are unable to affect the market price of goods through their production and consumption decisions. The two types of price-takers are:

1. Price-taking producers

A price-taking producer is a producer that cannot affect the market price of the product or service they are selling.

2. Price-taking consumer

A price-taking consumer is a consumer that cannot affect the market price of a good or service.

Prerequisites of Perfect Competition

1. No individual firm possesses a substantial market share

For an industry to be perfectly competitive, no individual producers must have a large market share. Market share is the proportion of the total industry's output that belongs to a single firm.
For example, consider the wheat market. Many farmers grow wheat, and market share is dispersed among them. There are no farmers that could potentially affect the price of wheat on the market.

2. The industry output is a standardized product

Perfect competition can only occur when consumers perceive the products of all producers to be equivalent. Therefore, it can only occur when the industry output is a commodity, otherwise known as a standardized product.
Since standardized products are homogenous, a single producer cannot increase the price of their good or service without losing all sales to the

competition. It implies that price-taking firms face perfect price-elasticity of demand.

3. Freedom of entry and exit

The majority of perfectly competitive industries allow firms to easily enter and exit the industry. The arrival of new firms into an industry is referred to as market entry. Market entry is enabled by the absence of obstacles posed by government regulation or low start-up costs. The departure of firms out of an industry is referred to as a market exit. Firms can easily exit the market if there are no additional costs attributable to shutting down the business. For example, consider the mining industry. In the mining industry, firms must recognize an Asset Retirement Obligation (ARO) to restore the property to its previous state after the desired metals are extracted. An ARO refers to a liability that is amortized throughout the investment horizon and exemplifies an exit cost for mining firms.

Optimal Production Output in a Perfect Competition

In order for firms to generate maximum profits, they must determine their optimal output to produce. In a perfect competition, firms produce an output quantity where the marginal cost of the last unit produced is equal to the marginal revenue of the product. For a price-taking firm, the marginal revenue is equal to the market price. It is because no firm can affect the market price; therefore, the additional revenue generated by producing one more unit is the market price. Consequently, an individual firm faces a perfectly elastic demand curve. The price-taking firm's demand curve is equal to its marginal revenue. The demand and marginal revenue curve can be illustrated by a horizontal line drawn at the market price.

Equilibrium of the Firm and Industry under Perfect Competition

1. Meaning of Firm and Industry
2. Equilibrium of the Firm
3. Equilibrium of the Industry under Perfect Competition

Meaning of Firm and Industry:

It is essential to know the meanings of firm and industry before analysing the two. A firm is an organisation which produces and supplies goods that are demanded by the people. According to Prof. S.E. Lands-bury, "Firm is an organisation that produces and sells goods with the goal of maximising its profits. In the words of Prof. R.L. Miller, "Firm is an organisation that buys and hires resources and sells goods and services."

Industry is a group of firms producing homogeneous products in a market. In the words of Prof. Miller, "Industry is a group of firms that produces a homogeneous product." For example, Raymond, Maffatlal, Arvind, etc., are cloth manufacturing firms, whereas a group of such firms is called the textile industry.

Equilibrium of the Firm:

Meaning:

A firm is in equilibrium when it has no tendency to change its level of output. It needs neither expansion nor contraction. It wants to earn maximum profits. In the words of A.W. Stonier and D.C. Hague, "A firm will be in equilibrium when it is earning maximum money profits." Equilibrium of the firm can be analysed in both short-run and long-run periods. A firm can earn the maximum profits in the short run or may incur the minimum loss. But in the long run, it can earn only normal profit.

Short-run Equilibrium of the Firm:

The short run is a period of time in which the firm can vary its output by changing the variable factors of production in order to earn maximum profits or to incur minimum losses. The number of firms in the industry is fixed because neither the existing firms can leave nor new firms can enter it.

It's Conditions:

The firm is in equilibrium when it is earning maximum profits as the difference between its total revenue and total cost.

For this, it essential that it must satisfy two conditions:

(1) MC = MR, and (2) the MC curve must cut the MR curve from below at the point of equality and then rise upwards. The price at which each firm sells its output is set by the market forces of demand and supply. Each firm will be able to sell as much as it chooses at that price. But due to competition, it will not be able to sell at all at a higher price than the market price. Thus the firm's demand curve will be horizontal at that price so that P = AR = MR for the firm.

1. Marginal Revenue and Marginal Cost Approach:

The short-run equilibrium of the firm can be explained with the help of the marginal analysis as well as with total cost-total revenue analysis. We first take the marginal analysis under identical cost conditions.

This analysis is based on the following assumptions:

1. All firms in an industry use homogeneous factors of production.
2. Their costs are equal. Therefore, all cost curves are uniform.
3. They use homogeneous plants so that their SAC curves are equal.
4. All firms are of equal efficiency.
5. All firms sell their products at the same price determined by demand and supply of the industry so that the price of each firm is equal to AR = MR.

Determination of Equilibrium:

Given these assumptions, suppose that price OP in the competitive market for the product of all the firms in the industry is determined by the equality of demand curve D and the supply curve S at point E in Figure 1(A) so that their average revenue curve (AR) coincides with the marginal revenue curve (MR). At this price, each firm is in equilibrium at point L in Panel (B) of the figure where (i) SMC equals MR and AR, and (ii) the SMC curve cuts the MR curve from below. Each firm would be producing OQ output and earning normal profits at the maximum average total costs QL. A firm earns normal profits when the MR curve is tangent to the SAC curve at its minimum point.

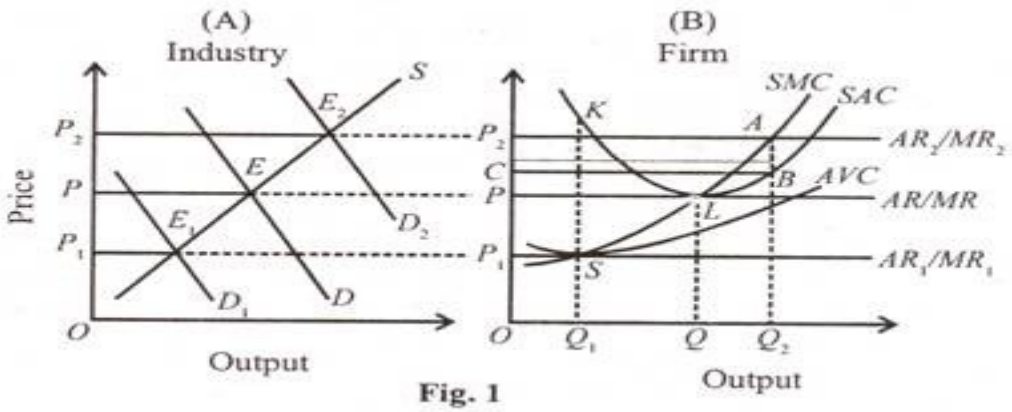


Fig. 1

If the price is higher than these minimum average total costs, each firm will be earning supernormal profits. Suppose the price rises to $0P_r$ where the SMC curve cuts the new marginal revenue curve MR_2 (=AR_2) from below at point A which now becomes the equilibrium point. In this situation, each firm produces OQ_2 output and earns supernormal profits equal to the area of the rectangle P_2 ABC.

If the price falls below $OP_1$the firm would make a loss because the SAC would be higher than the price. In the short-run, it would continue to produce and sell OQ_1 output at OP_1price so long as it covers its AVC. S is thus the shut-down point at which the firm is incurring the maximum loss equal to SK per unit of output. If the price falls below OP_1 the firm will close down because it would fail to cover even the minimum average variable cost. OP_1 is thus the shut-down price.

We may conclude from the above discussion that in the short-run each firm may be making either supernormal profits, or normal profits or losses depending upon the price of the product.

2. Total Cost Revenue Analysis:

The short-run equilibrium of the firm can also be shown with the help of total cost and total revenue curves. The firm is able to maximize its profits at that level of output where the difference between total revenue and total cost is the maximum. This is shown in Figure 2 where TR is the total revenue curve and TC total cost curve. The total revenue curve is an upward sloping straight line curve starting from O. This is because the firm sells

small or large quantities of its product at a constant price under perfect competition. If the firm produces nothing, total revenue will be zero. The more it produces, the larger is the increase in total revenue. Hence the TR curve is linear and slopes upward.

The firm will maximize its profits at that level of output where the gap between the TR curve and the 1C curve is the maximum. Geometrically, it is that level at which the slope of a tangent drawn to the total cost curve equals the slope of the total revenue curve. In Figure 2, the maximum amount of profit is measured by TP at OQ output. At outputs smaller or larger than OQ between A and B points, the firm's profits shrink. If the firm produces OQ_1 output, its losses are the maximum because the TC curve is i above the TR curve. At Q_1 its profits are zero. Similar situation prevails at Q_2.

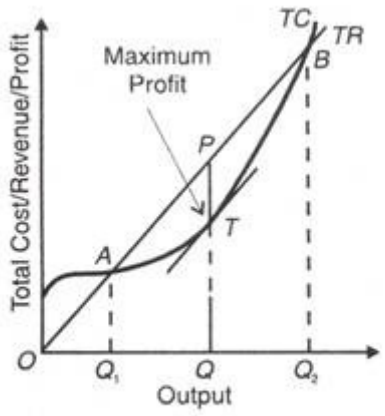


Fig. 2

Since the marginal revenue equals the slope of the total o revenue curve and the marginal cost equals the slope of the tangent to the total cost curve, it follows that where the slopes of the total cost and revenue curves are equal as at P and T, the marginal cost equals the marginal revenue. It should be clear of that the point of maximum profits lies in the region of rising marginal cost (when TC is below TR) and of maximum loss in the falling marginal cost region (where TC is above TR).

The explanation of the equilibrium of the firm by using total cost-revenue curves does not throw more light than is provided by the marginal cost-marginal revenue analysis. It is useful only in the case of certain marginal decisions where the total cost curve is also linear over a certain

range of output. But it makes the equilibrium of the firm a cumbersome and difficult analysis particularly when one has to compare the change in cost and revenue resulting from a change in the volume of output. Further, maximum profits cannot be known at once. For this, a number of tangents are required to be drawn which is a real difficulty.

Long-run Equilibrium of the Firm:

In the long-run, it is possible to make more adjustments than in the short-run. The firm can adjust its plant capacity and scale of operations to the changed circumstances. Therefore, all costs are variable. Firms must earn only normal profits. In case the price is above the long-run AC curve firms will be earning supernormal profits. Attracted by them, new firms will enter the industry and supernormal profits will be competed away. If the price is below the LAC curve firms will be incurring losses. As a result, some of the firms will leave the industry so that no firm earns more than normal profits. Thus "in the long-run firms are in equilibrium when they have adjusted their plant so as to produce at the minimum point of their long-run AC curve, which is tangent to the demand (AR) curve defined by the market price" so that they earn normal profits.

This analysis is based on the following assumptions:

1. Firms are free to enter into or leave the industry.
2. All firms are of equal efficiency.
3. All factors are homogeneous. They can be obtained at constant and uniform prices.
4. Cost curves of firms are uniform.
5. The plants of firm: are equal having given technology.
6. All firms have perfect knowledge about price and output.

Determination:

Given these assumptions, each firm of the industry will be in the following two conditions. (1) In equilibrium, its short-run marginal cost (SMC) must equal to its long-run marginal cost (LMC) as well as its short-run average cost (SAC) and its long-run average cost (LAC) and both should be equal to MR=AR=P. Thus the first equilibrium condition is:

SMC = LMC = MR = AR = P = SAC = LAC at its minimum point, and (2) LMC curve must cut MR curve from below.

Both these conditions of equilibrium are satisfied at point E in Figure 3 where SMC and LMC curves cut from below SAC and LAC curves at their minimum point E and SMC and LMC curves cut AR = MR curve from below. All curves meet at this point E and the firm produces OQ optimum quantity and sell it at OP price.

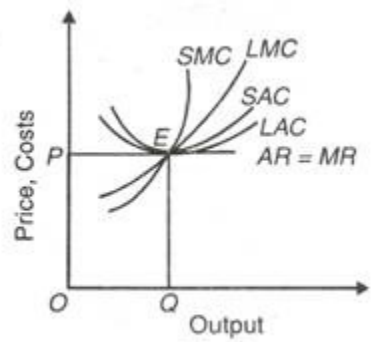


Fig. 3

Since we assume equal costs of all the firms of industry, all firms will be in equilibrium m the long-run. At OP price a firm will have neither a tendency to neither leave nor enter the industry and all firms will earn normal profit.

Equilibrium of the Industry under Perfect Competition:

An industry is in equilibrium:

(i) When there is no tendency for the firms either to leave or enter the industry, and (ii) when each firm is also in equilibrium. The first condition implies that the average cost curves coincide with the average revenue curve of all the firms in the industry. They are earning only normal profits, which are supposed to be included in the average cost curves of the firms. The second condition implies the equality of MC and MR. Under a perfectly competitive industry, these two conditions must be satisfied at the point of equilibrium, i.e.,

SMC = MR

SAC = AR

P = AR = MR

SMC = SAC = AR = P

Such a situation represents full equilibrium of the industry.

Short-Run Equilibrium of the Industry:

An industry is in equilibrium in the short run when its total output remains steady, there being no tendency to expand or contract its output. If all firms are in equilibrium, the industry is also in equilibrium. For full equilibrium of the industry in the short run, all firms must be earning only normal profits. The condition for this is SMC = MR = AR = SAC. But full equilibrium of the industry is by sheer accident because in the short run some firms may be earning supernormal profits and some incurring losses.

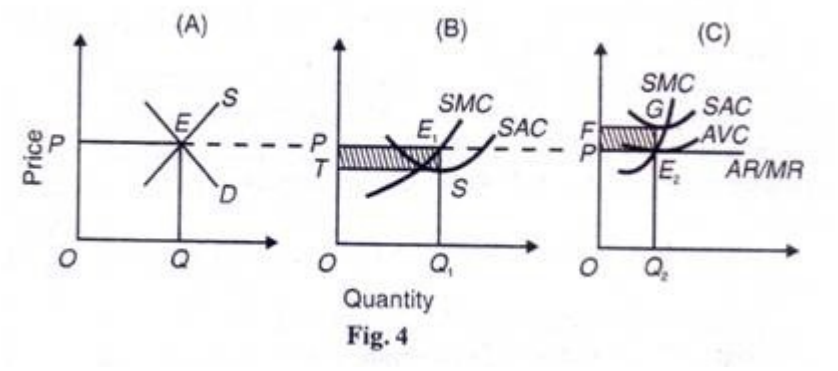


Fig. 4

Even then, the industry is in short- run equilibrium when its quantity demanded and quantity supplied are equal at the price which clears the market. This is illustrated in Figure 4, where in Panel (A), the industry is in equilibrium at point E where its demand curve D and supply curve S intersect which determine OP price at which its total output OQ is cleared. But at the prevailing price OP some firms are earning supernormal profits PE_1ST as shown in Panel (B), while some other firms are incurring FGE_2P losses as shown in Panel (C) of the figure.

Long-Run Equilibrium of the Industry:

The industry is in equilibrium in the long run when all firms earn normal profits. There is no incentive for firms to leave the industry or for new firms to enter it. With all factors homogeneous and given their prices and the same technology, each firm and industry as a whole are in full

equilibrium where LMC = MR =AR(=p) =LAC at its minimum. Such an equilibrium position is attained when the long-run price for the industry is determined by the equality of total demand and supply of the industry.

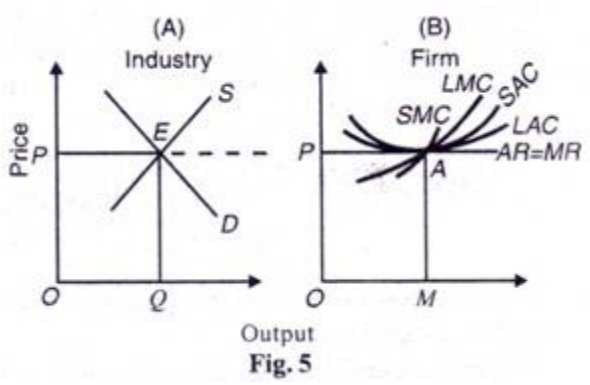


Fig. 5

The long-run equilibrium of the industry is illustrated in Figure 5(A) where the long-run price op and OQ output are determined by the intersection of the demand curve d and the supply curve s at point E. At this price op, the firms are in equilibrium at point A in Panel (B) at OM level of output where LMC =SMC= MR= p (=AR) =SAC= LAC at its minimum. At this level, the firms are earning normal profits and have no incentive to enter or leave the industry. It follows that when the industry is in long-run equilibrium, each firm in the industry is also in long-run equilibrium. If both the industry and the firms are in long-run equilibrium, they are also in short-run equilibrium.

Even though all firms in a perfectly competitive industry in the long run have the same cost curves, the firms can be of different efficiency. Firms using superior resources or inputs such as superior management must pay them higher rewards; otherwise they will shift to new firms which offer them higher prices. So the forces of competition will force the more efficient firms to pay superior resources higher prices at their opportunity cost. As a result, the lac curve of the more efficient firms will shift upwards and they will benefit in the form of higher output at the higher long-run equilibrium price set by the industry.

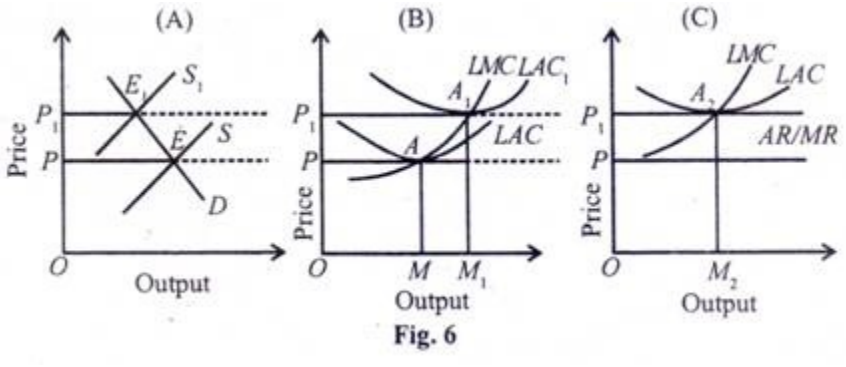


Fig. 6

Unable to pay higher prices to resources or inputs, less efficient firms will be competed away. New firms which are able to pay more and attracted by the new higher market price will enter the industry. But at the new long-run equilibrium price of the industry, all firms will be producing at the minimum LAC. This is illustrated in Figure 6 where the industry is in initial equilibrium at point E with price OP m Panel (A) and the more efficient firms like all other firms are in equilibrium at point A in Panel (B). As the industry is in equilibrium, the new firms do not exist as they are not in a position to cover their costs at OP price.

When the more efficient firms pay higher prices to resources or inputs, their LAC curve rises to LAC_1 At the new long-run equilibrium price of the industry set at OP $_1$ the more efficient firms are in equilibrium where $P_1 = LAC_1$ at its minimum point A_1 in Panel (B). They are now producing larger output OM_1 even though they earn normal profits. The new firms also earn normal profits at point A_2, as shown in Panel (C). But they produce less output OM_2 than OM_1 produced by the more efficient firms.

Supply Curve of a Firm and Industry: Short-Run and Long-Run Supply Curve

Supply curve indicates the relationship between price and quantity supplied. In other words, supply curve shows the quantities that a seller is willing to sell at different prices. According to Dorfman, “Supply curve is that curve which indicates various quantities supplied by the firm at different prices”. The concept of supply curve applies only under the conditions of perfect competition.

Supply curve can be divided into two parts as:

A. Short Run Supply Curve

B. Long Run Supply Curve

A. Short Run Supply Curve

(i) Short Run Supply Curve of a Firm:

Short run is a period in which supply can be changed by changing only the variable factors, fixed factors remaining the same. That way, if the firm shuts down, it has to bear fixed costs. That is why in the short run, the firm will supply commodity till price is either greater or equal to average variable cost. Thus a firm will continue supplying the commodity till marginal cost is equal to price or average revenue. Under perfect competition average revenue is equal to marginal revenue, so the firm will produce up to that point where marginal revenue and marginal cost are equal.

Short run supply curve of a perfectly competitive firm is that portion of marginal cost curve which is above average variable cost curve. According to C.E. Ferguson, "The short run supply curve of a firm in perfect competition is precisely its Marginal Cost Curve for all rates of output equal to or greater than the rate of output associated with minimum average variable cost."

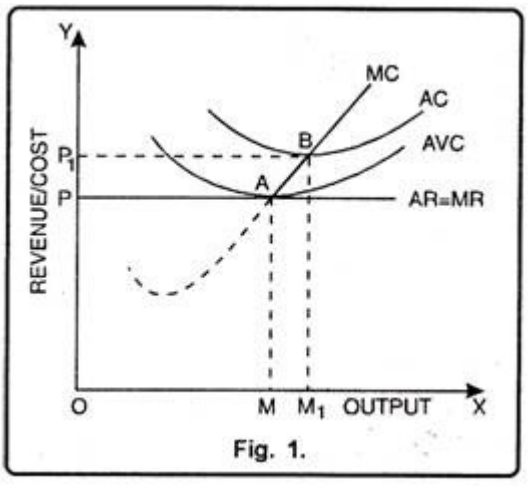


Fig. 1.

Prof. Bilas has defined it in simple words, "The Firm's short period supply curve is that portion of its marginal cost curve that lies-above the minimum

point of the average variable cost curve." However, short run supply curve of a firm can be shown with the help of fig. 1. From fig. 1 it is clear that there is no supply if price is below OP. At price less than OP, the firm will not be covering its average variable cost. At OP price, OM is the supply. In this case, firms' marginal revenue and marginal cost cut each other at A, OM is equilibrium output. If price goes up to OP1, the firm will produce OM1 output. This firm's short run supply curve starts from A upwards i.e., thick line AB.

(ii) Short Run Supply Curve of an Industry:

An industry is a blend of firms producing homogeneous goods. That way, supply curve of an industry is a lateral summation of all firms. This can be made clear with the help of a Fig. 2.

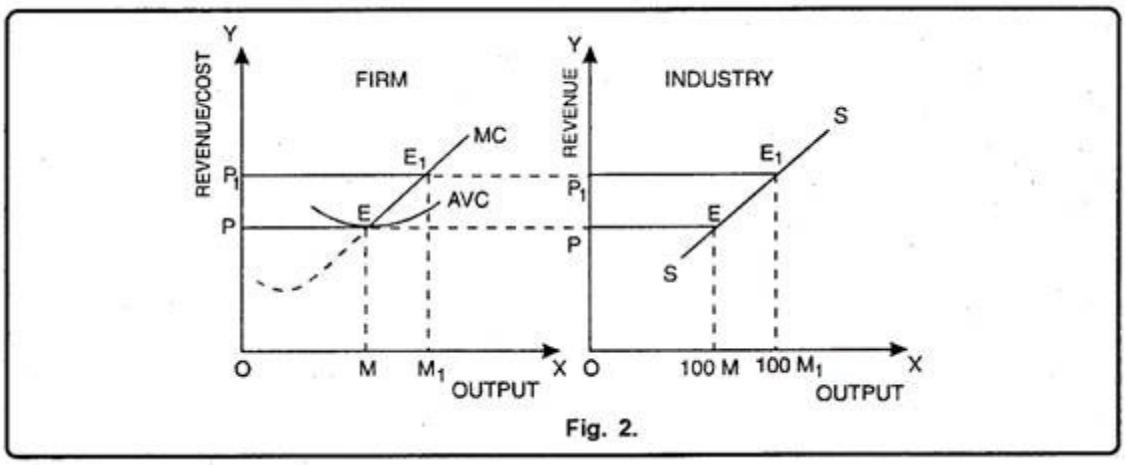


Fig. 2.

Here, we have assumed that different firms in the industry are producing identical products. Each firm at OP price is producing OM output. It is because all firms have identical costs. At OP price, supply of industry is 100 x M = 100M. Similarly at OP_1 price, all the firms of industry are producing 100 xM_1 =100M_1 quantity of output. These quantities will be called supply or output of industry. SS is the supply curve of industry. Point E shows that at OP price firm's supply is OM and an industry's total supply is 100 × M = 100M. At OP_1 price, firm's supply is OM1 and industry's supply is 100M). We get industry's supply curve by joining points E and E_1.

Thus, under perfect competition, lateral summation of that part of short run marginal cost curves of the firms which lie above the average

variable cost constitutes the supply curve of the industry. According to Stonier and Hague, "short run supply curve of a competitive industry will always slope upwards since the short run marginal cost curve of the industrial firms always slope upward."

B. Long Run Supply Curve:

Long run supply curve can also be analyzed from firm and industry's point of view:

1. Long Run Supply Curve of a Firm:

Long run is a period in which supply can be changed by changing all the factors of production. There is no distinction between fixed and variable factors. In the long run, firm produces only at minimum average cost. In this situation, long run marginal cost, marginal revenue, average revenue and long run average cost are equal i.e., LMCMR (= AR)LAC (minimum). The firm is enjoying only normal profits. So that position of marginal cost curve will determine the supply curve which is above the minimum average variable cost. The point where minimum average cost is equal to marginal cost is called optimum production. Thus Long Run Supply Curve of a firm is that portion of its marginal cost curve that lies above the minimum point of the average cost curve.

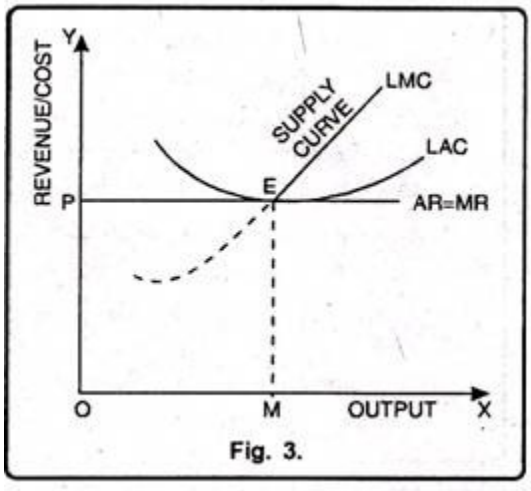


Fig. 3.

In figure 3 the firm is in equilibrium at point E where MRLMC (=AR). AC is minimum corresponding to this point. This point E is also called

optimum point because at this point MR=LMCAR minimum LAC. That portion of LMC which is above E is called long run supply curve.

2. Long Run Supply Curve of an Industry:

In the long run, industry's supply curve is determined by the supply curve of firms in the long run. Long run supply curve in the long run is not lateral summation of the short run supply curves. Industry's long run supply curve depends upon the change in the optimum size of firms and change in the number of firms.

It is on account of two reasons:

(i) In the long run, firms continue to enter into and exit from the industry,

(ii) Firms get economies and diseconomies of scale. This displaces the long run marginal cost (LMC). Due to these reasons, long run supply curve of industry is not the lateral summation of supply curve of firms. In reality, long run supply curve of industry can be known from the long run optimum production of firms multiplied by the number of firms in an industry. LRS_i, = Q x N, Where LRS_1 is long run supply curve of industry. Q is the optimum output of a firm and N, the number of firms.

Break Even Analysis – Definition, Formula, Examples

Break Even Analysis is one of the financial tools that help in the calculation of the margin of safety of a new company or new product. In this article, you will learn about Break Even Analysis in detail and will also learn how to use Break Even Analysis to determine the profitability of your company with the help of an example. A new company blossom every other day. Do you think people who start these companies work on a hunch? Or whether they put their money in these businesses after thorough planning and analysis? How they make sure that their companies remain profitable? These are all the questions that come into the mind of a person who is naïve in the business world. To answer these questions, we study one of the financial tools that help professionals to make calculated decisions.

Definition

Break Even Analysis can be defined as a calculation of point where revenue generation is equal to the cost incurred in the production. Break-even point is a point where a company is neither making profits nor is losing any money. Break-Even Analysis is a financial tool used by companies to determine at what point they will start making profits on entering a new market or launching a new product. The term “Break-even” is used to refer to a situation where a company is neither making any profits nor losing any money. That means whatever business they do is enough to cover all types of costs incurred in the business. Companies’ use Break Even Analysis for the calculation of the exact number of sales that the company is required to make to cover all the costs. In a Break Even Analysis, you study the relationship between the revenue generated, fixed costs, and variable costs and use this information to make important decisions about your company.

The Break Even Analysis tool is used not only for industrial purposes but also used by financial planners, marketers, managers, accountants, and entrepreneurs. Managers can use this tool for setting goals for their subordinates to achieve the required sales goal to generate profit. If a manager knows exactly how many sales they need to make, then he can push his employees to put efforts accordingly.

Break Even Analysis helps you in seeing the full picture of your business. Having clear information allows managers to decide what methods should be adopted to meet the goal. For example: If you are planning for a seasonal sale, you can determine the discount you can offer on your products by still making a profit using the Break Even Analysis. The value of break-even point is different for different businesses. For example, the break- even point for a company will be high if its initial fixed cost and variable cost are high. On the other hand, if the fixed cost of a company is zero, then it might reach its break-even point just after selling its first unit.

Break Even Analysis Formula

The formula of Break Even Analysis can be calculated by dividing the total fixed cost with a contribution per unit. The contribution per unit can

be calculated by subtracting the variable cost per unit from selling price per unit. The following are the formulas of break-even point and contribution per unit in equation form.

Contribution per unit = Selling Price per Unit – Variable Cost per Unit

And the break-even point can be calculated using the following formula.

Break-even Point = Fixed Cost / Contribution per Unit

OR

Break-even Point = Fixed Cost/ (Selling Price per Unit – Variable Cost per Unit)

Calculate the values such as fixed costs, which can be obtained by adding all the fixed expenses and contribution per unit and put those values in the above formula to get the break-even point for your business.

Components in Break Even Analysis Calculation

1. Fixed Costs

Fixed costs can be defined as the business costs, which are directly related to the business but not directly associated with the level of production. Therefore, whether your production level is zero or at its highest capacity, the fixed costs are going to be there. For example, you are supposed to pay the rent of your factory building, whether there is no production going on for about a month. Fixed costs can also be referred to as overhead costs. These costs start as soon as you set up your business or production unit. These costs remain the same, whether your business is growing or going backward. However, your long-term fixed costs change when you decide to expand your business. For instance, when you set up a new production unit.

The followings are examples of fixed costs.

1. Taxes, 2. Salaries and wages, 3. Rent of the building or lease charges, 4. Energy cost, 5. Depreciation cost, 6. Marketing costs, 7. Research and development expenses, 8. Administration cost

2. Variable Costs:

Variable costs are the costs that are directly associated with the level of production. That means the variable cost will reduce with the reduction in the production and will become zero when you cease the production process. For example, the cost of raw material required for the production of goods is directly related to the number of units produced in the production process. The variable cost can be divided into two types, such as direct variable cost and indirect variable cost. Direct variable cost: Direct variable costs are those costs that are directly related to the production of a particular product or a specific production centre.

The followings are examples of direct variable costs.

1. Cost of raw material, 2. Cost of wages of workers hired, especially for production work, 3. Fuel consumed 4. packaging cost

2. Indirect variable cost

Direct variable costs are the costs that are directly associated with the production of goods but does not get affected by the level of production. For example, depreciation cost, machine maintenance cost, and labour cost. That means you are required to pay same day wages to labours whether you are producing 200 units a day or 500 units a day. The same is true in the case of the maintenance cost of machines and production units.

3. Semi variable cost

Semi variable costs are the costs that have characteristics of both variables as well as fixed costs. Initially, these costs are fixed, but later these costs vary with the expansion of business or with the complex nature of the business.

Break Even Analysis

Break Even Analysis can be used by managers and accountants at any time to get an idea of total sales required to make to generate profit. However, it is crucial to use Break Even Analysis before you do any of the following.

1. Setting up a new business

Break Even Analysis is essential to be performed before investing money in new business. Using Break Even Analysis, you can decide whether your business is realistic or not. You will get a realistic idea of investing in the business. In addition to this, you can use information obtained from Break Even Analysis to prepare an effective pricing strategy to avoid loss.

2. When you decide to change your business model

Changing the business model of your business is quite similar to getting into a new business. You might have a few resources from the previous business, but you are required to consider other factors that might increase your expenses. For example, if you are planning to switch from a wholesale business to a retail business then a Break Even Analysis will be helpful for you. Because using that information, you can plan a new price strategy to stay in profit.

3. While launching a new product

A Break Even Analysis will be helpful when you choose to start a new product even when you are running a successful business. A Break Even Analysis will help you in determining whether you should invest in a new product or not especially when the expenditure is too high.

4. Before starting promotional activities:

Many businesses make the use of promotional activities to increase their sales. But if you do not well-plan your promotional activity and well-calculate the expenditure on the activity and calculate expected a return on investment, then you might end up adding an additional financial burden rather than making profits. A Break Even Analysis will provide you with a rough idea of whether your promotional activity will be beneficial for you or not.

For example, if you are planning to start a sale in your store for the festive season, the Break Even Analysis will help to calculate the right discount you can offer and remain in profit.

Benefits of Break Even Analysis

The followings are the benefits of using Break Even Analysis.

1. Break Even Analysis helps price your product efficiently. You can use this method to give the best price to your product without increasing the current price abruptly.
2. Break Even Analysis helps in covering all the fixed costs. Knowing all fixed costs is vital to calculate the profit generated from the business.
3. Using the information obtained from Break Even Analysis, you can make smart and calculated decisions about business investments rather than investing money in a business instinctively.
4. Break Even Analysis can help you in setting a revenue target. When you know the number of units you need to sell to generate profit, you can allocate target to your sales team and can motivate them to achieve those targets within the decided time.
5. Using Break Even Analysis, you can calculate the funds required to start your business. You can use this information to raise funds from outside. Break Even Analysis is considered one of the most critical tools when someone is trying to raise funds for their business. Investors will ask for this information to know whether their investment in your business will be profitable or not.
6. Break Even Analysis requires you to consider all the costs to calculate the right value. In this way, you will not forget about any expenses associated with the business.

Limitations of Break Even Analysis

1. The Break Even Analysis doesn't provide accurate analysis in multi-product companies, as it assumes that the proportion of each product is constant which is not right. Therefore, it does not help companies with the information that which product is more beneficial and profitable for the company.

2. It is wrong to assume that variable cost is constant for all the units produced, which is not valid. The variable cost can be different for products produced under different batches. Therefore, the variable cost can never be the same for all units of the product.
3. Break Even Analysis assumes that all the units produced will be equal to all the units sold. It does not consider all the units which might break during the delivery or might end up being part of the inventory. Therefore, it does not provide an accurate break-even point.
4. The calculation of break-even point depends mainly on fixed costs, and it is assumed that fixed costs are constant. But it is not true because fixed costs change with the change in the scale of the business.
5. The last limitation of Break Even Analysis is that it does not consider the sales of products at different prices. It assumes that all produced units will be sold at the same price.

Shutdown Point

A shutdown point is an operating level where a business does not benefit in continuing production operations in the short run when revenue from selling their product is unable to cover variable costs of production. The shutdown point represents a point where a firm will incur higher and increasing losses if it continues production, as opposed to reduced losses if production is ceased. The shutdown point occurs at a point where marginal profit reaches a negative scale.

Understanding Shutdown Points

A shutdown arises when price or average revenue (AR) falls below average variable cost (AVC) at the profit-maximizing output level. Continued production will incur additional variable costs but will not generate enough revenue to cover them. At the same time, the firm will still have fixed costs to pay, further increasing the losses.

A shutdown point is typically a short-run position; however, in the long run, the firm should shut down and leave the industry if its product price is less than its average total cost. Therefore, there are two shutdown points for a firm – in the short run and the long run. The decision to shut down is dependent on which costs the firm can avoid by shutting down production. The short run is a period where at least one of the firm's inputs is fixed, resulting in fixed costs incurred despite the decision to shut down.
In summary, the shutdown point has the following characteristics:

1. It is the output and price point where a firm is able to just cover its total variable cost.
2. The average variable cost (AVC) is at its minimum point.
3. It is where the marginal cost (MC) curve intercepts the average variable cost (AVC) curve.
4. The firm is indifferent between shutting down and continuing production where losses equal to the total fixed costs are incurred regardless of either decision.

   Where:

   MC – Marginal Cost, **ATC** – Average Total Cost, **AVC** – Average Variable Cost, **SP** – Shutdown Price, **BEP** – Break-even Price

Short-Run Shutdown Decision

The cost of production is divided into two parts – fixed costs and variable costs. The break-even point is a point where revenue generated from sales of a product is equal to the production cost (fixed cost plus variable cost). Zero profit is generated at the break-even point. On the graph above, it is the point where the average total cost (ATC) is equal to marginal cost (MC) (i.e., MC = ATC). Marginal cost equals a change in total costs for each additional unit produced. Fixed costs do not change in the short run; hence, the change in total costs refers to variable cost only.

The shutdown zone represents an area between the break-even point and the shutdown point. it is an area where production can continue, as average revenue (AR) will still be able to cover average variable cost (AVC). However, in the shutdown zone, the firm will be making losses as the price

is below average total cost (ATC). The firm operates at any level above the AVC curve as long as it is where MC = MR (price). The MC curve above the AVC is also the short-run supply curve of the firm.

The shutdown rule states that a firm should continue operations as long as the price (average revenue) is able to cover average variable costs. The firm can continue operating, as it will be producing where marginal revenue (price, average revenue) is equal to marginal cost, a condition that ensures profit maximization or loss minimization.

A continuation of the shutdown rule states that in the short run, fixed costs are considered as sunk costs. Hence, it should not be considered in the decision of whether to shut down or continue with operations. In addition, in the short run, if the firm's total revenue is less than variable costs, the firm should shut down. A short-run decision to shut down is not the same as exiting the industry. Several firms in seasonal industries – such as agriculture, fishing, etc. – shut down their firms during the offseason to avoid unnecessary operating costs. They will not be generating any revenue during the off-season; hence they are unable to cover variable costs arising. It makes sense to temporarily shut down until the upcoming season commences.

Shutdown Point Illustration

Ender by Manufacturing is operating at a loss of $2,800. The firm cannot avoid paying fixed costs, whether they operate or not. If they choose to shut down and cease operations, they will generate zero revenue, zero variable costs, and incur fixed costs of $10,000, which means the total loss will increase to $10,000. However, if the firm continues to operate, it will still generate revenue of $16,000, where $8,800 will be expended to cover variable costs, and the balance of $7,200 will meet part of the fixed costs. Therefore, by continuing operations, the firm will only make a loss of $2,800 instead of $10,000 if they decide to shut down in the short run. However, if the selling price falls below $11 per unit and costs remain the same, the firm will have reached the shutdown point (AR < AVC). Such a condition satisfies the shutdown rule where shutdown is recommended.

Calculation of the Short-Run Shutdown Point

As illustrated above, the shutdown point is the output level at the minimum of the average variable cost curve (AVC). The shutdown point can be calculated using the total cost (TC) function. Suppose the total cost function is as follows:

Long-Run Shutdown

As a rule of thumb, a decision to shut down in the long run – i.e., exiting the industry – should only be undertaken if revenues are unable to cover total costs. It means in the long run, a firm making losses should shut down permanently and exit the industry. The short run is defined as a period where at least one fixed input or cost is present in the business. Fixed expenses such as rentals are incurred whether the firm undertakes production or not. In the long run, all inputs and costs are variable.

However, in the long run, if the firm is unable to raise the selling price per unit (to increase overall revenue) to cover total costs, the losses will continue ballooning until average revenue (AR) is exceeded by the average total cost (ATC). The firm will have reached the shutdown point where the only viable option is to shut down. Shutdown, as indicated above, is a short-run decision to minimize losses. It is because a firm can shut down in the interim, but if market conditions permit, it can still resume production. Even if the firm shuts down, it will still have incurred sunk costs in terms of investment in plant and equipment. Hence, it is possible for a firm to shut down in the short run and resume production in the long run.

However, if conditions do not improve, firms will resultantly decide to exit the industry, which is a long-run decision. The long-run exit decision is guided by the relationship between the price (P) and the long-run average cost (LRAC). Firms will exit the industry if $P < LRAC$. In the long run, if the firm decides to operate, it will still operate where the long-run marginal cost (LRMC) is equal to marginal revenue (MR). The long-run shutdown point is defined by the output corresponding to the minimum average total cost (ATC). The long-run shutdown point can be calculated much the same way

we did for the short-run shutdown point. We take the derivative of the ATC and solve for Q by setting it to zero. We plug it into the ATC function to get the price.

Monopoly Market Structure Shutdown Point

In the short run, a monopolist market structure shutdown point is reached when average revenue (price) is below average variable cost (AVC) at every output level. In such a case, it means that the demand curve is completely below the average variable cost curve. Even though a firm may be producing where marginal revenue is equal to marginal cost (MR = MC: the profit-maximizing level of output), average revenue would be less than average variable cost. The monopolist would be wise to shut down at such a point.

Real-World Application of the Shutdown Point

There are circumstances where firms can reach a shutdown point where the price is below AVC, but they decide not to shut down and keep operating because of any of the following reasons:

1. To retain long-term customers of the business. When a firm thinks that they are in a passing period of falling demand, they can opt to keep producing.
2. Some financially-strong firms are able to ride out a period of loss-making due to readily available credit support or a healthy standby reserve fund.
3. Firms can also decide to cut costs or increase product prices if they reach the shutdown point.
4. Firms can also shut down and leave the industry if they perceive a gloomy forecast of the long-term performance of the industry. It can happen well before the firm reaches the shutdown point – i.e., where AR < AVC.

5. Firms can also continue operating past the shutdown point, as it may take time to realize that they are operating at a loss. They may find out through management accounts once they are released.

Natural Monopoly

Definition: A natural monopoly occurs when the most efficient number of firms in the industry is one. A natural monopoly will typically have very high fixed costs meaning that it is impractical to have more than one firm producing the good. An example of a natural monopoly is tap water. It makes sense to have just one company providing a network of water pipes and sewers because there are very high capital costs involved in setting up a national network of pipes and sewage systems. To have two different companies offering water wouldn't make sense as the average cost would be very high compared to just one firm and one network. There would also be the inconvenience of having two firms dig up the road to lay a duplicate set of water pipes.

Definition of Natural Monopoly

William Baumol (1977) stated a natural monopoly is "[a]n industry in which multiform production is more costly than production by a monopoly"

- Suppose the industry demand is 10,000 units.
- If a firm produces 10,000 units, it will get the lowest possible average costs – £9.
- If there were three firms producing 3,000 units. The firms would have average costs of £17.
- Therefore, the optimal number of firms in the industry will be one (one firm producing all 10,000 units)

Examples of Natural Monopolies

- Gas network
- Electricity grid
- Railway infrastructure

- National fibre-optic broadband network.

Examples of potential natural monopolies

- **Aeroplane manufacture** – At the moment, this is a duopoly so it is not a natural monopoly, but it is close. There are very high fixed costs associated with aeroplane manufacturing, but with the global industry, two main producers can be supported.
- **Digital platforms**. In some cities, a product like Uber becomes ubiquitous for that segment of private taxi hire via an app. The fixed costs are not particularly high, but the dominant firm benefits from network economies, improved information, lower average prices and
- **Bus services in one particular region**. – The most logical number of bus companies within a town is one. There are high fixed costs, but more importantly issues of practicality. Even on a busy route between two towns, it might be inefficient to have two bus companies competing over the same route and offering the same peak and off-peak services. One company can avoid:
- Duplication of services
- Congestion at peak times
- Too much supply at off-peak times

However, some cities do have multiple bus services. On the one hand, this is more competition, but on the other hand, there is duplication. In buying gas for domestic use, there is competition. There are several companies who use the one national network. Therefore, gas is a natural monopoly at the distribution stage, but at the retail stage, it is possible to have competition.

Regulation of Natural Monopolies

Natural monopolies are incontestable and firms have no real competition. Therefore, without government intervention, they could abuse their market power and set higher prices. Therefore, natural monopolies

often need government regulation. For example, OFWAT and OFGEM regulate the water and energy markets respectively.

Marshall and Sraffa's views regarding monopoly

Sraffa 1926 is a general critique on Marshallian partial equilibrium theory. Sraffa states that external economies are always present and therefore all markets are interrelated. Sraffa then focuses on the real time element in economic development. There are historical reasons to deny that consumers are indifferent between goods of like-types (that goods are perfect substitutes for each other as is assumed under Marshallian perfect competition). When a producer creates a new product for a new market the producer, to use modern language, has a first-mover advantage and builds brand loyalty. Sraffa states that the price elasticity of demand for this product grows with time and therefore this then logically implies that "monopoly profits" are not competed-away. Additionally, credit and cost barriers-to-entry prevent perfect competition from competing away the monopoly profits.

Degree of monopoly power

The degree of monopoly power is measured in terms of difference between Marginal cost and the price. In a perfectly competitive market, price equals marginal cost and firms earn an economic profit of zero. In a monopoly, the price is set above marginal cost and the firm earns a positive economic profit. The degree of monopoly. Market power is the ability to charge a price above marginal cost. A firm in a competitive market produces where P=MC. This shows us how much of a 'mark-up' the firm is charging above its marginal cost, as a proportion of its price.

Incorporate Social cost monopoly

Regulation of monopoly

The government may wish to regulate monopolies to protect the interests of consumers. For example, monopolies have the market power to set prices higher than in competitive markets. The government can regulate monopolies through:

- Price capping – limiting price increases
- Regulation of mergers
- Breaking up monopolies
- Investigations into cartels and unfair practises
- Nationalisation – government ownership.
- The Government regulates monopolies

1. **Prevent excess prices**. Without government regulation, monopolies could put prices above the competitive equilibrium. This would lead to allocative inefficiency and a decline in consumer welfare.
2. **Quality of service**. If a firm has a monopoly over the provision of a particular service, it may have little incentive to offer a good quality service. Government regulation can ensure the firm meets minimum standards of service.
3. **Monopsony power**. A firm with monopoly selling power may also be in a position to exploit monopsony buying power. For example, supermarkets may use their dominant market position to squeeze profit margins of farmers.
4. **Promote competition**. In some industries, it is possible to encourage competition, and therefore there will be less need for government regulation.
5. **Natural Monopolies**. Some industries are natural monopolies – due to high economies of scale, the most efficient number of firms is one. Therefore, we cannot encourage competition, and it is essential to regulate the firm to prevent the abuse of monopoly power.

1. Price capping by regulators RPI-X

For many newly privatised industries, such as water, electricity and gas, the government created regulatory bodies such as:

- OFGEM – gas and electricity markets
- OFWAT – tap water.
- ORR – Office of rail regulator.

Amongst their functions, they are able to limit price increases. They can do this with a formula RPI-X

- X is the amount by which they have to cut prices by in real terms.
- If inflation is 3% and X= 1%

- Then firms can increase actual nominal prices by 3-1 = 2%

  If the regulator thinks a firm can make efficiency savings and is charging too much to consumers, it can set a high level of X. In the early years of telecom regulation, the level of X was quite high because efficiency savings enabled big price cuts.

  RPI+/- K – for water industry

  In water, the price cap system is RPI -/+ K.

  K is the amount of investment that the water firm needs to implement. Thus, if water companies need to invest in better water pipes, they will be able to increase prices to finance this investment.

Advantages of RPI-X Regulation

1. The regulator can set price increases depending on the state of the industry and potential efficiency savings.
2. If a firm cut costs by more than X, they can increase their profits. Arguably there is an incentive to cut costs.
3. Surrogate competition. In the absence of competition, RPI-X is a way to increase competition and prevent the abuse of monopoly power.

Disadvantages of RPI-X Regulation

1. It is costly and difficult to decide what the level of X should be.
2. There is a danger of regulatory capture, where regulators become too soft on the firm and allow them to increase prices and make supernormal profits.
3. However, firms may argue regulators are too strict and don't allow them to make enough profit for investment.
4. If a firm becomes very efficient, it may be penalised by having higher levels of X, so it can't keep its efficiency saving.

2. Regulation of quality of service

Regulators can examine the quality of the service provided by the monopoly. For example, the rail regulator examines the safety record of rail firms to ensure that they don't cut corners.

In gas and electricity markets, regulators will make sure that old people are treated with concern, e.g. not allow a monopoly to cut off gas supplies in winter.

3. Merger policy

The government has a policy to investigate mergers which could create monopoly power. If a new merger creates a firm with more than 25% of market share, it is automatically referred to the Competition and Markets Authority (CMA). The CMA can decide to allow or block the merger depending on whether it believes it is in the public interest.

- For example, CMA blocked the merger between Sainsbury's and Asda as being against the public interest.

4. Breaking up a monopoly

In certain cases, the government may decide a monopoly needs to be broken up because the firm has become too powerful. This rarely occurs. For example, the US looked into breaking up Microsoft, but in the end, the action was dropped. This tends to be seen as an extreme step, and there is no guarantee the new firms won't collude.

5. Yardstick or 'Rate of Return' Regulation

This is a different way of regulating monopolies to the RPI-X price capping. Rate of return regulation looks at the size of the firm and evaluates what would make a reasonable level of profit from the capital base. If the firm is making too much profit compared to their relative size, the regulator may enforce price cuts or take one-off tax. A disadvantage of the rate of return regulation is that it can encourage 'cost padding'. This is when firms allow costs to increase so that profit levels are not deemed excessive. Rate of return regulation gives little incentive to be efficient and increase profits. Also, rate of return regulation may fail to evaluate how much profit is reasonable. If it is set too high, the firm can abuse its monopoly power.

6. Investigation of abuse of monopoly power

In the UK, the office of fair trading can investigate the abuse of monopoly power. This may include unfair trading practices such as:

- Collusion (firms agree to set higher prices)

- Collusive tendering. This occurs when firms enter into agreements to fix the bid at which they will tender for projects. Firms will take it in turns to get the contract and enable a much higher price for the contract.
- Predatory pricing (setting low prices to try and force rival firms out of business)
- Vertical restraints – prevent retailers stock rival products
- Selective distribution For example, in the UK car industry firms entered into selective and exclusive distribution networks to keep prices high. The competition commission report of 2000 found UK cars were at least 10% higher than European cars

Peak-Load Pricing

The below mentioned article provides quick notes on peak-load pricing.

It is a form of inter-temporal price discrimination based on efficiency.

For goods and services, demand peaks at particular times — for roads and public transport during commuter rush hours, for electricity during late afternoon and so on.

MC is also high during these peak periods because of capacity constraints. Prices should, thus, be higher during peak periods as Fig. 9.13 shows, where D_1 is the demand curve for the peak period, and D_2 is the demand curve for non-peak period.

The firm sets MC = MR for each period, such that price P_1 is high for the peak period, and the price P_2 is lower for the off-peak period, with corresponding quantities Q_1 and Q_2. This increases the firm's profit above what it would be if it charged one price for all periods. It is also efficient; the sum of producer and consumer's surplus is greater because prices are closer to MC.

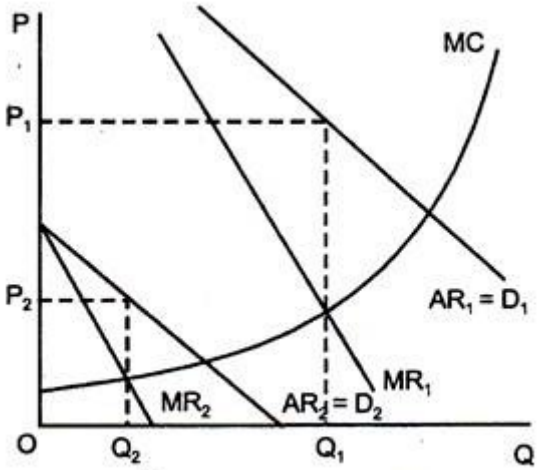


Fig. 9.13 : Peak-load Pricing

Peak-load pricing is different from third-degree price discrimination. With third-degree price discrimination, MR has to be equal for each group of consumers and equal to MC because the cost of serving the different groups is not independent. However, price and sales in peak and off-peak period can be determined independently by setting MC = MR for each period as Fig. 9.13 shows. For example, a movie theatre, which charges more for the evening show than for the matinee show because for theatres, the MC of serving customers during the matinee show is independent of the MC during the evening. The owner of a movie theatre can determine the optimal prices for the evening and matinee shows independently, using estimates of demand in each period and of MC.

Monopoly - Price Discrimination

Price discrimination happens when a firm charges a different price to different groups of consumers for an identical good or service, for reasons not associated with costs of supply.

Price discrimination takes us away from the standard assumption in that there is a single profit-maximising price for the same good or services. Make sure you have at least one applied example of each type of price discrimination in your notes. Nearly all businesses make use of dynamic pricing methods where prices are heavily determined by the strength of demand and consumers' willingness & ability to pay. Price discrimination is also known as yield management.

The main types of price discrimination

1st degree: Charging different prices for each individual unit purchased – where people pay their own individual willingness to pay

2nd degree: Prices varying by **quantity sold** such as bulk purchase discounts. Prices varying by **time of purchase** such as peak-time prices

3rd degree: Charging different prices to **groups of consumers** segmented by the coefficient of **price elasticity of demand**, income, age, sex

The main conditions required for a business to use price discrimination

1. Firms must have sufficient monopoly power: Monopolists always have pricing power – they are price makers not takers
2. Identifying different market segments: There must be groups of consumers with different price elasticity of demand
3. Ability to separate different groups: Requires information on the purchasing behaviour of consumers – often achieved by accumulating data on previous buying patterns
4. Ability to prevent re-sale (arbitrage): No secondary markets where arbitrage can take place at intermediate prices - limiting sales might be done by using age-restrictions, ID cards and so on

Price discrimination does not happen in perfectly competitive markets. It is only a feature of imperfect competition where firms have some discretion / power over the prices they charge.

Aims of price discrimination

Providing that extra units can be sold for a price above the marginal cost of supply, price discrimination is an effective way to increase revenue and profits

1. To increase total revenue by extracting consumer surplus and turning it into producer surplus
2. To increase total profit providing the marginal profit from selling to customers is positive
3. To generate cash-flow especially during a recession
4. To increase market share and build customer loyalty
5. To make more efficient use of a firm's spare capacity

6. To reduce waste and cut the cost of keeping products in stock / storage

Advantages from price discrimination

1. It makes fuller use of spare capacity leading to less waste and unsold stock. There are potential environmental benefits from this.
2. Helps generate extra cash flow for businesses which can ensure survival during a recession / tough economic times.
3. Can help fund the cross-subsidy of goods and services – for example premium prices for some can fund discounts for other groups perhaps living on lower incomes.
4. Higher monopoly profits can finance research and development spending which then drives improved dynamic efficiency.

Disadvantages from price discrimination

1. Price discrimination operates mainly in the interests of producers as they extract consumer surplus and turn it into extra supernormal profit
2. Can be used as a pricing tactic to reduce competition and reinforce the market dominance of leading firms
3. May lead to manipulation of groups with a price inelastic demand, not all of whom are on high incomes
4. Can be viewed as unfair to certain groups, for example there is some evidence of businesses using gender pricing on selected products

Price Discrimination under Monopoly: Types, Degrees and Other details

In monopoly, there is a single seller of a product called monopolist. The monopolist has control over pricing, demand, and supply decisions, thus, sets prices in a way, so that maximum profit can be earned. The monopolist often charges different prices from different consumers for the same product. This practice of charging different prices for identical product is called price discrimination.

According to Robinson, "Price discrimination is charging different prices for the same product or same price for the differentiated product." According to Stigler, "Price discrimination is the sale of various products at prices which are not proportional to their marginal costs." In the words of Dooley, "Discriminatory monopoly means charging different rates from different customers for the same good or service."

According to J.S. Bains, "Price discrimination refers strictly to the practice by a seller to charging different prices from different buyers for the same good."

Let us learn different types of price discrimination.

Types of Price Discrimination:

Price discrimination is a common pricing strategy' used by a monopolist having discretionary pricing power. This strategy is practiced by the monopolist to gain market advantage or to capture market position.

There are three types of price discrimination, which are shown in Figure-13:

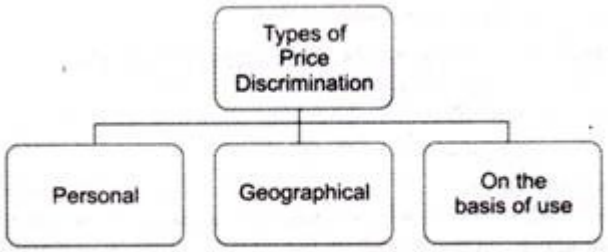


Figure-13: Price Discrimination

The different types of price discrimination (as shown in Figure-13) are explained as follows:

i. Personal:

Refers to price discrimination when different prices are charged from different individuals. The different prices are charged according to the level of income of consumers as well as their willingness to purchase a product. For example, a doctor charges different fees from poor and rich patients.

ii. Geographical:

Refers to price discrimination when the monopolist charges different prices at different places for the same product. This type of discrimination is also called dumping.

iii. On the basis of use:

Occurs when different prices are charged according to the use of a product. For instance, an electricity supply board charges lower rates for domestic consumption of electricity and higher rates for commercial consumption.

Degrees of Price Discrimination:

Price discrimination has become widespread in almost every market. In economic jargon, price discrimination is also called monopoly price discrimination or yield management. The degree of price discrimination vanes in different markets.

Figure-14 shows the degrees of price discrimination:

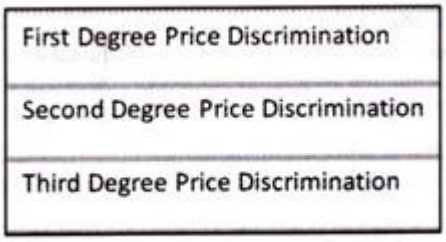


Figure-14: Degrees of Price Discrimination

These three degrees of price discrimination are explained as follows:

i. First-degree Price Discrimination:

Refers to a price discrimination in which a monopolist charges the maximum price that each buyer is willing to pay. This is also known as perfect price discrimination as it involves maximum exploitation of consumers. In this, consumers fail to enjoy any consumer surplus. First degree is practiced by lawyers and doctors.

ii. Second-degree Price Discrimination:

Refers to a price discrimination in which buyers are divided into different groups and different prices are charged from these groups depending upon what they are willing to pay. Railways and airlines practice this type of price discrimination.

iii. Third-degree Price Discrimination:

Refers to a price discrimination in which the monopolist divides the entire market into submarkets and different prices are charged in each submarket. Therefore, third-degree price discrimination is also termed as

market segmentation. In this type of price discrimination, the monopolist is required to segment market in a manner, so that products sold in one market cannot be resold in another market. Moreover, he/she should identify the price elasticity of demand of different submarkets. The groups are divided according to age, sex, and location. For instance, railways charge lower fares from senior citizens. Students get discount in cinemas, museums, and historical monuments.

Necessary Conditions for Price Discrimination:

Price discrimination implies charging different prices for identical goods.

It is possible under the following conditions:

i. Existence of Monopoly:

Implies that a supplier can discriminate prices only when there is monopoly. The degree of the price discrimination depends upon the degree of monopoly in the market.

ii. Separate Market:

Implies that there must be two or more markets that can be easily separated for discriminating prices. The buyer of one market cannot move to another market and goods sold in one market cannot be resold in another market.

iii. No Contact between Buyers:

Refers to one of the most important conditions for price discrimination. A supplier can discriminate prices if there is no contact between buyers of different markets. If buyers in one market come to know that prices charged in another market are lower, they will prefer to buy it in other market and sell in own market. The monopolists should be able to separate markets and avoid reselling in these markets.

iv. Different Elasticity of Demand:

Implies that the elasticity of demand in the markets should differ from each other. In markets with high elasticity of demand, low price will be charged, whereas in markets with low elasticity of demand, high prices will be charged. Price discrimination fails in case of markets having same elasticity- of demand.

Advantages and Disadvantages of Price Discrimination:

A monopolist practices price discrimination to gain profits. However, it acts as a loss for the consumers.

Following are some of the advantages of price discrimination:

i. Helps organizations to earn revenue and stabilize the business

ii. Facilitates the expansion plans of organizations as more revenue is generated

iii. Benefits customers, such as senior citizens and students, by providing them discounts

In spite of advantages, there are certain disadvantages of price discrimination.

Some of the disadvantages of price discrimination as follow:

i. Leads to losses as some consumers end up paying higher prices

ii. Involves administration costs for separating markets.

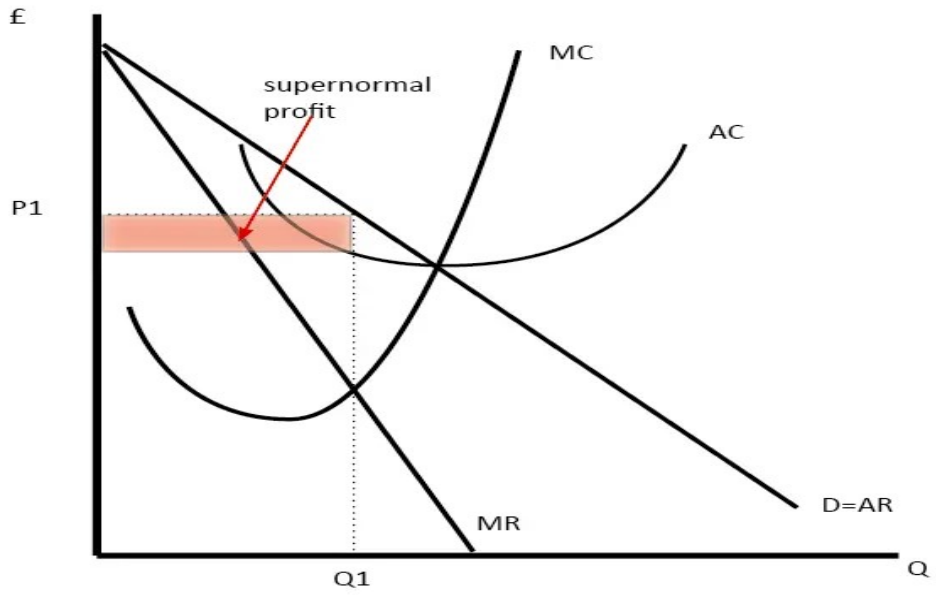
Monopolistic Competition – definition, diagram and examples

Definition: 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, but firms can differentiate their products. Therefore, they have an inelastic demand curve and so they can set prices. However, because there is freedom of entry, supernormal profits will encourage more firms to enter the market leading to normal profits in the long term.

A monopolistic competitive industry has the following features:

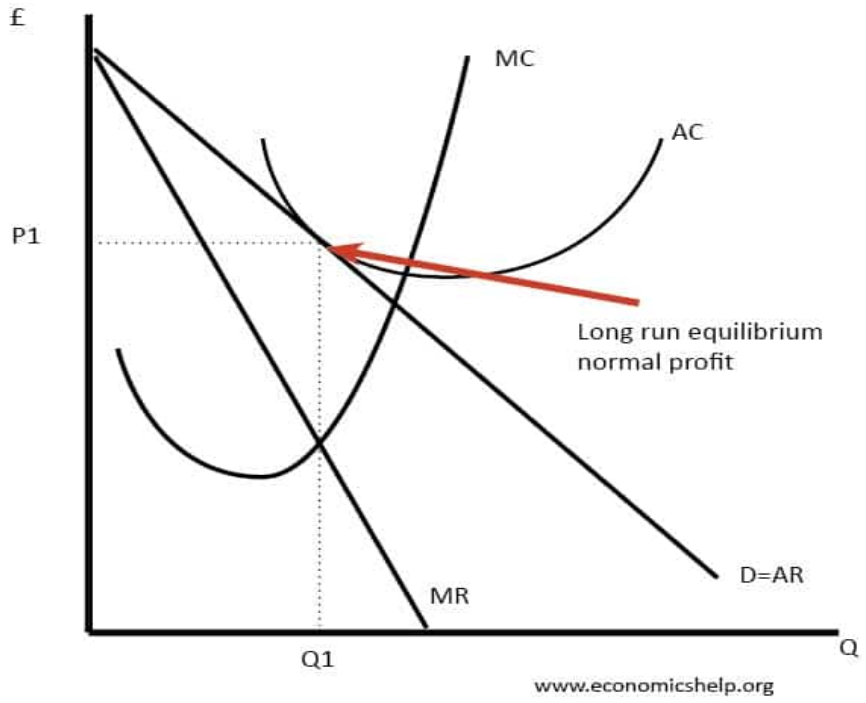
- Many firms.
- Freedom of entry and exit.
- Firms produce differentiated products.
- Firms have price inelastic demand; they are price makers because the good is highly differentiated
- Firms make normal profits in the long run but could make supernormal profits in the short term
- Firms are allocative and productively inefficient.

Diagram monopolistic competition short run



In the short run, the diagram for monopolistic competition is the same as for a monopoly. The firm maximises profit where MR=MC. This is at output Q1 and price P1, leading to supernormal profit

Monopolistic competition long run



Demand curve shifts to the left due to new firms entering the market.

In the long-run, supernormal profit encourages new firms to enter. This reduces demand for existing firms and leads to normal profit. I

Efficiency of firms in monopolistic competition

- Allocative inefficient. The above diagrams show a price set above marginal cost
- Productive inefficiency. The above diagram shows a firm not producing on the lowest point of AC curve
- Dynamic efficiency. This is possible as firms have profit to invest in research and development.
- X-efficiency. This is possible as the firm does face competitive pressures to cut cost and provide better products.

Examples of monopolistic competition

- Restaurants – restaurants compete on quality of food as much as price. Product differentiation is a key element of the business. There are relatively low barriers to entry in setting up a new restaurant.
- Hairdressers. A service which will give firms a reputation for the quality of their hair-cutting.
- Clothing. Designer label clothes are about the brand and product differentiation
- TV programmes – globalisation has increased the diversity of tv programmes from networks around the world. Consumers can choose between domestic channels but also imports from other countries and new services, such as Netflix.

Limitations of the model of monopolistic competition

- Some firms will be better at brand differentiation and therefore, in the real world, they will be able to make supernormal profit.
- New firms will not be seen as a close substitute.

- There is considerable overlap with oligopoly – except the model of monopolistic competition assumes no barriers to entry. In the real world, there are likely to be at least some barriers to entry
- If a firm has strong brand loyalty and product differentiation – this itself becomes a barrier to entry. A new firm can't easily capture the brand loyalty.
- Many industries, we may describe as monopolistically competitive are very profitable, so the assumption of normal profits is too simplistic.

Key difference with monopoly

In monopolistic competition there are no barriers to entry. Therefore in long run, the market will be competitive, with firms making normal profit.

Key difference with perfect competition

In Monopolistic competition, firms do produce differentiated products; therefore, they are not price takers (perfectly elastic demand). They have inelastic demand.

New trade theory and monopolistic competition

New trade theory places importance on the model of monopolistic competition for explaining trends in trade patterns. New trade theory suggests that a key element of product development is the drive for product differentiation – creating strong brands and new features for products. Therefore, specialisation doesn't need to be based on traditional theories of comparative advantage, but we can have countries both importing and exporting the same good. For example, we import Italian fashion labels and export British fashion labels. To consumers, the importance is the choice of goods.

Theory of Excess Capacity under Monopolistic Competition

The concept of excess capacity is found in the earlier works of Wicksell and Cairnes. P. Sraffa and Mrs. Joan Robinson also outlined it. But it was Chamberlin who expounded it in a most systematic manner followed by Kaldor, Kahn, Harrod and Cassels. The doctrine of excess (or unutilised) capacity is associated with monopolistic competition in the long- run and is

defined as “the difference between ideal (optimum) output and the output actually attained in the long-run.”

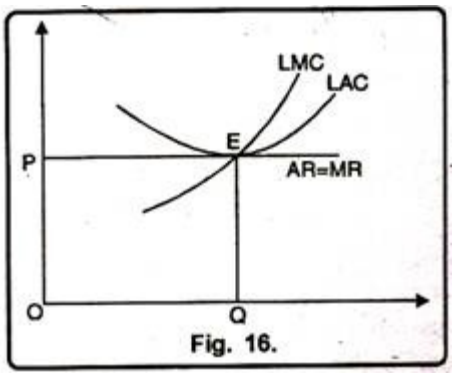


Fig. 16.

Under perfect competition, however, the demand curve (AR) is tangential to the long-run average cost curve (LAC) at its minimum point and conditions of full equilibrium are fulfilled: LMC = MR and AR (price) = Minimum LAC. This means that in the long-run the entry of new firms forces the existing firms to make the best use of their resources to produce at the point of lowest average total costs. At point E in Figure 16, abnormal profits will be competed away because MR = LMC = AR = LAC at its minimum point E and OQ will be the most efficient output which the society will be enjoying. This is the ideal or optimum output which firms produce in the long-run.

Under monopolistic competition the demand curve facing the individual firm is not horizontal as under perfect competition, but it is downward sloping. A downward sloping demand curve cannot be tangent to the LAC curve at its minimum point. The double condition of equilibrium LMC = MR = AR (d) = Minimum LAC will not be fulfilled. The firms will, therefore, be of less than the optimum size even when they are earning normal profits. No firm will have the incentive to produce the ideal output, since any effort to produce more than the equilibrium output would involve a higher long-run marginal cost than marginal revenue.

Thus each firm under monopolistic competition will be of less than the optimum size and work under excess capacity. This is illustrated in Figure 17 where the monopolistic competitive firm’s demand curve is d and MR_1 is its corresponding marginal revenue curve. LAC and LMC are the long-run average cost and marginal cost curves. The firm is in equilibrium at

E_1 where the LMC curve cuts the MR_1 curve from below and OQ_1 output is set at the price Q_1A_1. OQ_1 is the equilibrium output but not the ideal output because d is tangent to the LAC curve at A_1 to the left of the minimum point E. Any effort on the part of the firm to produce beyond OQ_1 will mean losses as beyond the equilibrium point E_1, LMC > MR_1. Thus the firm has negative excess capacity measured by OQ_1 which it cannot utilize working under monopolistic competition.

A comparison of the equilibrium positions under monopolistic competition and perfect competition with the help of Figure 17 reveals that the output of a firm under monopolistic competition is smaller and the price of its product is higher than under perfect competition. The monopolistic competition output OQ_1 is less than the perfectly competitive output OQ, and the monopolistic competitive price Q_1A_1 is higher than the competitive equilibrium price QE. This is because of the existence of excess capacity under monopolistic competition.

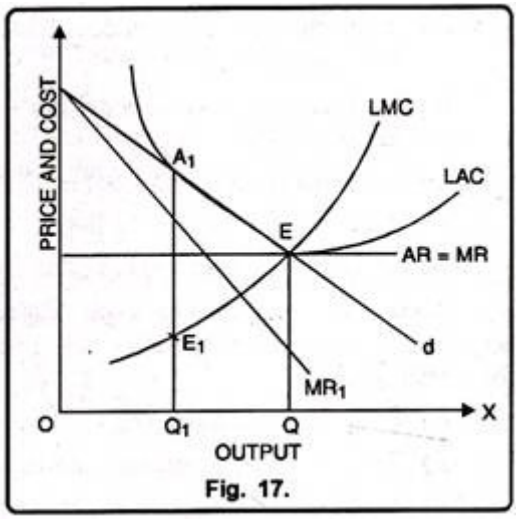


Fig. 17.

Chamberlin's Concept of Excess Capacity:

Prof. Chamberlin's explanation of the theory of excess capacity is different from that of ideal output under perfect competition. Under perfect competition, each firm produces at the minimum on its LAC curve and its horizontal demand curve is tangent to it at that point.

Its output is ideal and there is no excess capacity in the long-run. Since under monopolistic competition the demand curve of the firm is downward

sloping due to product differentiation, the long-run equilibrium of the firm is to the left of the minimum point on the LAC curve. According to Chamberlin, so long as there is freedom of entry and price competition in the product group under monopolistic competition, the tangency point between the firm's demand curve and the LAC curve would lead to the "ideal output" and no excess capacity.

Assumptions:

Chamberlin's concept of excess capacity assumes that:

(i) The number of firms is large;

(ii) Each produces a similar product independently of the others;

(iii) It can charge a lower price and attract other's customers and by raising its price will lose some of its customers;

(iv) 'Consumers' preferences are fairly evenly distributed among the different varieties of products;

(v) No firm has an institutional monopoly over the product;

(vi) Firms are free to enter its field of production;

(vii) The long-run cost curves of all the firms are identical and are U-shaped.

Reasons:

According to Chamberlin, excess capacity arises when there is no active price competition despite free entry of firms in a monopolistic competitive market.

He gives the following reasons for such a situation:

(i) Firms may consider costs rather than demand in fixing prices.

(ii) They may aim at ordinary profits rather than maximum profits,

(iii) They may follow a policy of 'live and let live' and may not resort to price reduction.

(iv) They may have formal or tacit agreements, open price associations, trading association activities in building up an esprit de corps and price maintenance.

(v) There may be the imposition of uniform prices on dealers by manufacturers.

(vi) Firms may resort to excessive differentiation of the product in an attempt to turn attention away from price cutting.

(vii) Business or professional ethics prevent firms from resorting to active price competition.

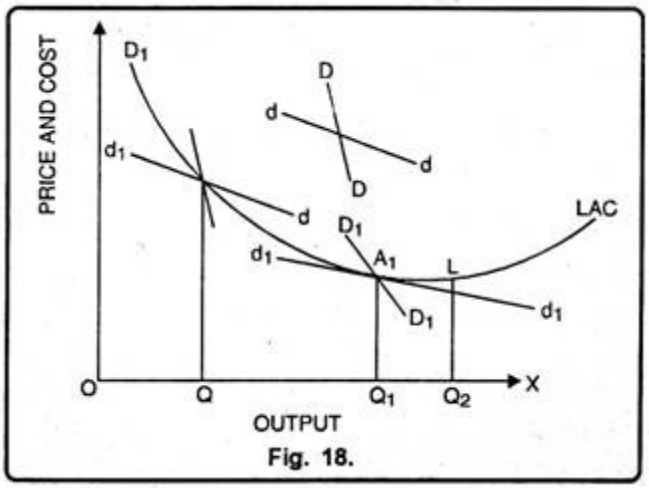


Fig. 18.

When there is no price competition due to the prevalence of these factors, the curve *dd* is of no significance and the firms are only concerned with the group DD curve. Suppose the initial short-run equilibrium is at S where the firms are earning supernormal profits because the price OP corresponding to point S is above the LAC curve.

With the entry of new firms in the group, super-normal profits will be competed away. The new firms will divide the market among themselves and the DD curve will be pushed to the left as d_1d_1 in Figure 18 where it becomes tangent to the LAC curve at point A_1, This point A_1 is of stable equilibrium in the absence of price competition for all firms in the group and they are earning only normal profits. Each firm is producing and selling OQ output at QA (= OP) price. In Chamberlin's analysis, O_1 is the 'ideal output'. But each firm in the group is producing OQ output in the absence of price competition. Thus OQ_1 represents excess capacity under non-price monopolistic competition.

Chamberlin concludes that when over long periods under non-price competition prices do not fall and costs rise, the two are equated by the development of excess productive capacity which does not possess

automatic corrective. Such excess capacity may develop under pure competition due to miscalculations on the part of producers or to sudden changes in demand or cost conditions. But under monopolistic competition it may develop over long periods with impunity, prices always covering costs, and may, in fact become permanent and normal through a failure of price competition to function. The surplus capacity is never abandoned and the result is high prices and wastes. They are the wastes of monopolistic competition.

Significance of Excess Capacity:

The concept of excess capacity is of much practical significance. Prof. Kaldor has characterised it as "intellectually striking", 'a highly ingenious' and 'revolutionary doctrine.'

1. It demonstrates an untraditional possibility that an increase in supply may lead to a rise in price. The 'wastes of competition' which were hitherto a mystery have been unfolded. They pertain to monopolistic competition rather than to perfect competition, as was strongly implied by the earlier economists.
2. It establishes the truth of the proposition that perfect competition and increasing returns are incompatible and proves that falling costs ultimately lead to monopoly or monopolistic competition. When monopolistic competition prevails, the number of firms will be large. But each firm will be of a smaller size than under perfect competition.
3. This entails a wasteful use of resources by bringing up firms with lower efficiency. Such firms may employ more manpower, equipment and raw materials than is necessary. This leads to excess or unutilized capacity.

Chamberlin's Concept of Excess Capacity | Imperfect Competition

Prof. Chamberlin's explanation of the theory of excess capacity is different from that of ideal output under perfect competition. Under perfect competition each firm produces at the minimum on its LAC curve and its

horizontal demand. Curve is tangent to it at that point. Its output is ideal and there is no excess capacity in the long-run.

Chamberlin's concept of excess capacity assumes that:

(i) The number firms be large,

(ii) Each should produce a similar product independently,

(iii) It should charge a lower price and attract other's customers and by raising its price will lose some of its own customers,

(iv) Customer's preferences be fairly evenly distributed among the different varieties of products,

(v) No firm should have an institutional monopoly over the product,

(vi) Firms are free to enter its field of production,

(vii) Long-run cost curves of all the firms are identical and are U-shaped.

As we are aware that this concept of excess capacity is associated with imperfect competition in the long period. This can be defined as the difference between the optimum level of output and the output actually obtained in the long-run.

Regarding this Prof. M. M. Bober has said:

"Under imperfect competition we are apt to witness too many firms all working under less than their optimum capacity all charging higher than competitive price, and all exercising their art of advertisement against each other in the efforts to retain customers and to take them away from each other."

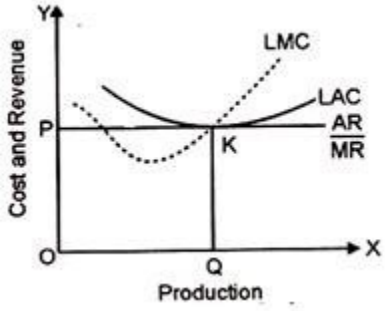
The above observation of Prof. Bober reveals three major aspects of excess capacity:

(i) In the imperfect competition the output of the firms is less than the optimum output which indicate that the monopolistic firm can produce more than the determine output and the cost of production can further fall.

(ii) The price is determined under imperfect competition is more than the price determined under perfect competition. It is simply because with the increase in sale the price falls.

(iii) The advertisement outlay of each firm influence the sale of his firm and affect the sale of other firms.

The analysis of Perfect Competition can be explained with the help of the following diagram:



In this figure the demand curve (AR) and marginal revenue (MR) are tangential to the long-run average cost curve (LAC) at minimum point and the condition of full equilibrium are fulfilled i.e., LMC = MR and AR = LAC, This means that in the long period the entry of new firms forces the existing firms to make the best use of their resources to produce at the point of lower average total cost. In this figure at point K the abnormal profits will be competed away and the firms will earn only normal profits because here LAC =AR = MR = LMC. OQ is the most efficient output and there is not excess capacity.

Excess Capacity in Imperfect Competition:

Under imperfect competition the demand curve facing the individual firm is downwards sloping curve. The downwards sloping demand curve cannot be tangent to the long-run average cost curve at the minimum point. The double condition of equilibrium LMC = MR = AR = minimum LAC will not be fulfilled. The firm will, therefore, be of less than optimum size even when they are earning normal profits. No firm will have the incentive to produce the optimum output because with the increase in the output the average revenue will fall and possibly the AC >AR. The firm will earn losses instead of profits or normal profits.

A comparison of two equilibrium (i.e., perfect competition and imperfect competition) reveals that the firm under imperfect competition will not be working to their full capacity. There will be chronic excess capacity and wastage. Each firm will be producing less than their full capacity, incurring higher costs and charging a higher price than the perfect

competition. Thus, in imperfect competition output will be OQ which is less than OQ_1 under perfect competition, the price (OP) under imperfect competition is more than the price OP_1 under perfect competition.

Definition of Monopsony

- A monopsony occurs when a firm has market power in employing factors of production (e.g. labour).
- A monopsony means there is one buyer and many sellers.
- It often refers to a monopsony employer – who has market power in hiring workers.
- This is a similar concept to monopoly where there is one seller and many buyers.

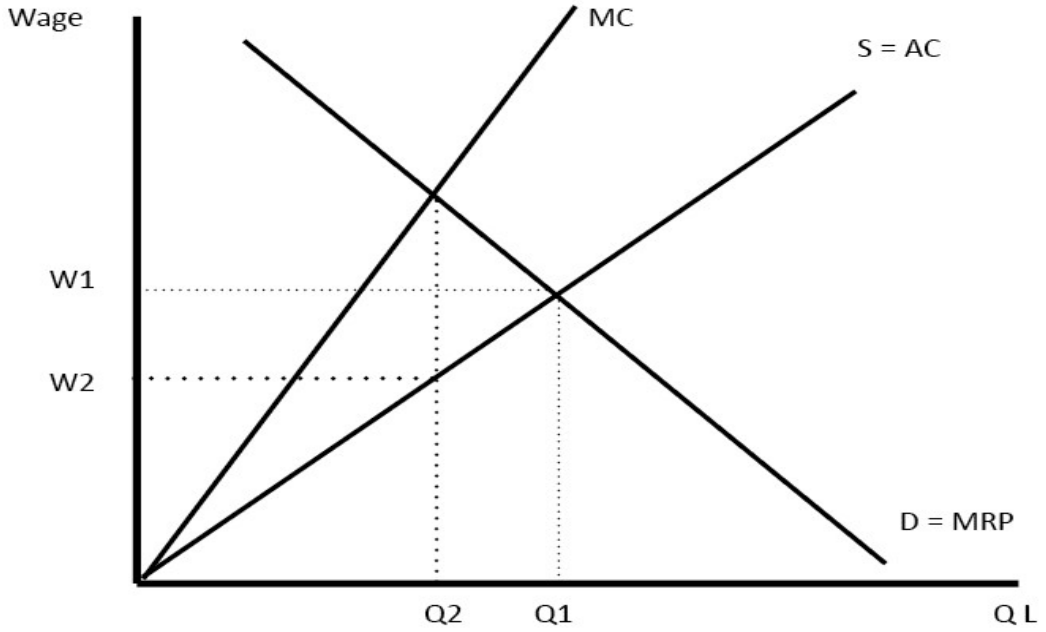
Monopsony in Labour Markets

An example of a monopsony occurs when there is one major employer and many workers seeking to gain employment.
If there is only one main employer of labour, then they have market power in setting wages and choosing how many workers to employ.

Examples of monopsony in labour markets

- Coal mine owner in town where coal mining is the primary source of employment.
- The government in the employment of civil servants, nurses, police and army officers.

Diagram of monopsony



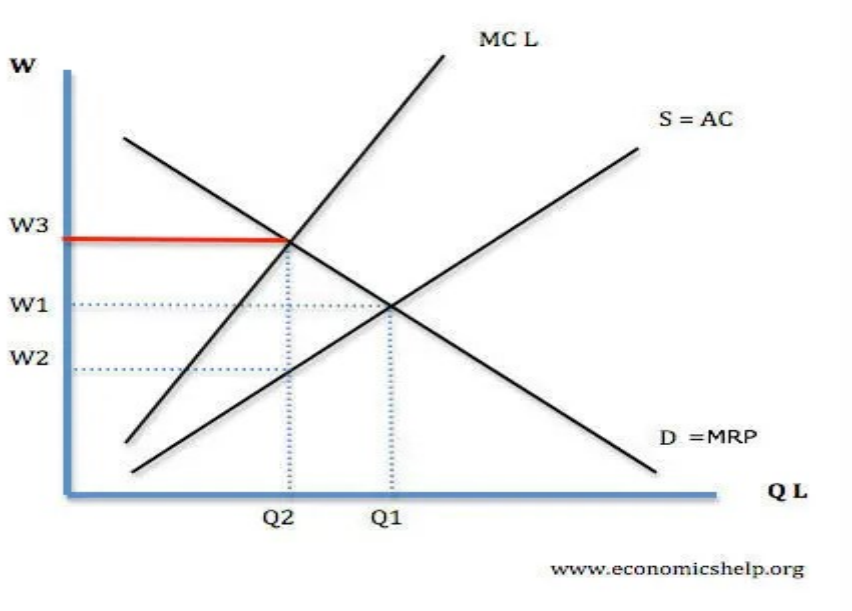
- In a competitive labour market, the equilibrium will be where D=S at Q1, W1.
- However, a monopsony can pay lower wages (W2) and employ fewer workers (Q2)

Profit Maximisation for a Monopsony

- The marginal cost of employing one more worker will be higher than the average cost – because to employ one extra worker the firm has to increase the wages of all workers.
- To maximise the level of profit, the firm employs Q2 of workers where the marginal cost of labour equals the marginal revenue product MRP = D
- In a competitive labour market, the firm would be a wage taker. If they tried to pay only W2, workers would go to other firms willing to pay a higher wage.

Minimum wage in a monopsony

In a monopsony, a minimum wage can increase wages without causing unemployment.



- A monopsony pays a wage of W2 and employs Q2.
- If a minimum wage was placed equal to W1, it would increase employment to Q1.
- A minimum wage of W3 would keep employment at Q2.

Monopsony in the real world

Even if a firm is not a pure monopsony, it may have a degree of monopsony power, due to geographical and occupational immobilities, which make it difficult for workers to switch jobs and find alternative employment. For example, there are several employers who might employ supermarket checkout workers. However, in practice, it is difficult for workers to switch jobs to take advantage of slightly higher wages in other supermarkets. There is a lack of information and barriers to moving jobs. Therefore, although there are several buyers of labour, in practice the big supermarkets have a degree of monopsony power in employing workers.

Monopsony and the gig economy

The gig economy refers to recent trends towards self-employment and very flexible labour practises. In practice, workers in the gig economy can easily face a monopsony employer. For example, Uber drivers have little control over rates of pay and have to meet strict criteria from Uber. In

theory, they could work elsewhere but in practise it is difficult to replicate that job.

Problems of monopsony in labour markets

- Monopsony can lead to lower wages for workers. This increases inequality in society.
- Workers are paid less than their marginal revenue product.
- Firms with monopsony power often have a degree of monopoly selling power. This enables them to make high profits at the expense of consumers and workers.
- Firms with monopsony power may also care less about working conditions because workers don't have many alternatives to the main firm.

Monopsony in product markets

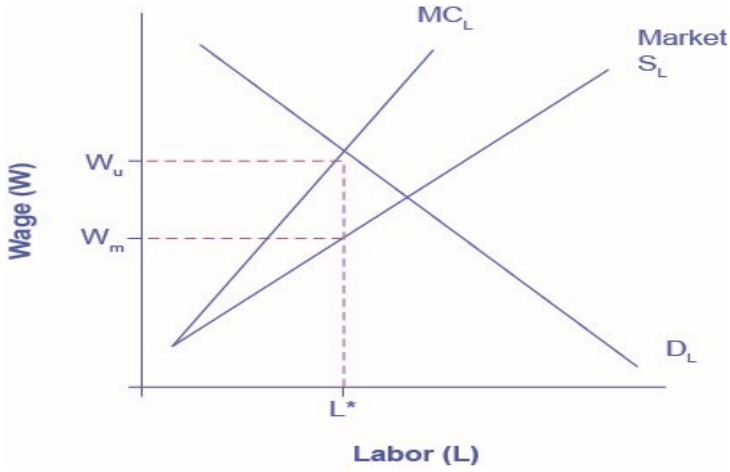
In several industries, there is one buyer and several sellers.

- Supermarkets have monopsony power in buying food from farmers. If farmers don't sell to the big supermarkets, there are few alternatives. This has led to farmer protests about the price of milk.
- Amazon.com is one of the biggest purchases of books. If publishers don't sell to Amazon at a discounted price, they will miss out on selling to the biggest distributor of books.

Bilateral Monopoly

How firms determine wages and employment when a specific labour market combines a union and a monopsony

What happens when there is market power on both sides of the labour market, in other words, when a union meets a monopsony? Economists call such a situation a bilateral monopoly. Bilateral Monopoly Employment, L*, will be lower in a bilateral monopoly than in a competitive labor market, but the equilibrium wage is indeterminate, somewhere in the range between Wu, what the union would choose, and Wm, what the monopsony would choose.



(Figure) is a combination of Figure 14.6 and Figure 14.11. A monopsony wants to reduce wages as well as employment, Wm and L* in the figure. A union wants to increase wages, but at the cost of lower employment, Wu and L* in the figure. Since both sides want to reduce employment, we can be sure that the outcome will be lower employment compared to a competitive labor market. What happens to the wage, though, is based on the monopsonist's relative bargaining power compared to the union. The actual outcome is indeterminate in the graph, but it will be closer to Wu if the union has more power and closer to Wm if the monopsonist has more power.

Key Concepts and Summary

A bilateral monopoly is a labor market with a union on the supply side and a monopsony on the demand side. Since both sides have monopoly power, the equilibrium level of employment will be lower than that for a competitive labor market, but the equilibrium wage could be higher or lower depending on which side negotiates better. The union favors a higher wage, while the monopsony favors a lower wage, but the outcome is indeterminate in the model.

Unit-IV

GENERAL EQUILIBRIUM ANALYSIS

General Equilibrium Theory

A. Interdependence in the Economy:

We have adopted a partial equilibrium approach, concentrating on decisions in a particular segment of the economy in isolation of what was happening in other segments, under the ceteris paribus assumption. We examined the utility-maximizing behaviour of a household under the assumption that its income was given, although income depends on the amount of labour and other factors of production that the consumer owns and on their prices. The ceteris paribus assumption was useful in that it enabled us to study the individual's demand for different commodities in isolation from influences arising from other parts of the economy. We studied the production decision of a firm on the assumption that factor prices, the state of technology and the prices of commodities were given. The ceteris paribus assumption allowed us to study the cost-minimization behaviour of a firm in isolation from such factors as the demands for the products, which in turn are influenced by the level of employment, income and tastes of consumers.

Product markets, where buyers and sellers interacted with each other and among themselves to determine prices and levels of outputs of various commodities, were studied under the ceteris paribus assumption; relationships with other markets were ignored. Factor markets, where firms and households as owners of productive resources interacted with each other and among themselves to determine prices and quantities of various factors employed, were also analysed on the basis of the ceteris paribus assumption. The interrelationship between the various factor markets and commodity markets were left out of the analysis. In summary, the basic characteristic of a partial equilibrium approach is the determination of the price and quantity in each market by demand and supply curves drawn on the ceteris paribus clause. Each market in the Marshallian methodology is regarded independently of the others.

However, a fundamental feature of any economic system is the interdependence among its constituent parts. The markets of all commodities and all productive factors are interrelated, and the prices in all markets are simultaneously determined. For example, consumers' demands for various goods and services depend on their tastes and incomes. This circular interdependence of the activity within an economic system can be illustrated with a simple economy composed of two sectors, a consumer sector, which includes households and a business sector, which includes firms.

It is assumed that:

(a) All production takes place in the business sector;

(b) All factors of production are owned by the households;

(c) All factors are fully employed;

(d) All incomes are spent.

The economic activity in the system takes the form of two flows between the consumer sector and the business sector: a real flow and a monetary flow (figure 22.1).

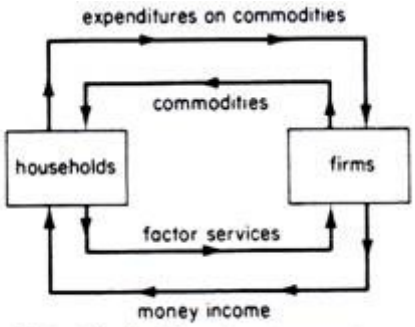


Figure 22.1 Circular flows in a two-sector economy

The real flow is the exchange of goods for the services of factors of production: firms produce and offer final goods to the household sector, and consumers offer to firms the services of factors which they own. The monetary flow is the real flow expressed in monetary terms. The consumers receive income payments from the firms for offering their factor services. These incomes are spent by consumers for the acquisition of the finished goods produced by the business sector. The expenditures of firms become the money incomes of the households. Similarly, the expenditures of

households become the receipts of firms, which they once again pay the households for the factor services which they supply.

The real flow and the monetary flow, which represent the transactions and the interdependence of the two sectors, move in opposite directions. They are linked by the prices of goods and factor services. The economic system is in equilibrium when a set of prices is attained at which the magnitude of the income flow from firms to households is equal to the magnitude of the money expenditure flow from households to firms. The interdependence of markets is concealed by the partial equilibrium approach. Markets consist of buyers and sellers. Thus an economic system consists of millions of economic decision-making units who are motivated by self-interest. Each one pursues his own goal and strives for his own equilibrium independently of the others. In traditional economic theory the goal of a decision-making agent, consumer or producer, is maximisation of something.

The consumer maximises satisfaction subject to a budget constraint. The firm maximizes profit, subject to the technological constraint of the production function. A worker determines his supply of labour by maximising satisfaction derived from work-leisure opportunities, subject to a given wage rate. The problem is to determine whether the independent, self-interest motivated behaviour of economic decision-makers is consistent with each individual agent's attaining equilibrium. All economic units, whether consumers, producers, or suppliers of factors, are interdependent. General equilibrium theory deals with the problem of whether the independent action by each decision-maker leads to a position in which equilibrium is reached by all. A general equilibrium is defined as a state in which all markets and all decision-making units are in simultaneous equilibrium. A general equilibrium exists if each market is cleared at a positive price, with each consumer maximising satisfaction and each firm maximising profit.

The scope of general equilibrium analysis is the examination of how this state can, if ever, be reached, that is, how prices are determined simultaneously in all markets, so that there is neither excess demand nor

excess supply, while at the same time the individual economic units attain their own goals. The interdependence between individuals and markets requires that equilibrium for all product and factor markets as well as for all participants in each market must be determined simultaneously in order to secure a consistent set of prices. General equilibrium emerges from the solution of a simultaneous equation model, of millions of equations in millions of unknowns.

The unknowns are the prices of all factors and all commodities and the quantities purchased and sold (of factors and commodities) by each consumer and each producer. The equations of the system are derived from the maximising behaviour of consumers and producers, and are of two types: behavioural equations describing the demand and supply functions in all markets by all individuals, and clearing-the-market equations. In principle a simultaneous-equation system has a solution if the number of independent equations is equal to the number of unknowns in the system. This approach has been followed by the founder of general equilibrium analysis Leon Walras.

B. The Walrasian System:

The most ambitious general equilibrium model was developed by the French economist Leon Walras (1834-1910). In his Elements of Pure Economics Walras argued that all prices and quantities in all markets are determined simultaneously through their interaction with one another. Walras used a system of simultaneous equations to describe the interaction of individual sellers and buyers in all markets, and he maintained that all the relevant magnitudes (prices and quantities of all commodities and all factor services) can be determined simultaneously by the solution of this system. In the Walrasian model the behaviour of each individual decision-maker is presented by a set of equations. For example, each consumer has a double role: he buys commodities and sells services of factors to firms. Thus for each consumer we have a set of equations consisting of two subsets: one describing his demands of the different commodities, and the other his supplies of factor inputs. Similarly, the behaviour of each firm is presented by a set of equations with two subsets one for the quantities of commodities

that it produces, and the other for the demand for factor inputs for each commodity produced. The important characteristic of these equations is their simultaneity or interdependence. The solution of this system of millions of simultaneous equations defines the 'unknowns' of the model, namely the prices and quantities of all commodities and all factor inputs.

In a general equilibrium system of the Walrasian type there are as many markets as there are commodities and factors of production. For each market there are three types of functions demand functions, supply functions and a 'clearing-the-market' equation, which stipulates that the quantities demanded be equal to the quantities supplied. In a commodity market the number of demand functions is equal to the number of consumers, and the number of the supply functions is equal to the number of firms which produce the commodity. In each factor market the number of demand functions is equal to the number of firms multiplied by the number of commodities they produce. The number of supply functions is equal to the number of consumers who own (ex hypothesis) the factors of production.

A necessary condition for the existence of a general equilibrium is that there must be in the system as many independent equations as the number of unknowns. Thus the first task is to describe the economy by means of a system of equations, defining how many equations are required to complete (and solve) the system. For example, assume that an economy consists of two consumers, A and B, who own two factors of production, K and L These factors are used by two firms to produce two commodities, X and Y. It is assumed that each firm produces one commodity, and each consumer buys some quantity of both. It is also assumed that both consumers own some quantity of both factors. Since the number of equations is equal to the number of unknowns, one should think that a general equilibrium solution exists. Unfortunately, the equality of numbers of equations and unknowns is neither a sufficient nor a necessary condition for the existence of a solution. In the Walrasian system one of the equations is not independent of the others there is a 'redundant equation' in the system which deprives the system of a solution, since the number of unknowns is larger than the

number of independent equations. In this model the absolute level of prices cannot be determined.

General equilibrium theorists have adopted the device of choosing arbitrarily the price of one commodity as a numeraire and express all other prices in terms of the price of the numeraire. With this device prices are determined only as ratios: each price is given relative to the price of the numeraire. If we assign unity to the price of the numeraire, we attain equality of the number of simultaneous equations and unknown variables. However, the absolute prices are still not determined: they are simply expressed in terms of the numeraire. This indeterminacy can be eliminated by the introduction explicitly in the model of a money market, in which money is not only the numeraire, but also the medium of exchange and store of wealth.

Even if there is equality of independent equations and unknowns, there is no guarantee that a general equilibrium solution exists. The proof of the existence of a general equilibrium solution is difficult. Leon Walras was never able to prove the existence of a general equilibrium. In 1954 Arrow and Debreu provided a proof of the existence of a general equilibrium in perfectly competitive markets, in which there are no indivisibilities and no increasing returns to scale. Furthermore, in 1971 Arrow and Hahn proved the existence of a general equilibrium for an economy with limited increasing returns and monopolistic competition, without indivisibilities. Both proofs are limited to specific market structures and are based on restrictive assumptions, regarding in particular the necessity of 'well-behaved' continuous production and demand functions.

Thus the available 'existence proofs' do not hold for the typical real world cases of discontinuities and indivisibilities in production processes. Our current state of knowledge does not enable us to be sure of the existence of a general equilibrium in the real world, which is dominated by oligopolistic firms and production processes which are characterised by indivisibilities. However, the proof of the existence of general equilibrium for a perfectly competitive economy (with no indivisibilities and no increasing returns to scale) is very important, because a perfectly competitive system

has certain ideal properties: it results in an efficient allocation of resources. Apart from the existence problem, two other problems are associated with equilibrium the problem of its stability and the problem of its uniqueness.

Three problems arise in connection with a general equilibrium

1. Does a general equilibrium solution exist? (Existence problem)
2. If an equilibrium solution exists, is it unique? (Uniqueness problem)
3. If an equilibrium solution exists, is it stable? (Stability problem)

These problems can best be illustrated with the partial-equilibrium example of a demand-supply model. Assume that a commodity is sold in a perfectly competitive market, so that from the utility-maximising behaviour of individual consumers there is a market demand function, and from the profit-maximising behaviour of firms there is a market supply function. An equilibrium exists when at a certain positive price the quantity demanded is equal to the quantity supplied. The price at which Qd = Qs is the equilibrium price. At such a price there is neither excess demand nor excess supply. (The latter is often called negative excess demand.) Thus an equilibrium price can be defined as the price at which the excess demand is zero the market is cleared and there is no excess demand. The equilibrium is stable if the demand function cuts the supply function from above. In this case an excess demand drives price up, while an excess supply (excess negative demand) drives the price down (figure 22.2).

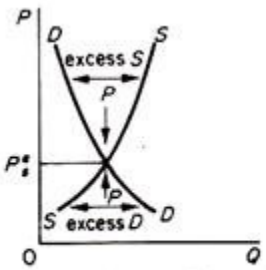


Figure 22.2 Unique, stable equilibrium

The equilibrium is unstable if the demand function cuts the supply function from below. In this case an excess demand drives the price down, and an excess supply drives the price up (figure 22.3).

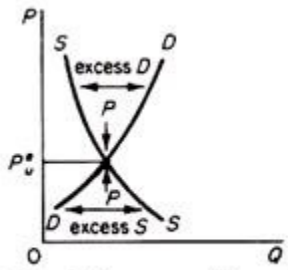


Figure 22.3 Unique, unstable equilibrium

In figure 22.4 we depict the case of multiple equilibria. It is obvious that at P^e_1 there is a stable equilibrium, while at P^e_2 the equilibrium is unstable. Finally in figure 22.5 an equilibrium (at a positive price) does not exist.

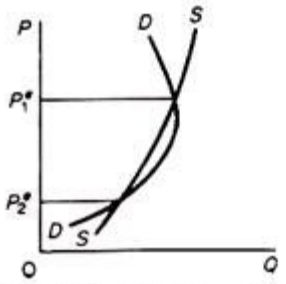


Figure 22.4 Multiple equilibria

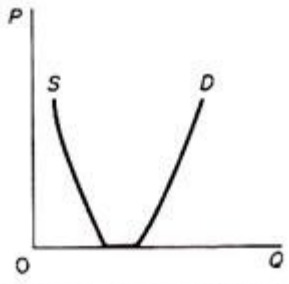


Figure 22.5 No equilibrium exists

It should be clear from the above discussion that (a) the existence of equilibrium is related to the problem of whether the consumers' and producers' behaviour ensures that the demand and supply curves intersect (at a positive price); (b) the stability of equilibrium depends on the relationship between the slopes of the demand and supply curves; (c) the uniqueness of equilibrium is related to the slope of the excess demand function, that is, the curve which shows the difference between Q_D and Q_s at any one price. In fact the three basic questions related to the existence, stability and uniqueness of an equilibrium can be expressed in terms of the excess demand function.

From the redrawn diagrams (in conjunction with the corresponding ones 22.2-22.5) we can draw the following conclusions:

1. The excess demand function, $E_{(P)}$, intersects the vertical (price)-axis when there is an equilibrium, that is, when the excess demand is zero. If $Q_D = Q_s$, then $E_{(P)} = 0$.

2. There are as many equilibria as the number of times that the excess demand curve $E_{(P)}$ intersects the vertical price-axis (figure 22.8).

3. The equilibrium is stable if the slope of the excess demand curve is negative at the point of its intersection with the price-axis (figure 22.6).

4. The equilibrium is unstable if the slope of the excess demand curve is positive at the point of its intersection with the price-axis (figure 22.7).

5. If the excess demand function does not intersect the vertical axis at any one price, an equilibrium does not exist (figure 22.9).

The above analysis of the existence, stability and uniqueness in terms of excess demand functions can be extended to general equilibrium analysis.

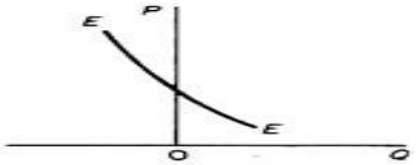


Figure 22.6 Stable equilibrium: slope of $E_{(P)} < 0$

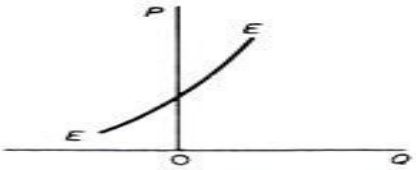


Figure 22.7 Unstable equilibrium: slope of $E_{(P)} > 0$

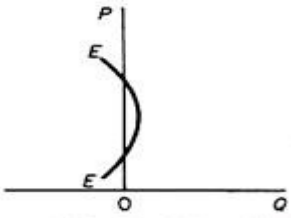


Figure 22.8 Multiple equilibria

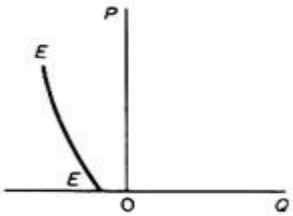


Figure 22.9 No equilibrium exists

D. A Graphical Treatment Of The Two-Factor, Two-Commodity, Two-Consumer (2 X 2 X 2) General Equilibrium Model:

Now we use graphical analysis to show the general equilibrium of a simple economy in which there are two factors of production, two commodities (each produced by a firm) and two consumers. This is known as the 2 x 2 x 2 general equilibrium model. We will restrict our analysis to the perfectly competitive market system, since with free competition it has been proved that a general equilibrium solution exists (given some additional assumptions about the form of the production and demand functions). Furthermore we will be concerned with the static properties of general equilibrium and not with the dynamic process of reaching the state

of such an equilibrium, the latter having been sketched in the preceding section.

Assumptions of the 2 X 2 X 2 Models:

1. There are two factors of production, labour (L) and capital (K), whose quantities are given exogenously. These factors are homogeneous and perfectly divisible.

2. Only two commodities are produced, X and Y. Technology is given. The production functions of the two commodities are represented by two isoquant maps, with the usual properties. The isoquants are smooth and convex to the origin, implying diminishing marginal rate of factor (technical) substitution along any isoquant. Each production function exhibits constant returns to scale. Finally, it is assumed that the two production functions are independent: there are no external economies or diseconomies for the production activity of one product arising from the production of the other.

3. There are two consumers in the economy, A and B, whose preferences are represented by ordinal indifference curves, which are convex to the origin, exhibiting diminishing marginal rate of substitution between the two commodities. It is assumed that consumer choices are independent: the consumption patterns of A do not affect B's utility, and vice versa. Bandwagon, snob, Veblenesque and other 'external' effects are ruled out. Finally, it is assumed that the consumers are sovereign, in the sense that their choice is not influenced by advertising or other activities of the firms.

4. The goal of each consumer is the maximisation of his own satisfaction (utility), subject to his income constraint.

5. The goal of each firm is profit maximisation, subject to the technological constraint of the production function.

6. The factors of production are owned by the consumers.

7. There is full employment of the factors of production, and all incomes received by their owners (A and B) are spent.

8. There is perfect competition in the commodity and factor markets. Consumers and firms pursue their goals faced by the same set of prices (P_x, P_y, w, r).

In this model a general equilibrium is reached when the four markets (two commodity markets and two factor markets) are cleared at a set of equilibrium prices (P_x, P_y, w, r) and each participant economic agent (two firms and two consumers) is simultaneously in equilibrium.

The general equilibrium solution thus requires the determination of the values of the following variables:

The total quantities of the two commodities X and Y, which will be produced by firms and bought by the consumers.

The allocation of the given K and L to the production of each commodity (K_x, K_y, L_x, L_y).

The quantities of X and Y which will be bought by the two consumers (X_A, X_B, Y_A, Y_B).

The prices of commodities (P_x and P_y) and of the factors of production (wage w, and rental of capital r).

The distribution of factor ownership between the two consumers (K_A, K_B, L_A, L_B). The quantities of factors multiplied by their prices define the income distribution between A and B, and hence their budget constraint.

General equilibrium and the allocation of resources:

In figure 22.26 the general equilibrium solution is shown by points T (on the production possibility curve) and T (on the Edgeworth contract curve). These points define six of the 'unknowns' of the system, namely the quantities to be produced of the two commodities (X_e and Y_e), and their distribution among the two consumers (X^A_e, X^B_e, Y^A_e, Y^B_e). We examine the determination of the allocation of resources between X and Y. The determination of the remaining unknowns (prices of factors and commodities, and the distribution of income between the two consumers) is examined in two separate sections below. Point T on the production transformation curve (figure 22.26) defines the equilibrium product mix Y_e and X_e. Recalling that the PPC is the locus of points of the Edgeworth contract curve of production mapped on the product space, point T corresponds to a given point on this contract curve, say T" in figure 22.28. Thus T" defines the allocation of the given resource endowments in the production of the general equilibrium commodity mix. The production of

X_e absorbs L_x of labour and K_x of capital, while Y_e employs the remaining quantities of factors of production; L_y and K_y. Thus four more 'unknowns' have been defined from the general equilibrium solution.

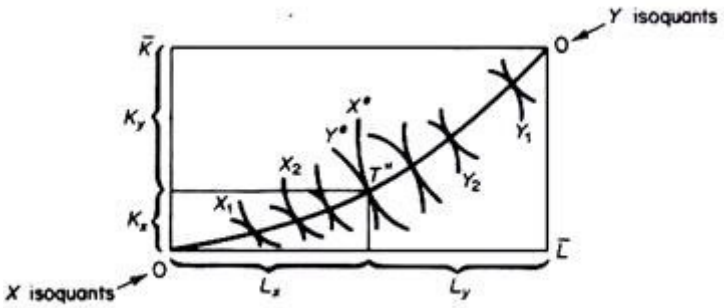


Figure 22.28 Allocation of resources to the production of X_e and Y_e

Prices of commodities and factors:

The next step in our analysis is to show the determination of prices in the general equilibrium model, under perfect competition. In the simple 2 x 2 x 2 model there are four prices to be determined, two commodity prices, P_x and P_y, and two factor prices, the wage rate w, and the rental of capital r. We thus need four independent equations. However, given the assumptions of the simple model, we can derive only three independent relations.

1. Profit maximisation by the individual firm implies least-cost production of the profit-maximising output. This requires that the producer adjusts his factor mix until the MRTS of labour for capital equals the w/r ratio, $MRTS^x_{L,k} = w/r = MRTS^y_{L,K}$ (5), In other words the individual producer maximises his profit at points of tangency between the isoquants and isocost lines whose slope equals the factor price ratio.

2. In perfect factor and output markets the individual profit-maximising producer will employ each factor up to the point where its marginal physical product times the price of the output it produces just equals the price of the factor

$w = (MPP_{L,X}).(P_X) = (MPP_{L,y}).(P_y)$ (6), $r = (MPP_{K,x}) . (P_x) = (MPP_{k,y}) . (P_y)$ (7)

3. The individual consumer maximises his utility by purchasing the output mix which puts him on the highest indifference curve, given his income constraint. In other words maximisation of utility if attained when the budget line, whose slope is equal to the ratio of commodity prices P_x/P_y, is

tangent to the highest utility curve, whose slope is the marginal rate of substitution of the two commodities, $MRS^{A}_{y,X} = P_y/P_x = MRS^{B}_{y,x}$ (8)
Although we have four relations between the four prices, one of them is not independent. Because, dividing (6) and (7), we obtain, Which is the same as expression (5). Thus we have three independent equations in four unknowns. Apparently the absolute values of w, r, P_x and P_y are not uniquely determined (although the general equilibrium solution is unique). Prices in the Walrasian system are determined only up to a ratio or a scale factor. We can express any three prices in terms of the fourth, which we choose arbitrarily as a numeraire or unit of account. For example assume that we choose P_x as the numeraire.

The terms in brackets are known values, that is, values determined by the general equilibrium solution and the maximising behaviour of economic decision-makers with a given state of technology and given tastes. Note that any good can serve as numeraire, and the change of numeraire leaves the relative prices unaffected. We can also assign any numerical value to the price of the numeraire. For convenience P_x is assigned the value of 1. But if, for example, we choose to set P_x = £b, then the price of y in £ will be, P_y = b. P_y /P_x (pounds), This, however, does not mean that the absolute level of the prices of the system is determined. It simply illustrates the fact that we can assign to the price of the numeraire any value we choose.

The reason that the prices are determined only up to a ratio is that money has not been introduced in the system as a commodity used for transactions or as a store of wealth. In a system with perfect certainty, where, for example, nobody would think of holding money, only relative prices matter. The three equations (13)—(15) establish the price ratios implied by the unique general equilibrium solution, and the absolute values of prices are of no importance. However, the general equilibrium model can be completed by adding one (or more) monetary equation. Then the absolute values of the four prices can be determined. Unless a market for money is explicitly introduced, the price side of the model depends on an endogenous numeraire.

Factor ownership and income distribution:

For the simultaneous equilibrium of production and consumption, consumers must earn the 'appropriate' incomes in order to be able to buy the quantities of the two commodities (X_A, X_B, Y_A, Y_B) implied by point T in figure 22.26. Consumers' income depends on the distribution of factor ownership (quantities of factors which they own) and on factor prices. We saw in the preceding paragraph that the prices of factors are determined only up to a ratio. This, however, is adequate for the required income distribution, if the ownership of factors by A and B is determined. For this purpose we require four independent relations, given that we have four unknowns (K_A, K_B, L_A, L_B).

From the assumption of constant returns to scale we can make use of Euler's 'product exhaustion theorem'. This postulates that, with constant returns to scale, the total factor income is equal to the total value of the product of the economy (in perfect factor markets, where inputs are paid their marginal product)

Thus we have three independent equations in four unknowns (K_A, K_B, L_A, L_B), whose values cannot be uniquely determined. The general equilibrium solution does not give absolute values for the distribution of ownership of the factors and money incomes between consumers A and B. This indeterminacy can be resolved only partly if one fixes arbitrarily the value of one of the four factor endowments, and then allocate the remaining three so as to make the individual incomes of A and B such as to lead them willingly to the consumption pattern implied by point T in figure 22.26. It should be clear that different distribution of resources among the two consumers can result in different product combinations, that is, different general equilibrium solutions.

The conclusion of this paragraph may be summarised as follows. The general equilibrium solution defines the total value of the product in the economy. With constant returns to scale this value is equal to the total income of the consumers. However, the individual incomes of A and B are not uniquely determined endogenously. One has to make a consistent assumption about the factor ownership distribution among the two

consumers, so that their incomes are compatible with the purchasing pattern of X_e and Y_e implied by the general equilibrium solution (T and T in figure 22.26). It should be stressed that the above result of factor and income distribution follows from the assumption of fixed amounts of L and K owned by the consumers and supplied to the firms irrespective of prices. The factor supplies did not depend (in this simple model) on the prices of factors and the prices of commodities. The model could be solved simultaneously for input allocations among X and Y, total output mix and commodity-distribution between the two consumers, and only subsequently could we superimpose on this solution the ownership of factors and money-income distribution problem.

E. Concluding Remarks:

There are several reasons why the study of general equilibrium theory is important.

First:

General equilibrium theory, despite its obvious shortcomings, is the most complete existing model of economic behaviour. General equilibrium theory, by viewing the economy as a vast system of mutually interdependent markets, makes the student aware of the tremendous complexity of the real world. At its present stage, general equilibrium theory is largely non-operational and unrealistic. However, the general equilibrium model can be improved so as to become more flexible, more realistic and, hence, more useful for analysing the real world.

Second:

Under certain assumptions the general equilibrium system has a solution: it yields a set of price ratios which lead to an optimal allocation of resources.

Third:

This solution and its optimality properties can be used as a norm to judge the significance and implications of deviations of the various markets from this 'ideal' state of equilibrium.

Fourth:

General equilibrium theory can be helpful in the resolution of macroeconomic controversies. If two macro models are both consistent with statistical data, then one might argue that the model which has closer links to individual optimising behaviour may be considered more nearly correct, since it has a better grounding in the wider body of traditional economic knowledge. At the present time the fourth of the above issues is more important, given the reopened debate between 'Keynesians' and 'Classics'. The participants in this controversy take recourse to general equilibrium theory in an attempt to give more credibility to their positions.

Pareto Optimality: Conditions and Composition

In this article we will discuss about:- 1. Introduction to Pareto Optimality 2. Efficiency in Production 3. Pareto Optimality in Production and Perfect Competition 4. Efficiency in Consumption or Exchange 5. Pareto Optimality in Consumption or Exchange and Perfect Competition 6. Pareto Optimality Conditions when the External Effects are Present and Other Details.

Contents:

1. Introduction to Pareto Optimality:

The welfare of a society depends, in the broadest sense, upon the satisfaction levels of all its consumers. But almost every change in the economic state of the society will have favourable effects on some members and unfavourable effects on others. Evaluation of such a social change is impossible unless the economist is ready to go into interpersonal

comparison of utility under some value judgement, which he may not be willing to do. Rather, he will be willing to evaluate such changes where at least one person has been better off and no one worse off.

The Italian economist Vilfredo Pareto (1848-1923) said that if a change in the economic state makes at least one individual better off without making anyone worse off, then the change is for the betterment of social welfare, i.e., the change is desirable. In that case, we say that the initial state was Pareto-non-optimal. On the other hand, if a change makes no one better off and at least one worse off, implying that the change will make the society worse off, then, from the point of view of welfare, the initial economic state is Pareto-optimal.

Therefore, the Pareto optimality criterion can be stated in this way:

A situation in which it is impossible to make any one better off without making someone worse off, is said to be Pareto optimal or Pareto-efficient. Obviously, the concept of Pareto optimality avoids interpersonal comparison of utility. Since most government policies involve changes in the economic state, which benefit some people and bring discomforts to others, it is obvious that the concept of Pareto optimality is of limited applicability in the real world situations.

Pareto Optimality Conditions:

For the attainment of Pareto-efficient situation in an economy, three marginal conditions must be satisfied.

These are:

1. Marginal condition for efficiency in the allocation of factors among firms (efficiency in production);
2. Marginal condition for efficiency of distribution of commodities among consumers (efficiency in consumption); and
3. Marginal condition for efficiency in the allocation of factors among commodities (efficiency in product-mix or composition of output).

Assumption:

In order to derive these three marginal conditions for the attainment of Pareto optimality, we shall assume, for the sake of simplicity, that there

are only two consumers (I and II), two factors of production (X_1 and X_2). and two commodities (Q_1 and Q_2), i.e., our model here would be a 2 x 2 x 2 model.

2. Efficiency in Production:

If we assume that the consumer goods are of "more is better" type and that external effects are absent in consumption, then an increment in the quantity produced of at least any one consumer good without a decrement in the quantity of any other, can lead to an improvement in utility level of at least one consumer without utility decrements for others.

Therefore, Pareto optimality in production requires that the output level of each consumer good be at a maximum, given the output levels of all other consumer goods. We may derive the marginal condition for Pareto-efficiency in production with the help of Fig. 21.1 which is called an Edge-worth box diagram. The dimensions of the rectangle in Fig. 21.1 represent the total available quantities, and x^0_2, of the inputs X_1 and X_2 that would all be used to produce the consumer goods Q_1 and Q_2. Any point in the box represents a particular allocation of the inputs over the production of the two goods.

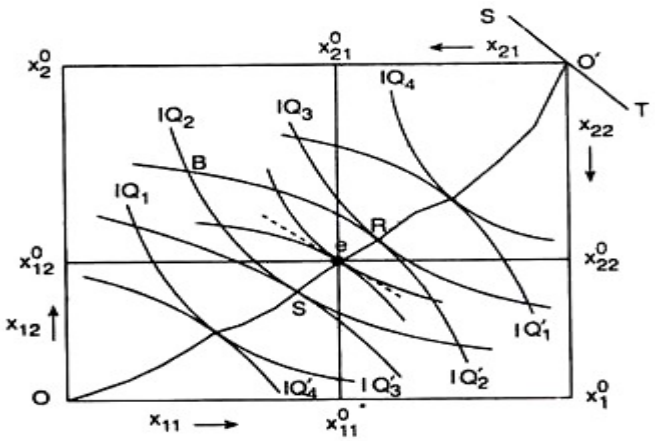


Fig. 21.1 Edgeworth contract curve for production

For example, if the allocation of the inputs is given by the point B, the quantities of X_1 and X_2 used in the production of good Q_1 are measured by the coordinates of B with reference to the origin O, and the quantities of X_1 and X_2 used in the production of good Q_2 are measured by the coordinates of point B with reference to the origin O'. The isoquant (IQ) maps for goods Q, and Q_2 are given in Fig. 21.1 with reference to the points of origin O and O', respectively.

Now, the marginal condition for Pareto efficiency in production would be obtained if we maximise the output of good Q_1 subject to a given output level of good Q_2. Such maximisation would occur at a point of tangency between the IQs for the two goods. For example, maximisation of output of Q_1 subject to the quantity of Q_2 as given by IQ3, would occur at the point of tangency S between the IQs for the goods. Similarly, maximisation of output of Q_2 subject to the quantity of Qi as given by IQ_3, would occur at the point of tangency R between the IQs for the two goods. However, at the point of tangency between the IQs for the two goods, we have numerical slope of IQ for good Q_1 = numerical slope of IQ for good $Q_{2.m}$ $MRTS_{X1,\ X2}$ or, in the production of Q_1 = $MRTS_{X1,X2}$ in the production of Q_2 (21.1).

Thus, the marginal condition for Pareto efficiency in production is given by (21.1) which states that the marginal rate of technical substitution (MRTS) between the two inputs should be the same in the production of the two goods. It is obvious from above that the Pareto efficiency point in production must necessarily be a point of tangency between the IQs for the two goods. If we join all the points of tangency between the IQs for the two goods, by a curve, we would obtain what is called the Edge-worth contract curve for production which we would denote by CCP. The CCP would run from the point O to the point O' in Fig. 21.1.

We have obtained then that all the points on the CCP are Pareto-efficient points in production. That is, if we are at some point on the CCP, then we are no longer able to effect by a change in the allocation of the inputs, an increase in the output of one of the goods without reducing the quantity of the other.

On the other hand, any point like B in Fig. 21.1, which does not lie on the CCP and which does not satisfy condition (21.1), is Pareto-non-optimal. At the point B, we are on IQ_2 for good Q_1 and on IQ'_2 for good Q_2. However, after a reallocation of the resources, if the economy reaches at some point on the CCP between R and S, then the quantities of both the goods would be larger, and if the economy reaches just at the point R or S, then the quantity of one of the goods would be larger and that of the other good would remain the same. This shows that any point B that does not lie on the CCP, is

Pareto-non-optimal, and, by a reallocation of the resources, if the economy is brought on to some point on the segment RS of the CCP, then at least one of the goods would be produced in a larger quantity, that of the other remaining the same.

We have seen that all the points on the CCP are Pareto-optimal. However, we cannot compare any two points, e.g., R and S, on the CCP because if the economy moves from S to R, the output of Q_1 would increase and that of Q_2 would decrease resulting in advantage for some people and disadvantage for some others, and since interpersonal comparison of utility is ruled out, we cannot compare the points R and S.

Mathematical Derivation of the Conditions:

We may also derive mathematically the marginal condition for Pareto efficiency in production.

Let us suppose that the production functions for the goods Q_1 and Q_2 are:

$q_1 = q_1 (x_{11}, x_{12})$, and $q_2 = q_2 (x_{21}, x_{22})$, (2.12), where q_1 and q_2 are the quantities produced of goods Q_1 and Q_2, x_{11} and x_{12} are the quantities of inputs X_1 and X_2 used in the production of Q_1, and x_{21} and x_{22} are the quantities of these inputs used in the production of good Q_2.

Since the total available quantities of the two inputs are x^0_1 and x^0_2, we may write:

$$\begin{aligned} x_{11} + x_{21} &= x_1^0 \\ x_{12} + x_{22} &= x_2^0 \end{aligned} \qquad (21.3)$$

As per the requirements of Pareto optimality, the efficiency conditions may be derived if we maximise q_1 as given by (21.2) subject to:

$$\begin{aligned} q_2 &= q_2^0 \\ \text{or, } q_2 - q_2^0 &= 0 \end{aligned} \qquad (21.4)$$

where q^0_2 is any given quantity of good Q_2.

The relevant Lagrange function for this constrained maximisation problem is:

$$\begin{aligned} Z &= q_1(x_{11}, x_{12}) + \mu \left[q_2(x_{21}, x_{22}) - q_2^0 \right] \\ &= q_1(x_{11}, x_{12}) + \mu \left[q_2(x_1^0 - x_{11}, x_2^0 - x_{12}) - q_2^0 \right] \end{aligned} \qquad (21.5)$$

Pareto efficiency condition (21.1) or (21.7) gives us that the available quantities of the two inputs, X_1 and X_2, should be allocated over the production of the two goods, Q_1 and Q_2, in such a way that the MRTS between the inputs may be the same in the production of the two goods. We may now see with the help of a simple example why condition (21.7) is necessary for Pareto efficiency in production. Let us suppose that in the production of Q_1, $MRTS_{X1},x_2 = 2$ and, in the production of Q_2, $MRTS_{X1},x_2 = 1$, i.e., the MRTS is not the same in the production of the two goods. It follows from above that we can substitute 1 unit of X_1 for 2 units of X_2 in the production of Q_1, and keep the output of Q_1 constant. Similarly, we can substitute 1 unit of X_1 for 1 unit of X_2 in the production of Q_2, and keep the output of Q_2 constant. So, all we have to do is to take 1 unit of X_1 out of the production of Q_2 and use it in the production of Q_1.

This releases 2 units of X_2 from the production of Q_1, 1 unit of which may be transferred to the production of Q_2 to keep its output at the initial level. If we do all this, the output of Q_1 and Q_2 would remain unchanged, and yet we are left with an extra unit of X_2. We can use this unit in the production of Q_1 (or Q_2) and get more of Q_1 (or of Q_2). Thus, one output is increased without reducing the other output. The above example shows that if the $MRTS_{X1, X2}$ in the production of the two goods are not equal, if MRTS in the production of Q_2 is lower, say, than that in the production of Q_1; then we have to take away the marginal unit of input X_1 from the production of Q_2 and transfer it to the production of Q_1 where the $MRTS_{X1,X2}$ is higher, and take away from the field the input X_2, in exchange.

As we continue the process, the MRTS in the production of Q_2 would rise as the quantity of X_1 falls, and the MRTS in the production of Q_1 would fall as the quantity of X_1 increases, and, as we have seen, the allocation becomes better in the Pareto sense. Therefore, if we are to reach the Pareto-efficient situation, we have to continue the process till the MRTS becomes equal in the production of the two goods. For when the MRTS in the production of both the goods becomes the same, no further reallocation will be able to increase the production of at least one of the goods without reducing the production of the other good.

To understand this, let us suppose that the MRTS between the two inputs are equal in the production of the two goods, and it is equal to 4. In that case, if we take away 1 unit of X, from the production of Q_2, and transfer it to the production of Q_1, the latter would release 4 units of X_2 in exchange, so that the output level of Q_1 might remain constant. These 4 units of X_2 should be transferred to the production of Q_2 because there the MRTS is 4, and when 4 units of X_2 are given to be used in the production of Q_2 in exchange for 1 unit of X_1, the output of Q_2 would remain unchanged at the initial level. Therefore, by means of a reallocation of the resources, we have not been able to increase the production of at least one of the goods. On the contrary, a reallocation of the inputs would keep the outputs of the two goods unchanged at their initial quantities.

3. Pareto Optimality in Production and Perfect Competition:

Pareto optimality in production is guaranteed under perfect competition. For, under perfect competition, the prices r_1 and r_2 of the two inputs, X_1 and X_2, are given to the firms that produce the goods Q_1 and Q_2, and each profit-maximising firm equates the $MRTS_{X1},x_2$ to the ratio of the prices of the inputs.

Since condition (21.9) is the same as condition (21.7), Pareto efficiency in production is a certainty under perfect competition. We may now obtain a graphical solution of equation (21.7) or (21.9) for the allocation of inputs X_1 and X_2 over the production of goods Q_1 and Q_2 and for the quantities produced of Q_1 and Q_2. The satisfaction of the marginal condition (21.7), is guaranteed under perfect competition. Let us suppose that in the competitive markets the prices of the inputs are given to be r_1 and r_2. Let us now draw a straight line ST of slope $– r_1/r_2$ through the point O' in Fig. 21.1, and pick up the point e on the contract curve for production (CCP) where the common slope of the isoquants has been equal to the slope of the line ST. That is, at the point e, we have numerical slopes of the IQs of two individuals = the numerical slope of the line ST = r_1/r_2. The marginal condition for efficiency of production has been satisfied. At this point

quantities of the two inputs, x^0_{11} and x^0_{21} would be used in the production of Q_1 and these quantities, when substituted in the production function for Q_1, would give us the output quantity of Similarly, quantities of the two inputs, x^0_{21} and x^0_{22}, would be used in the production of Q_2 and the output here would be q^0_2.

4. Efficiency in Consumption or Exchange:

A distribution of the given quantities of the two commodities Q_1 and Q_2 among two consumers I and II is said to be Pareto-efficient if it is impossible, by a redistribution of these goods, to increase the utility of one individual without $_0$ reducing the utility of the other. The marginal condition for efficiency in consumption or exchange can be derived with the help of the Edgeworth box diagram given in Fig. 21.2. The dimensions of the rectangle in Fig. 21.2 represent the total available quantities, q^0_1 and q^0_2, of the two goods in a pure- exchange economy. Any point in the box represents a particular distribution of the commodities between the two consumers. For example, if the distribution of commodities is given by point A, the quantities of Q_1 and Q_2 consumed by consumer I are measured by the coordinates of A with respect to the origin O and the quantities of the two goods consumed by II are measured by the coordinates of A w.r.t. the origin O'. The indifference map of consumer I has been given w.r.t. the origin O and that of II has been given w.r.t. the origin O'.

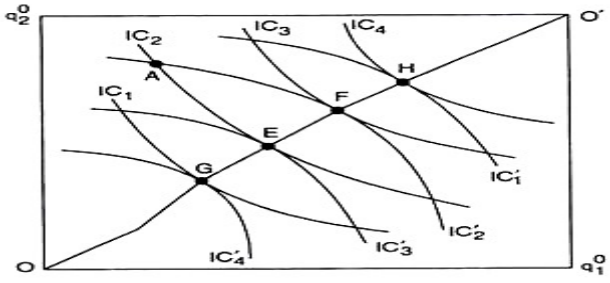


Fig. 21.2 Efficiency in consumption or exchange

Now, the marginal condition for Pareto efficiency in consumption or exchange would be obtained if we maximise the utility level of consumer I or II subject to the given utility level of consumer II or I. Such maximisation would occur at a point of tangency between the indifference curves (ICs) of the two consumers. For example, maximisation of utility of consumer I subject to the utility level of II as given by IC_1 of consumer II, would occur at

the point of tangency, E, between the ICs of two consumers. Similarly, maximisation of utility of consumer II subject to the utility level of I as given by IC_3 of consumer I would occur at the point of tangency, F, between the ICs of the two consumers. It may be added, therefore, that the exchange equilibrium is not unique.

Now, at the point of tangency between the ICs of the two consumers, we have numerical slope of IC of consumer I = numerical slope of IC of consumer II, => $MRS_{Q1,Q2}$ of consumer I = $MRS_{Q1,Q2}$ of consumer II, Thus, the marginal condition for Pareto efficiency in consumption is given by (21.11). It is obvious from above that any point of tangency between the ICs of two consumers is a Pareto efficiency point. If we join all such points of tangency by a curve in Fig. 21.2, we obtain what is known as the Edgeworth contract curve for consumption or Exchange (CCC or CCE), which would run from the point O to the point O'.

Therefore, all the points on the contract curve at which (21.11) is satisfied are Pareto-efficient points in consumption. For, if we are at some point on the contract curve, in Fig. (21.2), we are not able to effect, by a change in the distribution of the goods, an improvement in the utility of one consumer without reducing the utility of the other. Therefore, let us note again that the point of Pareto efficiency in exchange is not unique. On the other hand, any point like A, which does not lie on the contract curve and which does not satisfy (21.11), is Pareto-non-optimal. At the point, A, consumer I is on his IC_2 and consumer II is on his IC_2.

However, after a redistribution of the commodities, if the consumers are brought at some point on the contract curve between E and F, then both the consumers would benefit for both of them would reach now higher ICs, and if they are brought just at the point E or F, then one of them will benefit, while the utility level of the other will remain the same. This shows that any point A, which does not lie on the CCE, is Pareto-non-optimal and by a redistribution of the commodities, if we bring the consumers on to the EF segment of the CCE, then at least one of them would benefit, the utility level of the other remaining the same. We have seen that all points on the contract curve are Pareto-efficient. However, we cannot compare the points

on the contract curve because that will involve interpersonal comparison of utility, which is not possible without an explicit value judgement.

The above example shows us that if the MRS of the two individuals are not equal, if the MRS of II is lower, say, than that of I, then we have to take away the marginal unit of good Q, from individual II and give it to I whose MRS is higher, and take away from him good Q_2 in exchange. As we continue the process, the MRS of II would rise as the quantity of Q_1 with him decreases and the MRS of I would decrease as the quantity of Q_1 with him increases, and, as we have seen, the distribution becomes better in the Pareto sense. Therefore, if we are to reach the Pareto-efficient situation, we have to continue the process till MRS of the two persons become equal. For when the MRS of the two persons are equal, no further redistribution will be able to do good to at least one of them without harming the other. To understand this, let us suppose that MRS of both the persons are equal, and it is equal to 4.

In that case, if we take away 1 unit of Q_1 from consumer II and give it to consumer I, the latter would give us 4 units of Q_2 in exchange in order to keep his utility level intact. If we now give these 4 units of to individual II, his utility would assume the initial level. That is, by means of a redistribution of the goods, we have not been able to improve the utility level of at least one of the persons. On the contrary, a redistribution of the goods would keep the individuals on their initial utility levels.

5. Pareto Optimality in Consumption or Exchange and Perfect Competition:

It can be easily shown that Pareto optimality in consumption is automatically achieved under perfect competition. For under perfect competition, the prices P_1 and P_2 of the two goods are given to the consumers, and each utility-maximising consumer equates his MRS of Q_1 for Q_2 to the ratio of the prices of the goods.

6. Pareto Optimality Conditions when the External Effects are Present:

The marginal condition for a Pareto-efficient distribution of given amounts of two goods (Q_1 and Q_2) between the two individuals (I and II) as

given by (21.18) has been obtained on the basis of the assumption that externalities in consumption are absent. We shall now see that if the external effects are present, the Pareto optimality condition in consumption would generally be different from the marginal condition (21.18). Let us assume that the external effects are present in consumption in the sense that the utility level of one consumer depends also on the consumption of another.

In order to derive the conditions for this constrained maximisation, we have to form the

When referring to the "stability" of equilibrium, or the dynamics which establish equilibrium, most Neoclassical economists immediately think of the Walrasian tâtonnemen*t* process. This is, far and away, the most popular form of adjustment. It is what is taught immediately when the theory of "demand-and-supply" is first encountered.

The pioneer Léon Walras (1874) was obsessed with stability, in particular, in providing the mathematical conditions that would ensure that prices will gravitate towards equilibrium. Is such a proof necessary? Hopefully, as the discussion earlier indicated, it is. To the common claim that stability is already self-evident in the concept of equilibrium itself, Walras was to respond that "Nothing is self-evident except axioms, and this is not an axiom." (Walras, 1874: p.470).

It is also important for practical reasons. The entire defense of *laissez-faire* as an economic policy rests on the assumption that markets are "stable" in the sense given above, otherwise the economy might not move towards its celebrated equilibrium on its own. As Walras notes again: "Conversely, the fact that economists have often extended the principle of free competition beyond the limits of its true applicability is proof positive that the principle has not been demonstrated." (Walras, 1874: p.256-7)

Above everything, stability is theoretically important. The constant effort by economists to "solve" for the equilibrium of a system will turn out to be vacuous if that equilibrium is unstable, i.e. if there is no way of getting there, even theoretically. Walras was particularly insistent on this. As he argued: "Now there remains to show -- and this is the essential point -- that

the problem of exchange for which we have just given a theoretical solution is the selfsame problem that is solved empirically on the market by the mechanism of free competition." (Walras, 1874: p.162-3).

The "theoretical" problem Walras refers to in this passage is finding *what* the equilibrium is. For an economist, this is a matter of simple mathematics: when we have all demands and supplies laid out, it is straightforward enough to compute the equilibrium (although, in practice, that tends to be quite complicated). But how does "the market" find it? How does the market "move" the economy towards it? This is the "stability" question. Walras conceived of the "mechanism of free competition" in a precise, but very familiar manner: the Law of Excess Demand. This "Law" claims that prices adjust *in response* to excess quantity demanded. So, if there is an excess demand for apples, the price of apples increases; if there is an excess supply of apples, the price of apples declines. This law is depicted in Figure 1. If we begin at a price above equilibrium, then there is excess supply of that good, and thus its price will fall, if we begin at a price below equilibrium, then there is excess demand and its price will rise,

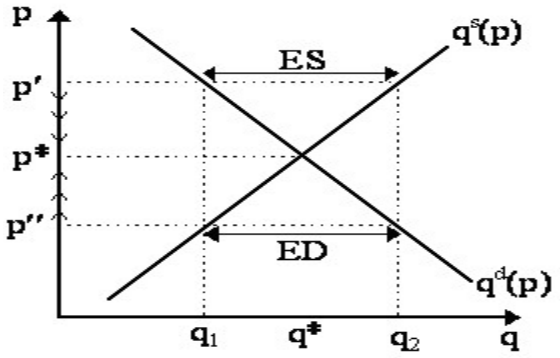


Fig. 1 - The Law of Excess Demand

Paul Samuelson (1941, 1947) has insisted that the "true dynamics" underlying the process must be clarified. In Figure 2, the familiar supply-and-demand diagram is depicted in the right quadrant, the price path over time, p (t), is depicted in the left quadrant. If the price of apples begins at p_0, the quantity demanded is q_0^d and the quantity supplied is q_0^s. There is excess supply in the market for apples, shown by the horizontal difference Z in Figure 2. As a result of this glut, the price begins to decline. As it falls,

the quantity demanded of apples begins to increase and the quantity supplied will be reduced. By time period t, price has gone down from p_0 to p_1, while the quantity demanded has increased from q_0^d to q_1^d and the quantity supplied decreased from q_0^s to q_1^s. However, p_1 is not good enough: there is still excess supply, so the price of apples will continue to fall below that. The price will only stop falling when it hits p*, as then the quantity demanded will be equal to

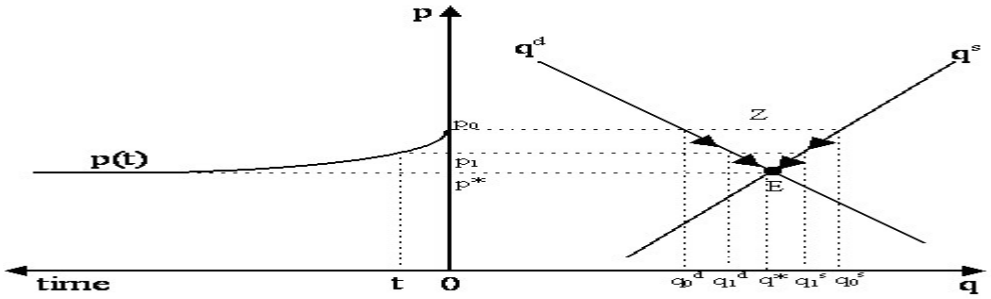


Fig. 2 - Walrasian Dynamics

There are important points not really clarified in this simple "Law". For instance, *who* does the adjusting of prices is not made clear. As Kenneth Arrow notes, "Each individual participant in the economy is supposed to take prices as given and determine his choices as to purchases and sales accordingly; there is no left over whose job it is to make the decision on price." (Arrow, 1959: p.43).

Léon Walras did not really answer Arrow's question directly. Rather Walras went on to argue that we could conceive of a market with free competition as *analogous* to a situation where there is an independent central "crier" or "auctioneer" announcing prices which traders take as "given" and react accordingly with their demands and supplies. Of course, this answer brings forth its own set of questions. Firstly, it is a simple fact that in most markets, this "auctioneer" does not exist. As explained by Kregel (1995), Walras's *tâtonnement* mechanism did have a "real world" counterpart in the operations of the 19th Century Paris *Bourse* (and, as Smith (1987) reminds us, in the modern London Gold Market). But these examples are few and far between. Extending the analogy to markets in general is by no means obvious.

Secondly, even if we did allow the *tâtonnement* analogy to be stretched, we still have not answered what are the underlying motivations and behavioral rules that make it work. As Tjalling Koopmans asks: "If, for instance, the net rate of increase in price is assumed to be proportional to the excess of demand over supply, whose behavior is thereby expressed? And how is that behavior motivated? And is the alternative hypothesis, that the rate of increase in supply is proportional to the excess of demand price over supply price any more plausible, or any better traceable to behaviour motivation?" (Koopmans, 1957: p.179)

Thirdly, pursuing the behavioural question further, if traders understand how the tâtonnement mechanism works, why do they reveal their true demands and supplies when the auctioneer makes his price announcement? A trader could act strategically, say, by holding back or falsely putting forth demands and supplies with the purpose of tricking the auctioneer into bringing down the prices for the goods the trader really wants and raising the prices of those goods he has to sell. Such manipulative behavior is not unknown in real markets. Yet Walras does not (really) discuss the market institutions that prevent strategic behavior.

Fourthly, even supposing that prices *could* be settled upon by such a mechanism, there is still nothing that implies that it is implementable in a decentralized manner, i.e. that there is a sequence of bilateral exchanges that would carry the equilibrium trades through. One would need the contrivance of money or some other institutional mechanism. But, if such details are introduced, would they not implicate "price-groping" phase as well? (cf. Starr, 1989; Ostroy and Starr, 1990).

A tremendous debate has also swirled around whether Walras himself believed that trade between agents occurred *before* or *only after* equilibrium prices were reached by the auctioneer. In the "no-trade-before-equilibrium" version, the "auctioneer" yells out a random set of prices. Traders react to these prices with demands and supplies of various articles. What they present to the auctioneer are *not* the goods themselves, but "*tickets*" (or "*bons*") representing the quantities they *wish* to trade. If the quantities demanded match the

quantities supplied in each market, then an equilibrium has been reached and, with auctioneer's assent, the agents can proceed to trade their articles. But if the quantities demanded and supplied do *not* match, then everything is cancelled and no trade occurs. The auctioneer must try again.

However, tickets were only introduced when Walras was discussing exchange in an economy with *production* (Walras, 1874: Leç. 20). In his discussion of stability in a *pure exchange* economy (Leç. 12), he did not use this artifact. This has led some commentators to argue, therefore, that at least his pure exchange process allowed for out-of-equilibrium trade -- what Hicks called trading at "false prices" (Hicks, 1939: p.128-9). For a flavor of the ensuing debate, see Goodwin (1951), Patinkin (1956), Morishima (1977), Jaffé (1967, 1981), Walker (1987, 1996) and Kompas (1992). Although the no-trade-before-equilibrium version of the story has dominated the modern discussion of stability in a Walrasian system, a small tradition of what are called *non-tâtonnement* mechanisms has pursued stability with out-of-equilibrium trade.

The time has come to begin making Walras's stability theory a bit more mathematically precise. Assuming that the Walrasian *tâtonnement* is the one and only adjustment mechanism in operation, does it actually yield stability? Does an auctioneer, groping around according to the Law of Excess Demand, eventually approach equilibrium?

Léon Walras (1874) did not demonstrate this "properly", i.e. with the mathematics of dynamical systems, but only intuitively. Most of his successors, such as Antonelli, Pareto, etc., failed or did not even attempt to prove it either. The first concrete step in the right direction was John Hicks (1939). The "true" dynamic conditions, in terms of differential equations, were only established later by Paul Samuelson (1941, 1942, 1947), Lloyd Metzler (1945), Kenneth Arrow, H.D. Block and Leonid Hurwicz (1959), Lionel McKenzie (1960) and others.

As this literature on both local stability and global stability in a multi-market scenario is quite complex, we review it separately. For the moment, it suffices perhaps to follow a simple demonstration, in a *single* market. Samuelson's (1941, 1947) "true" dynamics of differential

equations can make the intuition we have pursued thus far more precise. In a single market, let us have linear demand and supply curves:

$q^d(p) = a_0 + ap$

$q^s(p) = b_0 + bp$

where $a < 0$ and $b > 0$, naturally, so the demand curve is downward-sloping and the supply-curve is upward-sloping. Equilibrium price p* is defined where:

$q^d(p^*) = q^s(p^*)$

and this can be solved for our simple linear equations quite simply as:

$p^* = (a_0 - b_0)/(b - a)$

As $(b - a) > 0$ by assumption, then for p* to be positive, it must also be that $a_0 > b_0$ (i.e. horizontal demand intercept is greater than the horizontal supply intercept). Now let the Walras *tâtonnement* process, the Law of Excess Demand, be defined as a differential equation so:

$dp/dt = k[q^d(p) - q^s(p)]$

where k (a "speed of adjustment" parameter) is assumed positive. Thus, this equation relates price changes to excess demand in the market like Walras proposed. Plugging in our equations for demand and supply, we obtain:

$dp/dt + (b - a)kp = k[a_0 - b_0]$

which is a simple first order linear differential equation. The particular integral (i.e. setting $dp/dt = 0$ and solving for p) is actually our equilibrium price, $p^* = (a_0 - b_0)/(b - a)$. The complementary function is simply $(p^* - p(0))e^{-k(b-a)t}$ where p(0) is the initial price, p* the equilibrium price (thus $(p^* - p(0))$ expresses how far we are away from the equilibrium) and e is the natural base. Thus, the solution to the differential equation (the dynamic path of p) is merely: $p(t) = (p^* - p(0))e^{-k(b-a)t} + p^*$ Now, the market is "stable" if p(t) approaches p* over time, i.e. p(t) p* as t . For this to be true, it must be that the entire expression subscripted above the natural base is negative. We know k and t are positive, thus for $-k(b-a)t < 0$, it must be that (b-a) is also positive. But to say that $b > a$, or equivalently, in terms of slopes, $1/b < 1/a$. If our demand-and-supply curves take the "normal" shape we drew them in, then $a < 0$ and $b > 0$ and thus the condition is met. If both curves sloped

upwards, so b, a > 0, then the demand curve must be steeper than supply; if both slope downwards, so b, a < 0, then the supply curve must be steeper.

Social Welfare Function

In this article we will discuss about the social welfare function of economics, explained with the help of a suitable diagram.

The compensation principle only wants to know whether the losers could be compensated; it does not tell us that the losers should actually be compensated. It is argued that whether compensation should be provided and in what manner—all these are moral issues.

Another problem with the compensation criteria is that they only compare between a few alternatives to tell us what is the most desirable of these alternatives, but they do- not tell us the state that achieves the maximum possible welfare. Some economists thought, it was first suggested by Abram Bergson in 1938, that the problems with the compensation criteria could be solved by considering a social welfare function. The social welfare function (SWF) is a sort of social indifference map consisting of the social indifference curves (SICs). An SIC gives the various combinations of utilities of the two individuals that comprise the society, that result in the same level of social welfare (W). To show this diagrammatically, let us denote the utilities of the two consumers by U_I and U_{II}.

The SWF then is:

$W = f(U_I, U_{II})$, The SICs have been shown in Fig. 21.7. Each of them shows the different combinations of U_I and U_{II} that give a particular level of social welfare (W). SICs are negatively sloped because as I is made better off, II must be made worse off, to give us the same W. If, at the initial (U_I, U_{II}) combination, both the individuals, or any one of them, are made better off, the utility level of the other remaining the same, then that would result in a higher level of W and the society would move to a higher SIC. In Fig. 21.7, W_2 represents a higher level of social welfare than W_1. Let us note that the SICs need not be convex or concave to the origin. For there is no rule here that as U_I increases U_{II} would fall at a diminishing or at an increasing rate. Once we formulate the SWF and the SICs, we are well equipped to compare

different policies and find out the policies that maximise social welfare subject to the available economic resources.

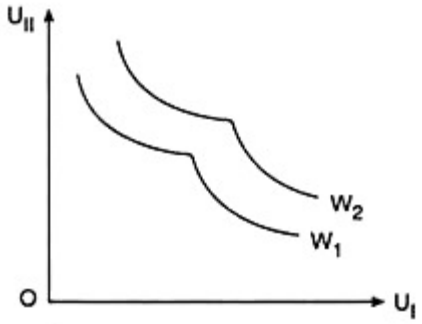


Fig. 21.7 The social indifference curves

But how are we to obtain, or who is to determine, the combinations of U_I and U_{II} that would give the society the same level of welfare, or, who is to determine the combinations that would not give the same—but a lower or higher—level of social welfare? In a dictatorship, the dictator performs this function. Here the SWF and SICs reflect the value judgements of the dictator.

In a democracy, the value judgements must be determined collectively by the members of the society. The individuals can express their value judgements by means of voting. But Arrow pointed out that social welfare could not be evaluated by a democratic vote. This is known as Arrow's Impossibility Theorem. According to Arrow, the social welfare choices should be transitive, i.e., if situation A is preferred to situation B and B is preferred to C, then A is preferred to C. Given the transitivity assumption, let us now consider the following rankings of three policies A, B and C by three individuals I, II and III (the lower number indicating a higher rank).

From the above rankings we obtain: Individuals I and II prefer the policy A to policy B. Thus, a majority vote between the policies A and B will lead to the choice of A. Again I and III prefer B to C. So a majority vote between B and C will lead to the choice of B. Thus, we obtain A is preferred to B and B is preferred C. This would imply, because of transitivity, that A is preferred to C. Therefore, if transitivity holds, we obtain A is preferred to B and A is preferred to C. Therefore, democratic vote gives us that A is the policy that should be selected. However, when we consider A and C, we find

that both II and III prefer C to A. So, the majority vote between A and C will lead to a choice of C, and transitivity will not hold. Thus, democratic votes lead to the choice of all the three policies, i.e., here we arrive at what is known as the voting paradox.

Individuals	Policies		
	A	B	C
I	1	2	3
II	2	3	1
III	3	1	2

The above method of voting by ranks is paradoxical and confusing, and we may come out of it, if account is taken of the intensity of the preferences of different individuals, and a scheme of compensation is made use of. This is actually the idea behind the compensation principle. For example, if the consumer I and II's preference for A is very intense and it is worth, say, Rs 1000 to each, and consumer I and Ill's preference for B, and II and Ill's preference for C, are not so intense, it is worth Rs 100, say, for each of them, then certainly a compensation scheme might be worked out and policy A might be implemented. However, the criterion of SWF and also the Kaldor-Hicks compensation criterion based on potential and not actual compensation require an assumption of omniscience on the part of the individuals evaluating the different policies.

But such an assumption is totally unrealistic, because individual's utilities are highly subjective, and it is very difficult for others to evaluate them. Only actual compensation will help an evaluation. In many instances, however, it is not clear to whom compensation is to be made.

Arrow's Theory of Social Choice

Prof. Arrow, in his monumental work, Social Choice and Individual Values, published in 1951, shows that the task of constructing a social welfare functions to reflect the aims and aspirations of a free democratic society is an impossible one. Arrow has proved a general theorem about the impossibility of constructing an ordering for society as a whole which will in some way reflect all the individual orderings of the members who make up

the society. Arrow's main concern is to consider if a social choice can be satisfactorily derived from individual decisions. The problem is easily solved by dictatorship under which social decisions are made by single individual in small group. In a democratic society, each and every individual will have their idea of social welfare function.

It is therefore difficult to construct a social welfare function which reflects the individual orderings. Therefore Arrow lays down five reasonable conditions which social choices must meet in order to reflect individual's preferences. The first condition may be called universality condition. It states that a definite social ordering is derivable from a reasonably wide range of individual orderings. This social ordering must have the properties of connexity and transitivity. By the axiom of connexity, any two alternatives must be related either by preference or indifference. Thus for any two alternatives X and Y, either X is preferred to Y, on Y is preferred to X or the two are indifferent. By the axiom of transitivity if X is preferred on indifferent to Y and Y is preferred on indifferent to Z, then X must be either preferred on indifferent to Z. These two axioms constitute the foundation of modern choice theory.

The second condition is called Responsiveness condition. It states that social ordering is positively related with the individual orderings. On simply social choices must move in the same direction as individual choices. The third condition is called by Arrow the Independence of Irrelevant Alternatives. It simply, states that the choice made by a society depends on the orderings of individuals in that environment and not on the orderings of alternatives outside that environment. Fourth is the non-imposition condition also called the conditions of citizens' sovereignty? It required that there should be no external control over a society's choice. The social welfare function is not to be imposed.

Condition number 5, called, the condition of Non-dictatorship is a part of the condition 4. It permits the construction of social choices by collective methods and not by dictatorial ones. Hence the condition is the social welfare function is not to be dictatorial. Condition 1 specifies the scope of social welfare function and other four are value judgements. Arrow next

considers whether a social ordering can be derived from any set of individual orderings. He demonstrates that impossibility of doing this without violating at-least one of the value judgements as expressed in five conditions. This is his “General Possibility Theorem”.

General Possibility Theorem:

Arrow first considers a simple case where there are only two alternatives and shows that in this case the method of majority decision yields a social welfare function satisfying all the five conditions. But when there are three or more alternatives difficulty emerges and no valued social welfare function can be derived therefore a social welfare function may be either imposed on dictatorial. Arrows offers three important deductions-consequences 1,2 and 3. The three alternatives are X,Y,Z and there are two individuals. Consequence I states that whenever both individuals prefer X to Y, society will prefer X to Y.

Consequence 2 states that, if in four given choice the will of individual 1 prevails against the opposition of 2, then individual 1 prevails against the opposition of 2, then individual 1’s view will prevail if 2 is indifferent on if he agrees with 1. Consequence III states that, if two individuals have opposing interests, then the society will be indifferent between the two alternatives. Based on these consequences, the General Possibility Theorem is stated in its simplest form. Let there be two individuals and three alternatives X,Y,A. If individual 1 prefers X to Y and individual 2, Y to X, the society is indifferent between the two.

If individual 1 has X,Y,Z and individual 2 has the ordering Z,X,Y. Since individual one prefers Y to Z and individual 2 prefers Z to Y, the society should be indifferent between the two. For both X is preferred to Y and society also prefers X to Y. By the axiom of transitivity society prefers X to Z. By since individual 1 prefers X to Z and 2 prefers Z to X we are to concluded that the society is indifferent as between X and Z. But this contradicts the earlier conclusion that X is preferred to Z. it can’t be that for society X is both preferred and also indifferent to Z.

The Theory of the Second Best

LEARNING OBJECTIVES

Understand the key features of the theory of the second best.

Distinguish between first-best and second-best equilibria.

Distinguish between first-best and second-best policies.

The theory of the second best was formalized by Richard Lipsey and Kelvin Lancaster in 1956. The primary focus of the theory is what happens when the optimum conditions are not satisfied in an economic model. Lipsey and Lancaster's results have important implications for the understanding of not only trade policies but also many other government policies. In this section, we will provide an overview of the main results and indicate some of the implications for trade policy analysis. We will then consider various applications of the theory to international trade policy issues.

First of all, one must note that economic models consist of exercises in which a set of assumptions is used to deduce a series of logical conclusions. The solution of a model is referred to as an equilibrium. An equilibrium is typically described by explaining the conditions or relationships that must be satisfied in order for the equilibrium to be realized. These are called the equilibrium conditions. In economic models, these conditions arise from the maximizing behavior of producers and consumers. Thus the solution is also called an optimum. For example, a standard perfectly competitive model includes the following equilibrium conditions: (1) the output price is equal to the marginal cost for each firm in an industry, (2) the ratio of prices between any two goods is equal to each consumer's marginal rate of substitution between the two goods, (3) the long-run profit of each firm is equal to zero, and (4) supply of all goods is equal to demand for all goods. In a general equilibrium model with many consumers, firms, industries, and markets, there will be numerous equilibrium conditions that must be satisfied simultaneously.

Lipsey and Lancaster's analysis asks the following simple question: What happens to the other optimal equilibrium conditions when one of the conditions cannot be satisfied for some reason? For example, what happens

if one of the markets does not clear—that is, supply does not equal demand in that one market? Would it still be appropriate for the firms to set the price equal to the marginal cost? Should consumers continue to set each price ratio equal to their marginal rate of substitution? Or would it be better if firms and consumers deviated from these conditions? Lipsey and Lancaster show that, generally, when one optimal equilibrium condition is not satisfied, for whatever reason, all the other equilibrium conditions will change. Thus if one market does not clear, it would no longer be optimal for firms to set the price equal to the marginal cost or for consumers to set the price ratio equal to the marginal rate of substitution.

First-Best versus Second-Best Equilibria

Consider a small perfectly competitive open economy that has no market imperfections or distortions, no externalities in production or consumption, and no public goods. This is an economy in which all resources are privately owned, the participants maximize their own well-being, firms maximize profit, and consumers maximize utility—always in the presence of perfect information. Markets always clear and there are no adjustment costs or unemployment of resources. The optimal government policy in this case is laissez-faire. With respect to trade policies, the optimal policy is free trade. Any type of tax or subsidy implemented by the government under these circumstances can only reduce economic efficiency and national welfare. Thus with a laissez-faire policy, the resulting equilibrium would be called *first best*. It is useful to think of this market condition as *economic nirvana* since there is no conceivable way of increasing economic efficiency at a first-best equilibrium.

Of course, the real world is unlikely to be so perfectly characterized. Instead, markets will likely have numerous distortions and imperfections. Some production and consumption activities have externality effects. Some goods have public good characteristics. Some markets have a small number of firms, each of which has some control over the price that prevails and makes positive economic profit. Governments invariably set taxes on consumption, profit, property and assets, and so on. Finally, information is rarely perfectly and costlessly available. Now imagine again a small, open,

perfectly competitive economy with no market imperfections or distortions. Suppose we introduce one distortion or imperfection into such an economy. The resulting equilibrium will now be less efficient from a national perspective than when the distortion was not present. In other words, the introduction of *one* distortion would reduce the optimal level of national welfare.

In terms of Lipsey and Lancaster's analysis, the introduction of the distortion into the system would sever one or more of the equilibrium conditions that must be satisfied to obtain economic nirvana. For example, suppose the imperfection that is introduced is the presence of a monopolistic firm in an industry. In this case, the firm's profit-maximizing equilibrium condition would be to set its price greater than the marginal cost rather than equal to the marginal cost as would be done by a profit-maximizing perfectly competitive firm. Since the economic optimum obtained in these circumstances would be less efficient than in economic nirvana, we would call this equilibrium a second-best equilibrium. Second-best equilibria arise whenever all the equilibrium conditions satisfying economic nirvana cannot occur simultaneously. In general, second-best equilibria arise whenever there are market imperfections or distortions present.

Welfare-Improving Policies in a Second-Best World

An economic rationale for government intervention in the private market arises whenever there are uncorrected market imperfections or distortions. In these circumstances, the economy is characterized by a second-best rather than a first-best equilibrium. In the best of cases, the government policy can correct the distortions completely and the economy would revert back to the state under economic nirvana. If the distortion is not corrected completely, then at least the new equilibrium conditions, altered by the presence of the distortion, can all be satisfied. In either case, an appropriate government policy can act to correct or reduce the detrimental effects of the market imperfection or distortion, raise economic efficiency, and improve national welfare.

It is for this reason that many types of trade policies can be shown to improve national welfare. Trade policies, chosen appropriate to the market circumstances, act to correct the imperfections or distortions. This remains true even though the trade policies themselves would act to reduce economic efficiency if applied starting from a state of economic nirvana. What happens is that the policy corrects the distortion or imperfection and thus raises national welfare by more than the loss in welfare arising from the application of the policy. Many different types of policies can be applied, even for the same distortion or imperfection. Governments can apply taxes, subsidies, or quantitative restrictions. They can apply these to production, to consumption, or to factor usage. Sometimes they even apply two or more of these policies simultaneously in the same market. Trade policies, like tariffs or export taxes, are designed to directly affect the flow of goods and services between countries. Domestic policies, like production subsidies or consumption taxes, are directed at a particular activity that occurs within the country but is not targeted directly at trade flows.

One prominent area of trade policy research focuses on identifying the optimal policy to be used in a particular second-best equilibrium situation. Invariably, this research has considered multiple policy options in any one situation and has attempted to rank order the potential policies in terms of their efficiency-enhancing capabilities.

Thus the ideal or optimal policy choice in the presence of a particular market distortion or imperfection is referred to as a first-best policy. The first-best policy will raise national welfare, or enhance aggregate economic efficiency, to the greatest extent possible in a particular situation. Many other policies can often be applied, some of which would improve welfare. If any such policy raises welfare to a lesser degree than a first-best policy, then it would be called a second-best policy. If there are many policy options that are inferior to the first-best policy, then it is common to refer to them all as second-best policies. Only if one can definitively rank three or more policy options would one ever refer to a third-best or fourth-best policy. Since these rankings are often difficult, third-best (and so on) policies are not commonly denoted.

Trade Policies in a Second-Best World

In a 1971 paper, Jagdish Bhagwati provided a framework for understanding the welfare implications of trade policies in the presence of market distortions.See J. N. Bhagwati, "The Generalized Theory of Distortions and Welfare," in *Trade, Balance of Payments and Growth*, ed. J. N. Bhagwati, R. W. Jones, R. A. Mundell, and J. Vanek (Amsterdam: North-Holland Publishing Co., 1971). This framework applied the theory of the second best to much of the welfare analysis that had been done in international trade theory up until that point. Bhagwati demonstrated the result that trade policies can improve national welfare if they occur in the presence of a market distortion and if they act to correct the detrimental effects caused by the distortion. However, Bhagwati also showed that in almost all circumstances a trade policy will be a second-best rather than a first-best policy choice. The first-best policy would likely be a purely domestic policy targeted directly at the distortion in the market. One exception to this rule occurs when a country is "large" in international markets and thus can affect international prices with its domestic policies. In this case, as was shown with optimal tariffs, quotas, voluntary export restraints (VERs), and export taxes, a trade policy is the first-best policy.

Since Bhagwati's paper, international trade policy analysis has advanced to include market imperfections such as monopolies, duopolies, and oligopolies. In many of these cases, it has been shown that appropriately chosen trade policies can improve national welfare. The reason trade policies can improve welfare, of course, is that the presence of the market imperfection means that the economy begins at a second-best equilibrium. The trade policy, if properly targeted, can reduce the negative aggregate effects caused by the imperfection and thus raise national welfare.

Summary of the Theory of the Second Best

In summary, the theory of the second best provides the theoretical underpinning to explain many of the reasons that trade policy can be shown to be welfare enhancing for an economy. In most (if not all) of the cases in which a trade policy is shown to improve national welfare, the economy begins at an equilibrium that can be characterized as second best. Second-

best equilibria arise whenever the market has distortions or imperfections present. In these cases, it is relatively straightforward to conceive of a trade policy that corrects the distortion or imperfection sufficiently to outweigh the detrimental effects of the policy itself. In other words, whenever market imperfections or distortions are present, it is always theoretically or conceptually possible to design a trade policy that would improve national welfare. As such, the theory of the second best provides a rationale for many different types of protection in an economy.

The main criticism suggested by the theory is that rarely is a trade policy the first-best policy choice to correct a market imperfection or distortion. Instead, a trade policy is second best. The first-best policy, generally, would be a purely domestic policy targeted directly at the market imperfection or distortion. In the remaining sections of this chapter, we use the theory of the second best to explain many of the justifications commonly given for protection or for government intervention with some form of trade policy. In each case, we also discuss the likely first-best policies.

KEY TAKEAWAYS

A first-best equilibrium occurs in a perfectly competitive market when no imperfections or distortions are present.

A second-best equilibrium arises whenever a market includes one or more imperfections or distortions.

A first-best policy is that policy that can improve national welfare to the greatest extent when beginning in a second-best equilibrium.

A second-best policy is one whose best national welfare effect is inferior to a first-best policy when beginning in a second-best equilibrium.

As a general rule of thumb, beginning in a second-best equilibrium, the first-best policy will be a policy that attacks the market imperfection or distortion as directly as possible.

As a general rule of thumb, domestic policies are usually first-best policies, whereas trade policies are usually second-best policies.

One exception to the previous rule of thumb is that a trade policy is the first-best policy choice to correct the imperfection of a large country in international markets.

Unit-V

THEORIES OF RISK AND UNCERTAINTY

Choice Under Uncertainty

1. Subject-matter of choice under uncertainty 2. Describing risk of choice under uncertainty 3. Preference towards Risk 4. Different Preferences towards Risk 5. Reducing Risk 6. Diversification 7. Insurance 8. Value of Information 9. Demand for Risky Assets 10. Assets and other things.

1. Subject-Matter:

Many of the choices that people make involve considerable uncertainty. Sometimes we need to choose between risky ventures. For example, what should we do with our savings? Should we invest in something safe, such as a bank savings account, or something riskier but more lucrative, such as the stock markets? Another example is the choice of a job or a career. Is it better to work for a large, stable company where job security is good but the chances of advancement are limited, or to join a new venture, which offers less job security but quicker advancement?

To answer these questions, we must be able to quantify risk so as to be able to compare the riskiness and alternative choices. Next, we will see how people can deal with risk or reduce risk — by diversification, by buying insurance, etc. or by investing in additional information. In different situations, people must choose the amount of risk they wish to bear. To analyse risk quantitatively, we need to know all possible outcomes of a particular action and the likelihood that each outcome will occur.

2. Describing Risk:

Probability refers to the likelihood that an outcome will occur. Suppose the probability that the oil exploration project is successful might be 1/4, and the probability that it is unsuccessful 3/4. Probability could be objective and subjective. Objective probability relies on the frequency with which certain events have occurred. Suppose we know from our experience that, of the last 100 offshore oil explorations, 1/4 have succeeded and 3/4 have failed. Then the probability of success of 1/4 is objective because it is based on the frequency of similar experiences. But what if there are no similar past experiences to help measure probability? In these cases,

objective measures of probability cannot be obtained, and a more subjective measure is needed. Subjective probability is the perception that an outcome will occur and the perception is based on a person's judgment or experience, but not on the frequency of outcome observed in the past.

Whatever be the interpretation of probability, it is used to calculate two important measures that help us describe and compare risky choices. One measure tells us the expected value and the other variability of the possible outcomes.

Expected Value:

The expected value of an uncertain event is a weighted average of the values associated with all possible outcomes, with the probabilities of each outcome used as weights. The expected value measures the central tendency. Suppose we are considering an investment proposal in an offshore oil company with two possible outcomes: success yields a payoff of £40 per share, while failure yields a payoff of £20 per share.

Variability:

Suppose we are choosing between two sales jobs that have the same expected income (£1,500). The first is based on commission. The second job is salaried. There are two equally likely incomes under the first job — £2,000 for a good sales effort and £1,000 for a moderate effort. The second job pays £ 1510 most of the time, but would pay £510 in severance pay if the business goes burst.

Table 5.1 summarizes these possibilities:

Procedure	*Conventional designations*	*Error variance*
1. Retest with same form on different occasions	Coefficient of stability	Temporal fluctuation
2. Retest with parallel form on different occasion	Coefficient of stability and equivalence	Temporal fluctuation and item specification
3. Retest with parallel form on same occasion	Coefficient of equivalence	Item specificity
4. Split half (odd-even or other parallel splits)	Coefficient of internal consistency	Item specificity
5. Kuder-Richardson (and other measures of inter-item consistency)	Coefficient of internal consistency	Item specificity and heterogeneity

The two jobs have the same expected income because .5 (£2,000) + .5 (£1,000) = .99 (£1,510) + 0.1 (£510) = £1,500. But the variability of the possible payoffs is different for the two jobs. The variability can be analysed by a measure that presumes that large differences between actual payoffs and the expected payoff, called deviations, signal greater risk.

The first job is, thus, substantially more risky than the second as the average deviation of £500 is much greater than the average deviation of £19.80 for the second job. The variability can be measured either by the variance which is the average of the squares of the deviations of the payoffs associated with each outcome from their expected value or by the standard deviation (σ^2) which is the square root of the variance. The standard deviation (a) is the square root of £9,900 or £99.50. We use variance or standard deviation to measure risk, the second job is less risky than the first. Both the variance and the standard deviation of the incomes earned are lower. The concept of variance applies equally well when there are many outcomes rather than just two.

Decision-making:

Suppose we are choosing between the two sales jobs described above. What job should we take? If we dislike risk; we will take the second job. It offers the same expected return as the first but with less risk. Now suppose we add £100 to each of the payoffs in the first job, so that the expected payoff increases from £1,500 to £1,600. Job 1 offers a higher expected income but is substantially riskier than job 2. Which job is preferred depends on us. If we are risk-lovers, we may opt for the higher expected income and higher variance, but a risk-averse person might opt for the second. We need to develop a consumer theory to see how people might decide between incomes that differ in both expected value and in riskiness.

3. Preference towards Risk:

We use the above job example to describe how people might evaluate risky outcomes, but the principles apply equally well to other choices. Here we concentrate on consumer choices generally, and on the utility that consumers derive from choosing among risky alternatives. To simplify

matters, we will consider the consumption of a single commodity, say, the consumer's income. We assume that consumers know probabilities and that payoff are now measured in terms of utility rather than money.

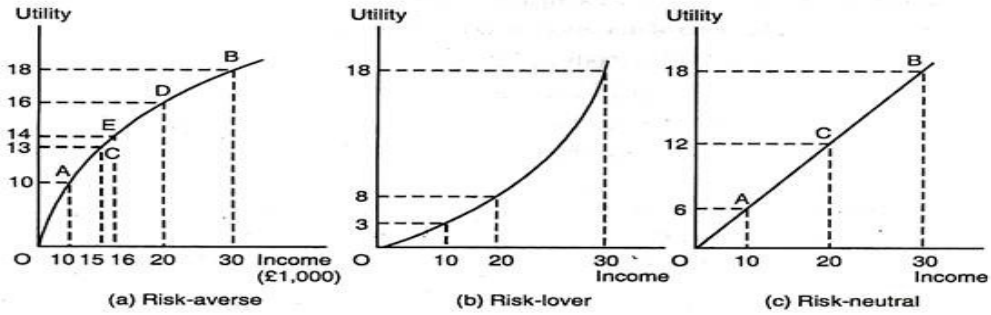


Fig. 5.1 : Risk-aversion

Now, suppose, we have an income of £15,000 and are considering a new but risky job that will either double our income to £30,000 or cause it to fall to £10,000. Each has a probability of 0.5. As Fig. 5.1(a) shows, the utility level associated with an income of £10,000 is 10 (point A), and the utility level associated with a level of £30,000 is 18 (point B). The risky job must be compared with the current job, for which utility is 13 (point C). To evaluate the new job, we can calculate the expected value of the resulting income. Because we are measuring value in terms of utility, we must calculate the expected utility we can get. The expected utility is the sum of the utilities associated with all possible outcomes, weighed by the probability that each outcome will occur. The new risky job is, thus, preferred to the old job because the expected utility of 14 is greater than the original utility of 13. The old job involved no risk — it guaranteed an income of £15,000 and a utility level of 13. The new job is risky, but it offers the prospect of both a higher expected income and a higher expected utility of 14. If we wished to increase our expected utility, we would take the risky job.

4. Different Preferences towards Risk:

People differ in their willingness to bear risk. Some are risk-averse, some risk-lovers and some risk-neutral. A person who prefers a certain given income to a-risky job with the same expected income is known as risk-averse which the most common attitude towards risk is. Most people not only insure against risks — such as, life insurance, health insurance, car

insurance, etc. but also seek occupation with relatively stable wages. Figure 5.1(a) applies to a person who is risk-averse. Suppose a person can have a certain income of £20,000 or a job yielding an income of £30,000 with probability 1/2 and an income of £10,000 with probability 1/2. As we have seen, the expected utility of the uncertain income is 14, an average of the utility at point A (10) and the utility at B (18), and is shown at E. Now we can compare the expected utility associated with the risky job to the utility generated if £20,000 were earned without risk which is given by D (16) in Fig. 5.1(a). It is definitely greater than the expected utility with the risky job E (14).

A person who is risk-neutral is indifferent between earning a certain income and an uncertain income with the same expected income. In Fig. 5.1(c) the utility associated with a job generating an income between £10,000 and £30,000 with equal probability is 12, as is the utility of receiving a certain income of £20,000. Fig. 5.1(b) shows the probability of risk-lover. In this case, the expected utility of an uncertain income that can be £10,000 with probability 1/2 or £30,000 with probability 1/2 is higher than the utility associated with a certain income of £20,000. As shown: E(U) = 1/2U(£10,000) + 1/2V(£30,000) = 1/2(3) + 1/2(18) =10.5 > U(£20,000) = 8. The main evidence of risk-loving is that people enjoy gambling. But very few people are risk-loving with respect to large amount of income or wealth. The risk premium is the amount that a risk-averse person would be willing to pay to avoid risk taking.

The magnitude of the risk premium depends on the risky alternatives that the person faces. The risk premium is determined in Fig. 5.2, which is the same utility function as in Fig. 5.1(a). An expected utility of 14 is achieved by a person who is going to take a risky job with an expected income of £20,000. This is shown in Fig. 5.2 by drawing a horizontal line to the vertical axis from point F, which bisects the straight line AB. But the utility level of 14 can also be achieved if the person has a certain income of £16,000. Thus, the risk premium of £4,000, given by line EF, is the amount of income one would give up to leave him indifferent between the risky job and the safe one.

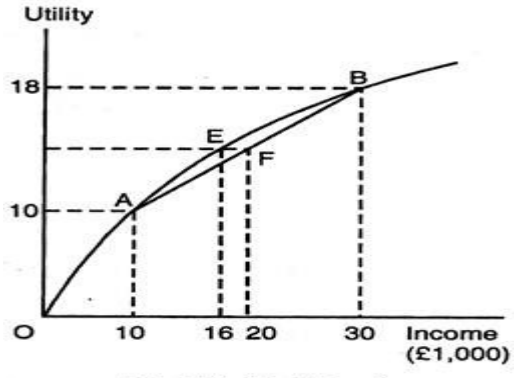


Fig. 5.2 : Risk Premium

How risk-averse a person is depends on the nature of the risk involved and on the person's income. Generally, risk-averse people prefer risks involving a smaller variability of outcomes. We saw that, when there are two outcomes, an income of £10,000 and £30,000 — the risk premium is £4,000. We now consider a second risky job, involving a 0.5 probability of receiving an income of £40,000 and a utility level of 20 and a 0.5 probability of getting an income of 0. The expected value is also £20,000, but the expected utility is only 10.

Expected utility = .5U (£0) + .5U (£40,000) = 0 + .5(20) = 10. Since the utility associated with having a certain income of £20,000 is 16, the person loses 6 units of utility if he is required to accept the job. The risk premium in this case is equal to £10,000 because the utility of a certain income of £10,000 is 10. He can, thus, afford to give up £10,000 of his £20,000 expected income to have a certain income of £10,000 and will have the same level of expected utility. Thus, the greater the variability, the more a person is willing to pay to avoid the risky situation.

5. Reducing Risk:

Sometimes consumers choose risky alternatives that suggest risk-loving rather than risk- adverse behaviour, as the recent growth in state lotteries suggest. Nevertheless, in the face of a broad variety of risky situations, consumers are generally risk-averse. Now we describe three ways in which consumers can reduce risks diversification, insurance, and obtaining more information about choices and payoffs.

6. Diversification:

Suppose that you are risk-averse and try to avoid risky situations as much as possible and you are planning to take a part-time selling job on a commission basis. You have a choice as to how to spend your time selling each appliance. Of course, you cannot be sure how hot or cold the weather will be next year. How should you apportion your time to minimize the risk involved in the sales job? The risk can be minimized by diversification — by allocating time towards selling two or more products, rather than a single product. For example, suppose that there is a fifty-fifty chance that it will be a relatively hot year, and a fifty-fifty chance that it will be relatively cold.
Table 5.3 gives the earnings you can make selling air-conditioners and heaters:

Table 5.3 : Income From Sale of Equipment

	Hot weather	Cold weather
Air-conditioner sales	£30,000	£12,000
Heater sales	£12,000	£30,000

If we decide to sell only air-conditioners or only heaters, our actual income will be either £12,000 or £30,000 and expected income will be £21,000 [.5(£30,000) + .5(£12,000)]. Suppose we diversify by dividing our time evenly between selling air-conditioners and heaters. Then our income will certainly be £21,000, whatever be the weather. If the weather is hot, we will earn £15,000 from air-conditioner sales and £6,000 from heater sales; if it is cold, we will earn £6,000 from air-conditioner sales and £ 15,000 from heater sales. In either case, by diversifying, we assure ourselves a certain income and eliminate all risks. Diversification is not always easy. In our example, whenever the sales of one were strong, the sales of the other were weak. But the principle of diversification has a general application. As long as we can allocate our effort or investment funds towards a variety of activities, whose outcomes are not closely related, we can eliminate some risk.

7. Insurance:

We have seen that risk-averse people will be willing to give up income to avoid risk. If, however, the cost of insurance is equal to the expected loss, risk-averse people will wish to buy enough insurance to offset losses they might suffer. The reasoning is implicit in our discussion of risk-aversion. Buying insurance means a person will have the same income whether or not there is a loss, because the insurance cost is equal to the expected loss. For a risk-averse person, the guarantee of the same income, whatever be the outcome, generates more utility than would be the case if that person had a high income when there is no loss and a low income when a loss occurred. Suppose a homeowner faces a 10% probability that his house will be burglarized and he will suffer a loss of £10,000. Let us assume that he has £50,000 worth of property.

Table 5.4 shows his wealth with two possibilities — to insure or not to insure:

Table 5.4 : Decision to Insure

Insurance	Burglary (Pr = .1)	No Burglary (Pr = .9)	Expected wealth
No	£40,000	£50,000	£49,000
Yes	£49,000	£49,000	£49,000

The decision to purchase insurance does not alter his expected wealth. It does smoothen it out over both possibilities. This generates a high level of expected utility to the house-owner, because the marginal utility in both situations is the same for the person who buys insurance. But when there is no insurance, the marginal utility in the event of a loss is higher than if no loss occurs. Thus, a transfer of wealth from the no-loss to the loss situation must increase total utility. And this transfer of wealth is exactly what is achieved through insurance. Persons usually buy insurance from companies that specialise in selling it. Generally, insurance companies are profit-maximising firms that offer insurance because they know that, when they pool risk, they face very little risk.

This avoidance of risk is based on the law of large numbers, which tells us that although single events may be random and difficult to predict,

the average outcome of many similar events may be predicted. For example, if one is selling automobile insurance, one cannot predict whether a particular driver will have an accident, but one can be reasonably sure, judging from past experience, about how many accidents a large group of drivers will have. By operating on a large scale, insurance companies can be sure that the total premiums paid in will be equal to the total amount of money paid out. In our burglary example, a man knows that there is a 10% probability of his house being burgled; if it is, he will suffer a £10,000 loss. Prior to facing this risk, he calculated his expected loss of £1,000 (£10,000 x 0.1), but this is a substantial risk of loss. Now suppose 100 people face this situation and all of them buy burglary insurance from a company. The insurance company charges each of them a premium of £1,000 which generates an insurance fund of £1, 00,000 from which losses can be paid.

The insurance company can rely on the law of large numbers which assures it that the expected loss for every individual is likely to be met. Thus, the total payout will be close to £1, 00,000 and the company need not worry about losing more than that amount. Insurance companies are likely to charge premiums higher than the expected loss because they need to cover their administrative costs. Thus, many people may prefer to self-insurance rather than buy from an insure company. One way to avoid risk is to self-insure by diversifying.

8. Value of Information:

The decision a consumer makes when outcomes are uncertain is based on limited information. If more information were available, the consumer could reduce risk. Since information is a valuable commodity, people will be prepared to pay for it. The value of complete information is the difference between the expected value with complete information and the expected value with incomplete information. To see the value of information, suppose you are a manager of a store and must decide how many suits to order for the fall season. If you order 100 suits, your cost is £180 per suit, but if you order 50 suits, your cost would be £200. You know you will be selling for £300 each, but you are not sure what total sales would be. All unsold suits could be returned but for half the price you paid for them.

Without further information, you will act on the belief that there is a 0.5 probability that 100 suits will be sold and a 0.5 probability that 50 will be sold.

Table 5.5 gives the profit that you could earn in each of the two cases:

Table 5.5 : Profits from Suits

	Sale of 50	Sale of 100	Expected Profit
1. Buy 50 suits	£5,000	£ 5,000	£5,000
2. Buy 100 suits	£1,500	£12,000	£6,750

Without more information, you would buy 100 suits if you were risk-neutral, taking the chance that your profit might be either £12,000 or £1,500. But if you were risk-averse, you might buy 50 suits for a guaranteed income of £5,000. With complete information, you can make the correct suit order, whatever the sales might be. If sales were going to be 50 suits and you order for 50, you make a profit of £5,000. On the other hand, if sales were going to be 100 and you order for 100, you make a profit of £12,000. Since both outcomes are equally likely, your expected profit with complete information would be £8,500.

The value of information is:

Expected value with complete information	£8,500.00
– Expected value with uncertainty	– £6,750.00
Value of complete information	£1,750.00

Thus, it is worth paying up to £1,750.00 to obtain as accurate information as possible.

9. Demand for Risky Assets:

People are generally risk-averse. Given a choice, they prefer a fixed income to one that is as large on average that fluctuates randomly. Yet many of these people will invest all or part of their savings in stocks, bonds and other assets that carry some risk. Why do risk-averse people invest in risky stocks either all or part of their investment? How do people decide how much risk to bear for the future? To answer these questions, we must examine the demand for risky assets.

10. Assets:

An asset is something that provides a monetary flow to its owner. The monetary flow from owning an asset can take the form of an explicit payment, such as the rental income from an apartment building. Another explicit payment is the dividend on shares. But sometimes the monetary flow from ownership of an asset is implicit; it takes the form of an increase or decrease in the price or value of the asset — a capital gain or a capital loss. A risky asset provides a monetary flow that is in part random, which means, the monetary flow is not known with certainty in advance. A share of a company is an obvious example of a risky asset — one cannot know whether the price of the stock will rise or fall over time, and one cannot even be sure that the company will continue to pay the same dividend per share. Although people often associate risk with the stock market, most other assets are also risky. The corporate bonds are example of this — the corporation that issued the bonds could go bankrupt and fail to pay bond owners their returns. Even long-term government bonds that mature in 10 or 20 years are risky.

Although it is unlikely that government will go bankrupt, the rate of inflation could increase and make future interest payments and the eventual repayment of principal worth less in real terms, and, thus, reduce the value of the bonds. In contrast to risky assets, we can call an asset riskless if it pays a monetary flow that is certain. Short-term government bonds known as Treasury Bills are risk-free assets because they mature within a short period, there is very little risk of an unexpected increase in inflation. And one can also be confident that government will not default on the bond. Other examples of riskless assets include passbook savings accounts in banks and building societies or short- term certificate of deposit.

11. Asset Returns:

People buy and hold assets because of the monetary flows they provide. Assets may be compared in terms of their monetary flow relative to the price of asset. The return on an asset is the total monetary flow it provides as a fraction of its value. For example, a bond worth £1,000 today that pays out £100 this year has a return of 10%. When people invest their

savings in stocks, bonds or other assets, they usually hope to earn a return that exceeds that rate of inflation, so that, by delaying consumption, they can consume more in the future. Thus, we often express the return on an asset in real terms which means return less the rate of inflation. For example, if the annual rate of inflation had been 5%, the bond would have yielded real return of 5%. Since most assets are risky, an investor cannot know in advance what return they are going to yield in future. However, one can compare assets by looking at their expected returns which is just the expected value of its return. In a particular year, the actual return may be higher or lower than expected, but over a long period the average return should be close to the expected return. Different assets have different expected returns.

Table 5.6 shows that the expected real return on Treasury Bills has been less than 1%, while the real return for a representative stock on the London Stock Market has been almost 9%. Why would a person buy a Treasury bill when the expected return on stocks is so much higher? The answer is that the demand for an asset depends not only on expected return, but also on its risk. One measure of risk, the standard deviation (σ) of the real return, is equal to 21.2% for common stock, but only 8.3% for corporate bonds, and 3.4% for Treasury Bills, as Table 5.6 shows. Clearly, the higher the expected return on investment, the greater the risk involved. As a result, a risk-averse investor must balance expected return against risk.

Table 5.6 : Investment-risk and Return

	Real Rate of Return (%)	Risk (Standard Deviation, σ, %)
Common Stocks	8.8	21.2
Long-term corporate bonds	2.1	8.3
Treasury Bills	0.4	3.4

12. Trade-Off between Risk and Return:

Suppose a person has to invest his savings in two assets — riskless Treasury Bills, and a risky representative group of stocks. He has to decide how much of his savings to invest in each of these two assets. This is

analogous to the consumer's problem of allocating a budget between two goods x and y. Let us denote the risk-free return on the Treasury Bill by R_f, where the expected and actual returns are the same. Also, assume the expected return from investing in the stock market is R_m, and the actual return is Y_m. The actual return is risky. At the time of investment decision, we know the likelihood of each possible outcome, but we do not know what particular outcome will occur. The risky asset will have a higher expected return than the risk-free asset ($R_m > R_f$) Otherwise, risk-averse investors would invest only in Treasury Bills and none at all in stocks.

To determine how much he will invest in each asset, let us assume b is the fraction of his savings placed in the stock market, and (1 – b) the fraction used to purchase Treasury Bills. The expected return on his total portfolio, R_p, is a weighted average of the expected return on the two assets $R_p = bR_m + (1 - b)R_f$..... (2), Suppose, the stock market's expected return is 12%. Treasury Bills pay 4%, and b = 1/2. Then R_p = 8%. How risky is this portfolio? The riskiness can be measured by the variance of the portfolio's return. Let us assume the variance of the risky stock market investment is σ^2_m and the standard deviation is σ_m. We can show that the σ of the portfolio is the fraction of the portfolio invested in the risky asset times the o of that asset: $\sigma_p = b\sigma_m$..... (3).

13. Investor's Choice Problem:

To determine how our investor should choose this fraction b, we must first show his risk- return trade-off analogous to the budget line of a consumer. To see this trade-off, we can rewrite equation (2) as

$$R_p = R_f + b(R_m - R_f).$$

From equation (3), we see that $b = \frac{\sigma_p}{\sigma_m}$, so that $R_p = R_f + \frac{(R_m - R_f)}{\sigma_m}\sigma_p$ .....(4)

This is the budget line because it explains the trade-off between risk (σ_p) and the expected return (R_p). The slope $R_m - R_f/\sigma_m$ is constant. The equation says that the expected return on the portfolio R_p increases as the standard deviation of that return σ_p increases. The slope of the budget line is $R_m - R/\sigma_m$, which is the price of risk as shown in Fig. 5.3. Three indifference curves are drawn; each curve shows combinations of risk and

return that have an investor equally satisfied. The curves are upward-sloping because a risk-averse investor will require a higher expected return if he is to bear a greater amount of risk. The utility-maximising investment portfolio is at the point where indifference curve U_2 is tangent to the budget line.

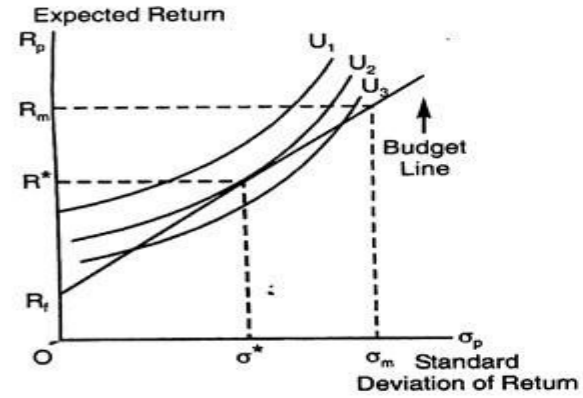


Fig. 5.3 : Choosing between Risk and Return

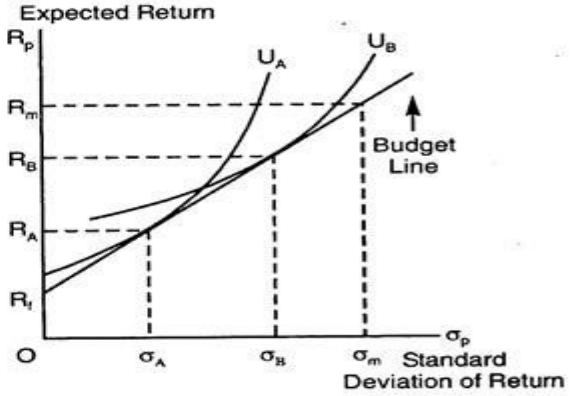


Fig. 5.4 : Choice of two Different Investors

14. Two Different Investors Choice with Different Attitudes to Risk:

Investor A is risk-averse. His portfolio will consist mostly of the risk-free asset, so his expected return, R_A, will be only slightly greater than the risk-free return, but the risk σ_A will be small. Investor B is less risk-averse. He will invest a large fraction of his funds in stocks. The expected return on his portfolio, R_B, will be larger, but the return will also be riskier.

Contingent Consumption

The consumer is presumably concerned with the probability distribution of getting different consumption bundles of goods. A probability distribution consists of a list of different outcomes—in this case, consumption bundles—and the probability associated with each outcome. When a consumer decides how much automobile insurance to buy or how much to invest in the stock market, he is in effect deciding on a pattern of probability distribution across different amounts of consumption. For example, suppose that you have $100 now and that you are contemplating Jouying lottery ticket number 13. If number 13 is drawn in the lottery, the holder will be paid $200. This ticket costs, say, $5. The two outcomes that are of interest are the event that the ticket is drawn and the event that it isn't.

Your original endowment of wealth—the amount that you would have if you did not purchase the lottery ticket—is $100 if 13 is drawn and $100 if it isn't drawn. But if you buy the lottery ticket for $5, you will have a wealth distribution consisting of $295 if the ticket is a winner and $95 if it is not a winner. The original endowment of probabilities of wealth in different circumstances has been changed by the purchase of the lottery ticket. Let us examine this point in more detail. In this discussion we'll restrict ourselves to examining monetary gambles for convenience of exposition. Of course, it is not money alone that matters; it is the consumption that money can buy that is the ultimate "good" being chosen. The same principles apply to gambles over goods, but restricting ourselves to monetary outcomes makes things simpler. Second, we will restrict ourselves to very simple situations where there are only a few possible outcomes. Again, this is only for reasons of simplicity. Above we described the case of gambling in a lottery; here we'll consider the case of insurance. Suppose that an individual initially has $35,000 worth of assets, but there is a possibility that he may lose $10,000. For example, his car may be stolen, or a storm may damage his house. Suppose that the probability of this event happening is p = .01. Then the probability distribution the person is facing is a 1 percent probability of having $25,000 of assets, and a 99 percent probability of having $35,000.

Insurance offers a way to change this probability distribution. Suppose that there is an insurance contract that will pay the person $100 if the loss occurs in exchange for a $1 premium. Of course the premium must be paid whether or not the loss occurs. If the person decides to purchase $10,000 dollars of insurance, it will cost him $100. In this case he will have a 1 percent chance of having $34,900 ($35,000 of other assets — $10,000 loss -f $10,000 payment from the insurance payment - $100 insurance premium) and a 99 percent chance of having $34,900 ($35,000 of assets - $100 insurance premium). Thus the consumer ends up with the same wealth no matter what happens. He is now fully insured against loss.

In general, if this person purchases K dollars of insurance and has to pay a premium 7K, then he will face the gamble: 1

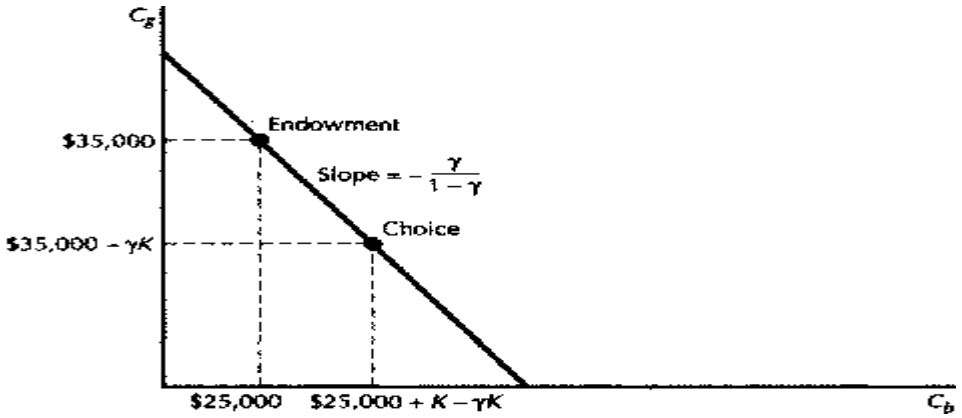
What kind of insurance will this person choose? Well, that depends on his preferences. He might be very conservative and choose to purchase a lot of insurance, or he might like to take risks and not purchase any insurance at all. People have different preferences over probability distributions in the same way that they have different preferences over the consumption of ordinary goods. In fact, one very fruitful way to look at decision making under uncertainty is just to think of the money available under different circumstances as different goods. A thousand dollars after a large loss has occurred may mean a very different thing from a thousand dollars when it hasn't. Of course, we don't have to apply this idea just to money: an ice cream cone if it happens to be hot and sunny tomorrow is a very different good from an ice cream cone if it is rainy and cold. In general, consumption goods will be of different value to a person depending upon the circumstances under which they become available.

Let us think of the different outcomes of some random event as being different states of nature. In the insurance example given above there were two states of nature: the loss occurs or it doesn't. But in general there could be many different states of nature. We can then think of a contingent consumption plan as being a specification of what will be consumed in each different state of nature—each different outcome of the random process. Contingent means depending on something not yet certain, so a contingent consumption plan means a plan that depends on the outcome of some event. In the case of insurance purchases, the contingent consumption was described by the terms of the insurance contract: how much money you would have if a loss occurred and how much you would have if it didn't. In the case of the rainy and sunny days, the contingent consumption would just be the plan of what would be consumed given the various outcomes of the weather.

People have preferences over different plans of consumption, just like they have preferences over actual consumption. It certainly might make you feel better now to know that you are fully insured. People make choices that reflect their preferences over consumption in different circumstances, and we can use the theory of choice that we have developed to analyze those

choices. If we think about a contingent consumption plan as being just an ordinary consumption bundle, we are right back in the framework described in the previous chapters. We can think of preferences as being defined over different consumption plans, with the "terms of trade" being given by the budget constraint. We can then model the consumer as choosing the best consumption plan he or she can afford, just as we have done all along.

Let's describe the insurance purchase in terms of the indifference-curve analysis we've been using. The two states of nature are the event that the loss occurs and the event that it doesn't. The contingent consumptions are the values of how much money you would have in each circumstance. We can plot this on a graph as in Figure 12.1.



The budget line associated with the purchase of insurance. The insurance premium y allows us to give up some consumption in the good outcome (Cg) in order to have more consumption to the bad outcome (C&). Your endowment of contingent consumption is $25,000 in the "bad" state^— if the loss occurs—and $35,000 in the "good" state—if it doesn't occur. Insurance offers you a way to move away from this endowment point. If you purchase K dollars' worth of insurance, you give up 7K dollars of consumption possibilities in the good state in exchange for K — 7K dollars of consumption possibilities in the bad state. Thus the consumption you lose in the good state, divided by the extra consumption you gain in the bad state, is

A Cb K-yK 1-7' this is the slope of the budget line through your endowment. It is just as if the price of consumption in the good state is 1 — 7 and the price in the bad state is 7. We can draw in the indifference curves

that a person might have for contingent consumption. Here again it is very natural for indifference curves to have a convex shape: this means that the person would rather have a constant amount of consumption in each state than a large amount in one state and a low amount in the other. Given the indifference curves for consumption in each state of nature, we can look at the choice of how much insurance to purchase. As usual, this will be characterized by a tangency condition: the marginal rate of substitution between consumption in each state of nature should be equal to the price at which you can trade off consumption in those states. Of course, once we have a model of optimal choice, we can apply all of the machinery developed in early chapters to its analysis. We can examine how the demand for insurance changes as the price of insurance changes, as the wealth of the consumer changes, and so on. The theory of consumer behaviour is perfectly adequate to model behaviour under uncertainty as well as certainty.

We have seen that insurance is a way to transfer wealth from good states of nature to bad states of nature. Of course there are two sides to these transactions: those who buy insurance and those who sell it. Here we focus on the sell side of insurance. The sell side of the insurance market is divided into a retail component, which deals directly with end buyers, and a wholesale component, in which insurers sell risks to other parties. The wholesale part of the market is known as the reinsurance market. Typically, the reinsurance market has relied on large investors such as pension funds to provide financial backing for risks. However, some reinsurers rely on large individual investors. Lloyd's of London, one of the most famous reinsurance consortia, generally uses private investors. Recently, the reinsurance industry has been experimenting with catastrophe bonds, which, according to some, are a more flexible way to provide reinsurance. These bonds, generally sold to large institutions, have typically been tied to natural disasters, like earthquakes or hurricanes.

A financial intermediary, such as a reinsurance company or an investment bank, issues a bond tied to a particular insurable event, such as an earthquake involving, say, at least $500 million in insurance claims. If

there is no earthquake, investors are paid a generous interest rate. But if the earthquake occurs and the Claris exceed the amount specified in the bond, investors sacrifice their principal and interest.

Catastrophe bonds have some attractive features. They can spread risks widely and can be subdivided indefinitely, allowing each investor to bear only a small part of the risk. The money backing up the insurance is paid in advance, so there is no default risk to the insured. Prom the economist's point of view, "cat bonds" are a form of state contingent security, that is, a security that pays off if and only if some particular event occurs. This concept was first introduced by Nobel laureate Kenneth J. Arrow in a paper published in 1952 and was long thought to be of only theoretical interest. But it turned out that all sorts of options and other derivatives could be best understood using contingent securities. Now Wall Street rocket scientists draw on this 50-year-old work when creating exotic new derivatives such as catastrophe bonds.

The Economic Properties of Utility Functions

In this section we discuss how knowledge of the properties of utility functions coupled with partial information on an investor's preferences can provide an insight into the process of rational choice. The expected utility theorem is based on a set of four axioms concerning investor behaviour. The first principle required of a utility function is that is consistent with more being preferred to less. This attribute, known as no satiation, states simply that the utility of more (X + 1) pounds is always higher than the utility of less (X) pounds. Thus, of a choice between alternative investments, an investor will always choose that with the largest expected payoff. Therefore, the first restriction placed on a utility function is that it has a positive first derivative.

The second principle of a utility function is an assumption of an investor's taste for risk. Three assumptions are possible: the investor is either averse to risk, neutral towards risk, or seeks risk. Risk-aversion means that an investor will reject a fair gamble. For example, a certain return of £1 will be preferred to an equal chance of £2 or £0. Risk-aversion implies that the second derivative of utility, with respect to wealth, is

negative. The assumption of risk-aversion means an investor will reject a fair gamble, because the decrease in utility caused by the loss is greater than the increase in utility of an equivalent gain. A risk-neutral investor is indifferent as to whether a fair gamble is undertaken or not and thus implies a zero second derivative. Risk-seeking means that an investor would select a fair gamble, unlike the risk adverse investor; see above. Functions that exhibit greater change in value for larger unit changes in an argument are functions with positive second derivatives. Thus the acceptance of a fair gamble implies a positive second derivative. These conditions may be summarised as follows:

Figure 3.8 below shows preference functions exhibiting alternative properties with respect to risk aversion. Figure 3.8a represents the shape of utility functions - in utility of wealth space - that exhibit risk aversion, risk neutrality, and risk preference. Figure 3.8b represents the shape of the indifference curves in expected return standard deviation space -that would be associated with each of these three types of utility functions. The third principle of a utility function is an assumption as to how investor preferences change with fluctuations in wealth. If an investor increases the amount invested in risky assets as wealth increases, that investor is said to exhibit decreasing absolute risk aversion. If an investor's investment in risky assets remains the same as wealth changes, that investor is said to exhibit constant absolute risk aversion. Finally, if an investor invests less in risky assets as wealth increases, that investor is said to exhibit increasing absolute risk aversion. As previously discussed, different degrees of risk aversion maybe associated with different derivatives of the utility function. A similar result applies to absolute risk aversion. If *U'(W)* and *U''(W)* are the first and second derivatives of the utility function at wealth (level) W, then it has been shown that

$$A(W) = \frac{-U''(W)}{U'(W)}$$

can be used to measure an investor's absolute risk aversion. Thus, *A'(W)*, the derivative of *A(W)* with respect to wealth, is a measure of how absolute

risk aversion behaves with changes in wealth. These conditions may be summarised as follows:

Therefore, if investor preferences towards absolute risk aversion can be defined, the number of possible options required to be considered can be further reduced. Furthermore, this assumption restricts the possible utility functions that can be used to describe preferences. The fourth and final principle used to restrict an investor's utility function is that of the percentage of wealth invested in risky assets - not nominal investment as described above - changing as wealth fluctuates. For example, an investor investing 60% of his wealth in risky assets, whether his wealth is W or $2W$. The investor's behaviour is then said to be characterised by constant relative risk aversion. If as his wealth increases an investor invests a greater percentage in risky investments, he is said to exhibit decreasing relative risk aversion: if he invests a smaller percentage, he is said to exhibit increasing relative risk aversion. Relative risk aversion is closely related to absolute risk aversion: it refers to the change in percentage of investment in risky assets as wealth changes. Absolute risk aversion refers to the change in the absolute amount invested in risky assets as wealth changes. The measure of relative risk aversion has been shown to be[7]

$$R(W) = \frac{-WU''(W)}{U'(W)} = WA(W)$$

If $R'(W)$ is the first derivative of W, then $R'(W) < 0$ indicates that the utility function exhibits decreasing relative risk aversion. If $R'(W) = 0$, than the utility function is said to exhibit constant relative risk aversion. Finally, if $R'(W) > 0$, then the function is said to exhibit increasing relative risk aversion. These conditions may be summarised as follows: There is general agreement in the literature that most investors exhibit decreasing absolute risk aversion. However, there is doubt concerning relative risk aversion. Generally, it is assumed that investors exhibit constant relative risk aversion. However, the justification for this is one of tractability rather than a belief in its descriptive validity. If an investor can nominate the state of relative risk aversion that best describes his preferences, he can again

reduce the number of portfolios to be considered, or further restrict the utility functions that may describe his behaviour.

Expected utility theory is a model that represents preference over risky objects, by weighted average of utility assigned to each possible outcome, where the weights are the probability of each outcome. The primary motivation for introducing expected utility, instead of taking the expected value of outcomes, is to explain attitudes toward risk. Consider for example a lottery, which gives $100 and $0 with even chances, and a sure receipt of $50. Here typically one chooses the sure receipt, whereas the two alternatives yield the same expected return. Another example is the Saint Petersburg paradox. Consider a game of flipping a fair coin until one has a tail. When the number of flips obtained is k, one receives 2^k, which happens with probability $(1/2)^k$. The expected return of this game is which infinity is. However, a typical decision maker is willing to pay only a finite amount for playing this game.

The theory resolves this problem by taking risk attitude into account. Here a risky object is a probability distribution over outcomes, denoted by p. Then the expected utility representation takes the form $U(p) = \Sigma\, u(x_k)\, p_k$ where p_k is the probability that outcome x_k is realized, and function u expresses the utility assigned to each outcome. Notice that $u(x)$ may not be x as it is, and the curvature of u explains the decision maker's risk attitude. When the graph of u is convex to the top, one has the formula $0.5u(100) + 0.5u(0) < u(50)$, which explains the first example (similarly for the second). When this is the case, the decision maker is said to be risk averse. Expected utility theory enables empirical analysis of choice under uncertainty such as financial decision, by quantifying the degree of curvature of u.

The theory originates from Daniel Bernoulli (1700–1782), an eighteenth–century mathematician, and was given an axiomatic foundation by John von Neumann and Oskar Morgenstern in the 1940s. They started from a preference ranking of probability distributions over outcomes, and provided the condition for its expected utility represent ability. The condition consists of three axioms: weak order, continuity, and independence. The

most prominent axiom is independence: when the decision maker prefers distribution p to distribution q, then he or she prefers the distribution made by mixing p and any another distribution r with proportion $\lambda : 1-\lambda$, that is $\lambda p + (1-\lambda) r$, to the distribution made by mixing q and r with the same proportion, that is $\lambda q + (1-\lambda) r$. Here $\lambda p + (1-\lambda) r$ refers to the distribution that assigns probability $\lambda p_k + (1-\lambda) r_k$ on each outcome x_k respectively. Informally speaking, when p is preferred to q then having "p with probability λ and r with probability $1-\lambda$" will be preferred to having "q with probability λ and r with probability $1-\lambda$," since the difference lies only in p and q.

The theory is extended to subjective expected utility theory, where the probabilities are not given objectively, but the decision maker is to hold a subjective belief over relevant events. Various criticisms to the expected utility theory motivate further developments, two of which are explained in this entry. The first criticism is that the independence axiom may be violated systematically, which is referred to as the Allais paradox. Consider for example a bet, which gives $120.00 with probability 0.9 and $0 with 0.1, and a sure receipt of $100.00. The typical choice here is to take the sure receipt. Now consider two bets, one gives $120.00 with probability 0.45 and $0 with 0.55, the other gives $100.00 with probability 0.5 and $0 with 0.5. Here the typical choice is to take the first bet. This violates independence since the second two bets are made by mixing the first two with the lottery that gives $0 for sure, with even proportion. One explanation of this is called certainty effect, that an outcome is overweighed when it is sure than when uncertain.

The second criticism is that risk attitudes may depend on status quo points, whereas the theory assumes that only the distributions over final outcomes matter. Suppose for example that the decision maker is given $1,000 initially and faces two alternatives, one gives $200 more and $0 (no change) with even chances, the other gives $100 more for sure. The typical choice here is to take the sure gain, which exhibits risk aversion. On the other hand, suppose one is given $1,200 initially and faces two alternatives, one yields a $200 loss and $0 with even chances, the other yields a $100

loss for sure. Now the typical choice is to take the risk, which exhibits risk loving, while the distributions over final outcomes are identical across the two comparisons.

These anomalies, together with other ones, motivate various models of no expected utility. Department of Economics and Finance, Universidad de Castilla-La Mancha, Albacete, Spain Throughout the financial literature, there is a great deal of debate about the nature of investors' risk preferences. In an ever-changing world, the main schools of knowledge discuss the constant or dynamic basis of these preferences. Based on an exhaustive review of the subject of risk aversion, this paper contributes to filling the gap that exists in the literature on the risk aversion parameter that best fits the investors' behaviour toward risk. The main determinants of risk attitude are examined and the different and most novel methodologies and perspectives are carefully analyzed.

Introduction

Risk aversion is one of the pillars of the theories used by economists. The theory of choice is also extensively analyzed by social sciences such as anthropology, psychology, political science, socio-biology, and sociology. Investor choice theory analyzes the behaviour of individuals when confronted with the task of ranking risky alternatives and the assumption of nonsatiation. Markowitz (1952) considers that an investor is risk averse when she\he receives more utility from the actuarial value of a gamble obtained with certainty than from taking the gamble itself. Markowitz (1952) and later Tobin (1958) consider risk attitude in the portfolio selection process, implementing the theory of games and economic behaviour described by Von Neumann and Morgenstern (1944). Beyond the characterization of a risk-averse utility function and how to measure a risk premium, Pratt (1964) and Arrow (1965) provide a specific definition of risk aversion. The Pratt-Arrow definition of risk aversion is useful because it provides much more insight into people's behaviour in the face of risk.

Risk preferences depend on a great deal of determinants; but, in order to make their implementation easier; the classical literature summarizes them by a single risk aversion coefficient. However, fixed risk attitude

coefficients could yield strategies that do not consider the variability in the agents' expectations. The financial literature considers risk aversion through a constant parameter or, alternatively, through a time-dependent parameter with respect to different macroeconomic and financial variables. As an additional component, recent developments take advantage of growing data processing capacity to reduce uncertainty and estimate ever more accurate changing risk aversion. The use of big data in financial markets enables faster and faster processing of data on many macro and financial variables. This better quality information provides insight into the expectations of modern investors and reduces their uncertainty about investment outcomes. In this context, the aim of this paper is to review the risk aversion literature by comparing the invariant and dynamic nature of risk preferences.

Background

Individual preferences are complex, depending on a variety of economic, political, human, or even cultural factors. Risk aversion plays a key role to understand the behaviour of different economic periods and, above all, economic recessions. This parameter amplifies the response of the most relevant macroeconomic variables to shocks of uncertainty and is, in short, the point of conciliation that makes it possible to relate finance, macroeconomics, and uncertainty. An extensive literature analyzes how fluctuations in economic uncertainty influence the dynamics of the economic cycle (see, e.g., Bernanke, 1983; Bertola and Caballero, 1994; Abel and Eberly, 1994, 1996; Caballero and Pindyck, 1996; Bloom, 2009; Bachmann and Bayer, 2013) and help explain the countercyclical behaviour of major economic variables (e.g., Campbell and Taksler, 2003; Storesletten et al., 2004; Eisfeldt and Rampini, 2006; Bloom, 2009). However, the traditional way in which macroeconomists explained economic fluctuations largely ignored the importance of risk aversion in understanding economic cycles. Thereunder, the new macro economy recognizes that financial shocks are relevant to the real economy. Jurado et al. (2015) observe a time-varying relationship between uncertainty and real activity based on a new measure of uncertainty linking financial risk aversion coefficients with macroeconomic variables. In this sense, market

risk premium and volatility reach their highest values in financial crises rather than in times of economic recession or war (Muir, 2016). Guiso et al. (2018) link changes in investor risk aversion to changes in wealth, expected income, and perceived probabilities and emotional changes in the utility function. Several authors highlight the interaction between political uncertainty and individual risk aversion. In this sense, Pástor and Veronesi (2013) incorporate this relationship into a general equilibrium model, while Brogaard and Detzel (2015) and Baker et al. (2016) examine this interaction by fitting political uncertainty through an index based on press reports.

Numerous studies show that risk aversion increases with age (e.g., Jianakoplos and Bernasek, 2006; Bucciol and Miniaci, 2011; Boyle et al., 2012; Brooks et al., 2018). Hansen et al. (1999) and Ilut and Schneider (2014) consider that consumers have pessimistic beliefs and, faced with a range of possible outcomes, act as if the worst outcomes occurred, displaying a behavior known as “ambiguity aversion.” This concept explains why many households do not invest in the stock market since the return on this investment is more ambiguous (i.e., they are not able to assign probabilities to possible outcomes) than the return on a bank deposit or a Treasury Bill (Dow and Werlang, 1992). Zhang et al. (2019) relate risk aversion with pessimism and rare disasters. Goetzmann et al. (2016) propose the availability heuristic theory in which the most recent observations have the greatest impact on an individual’s decision-making. Investors assign more probability to future stock market falls after a stock market crash. Kamstra et al. (2003) relate risk aversion to seasonal affective disorder, a depressive disorder associated with declining daylight hours. Nofsinger et al. (2018) examine the influence of behavioral biases as testosterone or stress on the individual’s risk aversion. Hoffmann and Post (2016, 2017) link up investor return experiences, confidence and investment beliefs and show why confident investors trade more than less confident investors. Falk et al. (2018) and Potrafke (2019) find a positive correlation between patience and intelligence. Suen (2018) suggests a possible discrepancy between individual and aggregate risk aversion in the context of

background risk. Finally, a branch of literature proofs the influence of social factors, ethical preferences, or religious beliefs on investor's risk attitude (e.g., Eisenhauer, 2008; Nielsen et al., 2017; Berrada et al., 2018).

Constant Risk Aversion

Although risk preferences depend on several factors, most theoretical literature fits the risk aversion as an invariant parameter that represents the personal level of risk attitude. Simple models are very relevant as they help us set ideas. Assuming constant risk aversion allows models to reach precise and relatively simple formulas for relationships between variables. Table 1 shows some applications of constant risk aversion parameters compared to other applications with time-varying coefficients. Empirical studies show contradictory evidence for this invariable parameter over time. For instance, the risk attitude parameter appears as stable for correlative periods of time in Chou (1988), or much more unstable in French et al. (1987).

TABLE 1

Safra and Segal (1998) define constant risk aversion as the invariant preference relation between outcomes of two distributions when adding or multiplying them by the same positive number. Quiggin and Chambers (2004) show that risk attitude is strongly linked with the family of generalized expected utility preferences which exhibit constant risk aversion. These expected utility preferences are constant only if the investor's utility function is quadratic, which is consistent with the capital asset pricing model (CAPM). In addition, these preferences are a generalization of both, invariant risk preferences (e.g., Quiggin and Chambers, 1998; Safra and Segal, 1998) and mean-standard deviation attitude. Other approaches link shape-invariant pricing kernels to the estimation of a constant risk aversion parameter (e.g., Lawton et al., 1972; Grith et al., 2013).

Recently, Dominiak and Tsjerengjimid (2018) generalized the preference structure in Gilboa and Schmeidler (1989) to allow for the decision maker's *ex post* preferences to be ambiguity averse, which implies constant risk appetite. Other studies assume that investor's risk preferences are constant and invariant to changes of unawareness and unforeseen

contingencies (e.g., Karni and Vierø, 2013, 2015; Mengel et al., 2016; Ma and Schipper, 2017). Baillon and Placido (2019) demonstrate that most ambiguity models forecast that risk aversion remains constant when individuals improve overall.

Time-Varying Risk Aversion

Considering the variability in agents' expectations, to model the risk aversion parameter has a cost in terms of complexity. Empirical papers document time-varying risk premia in several financial markets (e.g., Fama, 1984; Hodrick and Srivastava, 1986; Keim and Stambaugh, 1986; Harvey, 1989; Li et al., 2011). There are several studies in financial literature that refer to time-varying risk aversion as a dependent parameter of different macroeconomic and financial variables. In a seminal paper on asset pricing, Campbell and Cochrane (1999) consider that an individual is more or less risk averse according to the economic and political circumstances. Their "habit formation" model incorporates large and frequent variation of the risk aversion parameter. In the same vein, Brandt and Wang (2003) develop a consumption-based asset pricing model in which aggregate risk aversion responses to both consumption growth and inflation news. Eisenbach and Schmalz (2016) consider "anxious" investors, who are more risk averse to an imminent risk than to distant one and propose a theory that leads to a downward-sloping term structure of risk premia. In the same vein, Andries et al. (2018) propose a horizon-dependent risk aversion model involving term structures of risk premium consistent with the evidence that agents are more reluctant to immediate risks than to deferred risks. Behavioral approaches have also incorporated time-varying risk aversion by way of dynamic loss aversion or conditional disappointment aversion (e.g., Barberis et al., 2001; Routledge and Zin, 2010).

As mentioned, risk preferences are closely related to economic cycles. Many studies indicate that risk aversion is countercyclical. This way, Rosenberg and Engle (2002) observe a countercyclical investor risk aversion parameter by fitting a dynamic pricing kernel. Based on the consumption-based model of Campbell and Cochrane (1999), Li (2007) shows the influence of dynamic risk aversion on asset pricing by

observing that countercyclical changes in risk attitude lead to a procyclical time-varying risk premium. Furthermore, Cochrane (2017) notes that risk premia are countercyclical over time and are also coordinated across asset classes. Finally, González et al. (2018) observe the key role of time-varying risk aversion as a macroeconomic determinant of stock market betas.

Time-varying risk preferences have been modeled in several ways. A branch of the financial literature focuses on risk aversion implicit in option prices and realized returns. Option contracts offer several advantages when considering risk preferences (e.g., Bliss and Panigirtzoglou, 2004). To price options, it is only necessary to infer a discounted cash flow for a given horizon. In addition, there are options for different maturities. The multiplicity of prices for different payments on the same underlying asset provided by the options makes it possible to construct a density function for the distribution of the possible values of the underlying asset. The risk attitude implicit in option prices contains information of investors' behavior toward risk and, hence, its variability may be captured by the jumps in risk premia implicit in the market. From option prices and realized returns on the S&P500 index, Jackwerth (2000) derives investor's risk aversion functions and observes how shapes around financial crises change dramatically. As expected from the economic theory, these functions are positive and diminish in wealth during the pre-crisis period. On the other hand, their behavior is not consistent with the hypotheses after this event. Several authors, such as Aït-Sahalia and Lo (2000), Bedoui and Hamdi (2015), Yoon (2017), Kiesel and Rahe (2017), and Liao and Sung (2018), implement an implicit estimation of the individuals' risk attitude from the joint observations of the cross-section of option premiums and time series of underlying assets. They examine the risk preference of market participants in different states of the world and find that risk aversion level strongly increases during stressed market conditions.

Other approaches are related to the construction of indices or proxies that represent the time-varying aggregate investor sentiment in a given financial market. The main aversion indicators can be grouped into different types: indicators that use a principal component analysis (PCA) on several

financial variables (e.g., Baker and Wurgler, 2006, 2007; Han and Li, 2017; Cheema et al., 2018; Bekaert et al., 2019); indicators based on the correlation between volatilities and changes in asset prices (e.g., Kumar and Persaud, 2002); volatility indices, such as the "VIX" that uses the implied volatility of option prices on the Chicago Board Options Exchange (CBOE); and many others. For instance, Baker and Wurgler (2006) elaborate a composite index of investor sentiment derived from the first principal component of six basic proxies of investor sentiment based on various stock market indicators. On the basis of a dynamic asset pricing model with stochastic risk aversion, Bekaert et al. (2019) propose a measure of a time-varying risk aversion computed at a daily frequency that distinguishes the time variation in economic uncertainty (the amount of risk) from time variation in risk aversion (the price of risk). Most of these risk aversion indicators are used by other authors to test their ability to forecast financial crises. For example, Coudert and Gex (2008) use logit and multilogit models and observe that risk aversion indicators are good leading indicators of stock market crises.

There is a line of research linking investor risk aversion with the market risk premium derived from conditional heteroscedasticity models. The mean-variance Capital Asset Pricing Model (CAPM) of Sharpe (1964) and Lintner (1965) assumes constant second-order moments to arrive at its valuation expression, which is based on a linear relationship between expected return and risk. However, an extensive empirical evidence shows a conditional heteroscedasticity in the stock markets (e.g., Christie, 1982; Poterba and Summers, 1986; French et al., 1987)[1]. In parallel, models of Autoregressive Conditional Heteroscedasticity (ARCH) are developed (Engle, 1982) with a multitude of subsequent extensions (e.g., Bollerslev, 1986; Ding et al., 1993; Engle and Ng, 1993). Authors such as Giovannini and Jorion (1987, 1989) analyze the effects of conditional risk aversion and ARCH models for both market risk premium and performance in the static CAPM model. Given the instability of risk aversion coefficients and risk premiums over different time periods, Chou et al. (1992) improve the ARCH-in-mean (ARCH-M) model of Engle et al. (1987) with a rolling estimation

procedure in which the time-varying risk aversion is integrated by a Kalman filter method. This methodology is widely applied and expanded by the use of Generalized Autoregressive Conditional Heteroskedasticity in-mean models (GARCH-M) to test the validity of dynamic risk aversion parameters in the estimation of the market risk premium (e.g., Flannery et al., 1997; Elyasiani and Mansur, 1998; Devaney, 2001; Cotter and Hanly, 2010; Dias, 2017).

Recent literature proposes text-processing techniques based on Internet search volume of certain keywords to predict returns, rather than measures based on market trading volumes and returns. There is a debate about whether these high-frequency measures actually measure time-varying risk aversion (e.g., Vlastakis and Markellos, 2012; Kearney and Liu, 2014) or whether, on the contrary, they capture retail investors' attention toward the stock market (e.g., Da et al., 2011; Jacobs and Weber, 2012; Vlastakis and Markellos, 2012; Dimpfl and Jank, 2016) or even the investor sentiment (e.g., Da et al., 2015; Heston and Sinha, 2017). In any case, social networks can provide information on the collective behavior of investors, their state of mind, and thus allow an estimation of risk aversion at each instant of time. Financial analysis improves by increasing the speed of processing and the amount of data available. Big data and faster processors enhance investors' forecasts of future returns. As faster and wider access to information about assets reduces uncertainty, investors tend to perceive them as "safer" (Veldkamp, 2006) and makes portfolio selection results more predictable (Kacperczyk et al., 2016; Begenau et al., 2018). Given that traditional literature states that more risk-averse investors demand more information (e.g., Willinger, 1989; Eeckhoudt and Godfroid, 2000), future research can further explore the extent to which the greater current availability of information accessible to all investors could imply a reduction in risk aversion.

Conclusions

This literature review summarizes, critically examines, and clarifies alternative viewpoints of the most relevant contributions in each of the facets that affect the study and use of risk aversion in financial models. We

review the literature with a view to providing a clear understanding of both constant or invariable risk aversion and variable risk attitude over time in the context of investor behavior and investment decisions in an environment of uncertainty. Despite the influence of risk aversion in the investment context, most classical financial literature considers fixed values to reflect common levels of risk aversion over a full sample period. The use of a static estimate of the risk aversion coefficient over large timeframes may be desirable in simplifying models but could lead to investment decisions that do not reflect the investor's actual attitude toward risk. On this basis, an extensive literature both in economics and in many other disciplines shows the large number of determinants of risk aversion and proofs its changing nature over time. New macro economy and financial theory recognize the key role of risk aversion in economic cycles, finding a countercyclical relation between risk preferences and the economic period. Many asset pricing models exhibit countercyclical risk aversion, including a behavioural dimension by way of risk-averse utility functions. Furthermore, several approaches allow the inclusion of dynamic risk aversion, such as volatility or sentiment indices, implied methods based on financial option pricing, the rolling ARCH-M model and the Kalman filter methodology, or information technology and big data analysis.

Diversification

Diversification, in the context of insurance, is a risk management strategy wherein loss exposures are spread across a variety of areas, markets, or products. This technique recommends that to lower risks, a company or an individual should make different types of investments within a portfolio. Diversification reduces risk by investing in vehicles that span different financial instruments, industries, and other categories. Unsystematic risk can be mitigated through diversification while systematic or market risk is generally unavoidable.

Insuranceopedia Explains Diversification

The well-known aphorism says, "Don't put all your eggs in one basket," and this is the philosophy behind diversification. It suggests that a person should spread their risk among various investments so that a loss in

one can be offset by a return in another. A common example of diversification is investing in a combination of real estate, bonds, stocks, mutual funds, and treasury bills, instead of investing in just one of the three. By having a diverse portfolio, the investor does not have to worry as much about one of the investments' negative performance as the other investments' good performance can make up for a possible loss. Before investing, it is best to consider risks, charges, expenses, and investment objectives.

Risk Types

1. Market Risk. Exposure to uncertainty due to changes in rate or market price of an invested asset
2. Credit Risk.
3. Operational Risk.
4. Strategic Risk.
5. Liquidity Risk.
6. Event Risk.
7. Insurance and risk spreading

- Risk averse individuals want to avoid risks. But risks cannot simply be buried. When a house down, when someone is killed in an automobile accident, or when a hurricane tears through Florida-someone, somewhere, must bear the cost, Markets handle risks by risk spreading. This process takes risks that would be large for one person and spreads them around so that they are but small risks for a large number of people.
- The major form of risk spreading is insurance, which is a kind of gambling in reverse, I For example, in buying fire insurance on, a house, homeowners seem to be betting with the insurance company that the house win burn down. H it does DQ. The owner forfeits the charge. If it does burn down, the company must reimburse the owners for the loss at an agreed-upon rate. What is true of fire insurance is equally true of life, accident, automobile, or any kind of insurance.
- The insurance company is spreading risks by pooling many different risks: it may insure of houses or lives or cars. The advantage for the 10- assurance' company is that what is unpredictable for one individual is highly predictable for a population. Say that the Inland Fire Insurance Company insures I million homes, each worth $100,000. The chance that a house will

burn down is 1 in 1000 per year. The expected value of losses to Inland is then .001 X $100,000 = $ iOOper house per year. It charges each homeowner $100 plus another $100 for administration and for reserves.

- Each homeowner is faced with the choice between the certain loss of $200 for each year or the possible l-in-1000 catastrophic loss of $100,000. Because of risk aversion, the household will choose to buy insurance that costs more than. The expected value of the household's loss in order to avoid the small chance of a catastrophic loss. Insurance companies can set a premium that will earn the company a profit and at the same time produce .44 expected utility of individuals. Where does the economic gain come from? It arises from the law of diminishing marginal utility, Insurance transfers risks from those who are more risk-averse or who are exposed to disproportionately heavy risks to those who are less risk-averse or those who can more bear risks..Although insurance appears to be just a other form of gambling. it actually has exactly the opposite effect. Nature deals us risks, insurance helps reduce .individual risks by spreading them out.

Taxation in stock market

- The seller makes short-term capital gain when shares are sold at a price higher than the purchase price. Short-term capital gains are taxable at 15%. What if your tax slab rate is 10% or 20% or 30%? A special rate of tax of 15% is applicable to short-term capital gains, irrespective of your tax slab.
- Tax cuts boost demand by increasing disposable income and by encouraging businesses to hire and invest more. Tax increases do the reverse. These demand effects can be substantial when the economy is weak but smaller when it is operating near capacity.

Basic concept of Stock Market Trading and its impact on taxation, Taxation for Investing in Stock market – Equity Trading as Business and as Investment, Taxation for Investing in Stock market – F&O Trading as Non Speculation Business, How to compute turnover limit in F&O, Expenses which can be claimed against F&O turnover, carry forward of Loss in F&O, IT which can be filed by person in F&O Trading and due dates for ITR filling by Person in F&O Trade. Future and Options (F&O) Trading is a popular

activity amongst people during this pandemic period. Recent web series on Mr. Harshad Mehta added fuel in new comer with saying “Risk he to ishq he” In what can be good news for India’s stock market, 6.3 million demat account were opened between April-September as Covid drove millennials to enter and invest in stock markets due to availability of multiple online trading platforms. That’s why Income Tax provisions on F&O trading need to be analysed carefully.

1. Investing in Stock market – Equity Trading A. Business Speculative (Intra-day Trading) Non-Speculative (Frequent Delivery based Trading) B. Investment Short Term Long term

2. investing in Stock market – F&O Trading C. Business Normal business consider as Non – Speculative. Page Contents Basic concept for better understanding Taxation for Investing in Stock market – Equity Trading as Business Taxation for Investing in Stock market – Equity Trading as Investment Taxation for Investing in Stock market – F&O Trading as Non Speculation Business How to compute turnover limit in F&O? Any Expenses can be taken for set-off against F&O turnover? Can losses be carried forwarded in case of loss in F&O? Which ITR form to be filled? What is due dates for ITR filling? Basic concept for better understanding Derivatives is the instruments whose value is derived from an underlying asset. Its value is based on an underlying asset. The most popular derivatives are futures and options. Futures are a contract to buy or sale an underlying asset on a specified date at a pre-determined price. On expiry of contract, futures are executed by delivering the underlying asset or through payment. Options is a contract same as future, except in option, one party of the contract has an option (right). Advertisement Intra-day trading deals with buying and selling of stocks on the same day, during the trading hours such that all positions are closed before the market closes for the trading day. “Profits & Gains from Business or Profession” (section 43(5) of the Income Tax Act) “Profit” from Intraday Trading will be considered as Speculation Gain “Loss” from Intraday Trading will be considered as Speculation Loss Speculative income– Profits made from intraday trading of equity shares are classified as speculative income.

This is so because those investing in a stock for less than a day are presumably not investing in the company but only keen on speculating its price volatility to turn a profit Non-speculative income- (Intraday "F&O" trades) Profits made from intraday or overnight trading of Futures and Options are considered to be non-speculative income. This is so because certain F&O contracts still have a delivery clause whereby the underlying shares/commodities exchange hands between traders on the expiry of contracts. Taxation for Investing in Stock market – Equity Trading as Business Speculative business income/loss (intra-day equity trading) 1. Tax rate- normal rate of tax as per slabs 2. Loss in speculation business can be set off only against profit in speculation business Advertisement 3. Can be carried forward for 4 assessment year (subject to return filled within time limit as per section 139(1)) 4. Tax audit applicability – (i) where turnover exceed Rs. 1 Crores (ii) In case of books of accounts not maintained, turnover not exceed Rs. 2 crores and estimated income is below 6% of turnover and net total income is above basic exemption limit of Rs. 2.5 Lakh Non-speculative business income/loss

1. Tax rate- normal rate of tax as per slabs 2. Loss from a non-speculation business can be set off against income from speculation or non-speculation business (except salary income) 3. Can be carried forward for 8 assessment year (no time limit for unabsorbed depreciation, exp. on scientific research, loss from specified business u/s 35AD) (subject to return filled within time limit as per section 139(1)) 4. Tax audit applicability – (i) Where turnover exceed Rs. 1 Crores (ii) In case of books of accounts not maintained, turnover not exceed Rs. 2 crores and estimated income is below 6% of turnover and net total income is above basic exemption limit of Rs. 2.5 Lakh Advertisement Taxation for Investing in Stock market – Equity Trading as Investment Long term capital gain /loss- equity investment holding for more than 1 years 1. Tax rate -10% (gain upto Rs. 1 lakh not chargeable to tax) 2. Long term capital loss can be set off only against long term capital gain 3. Can be carried forward for 8 assessment year (subject to return filled within time limit as per section 139(1)) Short term capital gain /loss- holding period is more than 1 day and less than 1 year 1.

Tax rate -15% 2. Short term capital loss can be set off against any capital gain (long term / short term) 3. Can be carried forward for 8 assessment year (subject to return filled within time limit as per section 139(1)) Advertisement Taxation for Investing in Stock market – F&O Trading as Non Speculation Business Section 43 subsection 5 has excluded transaction of future and options as speculative transaction. However exemption is available only for equity. Thus if F&O for commodities are done the same will be termed as Speculative in Nature. Other then commodity trading profit or loss arising out of transaction is treated as Business Loss or profit in nature. In case of Profit from transactions of F&O Trading: 1. In the case of profit from derivative transactions, tax audit will be applicable if the turnover from such trading exceeds Rs. 1 crore. 2. If the turnover from such trading exceeds Rs. 1 crore but less than 2 crore then the audit can be avoided if we can show the profit at minimum 8% (6%, if all trades are digital).

3. Tax audit u/s 44AB r/w section 44AD will also be applicable, if the net profit from such transactions is less than 8% (6%, if all trades are digital) of the turnover from such transactions. 4. Further, please note that any turnover more than 2 crore then audit u/s 44AB will irrespective applicable In case of Loss from F&O Trading: 1. In case of Loss from derivative trading, since profit (Loss in this case) is less than 8% (6%, if all trades are digital) of the turnover, therefore Tax Audit will be applicable u/s 44AB r.w.s. 44AD. Advertisement How to compute turnover limit in F&O? Ans: In normal business turnover is based on sales and thus reaching the limit takes time. But in F&O it reached easily as each lot is valued high, Limit is reached easily. Therefore computation method need be different. Thus for computing turnover limit following things should be added: a. Profits from the trade b. Loss from the trade c. Premium received from sale of Options d. In case of Reverse Trade, difference should also be added To make it clearer let's take an example: Mr. Z buys 5000 units of Futures of ABC Ltd. at Rs. 200 and sells it at Rs. 180. He also buys 4000 units of futures of XYZ Ltd at Rs 220 and sells it at Rs 300. Loss made Mr. Z= 5000*(200-180) = 100,000(Negative is ignored in turnover) Profit made by MR. Z= 4000*(300-220) = 320,000

Therefore, total turnover shall be 4, 20,000 although income of Mr. Z is 2,20,000. For computation of turnover of futures, the total of positive and negative or favourable and unfavourable differences shall be taken as turnover. Any Expenses can be taken for set-off against F&O turnover? Ans: Expenses that can be claimed by F&O trader are the following associated expenses: 1. Broker's Commission 2. Subscription to trading journals 3. Internet and telephone charges 4. Depreciation on assets for e.g. computers used for trading 5. Consultancy expenses if any; can losses be carried forwarded in case of loss in F&O? Ans: Yes Losses can be carried forward subject to following conditions:

1. Return should be filed on or before due date: 2. Loss should be disclosed in the return 3. Set off is not allowed against Salary Income 4. Loss should not be of Commodity trading 5. Speculative Loss can be carried forward for 4 years. It can be set-off against Speculative Business Income only 6. Non-Speculative Loss can be carried forward for 8 years. It can be set-off against both Speculative Business Income and Non-Speculative Business Income. So, losses on F&O trading can be set off against income of interest income from bank, rental income or capital gains but only in the same year. Which ITR form to be filled? Since F&O trading is classified under business income we can use ITR 3 AND ITR4. ITR3 can be filed for F&O trading income and also if any capital gains are to be reported. ITR 4 is similar to ITR3 but with a presumptive scheme if you are using Section 44AD. It cannot be used to disclose any capital gains or if losses have to be carried forward under any head of income.

1. 31st July – If Tax Audit is not applicable 2. 30th September – If Tax Audit is applicable Conclusion: Based on the above discussion it can be said that F&O profit/loss consider as Normal business profit/loss and Equity Profit/loss differ as per case to case on basis maintaining books of accounts & separation of Trading equity and investment equity. However computation needs to be done carefully in order to avoid the litigation. Further profit margin also need to be identified as if it's below 8% (6%, if all trades are digital) same would be liable to tax Audit under section 44AB. About the Author: The Views Expressed in the article is personal opinion of author.

The author is Practicing CA having core expertise in Stock market transaction and one can reach him on mail queries on caparagdavda@gmail.com. The article can be said as reference material. However courts can take different opinion based on nature and circumstances of each case.

A risk asset is any asset that carries a degree of risk. Risk asset generally refers to assets that have a significant degree of price volatility, such as equities, commodities, high-yield bonds, real estate, and currencies.

Risks of assets

Risk assets are any assets that are not risk-free – they carry an element of risk. The term generally refers to any financial security or instrument, such as equities, commodities, high-yield bonds, and other financial products that are likely to fluctuate in price.

The Reality of Investment Risk

When it comes to risk, here's a reality check: All investments carry some degree of risk. Stocks, bonds, mutual funds and exchange-traded funds can lose value, even all their value, if market conditions sour. Even conservative, insured investments, such as certificates of deposit (CDs) issued by a bank or credit union, come with inflation risk. They may not earn enough over time to keep pace with the increasing cost of living.

What Is Risk?

When you invest, you make choices about what to do with your financial assets. Risk is any uncertainty with respect to your investments that has the potential to negatively affect your financial welfare. For example, your investment value might rise or fall because of market conditions (market risk). Corporate decisions, such as whether to expand into a new area of business or merge with another company, can affect the value of your investments (business risk). If you own an international investment, events within that country can affect your investment. There are other types of risk. How easy or hard it is to cash out of an investment when you need to is called liquidity risk. Another risk factor is tied to how many or how few investments you hold. Generally speaking, the more financial eggs

you have in one basket, say all your money in a single stock, the greater risk you take (concentration risk).

In short, risk is the possibility that a negative financial outcome that matters to you might occur. There are several key concepts you should understand when it comes to investment risk. Risk and Reward. The level of risk associated with a particular investment or asset class typically correlates with the level of return the investment might achieve. The rationale behind this relationship is that investors willing to take on risky investments and potentially lose money should be rewarded for their risk. In the context of investing, reward is the possibility of higher returns. Historically, stocks have enjoyed the most robust average annual returns over the long term (just over 10 percent per year), followed by corporate bonds (around 6 percent annually), Treasury bonds (5.5 percent per year) and cash/cash equivalents such as short-term Treasury bills (3.5 percent per year). The trade off is that with this higher return comes greater risk: as an asset class, stocks are riskier than corporate bonds, and corporate bonds are riskier than Treasury bonds or bank savings products.

Exceptions Abound

Although stocks have historically provided a higher return than bonds and cash investments (albeit, at a higher level of risk), it is not always the case that stocks outperform bonds or that bonds are lower risk than stocks. Both stocks and bonds involve risk, and their returns and risk levels can vary depending on the prevailing market and economic conditions and the manner in which they are used. So, even though target-date funds are generally designed to become more conservative as the target date approaches, investment risk exists throughout the lifespan of the fund.

While historic averages over long periods can guide decision-making about risk, it can be difficult to predict (and impossible to know) whether, given your specific circumstances and with your particular goals and needs, the historical averages will play in your favor. Even if you hold a broad, diversified portfolio of stocks such as the S&P 500 for an extended period of time, there is no guarantee that they will earn a rate of return equal to the

long-term historical average. The timing of both the purchase and sale of an investment are key determinants of your investment return (along with fees). But while we have all heard the adage, "buy low and sell high," the reality is that many investors do just the opposite. If you buy a stock or stock mutual fund when the market is hot and prices are high, you will have greater losses if the price drops for any reason compared with an investor who bought at a lower price. That means your average annualized returns will be less than theirs, and it will take you longer to recover.

Investors should also understand that holding a portfolio of stocks even for an extended period of time can result in negative returns. For example, on March 10, 2000, the NASDAQ composite closed at all-time high of 5,048.62. It has only been recently that the closing price has approached this record level, and for well over a decade the NASDAQ Composite was well off its historic high. In short, if you bought at or near the market's peak, you may still not be seeing a positive return on your investment. Investors holding individual stocks for an extended period of time also face the risk that the company they are invested in could enter a state of permanent decline or go bankrupt.

Time Can Be Your Friend or Foe

Based on historical data, holding a broad portfolio of stocks over an extended period of time (for instance a large-cap portfolio like the S&P 500 over a 20-year period) significantly reduces your chances of losing your principal. However, the historical data should not mislead investors into thinking that there is no risk in investing in stocks over a long period of time. For example, suppose an investor invests $10,000 in a broadly diversified stock portfolio and 19 years later sees that portfolio grow to $20,000. The following year, the investor's portfolio loses 20 percent of its value, or $4,000, during a market downturn. As a result, at the end of the 20-year period, the investor ends up with a $16,000 portfolio, rather than the $20,000 portfolio she held after 19 years. Money was made—but not as much as if shares were sold the previous year. That's why stocks are always risky investments, even over the long-term. They don't get safer the longer you hold them.

This is not a hypothetical risk. If you had planned to retire in the 2008 to 2009 timeframe—when stock prices dropped by 57 percent—and had the bulk of your retirement savings in stocks or stock mutual funds, you might have had to reconsider your retirement plan. Investors should also consider how realistic it will be for them to ride out the ups and downs of the market over the long-term. Will you have to sell stocks during an economic downturn to fill the gap caused by a job loss? Will you sell investments to pay for medical care or a child's college education? Predictable and unpredictable life events might make it difficult for some investors to stay invested in stocks over an extended period of time.

Managing Risk

You cannot eliminate investment risk. But two basic investment strategies can help manage both systemic risk (risk affecting the economy as a whole) and non-systemic risk.

Asset Allocation: By including different asset classes in your portfolio you increase the probability that some of your investments will provide satisfactory returns even if others are flat or losing value. Put another way, you're reducing the risk of major losses that can result from over-emphasizing a single asset class, however resilient you might expect that class to be.

Diversification: When you diversify, you divide the money you've allocated to a particular asset class, such as stocks, among various categories of investments that belong to that asset class. Diversification, with its emphasis on variety, allows you to spread you assets around. In short, you don't put all your investment eggs in one basket. Hedging and insurance can provide additional ways to manage risk. However, both strategies typically add to the costs of your investment, which eats away any returns. In addition, hedging typically involves speculative, higher risk activity such as short selling or investing in illiquid securities. The bottom line is all investments carry some degree of risk. By better understanding the nature of risk, and taking steps to manage those risks, you put yourself in a better position to meet your financial goals.