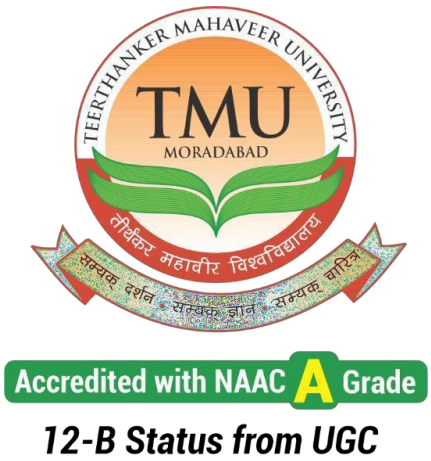
Macro Economics

CENTRE FOR DISTANCE & ONLINE EDUCATION



Programme: Master of Economics

Course: Macro Economics

Course Code: MAECC102

Semester-I

Programme Coordinator

Dr. Vibhor Jain

Associate Professor, Department of Arts

Center for Distance and Online Education,

Teerthanker Mahaveer University, Moradabad

Units Written by

Dr. Aditya Kumar Sharma

Editor

Professor Vipin Jain

Director,

Center For Distance and Online Education,
Teerthanker Mahaveer University,

Dr. Amit Kansal

Assistant Director,

Center for Distance and Online Education,
Teerthanker Mahaveer University, Moradabad
Moradabad

Edition : **2023 (Restricted Circulation)**

Published by : **Teerthanker Mahaveer University, Moradabad**

SYLLABUS

MACRO ECONOMICS

MAECC102

BLOCK I

Unit I: Review of Classical and Keynesian Economics, Classical theory of income, employment, wage rate and price level - macro, equilibrium in classical model - Classical dichotomy - Causes of, unemployment - Evaluation of classical model., Keynesian determination of income, output and employment:, Consumption function, investment multiplier and balanced budget, multiplier - Multiplier in an open economy; difference between Keynesian, and classical macro-equilibrium, Pigou's critique of Keynes's under, employment equilibrium - the Pigou effect.

Unit II: Neo-Classical and Keynesian Synthesis, Classical and Keynesian theories of interest: IS-LM model - Relative, effectiveness of monetary and fiscal policies - Extension of IS-LM model, to labour market, flexible price, Mundell-Fleming model of an open, economy.

BLOCK II

Unit III: Theories of Consumption, Extension of Keynesian consumption function to long run - Relative, Income Hypothesis - Life Cycle Hypothesis - Permanent Income, Hypothesis.

Unit IV: Theories of Investment, Investment demand: Present value criterion for investment, payback, period and internal rate of return, Marginal Efficiency of Investment, (MEI). Acceleration principle, its derivation from the profit maximizing, behaviour, flexible accelerator.

Unit V: Theories of Trade Cycle, Schumpeterian Theory: Multiplier-accelerator interaction - Kaldor's, theory - Monetary and fiscal policy for economic stabilization.

BLOCK III

Unit VI: Demand for and Supply of Money, The demand for money: Tobin's portfolio selection and Baumol's, transaction demand for money - Friedman's restatement of quantity theory, of money - Supply of money - Inside and outside money, Gurley-Shaw's, approach - Concept of exogenous money supply, Friedman's theory -, Endogenous money supply, Kaldor's theory.

Unit VII: Inflation, Monetarism and New-Classical Macroeconomics, Keynesian theory of inflation -

Phillips Curve - Monetarism: Basic, elements of monetarism - Monetarist theory of income determination, and inflation - Friedman-Phelp critique of Phillips Curve and money supply, rule., Price behaviour and expectation: Adaptive and rational expectations., New-classical economics, real business cycle and economic policy.

BLOCK IV

Unit VIII: Problems in Asset Pricing, Returns from assets and risk-Measure of risk, diversification-Risk in a, Portfolio, cost of capital, traditional theory and Modigliani-Miller theorem, - Capital asset pricing model and arbitrage pricing model.

Unit IX: Financial Market, Role, structure and characteristics of money market and capital market,, primary and secondary market for securities, treasury bill and commercial, bill market - Discount market, government securities, markets for, derivatives, futures and options - types, uses and pricing of derivatives, - depth, breadth and resilience of markets, construction of Sensex and, Nifty.

Unit X: Indian Financial Institutions, Working of capital market and SEBI: Commercial banks, co-operatives, regional rural banks (RRBs), problems of micro finance, self-help groups, (SHGs) - Non-bank financial intermediaries (NBFIs): Insurance, institutions: Life and general insurance, mutual benefit funds.

UNIT 1 REVIEW OF CLASSICAL AND KEYNESIAN ECONOMICS

NOTES

Structure

1.0 INTRODUCTION

The primary principle of classical economic theory is that the economy is self-regulating. The classical economists believed that an economy could achieve the natural level of real GDP or output only when the economy's resources were fully employed. For them, full employment in an economy was a normal situation and any deviation from this was regarded as abnormal. The classical economist Pigou stated that the tendency of the economic system is to automatically provide full employment in the labour market when the demand and supply of labour are equal.

The economist John Maynard Keynes in his work *General Theory of Employment, Interest and Money* criticized the fundamentals of classical theory of employment because of its reliance on unrealistic assumptions. Keynesian economics is fundamentally based on the criticism of classical economics. In this unit, you will learn about the classical model of employment and income and well as Keynes' criticism of the model.

1.1 UNIT OBJECTIVES

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

- Analyse the classical theory of income, employment, wage rate and price level
- Discuss the various causes of unemployment

- Criticise and evaluate the classical model
- Discuss the criticism of the Keynesian theory
- Describe the factors affecting consumption and multiplier analysis
- Analyse the concept of investment multiplier and balanced budget multiplier
- Differentiate between Keynesian and Classical macro-equilibrium

1.2 CLASSICAL THEORY OF INCOME, EMPLOYMENT, WAGE RATE AND PRICE LEVEL

Following the publication of Adam Smith's classic entitled *An Inquiry into the Nature and Causes of the Wealth of Nations* in 1776, a body of economic theory was gradually developed during the following century and a half. The chief architects of this theory, known as the classical economic theory, were David Ricardo, John Stuart Mill, Jean Baptiste Say and Alfred Marshall. The problem of unemployment was not the primary concern of this theory. Assuming that full employment exists in the economy in the long run, the classical economic theory was mainly concerned with the discussion of those factors which determined:

- What goods and services would be produced in the economy with its given resources
- The allocation of the economy's given resources between their different rival users
- The relative prices of different goods and services and of the factors of production
- The distribution of income earned from production between the different co-operating factors of production.

There are two assumptions of classical theory of employment: assumption of (i) full employment and (ii) flexibility of price and wages. Full employment is the employment level in which every individual, who is willing to work at the prevalent wage rate gets employed. Classical economists believed that three is always a condition of full employment of resources in an economy. However, they also advocated that the flexibility or adjustments in price of products and wages of individuals facilitates the condition of full employment.

In the classical economic theory, full employment is a rule in the long period. Deviations from it are viewed only as temporary exceptions. Full employment did not, however, rule out the existence of some unemployment in the economy. Even at the 'full employment' level, there would be some people in the economy who could be either frictionally or voluntarily unemployed. The frictional unemployment was temporary unemployment between job changes or on entry into the labour force while searching for jobs due to the lack of adequate knowledge on the part of workers about the available job opportunities in the economy. Voluntary unemployment was due to the reluctance or refusal on the part of workers to work at the going wage. Workers agitating for higher wages were an example of the voluntarily unemployed workers. While frictional unemployment would disappear with the workers getting acquainted with the available job opportunities in due course of time, voluntary unemployment was due to the workers' refusal to work at the current wage and did not worry the classical economists. In short, full employment only implied that involuntary unemployment—a state of being unemployed

in spite of the workers' willingness to work at the going wage rate—did not exist in the economy.

But what would happen if there were workers who were involuntarily unemployed in the economy? According to the classical economic theory, if there is unemployment in the economy, forces of correction will soon eliminate it and will restore full employment in the economy. The basic classical tenet was that in a free market economy, the aggregate demand for goods and services could not, except temporarily, fall short of the aggregate supply of goods and services. As long as the aggregate demand equalled the aggregate supply, there was no barrier to the production of goods and services corresponding to full employment in the economy. In the classical view, lapses from full employment were infrequent and short-lived. Depressions were, therefore, considered infrequent and short-lived occurrences. This conclusion is, however, puzzling to any serious student of economic history who knows about the severe and prolonged depressions of the 1870s, 1930s and other periods.

NOTES

Although the classical theory of employment, output and price level was attacked by a few dissenters in the 19th century—Thomas Robert Malthus, Jean Charles Leonard de Sismondi, Karl Marx, J A Hobson, Silvio Gesell and others—the attack was unsuccessful because no alternative theory was constructed to replace the classical theory. 'Since Malthus was unable to explain clearly (apart from an appeal to the facts of common observation) how and why effective demand could be deficient or excessive, he failed to furnish an alternative construction, Ricardo's theory was accepted by the city, by statesmen and by the academic world. But controversy ceased; the other point of view completely disappeared; it ceased to be discussed. The great puzzle of Effective Demand with which Malthus had wrestled vanished from economic literature.'

John Maynard Keynes successfully attacked the classical explanation of the determination of aggregate employment, output and general price level. It was the assumption of a given volume of total output, rather than its composition and technique of production, which was severely attacked by Keynes. The great depression of the 1930s gave a severe blow to the naïve classical economic theory.

The essential feature of classical macroeconomic analysis is that it presents a model of full employment in the economy in the long period. Underlying the analysis, are the assumptions of perfect competition in the factor and product markets and profit-maximization on the part of firms. There are three markets to study. First, there is the labour market which deals with the supply of and the demand for labour. The equilibrium condition for full employment in the labour market requires that the wage should be one corresponding to which the demand for and the supply of labour in the market are in equilibrium, i.e., there is neither an excess supply of nor an excess demand for labour in the market. In the labour market we are concerned with the analysis of the form of the aggregate demand and the aggregate supply functions of labour.

Second, there is the product market with its equilibrium flow condition which is equivalent in macroeconomic equilibrium to an equality between saving and investment. The equilibrium condition in the capital–bonds–market requires the equilibrium between the *ex ante* investment and *ex ante* saving. Third, there is the money market which is concerned with the demand for and the supply of money. The first two markets deal with the equilibrium of the real sector of the economy while the money market is concerned with the equilibrium of the monetary sector of the economy. The equilibrium in the monetary sector determines the absolute price level which does not influence the relative prices, aggregate employment and output which are determined in the real sector

NOTES

of the economy. In short, there is a dichotomy or separation between the real and monetary sectors of the economy in the classical economic system. This dichotomy arises from the argument of the classicists that 'money is a veil' (neutral).

In the classical economic theory, money does not matter and its function in the economy is merely to facilitate the real transactions by serving as a medium of exchange. It is neutral and does not interfere with the real processes of production and distribution in the economy; it only facilitates production, i.e., lubricates the wheels of the economic system. According to the classicists, changes in the money supply cause proportionate changes only in the equilibrium values of the *nominal* variables, leaving the equilibrium values of the *real* variables (output, employment, real wage, interest rate, etc.,) unchanged. The equilibrium values of these real variables are exclusively and solely determined in the real sector—in the labour, capital and commodity markets. In the classical macroeconomics, the economy's real sector can, therefore, be dichotomized from its monetary sector.

Money, however, does something more than merely act as a medium of exchange in the economy. In a dynamic world with uncertain future, money is also demanded for asset purposes. Consequently, it influences both the production and distribution in the economic system. In other words, changes which take place in the monetary sector also influence the real sector of the economy.

1.2.1 Aggregate Output and Employment

The aggregate production, labour supply and demand curves is used to illustrate the determination of full employment and the aggregate real output in the classical model. The determination of labour market equilibrim is shown in panel (a) of Figure 1.1. Labour demand and labour supply curves are represented by D_L and S_L curves, respectively. The abour demand and supply curves intersect at point *E*. The point of intersection of demand and supply curves determines simultaneously the equilibrium wage rate and full employment of labour. At point *E*, the equilibrium wage rate is determined at OW_r. At this wage rate, the demand for and supply of labour are equal at *OL* employment of labour. Given the short-run conditions, this is the level of full employment according to the classical theory of full employment.

The determination of output can now be shown by juxtaposing the production function with labour-market equilibrium. The short-run production function is reproduced in panel (b) of Figure 1.1 at the bottom of panel (a) on the same scale of labour. According to the classical theory, the national output in the short-run is the function of labour employment, capital remaining constant. As shown in panel (a), full employment of labour is determined at *OL* at real wage *EL*. The ordinate *EL* extended downward to the production function in panel (b) determines the equilibrium level of national output at *OY*. In panel (b), the extended ordinate *ERL* intersects the production function at point *R*. A line drawn from point *R* to the vertical axis determines the equilibrium level of national output at *OY*. Thus, employment and output are simultaneously determined in the classical model.

An ***important feature*** of the classical model is that factors operating on the supply side of the market determine the level of employment and output. Labour market equilibrium is determined by the demand for and supply of labour. The labour demand curve is, however, derived from the production function based on a given technology determined exogenously. The labour demand curve is therefore, in a sense, a datum, i.e.,

a given fact or law. According to classical theory, it is the labour supply, which is a function of real wages, that plays a more important role in the determination of the labour market equilibrium and employment. And, employment determines the level of output. Thus, *in the classical model, employment and output are determined solely by the factors operating on the supply side of the labour market.*

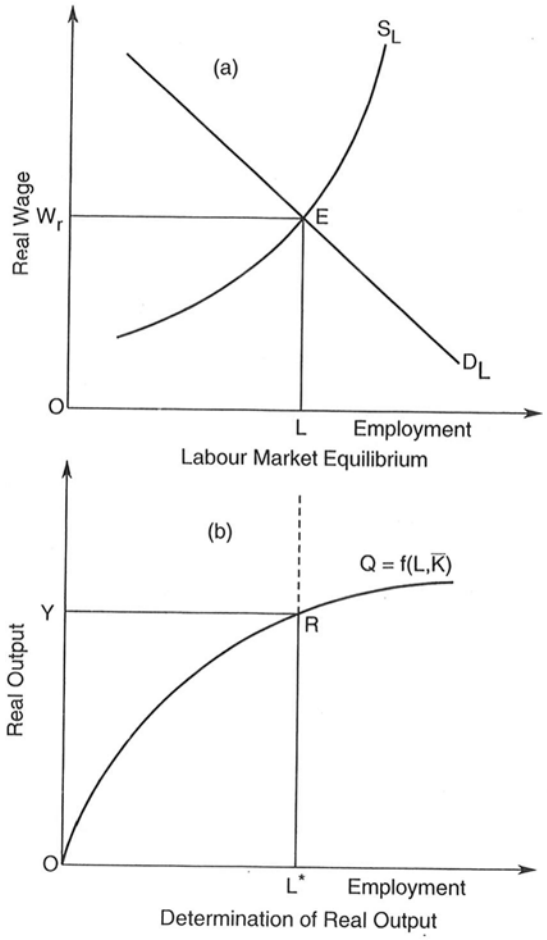


Fig. 1.1 *Determination of Equilibrium Output*

1.3 MACRO-EQUILIBRIUM IN CLASSICAL MODEL: CLASSICAL DICHOTOMY

In the classical theory, a change in the aggregate money supply will not affect the real wage, employment and output in the economy. The change in the money supply will affect only the general price level *P* and the money wage *W*. An increase in the money supply from M_1 to M_2 (with constant *V*) causes an upward shift in the *MV* curve from the position of M_1V to M_2V as shown in Figure 1.2(A). Since no one in the economy hoards money, an increase in the total money supply from M_1 to M_2 will mean an effective increase in total money supply of $V\Delta M$ with no increase in the supply of goods available for purchase. Consequently, people can get rid of the additional unwanted money supply $V\Delta M\ (=M_2 - M_1)$ by purchasing the same quantity of goods at the higher general price level $0P$. The increase in the general price level of $\Delta P(= 0P - 0P)$ should be sufficient

to make the additional aggregate money spending $Q\Delta P$ equal to the total effective increase in the aggregate money supply $V\Delta M$, i.e., $Q\Delta P = V\Delta M$ so that the stability condition $M_2V = 0P_2 \times 0Q_1$ is satisfied.

NOTES

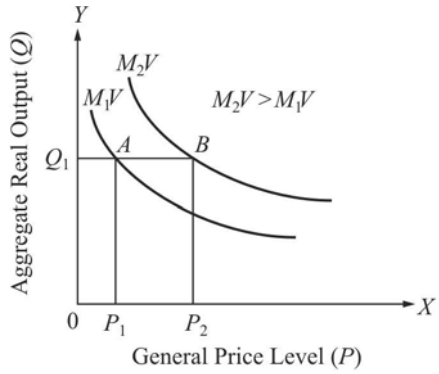


Fig. 1.2(A)

This conclusion is reached on the assumption that people in the economy demand money only as the medium of exchange. In other words, there is only the transactional demand for money and the asset demand for money is zero. Looked at in this way, the changes in the total money supply do not affect the aggregate real output and employment in the economy. Consequently, there is dichotomy between the real and the monetary sectors of the economy. John Maynard Keynes, Don Patinkin and others have criticized this classical dichotomy which exists between the real and the monetary sectors of the economy in the classical macroeconomic theory.

Figure 1.2(B) shows that when the general price level rises from $0P_1$ to $0P_2$ due to an increase in the aggregate money supply from M_1 to M_2, the money wage increases from $0W_1$ to $0W_2$ such that the real wage remains unchanged at the old W/P_1 level.

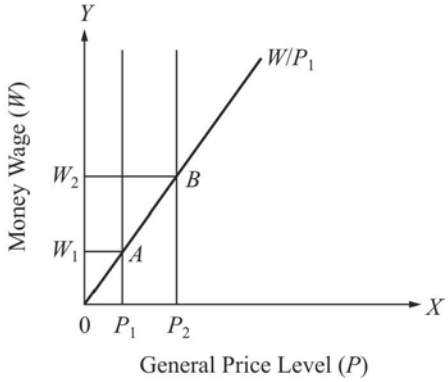


Fig. 1.2(B)

Check Your Progress

1. State the factors that the classical economic theory dealt with.
2. Name some of the thinkers who attacked the classical theory of employment, output and price.

1.4 CAUSES OF UNEMPLOYMENT

Classical economists believed that the way to maintain full employment was to cut wages and reduce taxes. Keynes did not agree with this. He stated: "The best way to destroy the capitalist system was to debauch the currency." Because the economy was determined by demand, the cut in wages would reduce employee income, decreasing

consumer spending. This reduces demand for products, leading to a reduction in production, forcing companies to not only cut wages, but lay off employees. Now the employees (and former employees) are now even more unable to spend, decreasing both consumer spending and demand. If it goes on like this, the recovery will be unforeseeable. Reducing taxes was not an option for the government when their budget was out of control due to the reduction of tax revenues. The way to recovery is to encourage spending. By encouraging consumers and firms alike to increase spending, demand will increase. This will increase quantity produced, leaving companies to need to hire more employees, increasing employee income, making them able to spend more.

In Keynes' classification of unemployment by its causes, unemployment due to downward-rigidity of money-wages (which for the "classical" economists was the chief type of cyclical unemployment and the only important type of secular or persistent unemployment) finds no place. As will be seen later, it is excluded on the ground that resistance to reductions in money wage-rates generally does not involve a reduction in the volume of employment and is favourable to employment rather than the reverse. The omission charged against the classical economists is their failure to note the lesser resistance of labour to reductions in real wages if unassociated with reductions in money wages per se, and their failure to recognize the existence of a large volume of unemployment for which the former is an available and practicable remedy, but not the latter. Keynes' reasoning points obviously to the superiority of inflationary remedies for unemployment over money-wage reductions.

According to Keynes,

> When employment increases, aggregate real income is increased. The psychology of the community is such that when aggregate real income is increased, aggregate consumption is increased, but not by so much as income. Hence employers would make a loss if the whole of the increased employment were to be devoted to satisfying the increased demand for immediate consumption. Thus, to justify any given amount of employment there must be an amount of current investment sufficient to absorb the excess of total output over what the community chooses to consume when employment is at the given level. For unless there is this amount of investment, the receipts of the entrepreneurs will be less than is required to induce them to offer the given amount of employment. It follows, therefore, that, given what we shall call the community's propensity to consume, the equilibrium level of employment, . . . will depend on the amount of current investment. The amount of current investment will depend, in turn, on what we shall call the inducement to invest; and [this] will . . . depend on the relation between the schedule of the marginal efficiency of capital and the complex of rates of interest. . .

Keynesian economics emphasizes the cyclical nature of unemployment and recommends interventions it claims will reduce unemployment during recessions. This theory focuses on recurrent supply shocks that suddenly reduce aggregate demand for goods and services and thus reduce demand for workers. Keynesian models recommend government interventions designed to increase demand for workers; these can include financial stimuli, publicly funded job creation, and expansionist monetary policies.

Marxian unemployment

Marxism focuses on the relations between the owners and the workers, whom, it claims, the owners pit against one another in a constant struggle for jobs and higher wages. The

NOTES

unemployment produced by this struggle is said to benefit the system by reducing wage costs for the owners. For Marxists the causes of and solutions to unemployment require abolishing capitalism and shifting to socialism or communism.

According to Karl Marx, unemployment is inherent within the unstable capitalist system and periodic crises of mass unemployment are to be expected. The function of the proletariat within the capitalist system is to provide a "reserve army of labour" that creates downward pressure on wages. This is accomplished by dividing the proletariat into surplus labour (employees) and under-employment (unemployed). This reserve army of labour fight among themselves for scarce jobs at lower and lower wages. At first glance, unemployment seems inefficient since unemployed workers do not increase profits. However, unemployment is profitable within the global capitalist system because unemployment lowers wages which are costs from the perspective of the owners. From this perspective low wages benefit the system by reducing economic rents. Yet, it does not benefit workers.

According to Marx, the only way to permanently eliminate unemployment would be to abolish capitalism and the system of forced competition for wages and then shift to a socialist or communist economic system. For contemporary Marxists, the existence of persistent unemployment is proof of the inability of capitalism to ensure full employment.

Full employment theory of Keynes

Theoretically, it is possible to do away with cyclical unemployment by increasing the aggregate demand (for workers and products). But eventually, the economy is bound to reach the 'inflation barrier', often considered to correspond with the natural rate of unemployment which is defined as the rate of unemployment existing when the labour market is in equilibrium. It is also called the Non-Accelerating Inflation Rate of Unemployment or NAIRU. It implies that if the unemployment rate falls too low, inflation will get worse if no wage or price controls are exercised. That is, inflation will rise with the fall in unemployment rate.

Full employment, also referred to as the ideal unemployment rate would exclude all inefficient forms of unemployment. This type of 'full employment' unemployment would correspond only to frictional unemployment but then, till the time hidden unemployment exists, official unemployment statistics cannot be completely relied on and it would be difficult to explain what unemployment rate would coincide with 'full employment'.

The classical macroeconomic theory which states that the free enterprise market economy automatically tends to move towards full employment equilibrium has been criticized by John Maynard Keynes and others on several grounds. The theory which at one time was held in high esteem among the economists lost its prestige during the great depression of the 30s on account of the following factors:

1. The Great Depression of the 1930s gave a severe blow to the unrealistic assumptions on which the classical macroeconomics was based. In the midst of mass unemployment, people found it difficult to believe the faulty classical view that full employment was a normal situation. Many, in fact, sorely felt that if anything was normal in the economy it was mass unemployment.
2. During the 20th century, a theory based on the assumption of perfect competition was an anachronism in an age of oligopolistic and monopolistic markets. Deviations

from the competitive market model resulted in the state assuming a stronger and positive role in the economy and consequently in invalidating the conclusions of the classical macroeconomic theory.

3. Keynes made a devastating attack on the classical theory. The classical contention was that unemployment in the economy was caused by a downwardly rigid money wage that was fixed too high to guarantee full employment. As against this contention of the classicists, Keynes argued that unemployment was due to the deficiency of aggregate effective demand resulting from instability of investment spending and the persistence of high saving propensity in the affluent economies combined with inadequate investment opportunities. Moreover, even if wages were high it was not practical to reduce them in the changed situation of the 1930s when workers were organized into strong trade unions. In place of the monetary policy which, according to the classical view, by raising the prices and lowering the real wage helps raise the level of employment in the economy, Keynes suggested the increasing use of fiscal policy to raise sufficiently the level of the aggregate effective demand to remove unemployment in the economy. The Keynesian liquidity trap caused complete emasculation of the monetary theory as an effective instrument of economic policy to cure the Depression.
4. The classical economists had overlooked an important point in their argument according to which to remove unemployment in the system, real wage (i.e., firms' costs) should be reduced. They forgot that a general wage-cut while reducing the firms' marginal costs will also reduce the factor incomes and consequently the total market demand for the product. Thus, if as a result of a general wage-cut the aggregate supply curve shifts to the right, the aggregate output (and consequently employment) could increase only if the aggregate demand curve did not shift leftward. Unfortunately, the same general wage-cut which shifts the aggregate supply curve to the right will also shift the aggregate demand curve to the left leaving output and unemployment unchanged. Moreover, if the leftward shift in the demand curve was more than the rightward shift in the supply curve, the aggregate output and employment may even fall rather than rise consequent upon a general wage-cut.

   The Keynesian criticism of this assumption of the classical theory is not altogether free from faults. In fact, it is difficult to see any direct relationship between the wage-cut and the aggregate demand. The demand for goods depends upon the level of income and even when the level of income falls due to a fall in the money wage, it cannot be said for certain that the aggregate demand will fall. It all depends on the income elasticity of demand. Moreover, by how much will the aggregate demand curve shift leftward due to a general wage-cut will depend upon, among other things, the proportion of wage-income in the total national income, the increase in the non-wage incomes when a general wage-cut takes place and the propensities to spend of the workers and employers.

   There is, however, another argument for concluding that a general money wage-cut will increase employment and output in the economy even if it means the fall in prices. This is known in the literature as the 'Pigou' or the real-balance effect. With the general fall in prices, idle cash balances result. Under the classical assumptions, these cash balances will be spent shifting the market demand curve to the right and preventing the prices from falling in the same proportion as the fall

NOTES

in wages. Thus, there will be a net increase in the aggregate output and employment in the economy.

Keynes seriously doubted that this would happen. He argued that under the expectation of persistently falling prices, wealth-holders would postpone spending indefinitely. He, however, conceded that a fall in wages and prices, by reducing the transactions demand for cash balances, would release some cash balances to cater to the speculative demand for cash balances which would lead to some fall in the rate of interest and consequently to an increase in the investment. But the important question is: by how much will the rate of interest fall and by how much will investment in the economy increase in response to the given fall in the rate of interest? It will depend upon the existing rate of interest, the interest-elasticity of the speculative demand for money at that rate of interest and the interest-elasticity of the investment demand. This reasoning in the literature is known as the Keynes-effect. Keynes did not think that the Keynes-effect was powerful enough to guarantee full employment through the wage-price deflation.

5. Keynes also seriously doubted that the real wage actually determined the labour's supply function (within some considerable range any way). He doubted that workers entered and left the labour market as the real wage rose and fell. Keynes asserted that a situation in which labour stipulated for the money wage rather than the real wage was the normal case. According to Keynes, workers suffered from money illusion and the supply of labour was a function of the nominal money wage and not of the real wage. If the nominal money wage increased (regardless of what happened to the general price level and, therefore, to the real wage) the supply of labour in the market would increase and *vice versa*.

   In Keynes' view, money wages moved more or less in line with the movement of the general price level and it was the aggregate effective demand, not the real wage, which determined the level of employment in the economy. Keynes argued that even if the classical theory demonstrating that a fall in the real wage would increase employment was correct, in real life wages and prices were generally sticky downward. Consequently, the solution prescribed by the classical theory was impracticable. Keynes argued that although the classical theory was logically neat and consistent, it was of no use in helping us to understand the real world.

6. The classical macroeconomic theory has also been criticized for ignoring the speculative or asset demand for money. According to the classical economists, individuals and businessmen hold money only for transactions purposes. They would never hold money as an asset since money as an asset was barren, yielding no return to its owners. Holding of idle cash balances indicates an irrational behaviour on the part of the wealth-holders because cash assets were barren since hoarded money earned no interest. Since a positive rate of interest could be earned by swapping money for some other assets such as the riskless government or the corporate bonds, people would willingly hold the fixed interest income yielding riskless government bonds rather than hold money and earn no interest. In short, the asset or speculative demand for money was ruled out in the classical theory. Consequently, Say's Law of Markets and the quantity theory of money, which ignores the demand for the speculative cash balances, were the two basic pillars on which the entire edifice of the classical macroeconomic theory rested.

Keynes argued that this was not always true. It was possible to envisage a situation in which an individual may choose to hold a part of his assets in the form of money although money was barren as it yielded no income. At some critically low rate of interest (around 2 per cent), people begin to expect that the interest rate will soon rise to a normal level. Since the interest rate and bond prices are inversely related, a rise in the rate of interest means capital losses for the bond-holders. Consequently, at some very low rate of interest, the prospective investors would weigh the probable low interest gains against the highly probable future capital losses and decide against making investment in bonds. This holding of money in the form of idle cash balances is termed as the speculative or asset demand for money. At some very low rate of interest which we may call the critical minimum or the liquidity trap interest rate, the asset demand for money becomes perfectly interest-elastic or infinite because money becomes a perfect substitute for the credit risk-free, although not market risk-free government bonds at this extremely low rate of interest.

Thus, the total demand for money depends on both the money value of the total real output (Y), which was recognized in the classical theory, and on the rate of interest, which was ignored in the classical theory. Money is demanded both for the transactions purpose and for the speculative purpose if the current rate of interest is very low to cause the strong expectations that it will rise in the future. Provided there exists a sufficiently low rate of interest which cannot fall any further, Keynes argued that the full employment investment demand and saving supply schedules would, under certain circumstances, not intersect each other at any interest rate either above this rate or even at this critically low rate of interest. In other words, these schedules would intersect only at an interest rate below this irreducible minimum rate of interest. Consequently, there would be an unfilled gap between the full employment investment and the full employment savings at this critically minimum rate of interest. In other words, at this low rate of interest the aggregate effective demand ($C + I$) will be less than the aggregate output or supply ($C + S$) causing an excess of the aggregate saving over the aggregate investment in the system. For the equilibrium to occur between the two, the aggregate saving will have to fall which is possible only if the aggregate income falls below the full employment income. In short, the aggregate saving and the aggregate investment will be in equilibrium at less than full employment income. This has been illustrated in Figure 1.3 which shows that corresponding to the minimum possible interest rate of two per cent (liquidity trap rate of interest), the full employment investment falls short of the full employment saving by the *DE* amount. In other words,

$C_F + I_F < C_F + S_F$; or $I_F < S_F$

Consequently, the equilibrium aggregate income cannot be maintained at the full employment level of income and it must fall unless the rate of interest can fall further to equate the saving and investment corresponding to the full employment income. But having already touched the floor at the 2 per cent, it cannot fall any more. The stickiness of the rate of interest at this low level under the impact of liquidity trap denies the possibility of the economy achieving the full employment in an automatic manner unless either the investment demand schedule shifts sufficiently to the right such that at the different rates of interest, there is more investment outlay or the saving supply schedule shifts sufficiently to the left such

that the people save less (consume more) at the different rates of interest or both the curves shift simultaneously in appropriate directions. Shifting of these schedules is not, however, possible because it involves the fundamental shifts in the psychological behaviour of the investors and consumers. This gap can be filled through the instrument of fiscal policy by increasing the autonomous expenditure incurred by the government on the public works programmes.

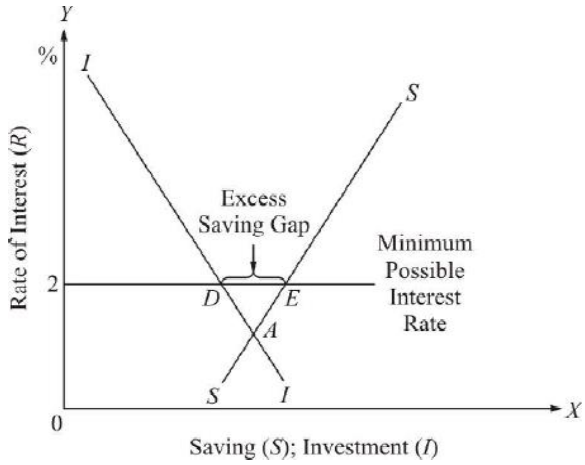


Fig. 1.3

Thus, Keynes argued that the speculative demand for money and its infinitely interest-elastic nature at some critically low rate of interest was really the crucial cause of unemployment in a free enterprise economy. The classical economists were guilty of ignoring this crucial factor. It was the speculative demand for money which, by preventing the interest rate from stabilizing the aggregate demand, threw a great burden on the wage-price flexibility which was imperfect and weak for the job. Moreover, since the speculative demand schedule for money was highly elastic no amount of deflation would work. Rigid money wage is not, therefore, the cause of unemployment; on the other hand, it prevents unemployment from creating a painful and unnecessary fathomless deflation.

The concept of liquidity trap has often been regarded as an article of faith and a bitter controversy has arisen about whether it constitutes the fundamental difference between the classical and the Keynesian economic analysis. In other words, the question is: will Keynes' criticism of the classical theory become meaningless in the absence of liquidity trap? The existence of the liquidity trap is not, however necessary to show that underemployment equilibrium in the economy may exist in certain circumstances. Keynes was somewhat overanxious to criticize the classical theory by seeking resort in the liquidity trap. In fact, even in the classical analysis there is a built-in liquidity trap present at the zero rate of interest since the rate of interest cannot fall below zero. It may well be that under certain circumstances full employment investment demand and saving supply schedules will intersect only at a negative rate of interest. This is likely to be true if the investment demand and the saving supply schedules are highly interest-inelastic and the two are also so situated, as shown in Figure 1.4, that both the investment demand and the saving supply schedules intersect each other below the zero rate of interest, i.e., at the $-R_1$ negative rate of interest. The rate of interest in the classical system, however, cannot fall below zero.

Consequently, at the zero rate of interest saving exceeds investment by *AB* amount (shown as gap) and, therefore, the full employment aggregate supply exceeds the full employment aggregate demand by this amount. Consequently, the Say's Law of Markets is invalidated. In consequence of this disequilibrium between saving and investment, the aggregate income and employment must fall until *I* and *S* are in equilibrium at less than full employment. In this case, the crucial role of the liquidity trap disappears and the issue whether such a trap exists and the interest rate at which it occurs loses practical significance.

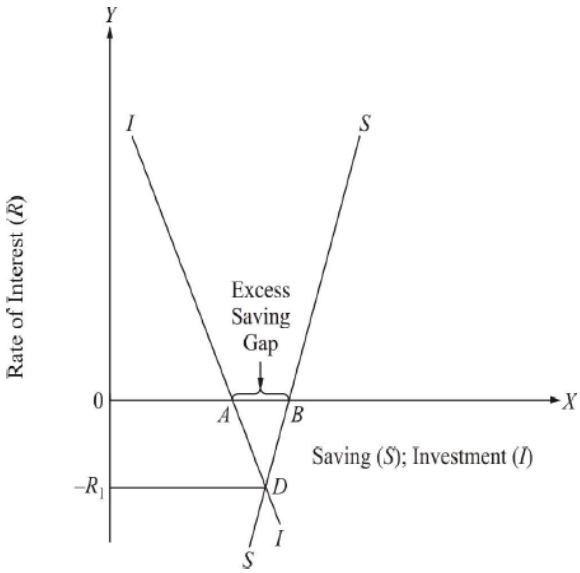


Fig. 1.4

It should be noted that we cannot continue to use the aggregate saving and investment schedules of Figure 1.4 without making substantial modifications. The rate of interest must always be positive but the *ex post* saving and investment are equal by definition. The problem can be solved only by shifting the investment and saving curves appropriately leftward so that they intersect at some positive rate of interest. In the Keynesian system, investment and saving are both functions of the aggregate income while in the classical system, the aggregate income having been regarded as fixed or given, these were regarded as functions of the rate of interest. In the Keynesian system, the investment and saving schedules shift every time the income changes. The amount by which both these two schedules will shift will depend on the marginal propensity in save (*MPS*) and the marginal propensity to invest (MP_1). In the simple Keynesian system, the stability condition requires that the marginal propensity to invest is less than the marginal propensity to save, i.e., $MPI < MPS$ or that the $MPC + MPI < 1$ or that the simple investment multiplier is finite.

If the *ex ante* aggregate investment is less than the *ex ante* aggregate saving, the aggregate income will fall. Consequently, both the aggregate investment demand and the aggregate saving supply schedules will shift to the left.

Since we have assumed that the marginal propensity to save (*MPS*) is more than the marginal propensity to invest (*MPI*) i.e., $MPS > MPI$, for any given fall in the aggregate income, the leftward shift in the saving supply schedule will be more than the leftward shift in the investment demand schedule. Consequently, the two

curves will approach nearer. With the fall in the aggregate income each time, the two curves come nearer until income falls sufficiently to make their intersection at some given positive interest rate possible. Figure 1.5 shows the process of fall in the aggregate income and the resulting leftward shift in the aggregate investment demand and the saving supply schedules. In Figure 1.5, when the aggregate income falls from Y_1 to Y_2 ($Y_1 > Y_2$), the aggregate saving supply schedule $S(Y_1)$ shifts leftward to the position of the dotted $S(Y_2)$ saving supply schedule. The aggregate investment demand schedule $I(Y_2)$ also shifts leftward to the position of the dotted investment demand schedule $I(Y_2)$. But in spite of both the curves shifting to the left, they intersect each other corresponding to the R_2 positive rate of interest because the saving supply curve shifts more to the left than does the investment demand curve. Consequently, Y_2 is the equilibrium aggregate income which is less than the full employment income.

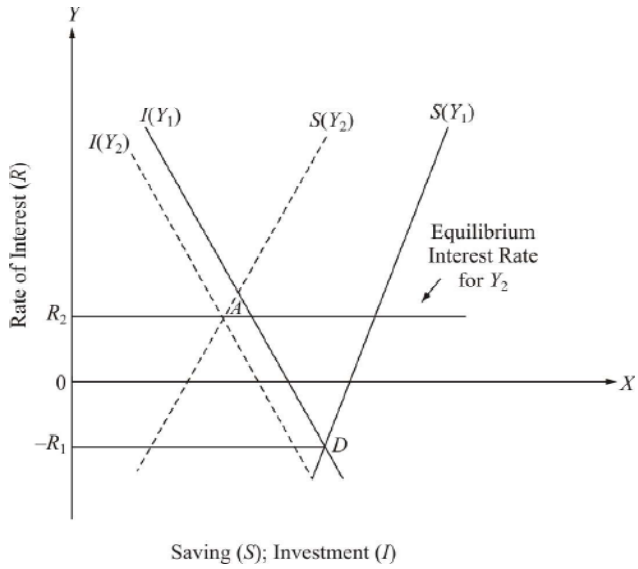


Fig. 1.5

In short, the crucial argument in the invalidation of the classical theoretical economic system is that the *ex ante* investment is not always equal to the *ex ante* saving at some positive rate of interest. Once this follows, there will be idle cash balances, the velocity of money will fluctuate cyclically and the money stock will no longer be proportional to the money output even though the prices are flexible. Consequently, the quantity theory of money will be invalidated and the aggregate demand function depending on the level of income will be necessary to determine the equilibrium level of the national product.

7. Keynes also criticized the classical dichotomy between the real and the monetary sectors of the economy. According to the classical economists, money was neutral and changes in the supply of money and its velocity, i.e., monetary changes did not exert any influence whatsoever on the relative prices of commodities although such changes significantly affected the absolute or general price level in the economy. Consequently, changes in the quantity of money did not affect the aggregate real output and employment in the economy. In the classical macroeconomic theory, money was nothing more than a convenient measuring yardstick in terms of which the relative values were stated and real flows were

measured. In his *General Theory*, Keynes related money and the aggregate demand. He introduced the 'causal nexus' between the aggregate supply of money and the rate of interest. According to Keynes, by influencing the aggregate investment spending the rate of interest influenced the aggregate effective demand in the economy. Consequently, Keynes successfully integrated the monetary and real sectors of the economy which were treated mutually exclusive in the classical macroeconomic theory. While the classical theory was concerned with a world which was undisturbed by uncertainty regarding the future, Keynes was concerned with an uncertain world in which money served as an important link between the present and the future.

1.5 EVALUATION OF CLASSICAL MODEL

Notwithstanding the scathing attacks of John Maynard Keynes and others on the classical theory of output, employment and price level, it is not completely dead. The so-called 'Keynesian Revolution' has not been able to wipe out the 'old order' completely. The disputes between Keynes and the classicists have at times been overstated. For example, the interest rate controversy whether interest rate is a real or a purely monetary phenomenon has been blown out of all legitimate proportions. We still have the monetarists with their modern quantity theory of money based upon the foundation which has its links with and in general outline resembles the old quantity theory of money. As a result of the monetarists' persuasive arguments and the serious empirical studies undertaken supporting the view that 'money does matter', the monetary policy has regained its lost position as an effective tool of economic policy both in depression and inflation. It has now regained parity with the fiscal policy as an area of concern and field of research.

Monetary policy has crept back towards the centrestage as an economic stabilization technique. It is not altogether correct to describe the classical theory of output, employment and price level as faulty and despite the great popularity and dramatic success of the Keynesian theory over the past seven decades, not few in positions of great responsibility, both in government and in business, have been raised on the teachings of the old theory. Alexander Gray has correctly stated that 'no point of view, once expressed, ever seems wholly to die; and in periods of transition like the present our ears are full of the whisperings of dead men.' In short, for a proper understanding of the complete macroeconomic theory, it is essential on our part to understand and acquire a thorough grasp of the classical macroeconomic theory.

1.6 KEYNESIAN DETERMINATION OF INCOME, OUTPUT AND EMPLOYMENT

Unlike the classical macroeconomic theory in which the real and monetary sectors of the economy remain dichotomized, in the Keynesian theory these are integrated forming parts of a compact whole. In the Keynesian theory, all the different sectors of the system remain tied together and all the variables are determined together. Notwithstanding the many crudities and rough edges present in the Keynesian macroeconomic theory, we owe it to Keynes' original pioneering work for the tremendous advances that have taken place in the field of macroeconomic analysis over the past seven decades following

Check Your Progress

3. What are the seven categories of unemployment?
4. State the impact of cyclical unemployment.
5. State one crucial argument in the invalidation of the classical theoretical economic system.

NOTES

the publication of Keynes' epoch-making book entitled *The General Theory of Employment, Interest and Money* in 1936. Although Keynes' apparatus was incomplete in certain respects and imperfect in certain others and although he employed it for drawing some wrong conclusions, these in no way diminish the excellence of his achievement.

Economists differ in making a partial appraisal of Keynes' contribution. Some mention the consumption function, others mention the investment multiplier while still some others mention the liquidity preference function as Keynes' single major contribution. For example, Alvin H. Hansen, a widely read author on current economic issues, considers the concept of consumption function as the 'heart of the Keynesian analysis' and an epoch-making contribution to the tools of economic analysis, analogous to but even more important than Marshall's discovery of the demand function. However, in making any piecemeal appraisal of Keynes' contribution, the essential nature of his originality is lost because it is not so much in the uniqueness of the individual pieces of his theoretical apparatus that Keynes' originality rests. In fact, antecedents for practically every component of the Keynesian model can be traced in the ideas of earlier economists. What is, however, striking about Keynes' genius is his deep insight which enabled him to combine all the individual components, creating in the process a different technique of quasi-general equilibrium analysis. Predating Keynes, onlyAdam Smith and Leon Walras were gifted with a similar rare vision and we can rank Keynes with them as one of the truly original economic thinkers of all times.

Simple Keynesian Theory

The simple Keynesian theory of income, output and employment determination can be studied either through the aggregate income-expenditure approach in the form of $Y = C + I$ or through the aggregate saving-investment approach in the form of $S = I$. Although it hardly matters as to which one of these two alternative approaches is adopted for the purpose of analysis since both have the same purpose, we shall, however, adopt the aggregate saving-investment approach in the form of $S = I$ to study the simple Keynesian model of income, output and employment determination.

The classical theory of income, output and employment determination focusses attention on the equilibrium between the demand for labour (in the production of consumption and capital goods) and the supply of labour in the labour market under perfect competition. The theory has been so designed that the outcome is full employment of resources, particularly of labour in the economy. Once full employment is determined, the general price level is determined by the naive 'quantity theory equations' approach quite independently of the real variables while the rate of interest is determined through the balancing mechanism between saving and investment.

The Keynesian system, however, has a different setting. The saving-investment relationship is no longer immaterial for determining the aggregate output. In fact, as a determinant of income, the saving-investment relationship exerts a dominating influence. According to Keynes, saving is a function of income and is related to income through the marginal propensity to save ($MPS = dS/dY$). The marginal propensity to save is positive but less than one, i.e., $0 < dS/dY < 1$. The saving supply function for the simple aggregate saving-investment Keynesian model may, therefore, be written as–

$$S = s(Y); \text{ and } dS/dY > 0 \qquad \ldots (1.1)$$

The investment demand may be treated as wholly autonomously determined being unrelated to the level of income, i.e., $I = I_A$. Alternatively, it may be treated as functionally dependent on income increasing as income increases and *vice versa*. Accordingly, the investment demand function can be written as–

$$I = i(Y); \text{ and } dI/dY > 0 \qquad ...(1.2)$$

For the stability of the system it is also necessary to assume that $dI/dY < dS/dY$, i.e., the investment demand function cuts the saving supply function from above as has been shown earlier in Figure 1.6(B) and Figure 1.7. There is also the following equilibrium condition of–

$$S = I \qquad ...(1.3)$$

Equations (1.1); (1.2) and (1.3) form a complete system and can be solved for the equilibrium values of *S*, *I* and *Y*. The determination of equilibrium income satisfying the equilibrium condition $S = I$ can be shown by means of Fig. 1.6 which has two parts *A* and *B*. In part *A* of the figure, investment as shown by the investment demand function $II = \bar{I}_A$ is wholly autonomously determined at the constant amount $\bar{I}_A$ while the saving is positively related to income. The equilibrium income is $0Y_e$. In part *B* of the figure, both the saving supply and the investment-demand are positively related to income. The investment demand function cuts the saving-supply function from above. In other words, the slope of the investment demand curve is less than the slope of the saving supply curve, i.e., $dI/dY < dS/dY$. Here also the equilibrium income is $0Y_e$.

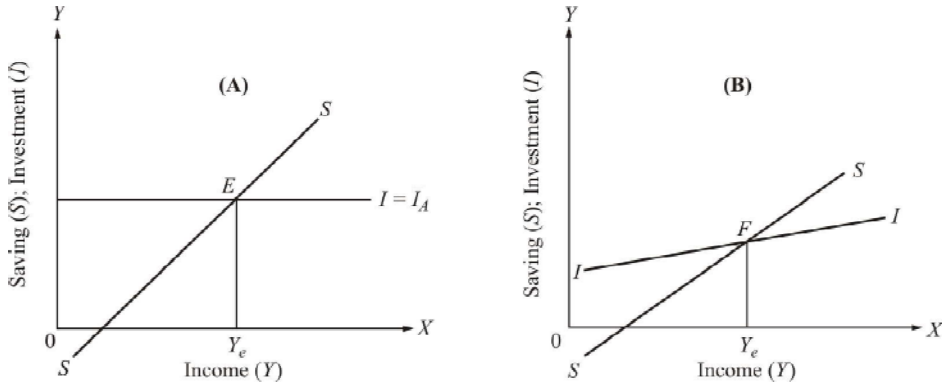


Fig. 1.6

Treating investment as wholly autonomous, consequent upon any given increase of $\Delta \bar{I}_A$ in the autonomous investment outlay the increase in the equilibrium income will be $\Delta Y = \Delta \bar{I}_{A/S}$ and the simple investment multiplier ($K = I/s$) operates with full force. Where, however, the investment is partly autonomous and partly induced, being positively related to income, the increase in the equilibrium income due to the given increase of ΔI_A amount in the autonomous investment will be equal to the product of the initial increase in the autonomous investment and the super-multiplier (K') which is the inverse of one minus the marginal propensity to spend, i.e., $\Delta Y = \Delta \bar{I}_A / 1 - b - e$.

This simple saving-investment Keynesian model of income and output determination focusses attention only on the equilibrium of the product market, with no consideration of what happens in the money market. It is not, however, realistic to treat investment either as wholly autonomous or related to only income since investment is also negatively related to the rate of interest as was stressed by the classicists. Keynes also did not deny that *ceteris paribus* investment would increase as a consequence of the fall in the rate of interest although he disagreed with the classicists on the issue of high interest elasticity of investment. According to him, the interest elasticity of investment was small with the result that a relatively large fall in the rate of interest was necessary to cause a given increase in investment demand which was necessary to generate full employment income in the economy. The rate of interest was, however, prevented from falling very low by the phenomenon of the liquidity trap where the asset demand for money becomes perfectly interest-elastic. Treating investment as a function of the rate of interest alone, the investment demand function may be written as–

$$I = i\,(r);\ \text{an}\ dI/dr < 0 \qquad ...(1.4)$$

The saving supply function which is, as before, a positive function of income may again be written as the equation in the following page.

$$S = s(Y);\ \text{and}\ dS/dY > 0 \qquad ...(1.5)$$

The equilibrium condition is once again given by the equation

$$S = I \qquad ...(1.6)$$

It is obvious that equations (1.4); (1.5) and (1.6) do not form a complete model and cannot be solved for the equilibrium values of the variables *S, I, Y* and *r* as there are four unknowns in only three equations. If either the rate of interest or income was known, it would be possible to solve either for the equilibrium values of *I, S* and *Y* or for the equilibrium values of *I, S* and *r.* In short, given the level of income *Y,* it would be possible to find out that rate of interest *r* at which the equilibrium condition $S = I$ would hold. Similarly, if the rate of interest is given, the value of *Y* corresponding to which saving equals investment can be found out. It is obvious that there is no single combination but numerous pairs of *Y* and *r* at which *S* and *I* are equal, i.e., at which the equilibrium condition $S = I$ holds. All such pairs of the aggregate income *Y* and rate of interest *r* when plotted on a graph would give us the *IS* curve.

The *IS* curve is negatively sloping showing that a higher *Y* requires a lower *r* for the equilibrium condition between *S* and *I* to hold. The reason for this relationship between *Y* and *r* is easy to understand. Since investment *I* is negatively related to interest rate *r,* a high *r* would mean a low *I* which via the investment multiplier would mean a low *Y* and *vice versa.* Alternatively, if *Y* is high, *S* must also be high since saving is a positive function of income. For *I* to be equal to a high *S,* the rate of interest has to be low since investment is inversely related to the rate of interest. Conversely, low *Y* means low *S* and for *I* to be equal to low *S,* the rate of interest *r* has to be high.

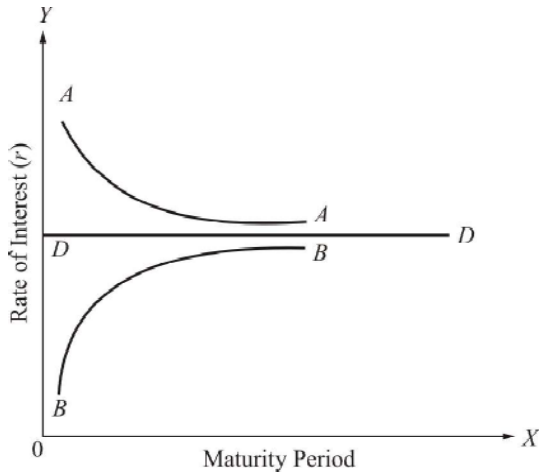


Fig. 1.7 *IS Curve*

The derivation of the *IS* curve alone does not, however, help us for it does not provide us with the unique combination of *Y* and *r* for which the equilibrium condition *S* = *I* holds. Being the locus of innumerable pairs of *Y* and *r* corresponding to which the aggregate saving *S* is equal to the aggregate investment *I*, each one of these pairs of *Y* and *r* is equally good. Unless *Y* is known beforehand, we cannot know the appropriate *r* or unless *r* is known we cannot know the relevant *Y*. While the equilibrium *Y* can be known from the saving-investment relationship in the product market, in the Keynesian system the rate of interest being a purely monetary phenomenon, is determined by the demand for and the supply of money in the money market. This requires the consideration of the supply of and the demand for money as an essential part of the Keynesian theory of income, output and employment.

Demand for and Supply of Money

The most distinguished feature of the Keynesian approach to the demand for money is the speculative demand for money M_2 which is inversely related to the rate of interest. Moreover, the speculative demand for money becomes perfectly interest-elastic at some critically low enough (around 2 per cent) rate of interest as a result of which the rate of interest becomes sticky in the downward direction. This situation has been termed the *liquidity trap*. It completely emasculates the monetary policy as an effective remedy for removing unemployment in the economy. The transactions demand and the precautionary demand for money (M_1) are positively related to the level of income. The composite demand function for money may, therefore, be written as:

$$M_d = kY + L(r) \qquad ...(1.7)$$

where $Y(= PQ)$ is the aggregate money income and k is the proportionality factor showing that the transactions and precautionary demand for money M_1 is k times of the aggregate money income Y. The constant value of k is more than zero but less than one, i.e., $0 < k < 1$. In the above equation, $L(r)$ is the speculative demand for money. The total demand for money M_d is the composite of the transactions demand, precautionary demand and the speculative demand for money. The total supply of money M_s is exogenously given, being autonomously determined by the monetary authority. We have the following equilibrium condition:

$$M_s = M_d = kY + L(r) \qquad ...(1.8)$$

Here also there are many combinations of *Y* and *r* corresponding to which the demand for and supply of money are equal, i.e., the money market is in equilibrium. Since the transactions demand for money is a positive function of income and the speculative demand for money is a negative function of the rate of interest, corresponding to a higher income *Y,* the rate of interest *r* must also be higher if the equilibrium condition $M_s = M_d$ is to be satisfied. The explanation for this is simple. Higher income *Y* means higher transactions demand for money which, with the given money supply, can be met only by releasing a part of the speculative cash balances which will be possible only if the rate of interest rises and *vice versa.* Plotting all such equilibrium pairs of *Y* and gives us the *LM* curve.

Taking the two equilibrium conditions as shown in equations (1.6) and (1.8) together, we obtain the two relationships between *Y* and *r.* From the equilibrium equation (1.6) *Y* is a decreasing function of *r* and from the equilibrium equation (1.8) *Y* is an increasing function of *r* such that *Y* approaches the maximum income as *r* approaches infinity and *r* approaches the minimum rate of interest as *Y* approaches zero. It is, therefore, obvious that the negatively sloping *IS* curve and the positively sloping *LM* curve will intersect each other only once, as has been shown in Figure 1.8, giving a unique solution for the equilibrium values of *r* and *Y* . In Figure 1.8, the *IS* and the *LM* curves intersect each other at point *A* corresponding to which we have the equilibrium values for *r* and *Y* in $0r_e$ and $0Y_e$ respectively. The combination of $0Y_e$ aggregate income and $0r_e$ rate of interest ensures the simultaneous equilibrium in the goods market (real sector) and the money market (monetary sector) of the economy. For any other values of *r* and *Y,* the two sectors will not be in simultaneous equilibrium.

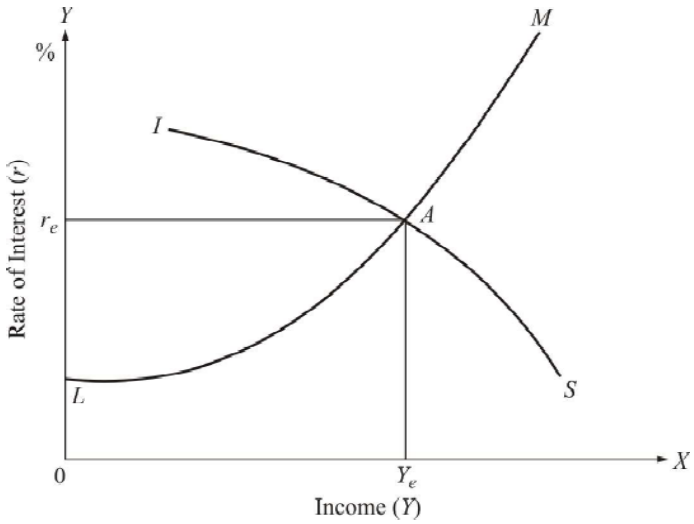


Fig. 1.8

In this two-sector Keynesian model, the money supply is treated as autonomously given at $\bar{M}_s$, i.e., M_s is a parameter. An increase in the total money supply M_s will increase the equilibrium aggregate income Y_e and reduce the equilibrium rate of interest r_e assuming no change, i.e., no shift in the position of the *IS* curve. This conclusion is in sharp contrast

to the classical model where an increase in the money supply increases only the general price level making no dent on the rate of interest. In the classical macroeconomic analysis, changes in the aggregate money supply do not affect the real sector of the economy in which total employment and output are determined.

The Keynesian model outlined here assumes a given general price level *P*. Ashift in either the *IS* curve or the *LM* curve or in both these curves would shift the equilibrium position affecting the equilibrium values of *r* and *Y*. The effect of any given shift in the *IS* curve (assuming no change in the position of *LM* curve) on *r* and *Y* would depend on whether the *IS* curve intersects the *LM* curve on a steeply rising or a relatively flat portion. If the intersection between these two curves takes place on the steeply rising portion of the *LM* curve, income would rise or fall only moderately while the rate of interest would increase or decrease substantially. This is the classical range. On the other hand, if the intersection takes place on the flat part of the *LM* curve, income would rise or fall substantially with little or no effect on the rate of interest. This is the Keynesian range. If the *IS* curve intersects the *LM* curve in the intermediate range where it has a marked curvature, the effect on *r* and *Y* would be more or less evenly felt with both *r* and *Y* rising or falling in the process as the *IS* curve shifts either upward to the right or downward to the left. Ashift to the right in the *LM* curve (with no change in the *IS* curve) in the classical range will raise the level of income substantially while at the same time causing a substantial fall in the rate of interest. In the Keynesian range there is, however, no effect either on the level of income *Y* or on the rate of interest *r* of such a rightward shift in the *LM* curve which is brought about by an increase in the money supply.

In this analysis, saving is dependent only on the income *Y* and it is regarded as completely interest-inelastic. Investment depends only on the rate of interest *r* and is unrelated to income. When the investment is interest-inelastic so that changes in *r* either do not effect *I* or effect it only very slightly, the *IS* curve will be either vertical or near-vertical. Consequently, shifts in the *LM* curve caused by changes in the money supply will have no or very little effect on the equilibrium income *Y* and will affect only the equilibrium rate of interest *r*.

Alternative Diagrammatical Analysis

The Keynesian model of income determination explained through the Hicks-Hansen *IS-LM* curves diagram is very simple and on account of its simplicity it commands wide appeal. It has, however, the serious disadvantage of hiding from sight the working of those functions or factors which lie behind the *IS* and the *LM* curves. We may, therefore, adopt an alternative diagrammatical approach which exposes to a greater extent the actual relationships involved in the analysis. As before, we assume that saving depends only on income, i.e., $S = s(Y)$ and investment depends only on the rate of interest, i.e., $I = i(r)$. While this kind of assumed relationship simplifies matters greatly, the results would not change materially if we make saving also dependent on the rate of interest and investment also dependent on income; only the analysis would become too cumbersome to handle. The relationship of saving with the rate of interest and the relationship of investment with income which have been omitted are quantitatively insignificant.

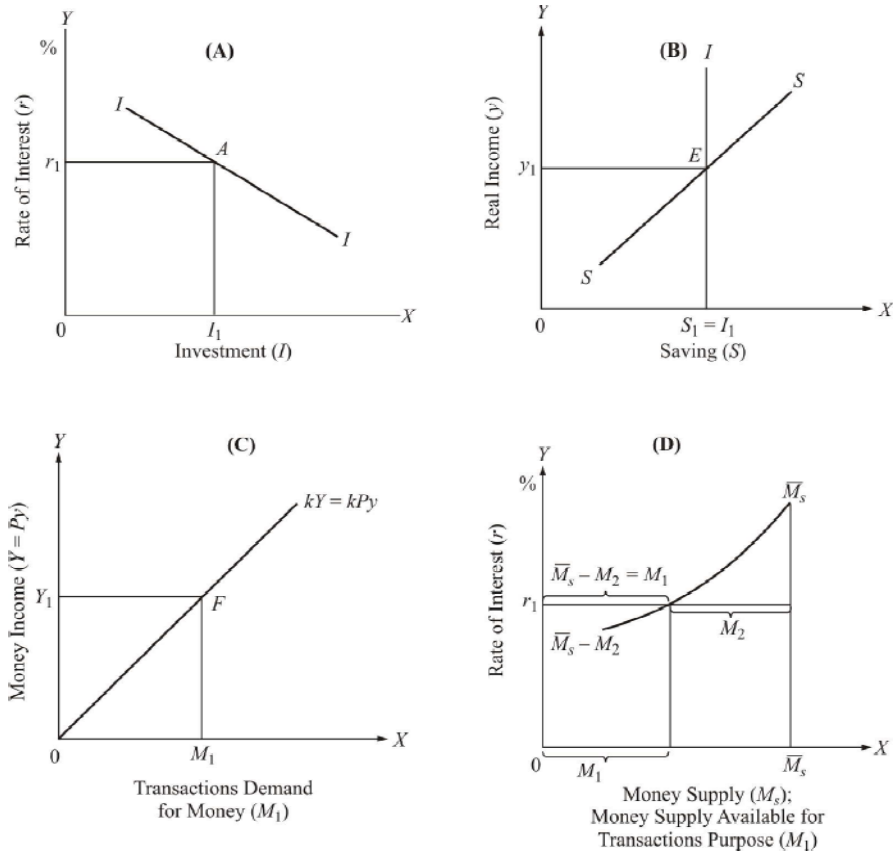


Fig. 1.9

The analysis can be explained diagrammatically by means of Figure 1.9 which has four parts. Part *A* of the figure shows the investment demand function which is inversely related to the rate of interest. Part *D* relates the rate of interest *r* and the speculative demand for money M_2 in such a way that with the given money supply M_s, it shows the total money supply available to satisfy the transactions demand at each different rate of interest such that the total demand for money equals the total money supply at that rate. Part *B* relates saving with the real income in a positive manner. Part *C* shows the relationship between the total money income and the transactions demand for money M_1. Parts *A* and *B* of the figure together show the equilibrium of the real sector or the goods market and furnish us with the equilibrium combination of the rate of interest *r* and the real income *y* such that corresponding to this particular combination of *r* and *y* the aggregate saving equals the aggregate investment, i.e., $S = I$. Parts *C* and *D* of the figure show the equilibrium of the economy's monetary sector giving that combination of the rate of interest *r* and the money income *Y* corresponding to which the monetary sector of the economy is in equilibrium in the sense that the total demand for money equals the total supply of money, i.e., $M_d = M_s$.

All the four parts of the figure put together give us a unique combination of income and rate of interest corresponding to which the real and the monetary sectors of the economy are simultaneously in equilibrium, satisfying the double condition of equilibrium $S = I$ and $M_d = M_s$. Such a unique combination of the aggregate money income and rate of interest is $Y_1 (= P_1 y_1)$ and $0r_1$. The analysis assumes a given general price level P_1 such that any given change in the aggregate real income *y* produces an equi-proportionate change in the aggregate money income $Y (= Py)$. For simplicity, if we assume $P = 1$, the aggregate real income (y) and the aggregate money income (Y) would be identical in the

sense that any given distance on the two income axes would measure the same units both in the real and money terms, Furthermore, in the diagram the income is compressed, i.e., a given distance on the income scale measures more than the same distance on the investment and saving and the transactions demand for money scales. Part B shows the equilibrium between saving an investment. While saving is a positive function of the aggregate real income, investment is autonomously determined. In part *D*, the $\bar{M}_s - M_2$ curve shows the total quantity of money which is available out of the given money supply $\bar{M}_s$ at different rates of interest for the transactions purposes (M_1). The total money supply remaining unchanged at $\bar{M}_s$, as the rate of interest increases so also increases the total supply of money available to meet the transactions demand for money. This is understandable since at the higher rate of interest less amounts of cash balances are held for the speculative purposes thereby making a larger part of the given money supply $\bar{M}_s$ available for meeting the transactions demand for money M_1 which is determined by the level of money income *Y* and the proportionality factor *k* as shown in part *C* of the figure. From the composite figure we settle down at the equilibrium values of money income, rate of interest, investment and saving as given by $0Y_1$, $0r_1$, $0I_1$ and $0S_1$. Any other set of values for these variables would be inconsistent with the general equilibrium. All these values are, however, valid on the assumption of given general price level P_1 which also gives the equilibrium aggregate real income $0y_1$ (which equals the aggregate money income $0Y_1$ divided by the general price level P_1) i.e., $0y_1 = 0Y_1/P_1$.

Effects of Shift in the Investment Demand Function

Any given shift in the aggregate investment demand function *II* in Figure 1.9(A) will change the equilibrium values of all the variables by setting in motion a whole chain of reactions. Ashift in the aggregate investment demand function might either be reflected in the leftward or rightward shift in the entire investment demand function. We may assume the shift to come about as a result of a given increase of ΔI amount in the autonomous investment spending which shifts the entire investment demand curve *II* rightward to the position of new investment demand curve *I′I′*as shown in Figure 1.10(A) (at the old $0r_1$ rate of interest the aggregate investment increases from $0I_1$ to $0I_3$, i.e., by I_1I_3 (= ΔI) showing that at each rate of interest investment increases by ΔI. Thus, the aggregate real income *y* also increases from $0y_1$ to $0y_3$ i.e., by y_1y_3 (= $K.\Delta I$) through the multiplier effect. The increase in the aggregate real income causes an equi-proportionate increase in the aggregate money income *Y* (given the general price level P_1) from $0Y_1$ to $0Y_3$. This raises the transactions demand for money $0M_1$ from $0M_1$ to $0M_1''$. With the given money supply $0\bar{M}_s$, the increased transactions demand for money can be met only by releasing a part of the speculative cash balances which would require an increase in the rate of interest *r* from $0r_1$ to $0r_3$ as shown in Figure 1.10(D). The rise in the rate of interest rebounds on investment cutting back the investment; the size of the actual cut in investment will depend on the interest elasticity of the investment demand function. The above entire chain of effects assumes no change in the rate of interest while investment increases by ΔI. The rate of interest, however, rises in the process of meeting the increased transactions demand for money M_1 out of the given money supply $\bar{M}_s$. This means that the entire chain of effects traced above could not take place. The new equilibrium rate of interest will be higher than $0r_1$. Consequently, the actual effective increase in investment will be smaller than ΔI due to which the new equilibrium investment will be smaller than $0I_3$ and the new equilibrium real income will be smaller than $0y_3$. The new equilibrium money income and the transactions demand for money will also be

NOTES

smaller than Y_3 and M_1 respectively. The new equilibrium variables shown by $0r_2$, $0y_2$, $0I_2$, $0Y_2$, and $0M'_1$ show that the actual increase in the aggregate real income is less than what the simple investment multiplier effect would have given.

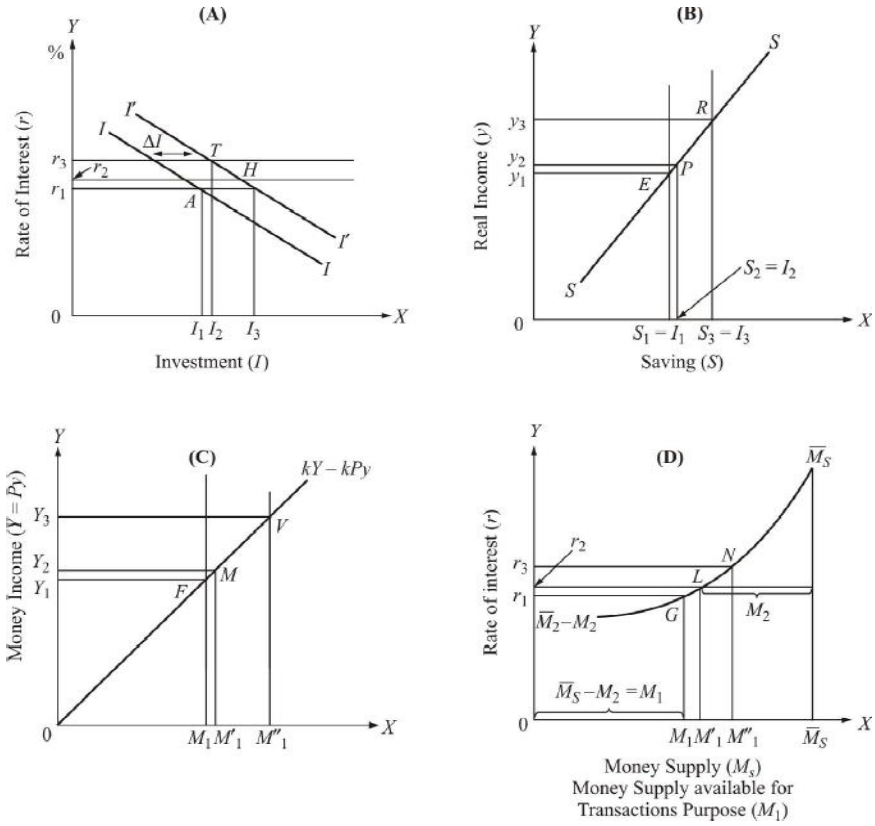


Fig. 1.10

The dampening of the multiplier effect is due to the rise in the rate of interest which adversely affects the investment outlay which in turn reduces the aggregate real income. The extent to which the rise in the rate of interest will check the increase in income will depend on the slopes of the investment demand schedule, saving supply schedule, transactions demand and the speculative demand for money schedules. If the investment demand schedule has a steep slope, a given rise in the rate of interest will cause a relatively small decrease in investment. Combined with a high MPS, i.e., steeply sloping saving supply function, the dampening of the increase in income will be relatively small. Similarly, a low *k* and a gently sloping speculative demand for money curve would mean a smaller rise in the rate of interest for a given increase in the money income.

It is obvious that in this model the conclusions of the simple Keynesian model have been substantially modified. No longer is the monetary policy sterile. In fact, it now assumes a new role of affecting the investment and income via influencing the rate of interest. No longer does the increase in income consequent upon a given increase in the autonomous investment or consumption spending take place according to the simple investment multiplier except in the limiting Keynesian case within the range of the liquidity trap interest rate. Similarly, monetary policy alone does not matter in determining the equilibrium aggregate income except in the other limiting situation known as the classical range where the rate of interest is so high that the speculative demand for money falls to zero. In this classical range, the conclusions of the simple Keynesian model are completely

invalidated and the investment multiplier becomes zero. In the classical range, monetary policy alone is powerful and fiscal policy does not matter. On the other hand, in the Keynesian range fiscal policy alone matters and the simple investment multiplier operates with the monetary policy becoming completely ineffective.

Employment and Flexible Wage

A minimum money wage fixed above the equilibrium wage in a competitive labour market will lead to unemployment. However, if this was the only thing which Keynes had demonstrated, there would hardly have been any controversy between him and the classicists. In fact, if Keynes had proved the possibility of competitive under-employment equilibrium only by invoking the rigidity of money wage in the downward direction, it would have hardly justified his break with the classical economics. On the issue of rigid money wage being responsible for unemployment, there is no difference between Keynes and the classicists. In fact, there is a full chapter in Pigou's work entitled *Industrial Fluctuations* published in 1927 in which it has been asserted that institutional wage rigidities were capable of causing unemployment. It is, therefore, obvious that Keynes' break with the classical economics had something more serious to it. Keynes denied the classical presumption that general wage-cutting, even if it was feasible, would increase the aggregate effective demand. Keynes' main message, as Don Patinkin has rightly stated, was to show that the 'automatic adjustment process of the market (even with the real-balance effect and even when supplemented by monetary policy) is too unreliable to serve as a practical basis of full employment policy.' The main thrust of Keynes' indictment was that his predecessors had suggested wage-cutting and the cheap money policy as the only remedy to eliminate unemployment in the economy. Keynes cast serious doubts on the efficacy of such policy measures and instead suggested that the government should take to deficit budgeting by means of massive spending on public works. This is the core of contrast between the classical and the Keynesian economics.

Keynes' argument against the general wage-cutting (as an effective and practical remedy against unemployment) was based on the fact that while a policy of wage-cut in one single industry might help to increase the demand for the product of that industry, but to jump from this to the conclusion that an economy-wide general cut in wages would likewise also increase the aggregate effective demand was erroneous. When the wage is reduced in any single industry, the demand curve for the product of that industry does not shift downward to the left following the wage-cut because a wage-cut in any one industry or firm does not significantly affect the total wage income in the economy. This is not, however, so with the aggregate demand curve. When we consider the economics of general wage-cutting, the aggregate demand function shifts downward with each dose of general cut in money wages. Consequently, the aggregate effective demand does not increase as a consequence of pursuing a policy of general wage-cut. On the contrary, it may well be that the aggregate effective demand may fall and consequently employment may decrease. Moreover, the fall in the interest rate occasioned through the fall in the transactions demand for money as a consequence of the fall in money wages and prices, may fail to generate sufficient increase in investment demand due to the interest-inelastic nature of the investment demand schedule. It may as well be that the rate of interest already being in the region of the 'liquidity trap' may not fall at all shattering all our hopes to remove unemployment by resorting to the policy of wage-price deflation. Moreover, the policy of general wage-cut was harsh and left behind a trail of social bitterness even if it could be economically defended.

NOTES

Criticism of the Keynesian Theory

Although decidedly an improvement over the classical theory of output and employment, the Keynesian theory of income, output and employment determination has been criticized on several grounds. *First*, it has been argued by the critics that the Keynesian model is 'too aggregative'. It has fewer than necessary variables and relationships. It does not segregate the variables involved into their various component parts. For example, investment has been discussed as a single entity without breaking it into fixed plant investment, inventory investment, residential investment, etc. Had this been done, we would have been able to understand better the effect of the autonomous factors—for example, of the government measures aiming at affecting the house-building activity in the economy or the supply position of capital goods—on these different categories of investment. Similarly, consumption outlay has not been broken into the expenditure on the consumer durables and non-durables.

According to some critics, the use of the aggregate expenditure concepts such as the aggregate consumption and investment dooms the model to give wrong or seriously misleading advice. The unit of analysis must be the individual commodity, or commodities grouped in some other manner, *e.g.*, by degree of elasticity of supply (contrasting, perhaps, agricultural products, whose output has little short run elasticity; many services, which may have great elasticity of supply but almost zero employment effects; some manufactured goods, production of which can be expanded easily but in proportion to labour input; manufactured goods using imported raw materials, etc.). The issue involved here is a complex one. It does appear that, *particularly for some economies*, the criticism has merit.

Moreover, the Keynesian theory does not consider the problem of income distribution. No consideration has been given in the analysis to the study of consumption patterns of different functional groups in the community. This assumes that another analysis should be made of the study of consumption patterns of the different functional groups in the community. The marginal propensities to consume of the different functional groups have been lumped together. Moreover, the real consumption spending has been made a stable function of real income alone. Consequently, the general price level, rate of interest, quantity of money, total wealth, distribution of income, etc., have not been given their due importance as determinants of consumption. Similarly, saving has not been disaggregated into personal saving and corporate saving.

Second, the Keynesian theory has been criticized as being 'too static' in the sense of being concerned with equilibrium conditions during the short period in which technology and capital stock are given and are not likely to change. 'On the one hand, the model cannot deal with the short-run dynamics of income change; on the other, it is not suited to the analysis of problems of long-term growth.' The static Keynesian analysis offers very little by way of increasing our understanding of short-run cyclical income fluctuations. The most crucial factor in the short-run dynamics of business cycle is the change in inventory investment. In the static Keynesian model, there is no discussion of changes in inventory investment. This criticism becomes all the more serious since the Keynesian model by its own assumption is related to the short period analysis. It has been assumed that the aggregate output is a function of only the employment. This assumes that the other important variables such as capital stock, technology, etc., are assumed as parameters.

The Keynesian theory also suffers from another self-contradiction. If it is a short period analysis model, then the analysis should not deal with those situations in which net

NOTES

investment may be positive or negative, i.e., situations in which the economy's total capital stock is changing. The Keynesian analysis, however, does deal with such situations of growth in the total-capital stock.

Third, the Keynesian assumption of autonomously determined rigid money wage is analytically unsatisfactory and practically untenable.

The *fourth* criticism of the Keynesian analysis of money and interest rate is that it is defective. According to Gardner Ackley, 'Perhaps the most valid criticism of the speculative demand for money analysis is the implication sometimes supplied that the speculative demand schedule possesses any degree of permanence or stability. Its position and shape clearly depend upon the level and dispersion of the interest rate expectations of wealth-holders. While it is conceivable that these might be stable and unchanging, this is hardly plausible. Major revisions in the actual level of interest rates (as, for instance, between 1929 and the mid-30s, or between the end of World War II and the 1950s) must certainly give rise to entirely different speculative demand curves. And even in the short run, one might well suppose that the level of interest rate expectations trails along after the movement of actual rates. The idea that wealth-holders speculate with respect to the interest rate, (i.e., bond prices), and that their expectations prevent the adjustment of the rate to changes in the level which would equate current saving and investment is entirely reasonable, and surely calls attention to a phenomenon of crucial importance.' However, to reduce this idea to an almost stable or permanent schedule is tantamount to carrying the idea further than is reasonable or necessary. Taking cognisance of this criticism increases the complexity of the analysis considerably. It is something which the actual importance of the interest rate for aggregate demand may not justify.

The *fifth* criticism is that Keynes limits the composition of the asset-portfolios of the wealth-holders to money, bonds and goods completely ignoring the fact that there are corporate shares, debentures and a myriad of other securities which the wealth-holders acquire as a part of their asset-portfolios. Even when we confine ourselves to only bonds, these bonds are of differing maturities issued by the various government and semi-government institutions carrying with them different risk component. It is, therefore, obvious that investment may be affected as much by the differential availability of the different forms of finance as by the rate of interest.

The *sixth* criticism is that the Keynesian theory ignores lagged relationships with the result that it excludes discussion of both the dynamic multiplier and the accelerator. Moreover, the theory was drawn in terms which considered it much more relevant to the demand-deficiency situations than to the conditions of excess demand or inflation. Keynes did not distinguish clearly enough between *ex post* and *ex ante* magnitudes and a fierce controversy raged for many years over the issue of *ex post* saving and investment.

Finally, Keynes offered no empirical evidence for making some crucial assumptions in his model. For example, he offered no empirical evidence about the characteristic shapes of the consumption and liquidity preference functions employed by him in the model.

Yet another criticism of the Keynesian model arises from the failure of the model to consider the impact of the Pigou or the real-balance effect. Keynes rejected the general wage-cut as a policy for reducing the unemployment largely on the ground that whatever good the general wage-cut might accomplish could as well be achieved through the other quicker and more efficient ways. In this, however, it can be said in defence of Keynes that unless the real-balance effect was very strong, wages cannot be made flexible in the downward direction to the extent required by the real-balance effect to be

NOTES

effective. Empirical evidence is lacking in support of the view that the real-balance effect is of any appreciable significance for the economy as a remedy against unemployment.

Evaluation

All the criticisms notwithstanding, it is, however, still true to say that the Keynesian variables are generally operationally significant and lend strength to the Keynesian system. It should at least be recognized that the Keynesian system is internally complete and comprehensive. Given Keynes' assumptions, the analysis proceeds logically to the conclusions it reaches. Moreover, the Keynesian economics has stimulated great interest in the collection, organization and analysis of the data pertaining to national income, aggregate investment, consumption, output, employment, saving, etc. One might, however, say that the Keynesian concepts call for intended consumption, saving and investment. It may also be argued that the Keynesian key investment concept—the marginal efficiency of capital—is not operationally important.

However, all told, the greatest strength of the Keynesian theory of income, output and employment depends upon the operational nature of the Keynesian variables. The consumption function has a vast empirical content and has stimulated great strides in this direction. It has direct relevance to unemployment which today remains a problem of great social concern. Keynes successfully set out his system in terms of the statistically testable relationships that have stimulated empirical investigations and have given great impetus to the growth of econometrics. The Keynesian analysis also nicely isolates those variables which are strategic for the monetary and fiscal policy. Although economic growth is not the core concern of the Keynesian analysis, 'yet with all its acknowledged deficiencies, the Keynesian analysis still stands as the most useful point of departure in macroeconomic theory. Itself incomplete arid imperfect, it remains the foundation of the great majority of the significant theoretical works in macroeconomics of the past two decades. It has also long provided the basic framework for most governmental analyses of economic conditions and forecasts, and, increasingly, of the analyses and forecasts made by private groups and firms.'

Although the Keynesian analysis is inadequate in providing us with a complete understanding of the phenomenon of inflation and cannot furnish till guidance to frame appropriate policies to deal with inflation, it can nevertheless be applied to inflation and provides significant insights. The fact remains that the Keynesian macroeconomic analysis with all its known pitfalls remains the bedrock of many of the important theoretical studies which have been undertaken and completed in macroeconomics over the past seven decades.

1.6.1 Consumption Function

Consumption plays a crucial role in the determination of income and employment generation. If the consumption increases income and employment will also increase.

Consumption function is also known as propensity to consume. The meaning of consumption is the amount spent on consumption at a given level of income, while consumption function or propensity to consume means the schedule showing consumption expenditure at various levels of income.

In other words, it tells you how the consumption increase or decrease with increase or decrease in income of the consumer. We can say that propensity to consume or

consumption function shows a relationship between consumption expenditure and income. With the increase income, expenditure also increase but less than the increase in income (according to Keynesian Psychological Law).

For example, at the level of national income equal to ₹ 1300 crores, the amount of consumption is ₹ 950 crores. Now the national income increases to ₹ 1600 crores, the consumption rises to 1100 crores. Now it is clear that the amount of consumption is different at different levels of national income.

According to Keynesian Psychological Law, 'when income increases consumption also increases, but not as much as the income, because a part of increase in income is saved'.

From the above example, it can be inferred that as the income increases from 1300 crores to 1600 crores, the amount of consumption increases from 950 crores to 1100 crores. Thus, the income increases by 300 crores, while consumption increases by 150 crores, the remaining ₹ 150 crores are saved.

Propensity to consume depends on the several factors like, interest rate, price level and stock of wealth and on some other subjective factors. Keynesian consumption function deals with short-run, so he assumed price level interest rate and stock of wealth constant.

According to Keynes consumption is a function of current income. Thus,

$$C = f(Y)$$

Or, in a specific form, Keynesian consumption function can be written as:

$$C = a + bY$$

Where a and b are constant. 'a' is an intercept term of the consumption function and b is the slope of consumption function and it also shows the marginal propensity to consume and Y represents the level of current income. Keynesian consumption function is shown in Figure 1.11.

In Figure 1.11, national income is measured on X axis and consumption is measured on Y axis. A line OM making 45° angles with X axis has been drawn. This angle shows that with increase in income, consumption also increases by the same amount.

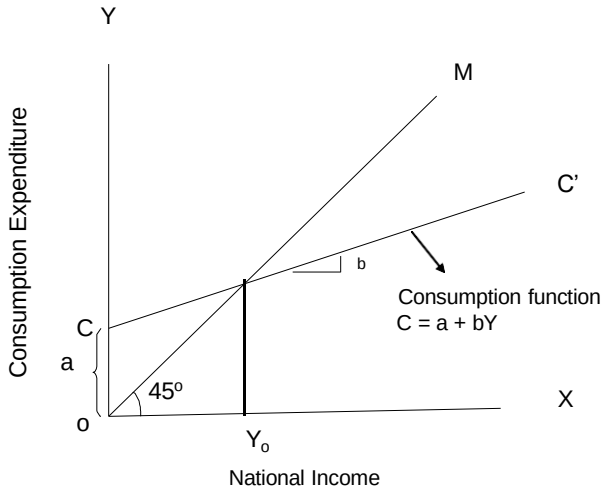


Fig. 1.11 *Keynesian Consumption Function*

But, according to Keynesian consumption function, the actual propensity to consume will deviate from 450º line, that is CC' in Figure 1.11. It is clear from the Figure that at a lower level of income, consumption is more than income. In this condition, people may draw upon their accumulated savings to maintain their consumption level. With the increase in income consumption also increase, at the income level of OYO, consumption is equal to the income. After this point, consumption increases with income, but less than income.

Average and Marginal Propensity to Consume

There are two concepts of propensity to consume, the first one is average propensity to consume (APC) and the second one is marginal propensity to consume (MPC).

Average propensity to consume is the ratio of the amount of consumption to total income.

$$APC = C/Y$$

where APC = Average propensity to consume

C = Amount of consumption

Y = Level of income

Marginal propensity to consume is a very important concept because it tells us how much part of the increased income is consumed and how much saved. Marginal propensity to consume is the ratio of change in consumption to the change in income.

$$MPC = \Delta C / \Delta Y$$

Where MPC = Marginal propensity to Consume

ΔC = Change in Income

ΔY = Change in Consumption

Factors Affecting Consumption and Multiplier Analysis

Now the question is, what are the factors which affect the consumption or propensity to consume. Keynes has divided these factors into two parts, the first one is subjective factors and the second one is objective factors. Now we shall discuss these factors in detail.

Subjective factors

Subjective factors are those which encourage and prompt people to postpone the consumption and to save some part of their income. The major subjective factors which influence the propensity to consume are as follows:

- People save for the future, for example illness, accidents and unemployment. People want to provide for unforeseen contingencies, that's why they save. For example for the education of their children, marriages of their children etc.
- Several people save from their current income, for investment. From this investment they can get more income in terms of profits and investment.
- To increase their social status, people want to more and more wealth, to get large wealth, people are motivated to save.

- For speculative purpose, some people also save to use that amount in making the profit from market fluctuations.
- Because of their miserly instinct and habits, some people also save.

All the above factors decrease the propensity to consume, and hence they increase the propensity to save. These subjective factors play an important role in determining the level of consumption function.

Objective factors

According to Keynes, the following six factors influence the consumption function:

1. **Fiscal policy of the government:** Fiscal policy can affect the consumption function or propensity to consume. Atool of fiscal policy especially various taxation policies like excise duties, direct taxes, sales tax can affect the consumption function. If the government increases the rate of taxation, it cuts down the consumption and hence increases saving of the country. On the other hand if the government cuts the rate of taxes, it increases the propensity to consume and decreases the propensity to save.
2. **Changes in the general price level:** With an increase in the general price level, the consumption function shift downward. Because of the increase in the general price level the real income or purchasing power of the people decrease and hence the propensity to consume also decreases. On the other hand with a decrease in the general price level, propensity to consume or consumption function shift upward.
3. **Rate of interest:** Rate of interest also has a crucial role in affecting the consumption level or propensity to consume. With a higher rate of interest people want to save more to get maximum profit in-terms of interest hence, the propensity to consume or consumption function decrease. On the other hand with a lower interest rate, people feel that it is not the right time to invest, hence they consume their income instead to save. The lower the interest rates higher the propensity to consume. In other words there is a negative relationship between rate of interest and propensity to consume.
4. **Credit conditions:** Availability of credit also plays a crucial role in determining the consumption level of the country. Easy availability of the credit increases the consumption level or propensity to consume of the country. The main reason of global meltdown in America is mainly because of the easy availability of credit to the consumers. In recent years government of India decreases the interest rate of home loan, car loan and on other consumer durables, because of this the consumption level of the citizens of India increased. The emergence of credit cards also increases the propensity to consume of the people because with the help of these credit cards the credit is easily available. On the other hand tightening of the availability of credit shifts the consumption function downward or it decreases the propensity to consume.
5. **Income distribution:** Distribution of income in an economy plays a crucial role in determining the level of the consumption. The unequal distribution of the income has a negative impact on consumption level, because the propensity to consume of the rich people is less in comparison with the propensity to consume of poor people. On the other hand if the distribution of income is equal it has a positive

impact on consumption level, or it increases the propensity to consume of the people.

6. **Stock of wealth:** Stock of wealth include both real assets like land, building, automobiles and financial assets, like cash in hand, savings with banks, shares, bonds etc. More, the stock of wealth more the consumption level while less the stock of wealth less the consumption level or propensity to consume. Because with a more stock of wealth people feel weaker incentive in investment, they prefer to consume rather than save.
7. **Multiplier analysis:** Multiplier is the ratio of increment in income and increment in investment. Here ΔY represents increment in income and ΔI represents the increment in investment

   $\kappa = \Delta Y / \Delta I$, where κ stands for multiplier,

   In other words, we can say that because of increment in investment, more increment in income is created than the increment in investment because of propensity to consume.

$$\Delta Y = \Delta I \frac{1}{1 - MPC}$$

$$\frac{\Delta Y}{\Delta I} = \frac{1}{1 - MPC}$$

or

$$\kappa = \frac{1}{1 - MPC}$$

It is clear from above that the size of multiplier depends upon the marginal propensity to consume of the people.

Calculation of consumption, saving, marginal propensity to consume (MPC), and marginal propensity to save (MPS)

With the help of Table 1.1, you can easily understand the determination of MPC, and MPS.

Table 1.1 *Consumption, Saving, MPC and MPS*

Disposable Income (In ,000₹)	Consumption	Saving	MPC	MPS
0	1	-1	-	-
1	1.75	-.75	.75	.25
2	2.50	-.50	.75	.25
3	3.25	-.25	.75	.25
4	4.00	0	.75	.25
5	4.75	.25	.75	.25
6	5.50	.50	.75	.25
7	6.25	.75	.75	.25
8	7.00	1	.75	.25

You can determine the disposable income of a person, by deducting the amount of direct tax from his or her gross income. Gross income is that income which a consumer can consume. He or she can consume the 100 per cent of it or he or she can save some of the part of it.

As you can see in Table 1.1, when the disposable income of a person is zero, his consumption is one. This is possible as he can consume from his previous savings. When his income reaches ₹ 4,000 his disposable income becomes just equal to his expenditure.

As you know, $MPC = \Delta C / \Delta Y$

In first case it is, $0.75/1 = 0.75$,

When $MPC = 0.75$, $MPS = 1 - MPC$

or $1 - 0.75 = 0.25$

$MPS = 0.25$,

In the whole table,

$MPC = 0.75$

$MPS = 0.25$

1.7 INVESTMENT MULTIPLIER AND BALANCED BUDGET MULTIPLIER

Although the economy's actual aggregate real output may differ from the equilibrium aggregate output due to the businessmen erring by over-producing or under-producing, it will, however, always tend to attain the level of equilibrium aggregate output. The level of equilibrium aggregate income and output changes due to shifts in the aggregate demand function. The shifts in the aggregate demand function can be caused either by the shifts in the aggregate investment demand function or by the shifts in the aggregate consumption demand function or by the simultaneous shifting of both these demand functions. The economy's equilibrium aggregate income and output can also change due to changes in government spending and/or taxes and changes in the imports and/or exports. These shifts are like the shifts in either the market supply and/or market demand schedules of a commodity which cause changes in the equilibrium market price of a commodity.

Although both the aggregate consumption demand function and the aggregate investment demand function can shift upward or downward, however, substantial evidence indicates that consumption function is stable, i.e., although the total consumption spending varies directly with changes in the aggregate income, the consumption function itself is stable. On the other hand, the investment demand function has been seen to be unstable and sudden shifts in the function have caused violent fluctuations in the aggregate income, output and employment in the economy. In short, the amount of consumption outlay corresponding to any given income level is relatively stable and a change in the level of equilibrium aggregate income generally comes about as a result of the shift in the autonomous investment demand function.

Check Your Progress

6. State the most distinguished feature of the Keynesian approach.
7. How is the Keynesian theory of income, output and employment determined?
8. State one ground on which the Keynesian theory is criticised.
9. What is consumption function also known as? What crucial role does it play?

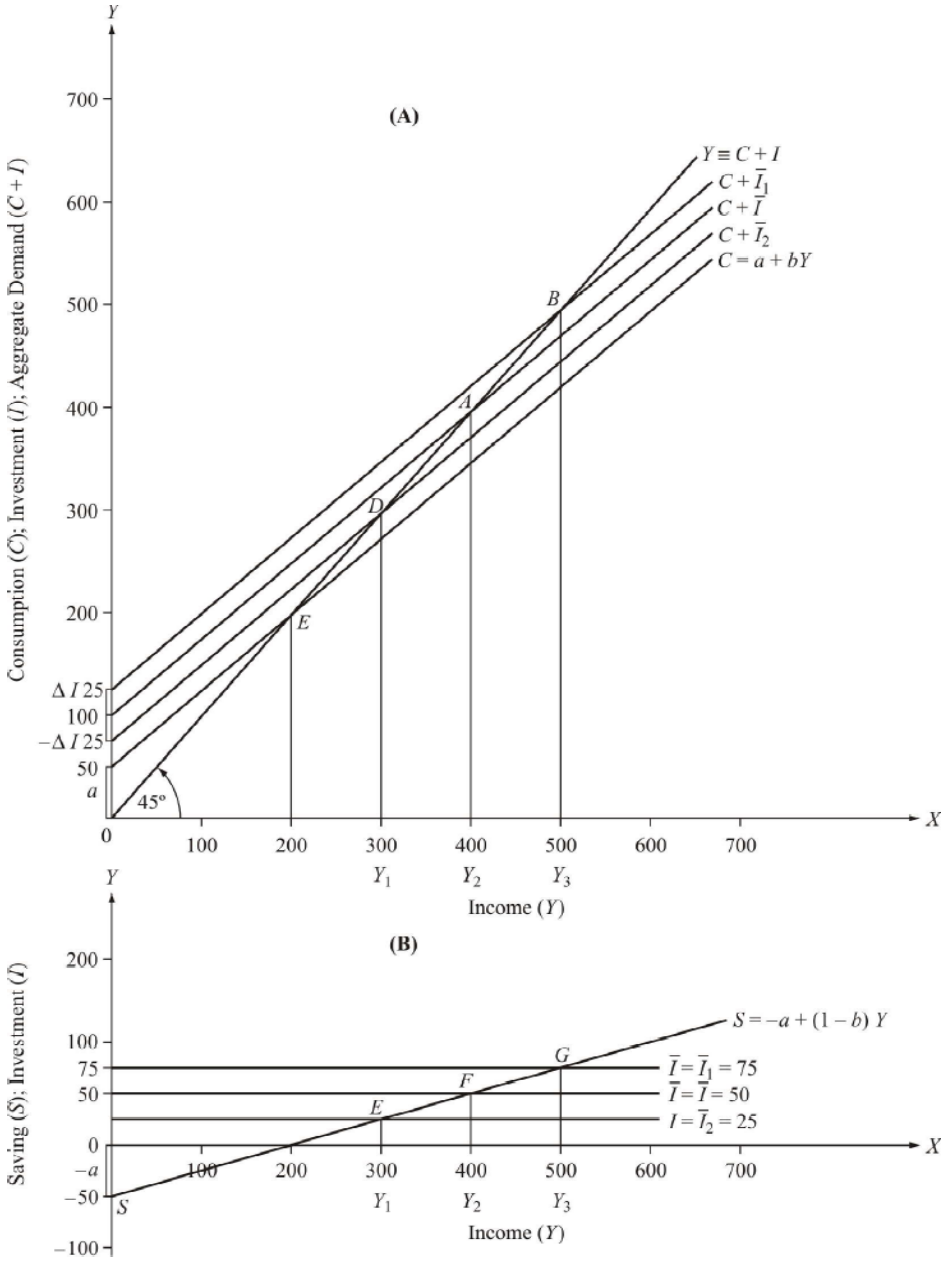


Fig. 1.12

In Figure 1.12(A), the stability of the consumption function has been shown by the fact that the entire $C = a + bY$ consumption function does not move up and down over the short period although the composite $C + I$ curve derived by aggregating the consumption and investment outlays at each level of income bounces up and down over the short period. The upward or downward shifts in the aggregate demand curve $C + I$ are wholly due to instability—shifts in the autonomous investment component of the aggregate demand function which is represented by the $C + I$ curve in Figure 1.12(A).

When the autonomous investment demand curve $I = \bar{I}$ in Figure 1.12(B) shifts upward to the position of $I = \bar{I}_1$ investment demand curve, the entire aggregate demand curve $C + \bar{I}$ also shifts upward in a parallel way to the position of $C + \bar{I}_1$ aggregate demand curve. Consequently, the equilibrium aggregate income and output increases from $0Y_2$ (= 400 units) to $0Y_3$ (= 500 units). Conversely, when the initial autonomous investment

demand curve $I = \bar{I}$ shifts downward to the position of $I = \bar{I}_2$ curve, the entire aggregate demand curve $C + \bar{I}$ also shifts downward to the position of $C + \bar{I}_2$ aggregate demand curve. Consequently, the equilibrium aggregate income and output decreases from $0Y_2$ (= 400 units) to $0Y_1$ (= 300 units). The figure shows that an initial given change of $\Delta \bar{I}$ amount in the autonomous investment spending causes a magnified (larger) change of ΔY amount in the equilibrium aggregate income and output. For example, when the autonomous investment increases by 25 units and remains stable at this higher level, the equilibrium aggregate income increases by 100 units from 400 units to 500 units. Similarly, when the autonomous investment decreases by 25 units, the equilibrium, aggregate income decreases by 100 units from 400 units to 300 units. Thus, an increase in the autonomous investment outlay of 25 units from 50 units to 75 units causes the equilibrium aggregate income to increase by 100 units from 400 units to 500 units. Conversely, a fall of 25 units in the autonomous investment outlay from 50 units to 25 units causes the equilibrium aggregate income to fall by 100 units from 400 units to 300 units.

The equilibrium aggregate income as a result of 25 units increase in the autonomous investment outlay may be derived algebraically. Assuming that $C = 50 + 0.75\ Y$ and $I = \bar{I}_1 = 75 = 50 + 25$, we have–

$$Y = C + \underline{I} \qquad ...(1.9)$$

$$Y = 50 + 0.75Y + 75 \qquad ...(1.10)$$

$$Y - 0.75Y = 125 \qquad ...(1.11)$$

$$Y = 500 \qquad ...(1.12)$$

Alternatively, we can write our derived equilibrium aggregate income equation as–

$$Y = \frac{a + \bar{I}}{1 - b} \qquad ...(1.13)$$

$$Y = \frac{50 + 75}{1 - 0.75} = 500 \qquad ...(1.14)$$

The linear investment demand curve $I = I_2$ in Figure 1.13(B) shows a decrease of 25 units in the autonomous investment. Consequent upon the downward shift in the autonomous investment demand curve, there is a magnified negative change in the aggregate income which causes the equilibrium aggregate income to fall by 100 units from 400 units to 300 units. The fact of the equilibrium aggregate income changing by more than the given change in the autonomous investment is due to the operation of the investment multiplier whose value is 4. The investment multiplier is the ratio or coefficient of the change in total income (ΔY) in response to a given change (ΔI) in the aggregate autonomous investment, i.e., $K = \Delta Y/\Delta I$. The investment multiplier may also be expressed as the marginal output-capital ratio.

Investment Multiplier and Pump-priming

Before John Maynard Keynes, the importance of increase in the new investment for causing the increase in aggregate income had been stressed by the business cycle theorists, more prominently by Knut Wicksell, Michael Tugan Baranowsky and Arthur Spiethoff. It was not, however, until the clear formulation of the multiplier theory in the 30s that the

NOTES

expansionary effect of an increase in the autonomous investment was understood and appreciated.

Prior to the great depression of the 30s, John Maynard Keynes believed in the pump-priming thesis according to which a temporary injection of new government spending would set the wheels of private enterprise in motion and once the private enterprise was back on its feet, government spending could be withdrawn without causing any relapse in the total economic activity. The pump-priming thesis assumes that a temporary new expenditure will have a lasting tendency to raise the level of economic activity in the system. Pump-priming implies that the system has been in unstable equilibrium before it is pushed back by new spending on the track from which it had been derailed by some fortuitous event.

The multiplier theory is not pump-priming. According to it, the income-generating effects of new expenditure will continue only as long as the expenditure continues with some time lag after which it will peter out. Consequently, repeated shoves and not just a single shove are required to move the economy towards prosperity. By the time he was writing *The General Theory of Employment, Interest and Money* in 1935, Keynes had abandoned the pump-priming thesis and his belief in the investment multiplier was reinforced by the experience of the New Deal economic policy undertaken by the late US President Franklin Delano Roosevelt in the early 30s. The operation of the New Deal policy suggested that the empirical value of the investment multiplier was in the range of 3 to 4.

Great Depression of the 30s and the Multiplier Theory

The origins of the multiplier theory are traceable in the Great Depression of the 30s. The British general elections of 1929 provided an occasion for a new development in Keynes' practical outlook. Unemployment was the leading issue in the election campaign in which Lloyd George was the leader of the Liberal party while Stanley Baldwin and Ramsay Macdonald were the leaders of the Conservative and Labour parties respectively. Lloyd George offered public works programme, as the remedy for mass unemployment from which England was suffering at this time. Keynes in collaboration with H D Henderson wrote a pamphlet entitled *Can Lloyd George Do It? An Examination of the Liberal Pledge* in support of the Liberal party's programme.

Keynes supported Lloyd George's promise that his proposed public works programme would involve no increase in taxation since the increased primary and secondary employment would augment the taxable national income while the decrease in expenditure on unemployment relief in sufficient amount would offset the additional outlays incurred on public works. In attempting to establish the validity of these promises, Keynes and Henderson made the quantitative estimates of the relationship between the initial outlay incurred on the public works and the final increase in the national income. This marks the genesis of Keynes' theory of the multiplier which is a refinement of the commonsense insight that in a situation of mass unused resources in the economy, an expenditure on public works will increase national income not only by the amount of direct government outlay but by some multiple of this outlay. While Richard F Kahn first gave a refined technical formulation to the multiplier theory, the basic insight and clear explanation of its practical significance was clearly expressed in the Keynes-Henderson pamphlet of 1929 whereas Kahn's article in *The Economic Journal* did not appear until 1931.

The concept of simple investment multiplier constitutes an important pillar of the whole edifice of the Keynesian theory of income and employment. The concept of investment multiplier, in the context of Keynes' theory of income and employment, refers

to the increase (decrease) in the equilibrium level of national income consequent upon a given increase (decrease) in the autonomous investment spending. The economic importance of the multiplier for economic policy purposes stems from the assumption that a given initial change (increase) in the autonomous investment spending causes a magnified change (increase) in the equilibrium aggregate income in the economy.

John Maynard Keynes has discussed the investment multiplier in his magnum opus, *The General Theory of Employment, Interest and Money*. Keynes' discussion of investment multiplier is based on Richard F Kahn's work who discussed the concept of employment multiplier in his well-known article entitled 'The Relation of Home Investment to Unemployment' published in *The Economic Journal* in June 1931. Keynes has paid rich encomiums to Kahn for this in the preface of his book *The General Theory*.

The theory of multiplier is the by-product of the Great Depression of the 30s. The multiplier theory states that the increase in total income occasioned by any given increase in autonomous investment (or consumption) outlay is a certain multiple of the original increase in autonomous expenditure and the magnitude of the increase in total income depends upon the value of the marginal propensity to consume. The investment multiplier and the marginal propensity to consume are related in such a manner that higher the marginal propensity to consume, higher is the investment multiplier and *vice versa.*

Keynes has used the concept of multiplier as a policy tool to analyse the income effect of autonomous investment in the economy. The enunciation of the multiplier principle highlighted the following two fundamental features that were not understood before Keynes' enunciation of the investment multiplier.

1. The incurring of new expenditure will have an expansionary effect on the economy with unemployed resources which will be larger in magnitude than the size of the new expenditure itself
2. The expansionary process is necessarily limited and loses force on account of the leakages from the expenditure flow stream.

Multipliers are not unique and there are hundreds of them depending upon the model we use. As the specifications of each model differ, so will the multipliers differ. In fact, it is misleading to talk about any one single unique multiplier.

Algebraic Derivation

The investment multiplier is the direct function of the marginal propensity to consume ($\Delta C/\Delta Y = b$) such that given the value of the latter, the value of the former can be easily derived by using the formula

$$K = \frac{1}{1 - b} \qquad ...(1.15)$$

where K is the simple investment multiplier and b is the familiar marginal propensity to consume (MPC). The above formula for the simple investment multiplier has been derived in the following manner–

$$Y = C + I \qquad ...(1.16)$$

$$C = a + bY \qquad ...(1.17)$$

$$I = \bar{I} \qquad ...(1.18)$$

Substituting for C and I in the equilibrium income equation (1.16) we get–

$$Y = a + bY + \bar{I} \qquad ...(1.19)$$

$$Y(1-b) = a + \bar{I} \qquad ...(1.21)$$

$$Y = \frac{a+\bar{I}}{1-b} \qquad ...(1.22)$$

NOTES

When the total autonomous investment $\bar{I}$ increases by a certain given amount $\Delta\bar{I}$, the total autonomous investment outlay becomes $\bar{I} + \Delta\bar{I}$. Consequently, the new equilibrium aggregate income $Y + \Delta Y$ will be equal to–

$$Y + \Delta Y^3 = \frac{a+\bar{I}+\Delta\bar{I}}{1-b} = \frac{a+\bar{I}}{1-b} + \frac{\Delta\bar{I}}{1+b} \qquad ...(1.23)$$

By subtracting equation (1.22) from equation (1.23) we get–

$$Y + \Delta Y - Y = \left[\frac{a+\bar{I}}{1-b} + \frac{\Delta\bar{I}}{1-b}\right] - \left[\frac{a+\bar{I}}{1-b}\right] \qquad ...(1.24)$$

$$\Delta Y = \frac{\bar{\Delta I}}{1-b} \qquad ...(1.25)$$

It, therefore, follows that the change in the equilibrium aggregate income (ΔY) equals the initial given change in the autonomous investment (ΔI) times $1/1 - b$. The term $1/1 - b$ is the simple investment multiplier. By dividing both sides of equation (15.17) by $\Delta\bar{I}$ we get–

$$\frac{\Delta Y}{\Delta\bar{I}} = \frac{1}{1-b} = K \qquad ...(1.26)$$

The term $\Delta Y/\Delta\bar{I}$ is the ratio of the total change in the aggregate income to the given change in the autonomous investment. This ratio is the definition of the investment multiplier. The multiplier is the inverse of one minus the marginal propensity to consume (MPC). For example, if the MPC (b) is 0.8, the value of the simple investment multiplier (K) will be $1/1 - 0.8 = 5$.

The investment multiplier is also related to the marginal propensity to save (MPS = $\Delta S/\Delta Y$). Since the sum of the marginal propensity to consume and the marginal propensity to save is equal to one, the marginal propensity to save is equal to one minus the marginal propensity to consume, i.e., the MPS is $1 - b$ where b is the marginal propensity to consume. Thus, we can substitute MPS for $1 - b$ in equation (1.26). Accordingly, the simple investment multiplier $K = 1/\text{MPS}$. By writing s for the MPS we can write $K = 1/s$. From this it follows that the investment multiplier and the marginal propensity to save are inversely related such that a high marginal propensity to save (MPS) denotes a low investment multiplier (K) and *vice versa*.The investment multiplier analysis tells us that the increase in the equilibrium aggregate income (ΔY) equals the given increase in the autonomous investment times the multiplier,i.e., $\Delta Y = \Delta I . K$. The only assumption made here is that the marginal propensity to consume (and, therefore, also the marginal propensity to save) is positive, constant and less than one, i.e., $0 < b < 1$. This assumption is necessary for the stability of the system. Since b is a positive fraction, it follows from

either zero or one, the multiplier will be either one or infinite. Since in our assumption *b* is finite, being more than zero and less than one, it follows that the value of the investment multiplier is finite, being more than one. Consequently, the change in the equilibrium aggregate income is greater than the given change in the autonomous investment outlay of which the change in the equilibrium aggregate income is the consequence.

Multiplier Action

But by what process does the multiplier increase in the aggregate income, which is due to a given increase in the autonomous investment outlay, take place? In other words, does the increase in the aggregate income take place simultaneously with the increase in the autonomous investment or does it take place after a certain time interval? While discussing the process of income propagation, we might consider either the income increasing simultaneously with the increase in the investment outlay so that there is no problem of time-lag between the increase in autonomous investment and the resulting increase in income or the increase in income may be regarded as lagging behind the increase in autonomous investment. The first mode of income expansion is the simultaneous or static multiplier analysis allowing for no time-lag between the increase in investment and the resulting increase in income while the second is the period or dynamic or sequence multiplier analysis. Both these multiplier analyses may be discussed with reference to either the discontinuous or single-dose of increase in the autonomous investment or the repeated constant doses of increase in the autonomous investment, i.e., the initial given increase in autonomous investment continues in each subsequent time period. In other words, we might consider the question: what happens to the equilibrium income if (i) the investment which increases in any given time period is cut-back to its original level in the subsequent next time period; and (ii) if the given increase in the autonomous investment continues constantly in each succeeding time period?

Simultaneous or Logical Multiplier

The simultaneous or logical multiplier analysis assumes that the change in all the three variables—autonomous investment, induced consumption and income—which constitute the core of the multiplier analysis takes place simultaneously so that there is no time-lag between the change in autonomous investment and the resulting induced changes in the consumption spending and equilibrium income. Thus, the logical theory of the multiplier rests on the assumption that the initial given change in the aggregate autonomous investment outlay and the resulting total change in the equilibrium aggregate income occur simultaneously.

Keynes' discussion of multiplier runs mainly in terms of the 'logical theory of the multiplier which holds good continuously, without time-lag, at all moments of time'. Throughout the book *The General Theory,* except section IV where a mention has been made of the time-lag or period analysis, the multiplier has been discussed 'on the basis of a change in aggregate investment which has been foreseen sufficiently in advance for the consumption industries to advance *part passu* with the capital-goods industries...'

Martin J Bailey has elaborated the simultaneous multiplier theory through this imaginary example: '...suppose some added investment activity is announced on a Saturday evening, after the end of the work week, to commence on Monday morning. Unemployed workers are notified to return to work on Monday morning on the new investment projects. Expecting their new paychecks, they plan higher purchases than before, beginning immediately: they call their grocers and other retailers to assure credit, based on their

NOTES

expected paychecks. These retailers shrewdly place new orders over the weekend with their wholesalers; the wholesalers similarly place increased orders with manufacturers, and the manufacturers increase their planned production runs while notifying previously laid-off workers to report to work Monday morning. These workers in turn plan increased spending, get in touch with retailers, who call their wholesalers, who call manufacturers, who rehire workers and so on *ad infinitum*. The whole process can be completed by Monday morning, so that at one stroke, income rises by the full amount of the multiplier times new investment. Indeed all this could happen without any action by anyone to get in touch with his suppliers. Each retailer, wholesaler, and manufacturer could anticipate, without being told, what the effect on him would be, and the workers could likewise show up at their plants without being notified, in the correct expectation that work would be available.'

In a two-sector economy, the aggregate demand is composed of the aggregate consumption and investment outlays, i.e., $Y = C + \bar{I}$. Starting from the position of equilibrium in which the aggregate demand is equal to the aggregate output or income, i.e., $Y = C + I$, the question is: what happens to the equilibrium aggregate income and output if the aggregate autonomous investment is increased by some given amount $\Delta\bar{I}$? With the given increase in the planned investment, the initial equilibrium between the planned saving and planned investment at the old equilibrium income is disturbed since at this income level investment exceeds saving, i.e., $I > S$. Saving has, therefore, to increase to become equal to investment which is now higher than before. Being, however, a function of income, saving cannot increase unless income increases sufficiently to allow for the required increase in saving to be equal to the given increase of ΔI amount in autonomous investment. In other words, the aggregate income must increase sufficiently so that the increase in total saving of ΔS amount out of this higher income is equal to the initial increase in the autonomous investment, i.e., $\Delta S = \Delta\bar{I}$. The increase in the aggregate income (ΔY) which will generate that increase in savings (ΔS) which is equal to the initial increase in autonomous investment ($\Delta\bar{I}$) will be equal to the inverse of the marginal propensity to save (MPS) times the given increase in the autonomous investment, i.e., $\Delta Y = I/\text{MPS} \times \Delta\bar{I}$. Thus, if the marginal propensity to save (it is equal to one minus the marginal propensity to consume) is known, the total change of ΔY amount in the equilibrium aggregate income resulting from a given change of $\Delta\bar{I}$ amount in the autonomous spending (consumption or investment) can be known.

Diagrammatical Illustration

The logical or simultaneous multiplier has been algebraically derived. It is now explained graphically in Figure 1.13 where the original equilibrium has been shown at point A corresponding to which the aggregate income is $0Y_0$ ($= AY_0$). It is assumed that the increase in the aggregate demand ($C + I$) comes about through a given increase of $\Delta\bar{I}$ amount in the autonomous investment outlay and that this increase in the autonomous investment outlay continues in each succeeding time period. In other words, the assumption is that once the autonomous investment is increased from its original level of $\bar{I}$ to the new higher level of $\bar{I} + \Delta\bar{I}$, it stays there. Consequently, upon a continuous given increase of ΔI amount in the autonomous investment, the aggregate demand function $C + \bar{I}$ shifts upward in a parallel manner to the position of the new aggregate demand function $C + \bar{I} + \Delta\bar{I}$. The vertical distance between the old aggregate demand function

$C + \bar{I}$ and the new aggregate demand function $C + \bar{I} + \Delta\bar{I}$ shows the given increase of $\Delta\bar{I}$ amount in the autonomous investment. The new aggregate demand curve intersects the 45° or the equality-line at point *B* where the aggregate supply (output) equals the aggregate demand, i.e., $BY_1 = 0Y_1$. Consequently, the aggregate income corresponding to the point of intersection *B* is the equilibrium aggregate income since at this aggregate income, the equilibrium condition $Y = C + I$ between the aggregate supply and the aggregate demand is satisfied. The total increase in income (ΔY) can be obtained by measuring the difference between the new higher equilibrium income ($0Y_1$) and the old equilibrium income ($0Y_0$). This is equal to Y_0Y_1, i.e., $\Delta Y = Y_0Y_1$. The increase in income being greater than the given increase of $\Delta\bar{I}$ amount in the autonomous investment equals the increase of $\Delta\bar{I}$ in the autonomous investment spending times the investment multiplier, i.e., $\Delta Y = \Delta\bar{I}\ .K$. Thus, the equilibrium aggregate income rises permanently to a higher level of $0Y_1$ (= BY_1) from its old lower level of $0Y_0$ (= AY_0) when the aggregate autonomous investment increases permanently by $\Delta\bar{I}$ amount from $\bar{I}$ to $\bar{I} + \Delta\bar{I}$.

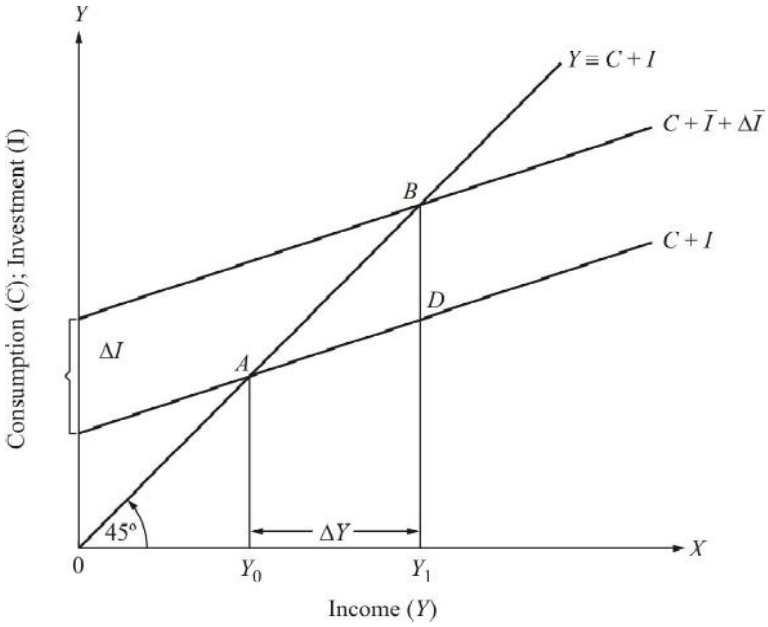


Fig. 1.13

Criticism

The logical or instantaneous multiplier theory has been criticized on the ground that it assumes instantaneous responses in the consumption spending to the given change in autonomous investment outlay. In practice, however, changes do take time to work themselves out in the system. Consequent upon a change in their income, consumers take some time to adjust their consumption spending to their new income. Even if it is granted that adjustment in consumption spending takes place simultaneously to a given change in investment so that there is no time-lag between the increased income receipts and the decision of the income receivers to spend more on consumption, it would not have the desired effect unless it is also assumed, unrealistically though, that the consumer goods industries, correctly foreseeing the possibility of increased consumer spending, expanded their production simultaneously. However, in a dynamic world with uncertain future, it is difficult to make any perfectly accurate forecasting.

NOTES

Second, the logical multiplier theory has been criticized on the ground that it does not explain the path which is followed by income as it moves awayfrom the original equilibrium to the new equilibrium position. Consequently, the logical multiplier analysis is static. The analysis only tells us either with the help of a diagram or through the algebraic equations, the new equilibrium level of income after a given change in investment spending has taken place. It does not say anything about the sequence of interactions of different peoples' decisions and behaviour that take place in the economy during the process of change from one equilibrium position to the other. Thus, the relationship between the new investment, consumption and income expressed by the logical multiplier which abstracts the transition period is at best only an approximation which holds true over a long period during which complete adjustment to the new equilibrium has taken place.

Dynamic or Period or Sequence Multiplier

Since the aggregate income does not change simultaneously with a given change in the autonomous investment spending but changes after a certain time-lag, it is important to take this fact into account. Consequently, the multiplier increase in the aggregate income comes through a series of lagged induced increases in the aggregate consumption spending. According to the static equilibrium analysis, as a consequence of any given initial change in the autonomous investment, the aggregate income moves on to the new equilibrium at which income and expenditure are equal, provided the two variables are related in a manner which affords stability. While the static analysis ignores the time needed for the adjustment of income to a given increase in expenditure, the fact is that a certain time interval occurs between the given increase in the autonomous investment outlay and the resulting increase in the aggregate income in consequence of the given increase in the autonomous investment. It is, therefore, appropriate to consider the effect of investment changes on income through time. In order to do this, it is necessary to focus attention explicitly on the timing of events. As we did in the preceding static analysis, we assume that investment is a part of total expenditure and that it changes quite independently of the changes in income. As a consequence, the aggregate income also ultimately increases by the amount of the given increase in the autonomous investment outlay times the investment multiplier.

The basic purpose of the present discussion is to show or trace the path of income propagation through time following the injection of (i) a single unrepeated dose of autonomous investment and (ii) a fixed amount constant dose of autonomous investment which is repeated in each succeeding time period.

When the aggregate investment demand increases consequent upon the given increase in the autonomous investment outlay, the entrepreneurs will want to increase their output to meet the increased demand. The output cannot, however, increase instantaneously and some time interval must elapse before the higher output and income can be achieved. This is necessarily so because the various steps, such as arranging for the raw material and hiring of labour which involves time, have to be taken up before production can increase. In short, while the appearance of additional demand created by a given increase in the autonomous investment outlay causes the aggregate income to rise, the additional output and corresponding income can, however, be generated only after the lapse of a certain time interval. Additional demand in one time period generates corresponding income in the succeeding time period, i.e., an excess of aggregate demand over the aggregate supply in one time period generates a corresponding increase in the

aggregate income in the immediately succeeding time period. We may now formulate this hypothesis in the form of symbols.

The total spending in the economy, which is the sum of the total consumption and investment outlays, can be designated as *E*. The aggregate supply and aggregate demand equilibrium condition can be symbolically stated as:

$$Y = C + I$$

$$Y = E \text{ (since } C + I = E)$$

In the dynamic or sequence multiplier analysis of income propagation, variables *Y* and *E* must be dated. This is done by the use of subscript *t*, taking the integral values 1, 2, 3, 4,... which denote the corresponding time periods. Now the above hypothesis can be stated as–

$$E_t - E_{t-1} = \Delta E_t = Y_{t+1} - Y_t = \Delta Y_{t+1}$$

This implies that–

$$\Delta E_t = \Delta Y_{t+1}$$

With this interpretation of time-lag of the change in income lagging one time period behind the change in expenditure (demand), we can analyse the simple dynamics of income propagation following a single one time period given increase of ΔI amount in the autonomous investment outlay.

Starting from the position of equilibrium, let us suppose that the autonomous investment increases in one single time period by ΔI amount and thereafter, it falls back to its original level. Let us also assume that the marginal propensity to consume (MPC) is the constant positive fraction *b* such that $1 > b > 0$ = constant. It means that people spend *b* part of their additional income (ΔY) on consumption, i.e., $\Delta C = b\Delta Y$. The production of additional consumer goods to meet this additional consumer demand will itself generate additional income equal to the additional consumption. It is, therefore, possible to view the entire process of income propagation in the economy as the sum of additional income generated in each time period starting from the initial given increase of ΔI in the autonomous investment outlay in time period 0 and ending in time period *t* with *t* tending to approach infinity. Assuming one time period lag to exist between the expenditure and income, the entire process of income propagation may be explained in the following manner.

When the autonomous investment increases by ΔI in time period 0, the aggregate income in time period 1 increases by the full amount of additional autonomous investment outlay of ΔI amount which was incurred in the immediately preceding time period 0, i.e., $\Delta Y_1 = \Delta I$. Consequent upon the increase in aggregate income in time period 1, there will be a corresponding increase in the aggregate consumption spending in time period *I* (ΔC_1) equal to the product of the MPC (= *b*) and the additional income ΔY_1, i.e.,

$$\Delta C_1 = b\Delta Y_1 = b\Delta I \text{ (since } \Delta Y_1 = \Delta I)$$

Since the additional expenditure incurred in any given time period causes a corresponding increase in the income of the following time period, additional expenditure incurred on consumption in time period 1 causes an equivalent increase in income in the subsequent time period 2. In other words, $\Delta Y_2 = \Delta C_1$. But $\Delta C_1 = b\Delta I$ and, therefore, $\Delta Y_2 = b\Delta I$. The additional income in time period 2 will induce additional consumption spending equal to *b* time of the increase in aggregate income in time period 2, i.e.,

$$\Delta C_2 = b\Delta Y_2 = b\Delta C_1 = b\,(b\Delta Y_1) = b\,(b\Delta I) = b^2\Delta I$$

This additional consumption expenditure will cause the corresponding increase in the aggregate income in time period 3, i.e.,

$$\Delta Y_3 = \Delta C_2 = b^2 \Delta I$$

Similarly, $$\Delta Y_4 = \Delta C_3 = b^3 \Delta I$$

and so on for each succeeding time period. Finally, at the end in time period t, the additional income generated will be equal to the additional consumption expenditure incurred in the immediately preceding time period $t-1$ which itself will be equal to $b^{t-1}\Delta I$. Thus,

$$\Delta Y_t = \Delta C_{t-1} = b^{t-1} \Delta I$$

The total cumulative increase in the aggregate income (ΔY) resulting from the given initial increase of ΔI amount in the autonomous investment will be equal to the sum of the whole series of additional income generated over infinite time periods beginning from time period 1 and ending with time period t (limit $t \to \infty$). The following table shows the process of multiplier income propagation spread over an infinite series of time periods for a single dose increase in the aggregate autonomous investment.

Table 1.2

Excess of the Variable over the Base Period Value	***Time Period***						
	0	1	2	3	4	...$t-1$	$t\to\infty$
E	ΔI	$b\Delta I$	$b^2\Delta I$	$b^3\Delta I$	$b^4\Delta I$	...$b^{t-1}\Delta I$	$b^t\Delta I$
Y	0	ΔI	$b\Delta I$	$b^2\Delta I$	$b^4\Delta I$	...$b^{t-2}\Delta I$	$b^{t-1}\Delta I$
C	0	$b\Delta I$	$b^2\Delta I$	$b^3\Delta I$	$b^4\Delta I$	...$b^{t-1}\Delta I$	$b^t\Delta I$
I	ΔI	0	0	0	0	0	0

It follows from the above table that–

$$\Delta Y = \Delta Y_1 + \Delta Y_2 + \Delta Y_3 + \ldots\ldots\ldots + \Delta Y_t \quad \ldots(1.27)$$

$$\Delta Y = \Delta I + b\Delta I + b^2\Delta I + \ldots\ldots\ldots + b^{t-1}\Delta I \quad \ldots(1.28)$$

$$\Delta Y = \Delta I\,(1 + b + b^2 + b^3 + \ldots\ldots\ldots + b^{t-1}) \quad \ldots(1.29)$$

The infinitely long sum in the parenthesis is a converging geometric series and is equal to $\frac{1}{1-b}$

$$\therefore \quad \Delta Y = \frac{1}{1-b}\Delta I = K\Delta I \text{ (since } \frac{1}{1-b} \text{ is the investment multiplier } K) \quad \ldots(1.30)$$

Dividing both sides of equation (1.30) by ΔI, we get

$$\frac{\Delta Y}{\Delta I} = \frac{1}{1-b} = K \quad \ldots(1.31)$$

We obtain the same result through the sequence or dynamic multiplier analysis as was obtained in the static multiplier analysis in equation (1.26) namely, that the total increase in the aggregate income ΔY is equal to the investment multiplier $\frac{1}{1-b}$ times the given initial increase of ΔI amount in the autonomous investment.

Superiority of the Dynamic or Period Multiplier Analysis

The dynamic or period multiplier analysis of the effect on income of a single dose given increase in autonomous investment, however, shows that in each succeeding time period, the increase in income becomes smaller and eventually peters out. From the standpoint of economic policy, this means that unless the increase in autonomous investment is continued period after period, equilibrium income will fall to its old level after some time. In terms of the dynamic analysis contained in the above table and equations, in time period t (when t approaches infinity) the increase in aggregate income $\Delta Y_t (= b^{t-1}\Delta I)$ approaches zero. Consequently, the equilibrium income in time period t (Y_t) approaches the original equilibrium income Y_0 in time period zero before the increase in autonomous investment outlay had occurred. Adynamic multiplier may be defined as the ratio of the total cumulative change in the aggregate income over and above its initial level and the associated initial change in the aggregate autonomous expenditure when the change in expenditure is a one-shot (or one single period) event. Assuming that the once-over increase in the autonomous aggregate investment outlay ΔI = `100 crore and the marginal propensity to consume b is 0.8, the dynamic multiplier process of the total income propagation can be summarized below:

$$\Delta Y = \Delta Y_1 + \Delta Y_2 + \Delta Y_3 + \Delta Y_4 + \Delta Y_5 + \ldots \ldots \ldots + \Delta Y_t \quad \ldots(1.32)$$

$$\Delta Y = \Delta I + b\Delta I + b^2\Delta I + b^3\Delta I + b^4\Delta I + \ldots \ldots \ldots + b^{t-1}\Delta I \quad \ldots(1.33)$$

$$\Delta Y = 100 + 0.8(100) + 0.8^2(100) + 0.8^3(100) + \ldots \ldots + 0.8^{t-1}(100) \quad \ldots(1.34)$$

$$\Delta Y = \text{`}500 \text{ crore}$$

The path of income propagation through time following the given increase of a single unrepeated dose of ΔI amount in the autonomous investment can be shown with the help of Figures 1.14 and 1.15.

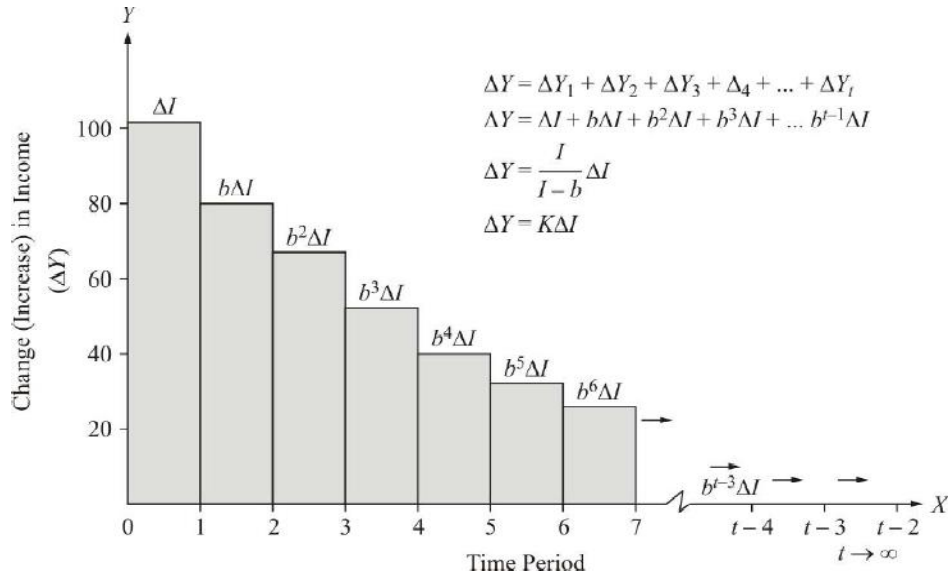


Fig. 1.14

Figure 1.14 shows that consequent upon the one-shot given increase of ΔI amount in the autonomous investment in time period 0, which is discontinued in the succeeding time periods, aggregate income in the immediately following time period 1 increases by the corresponding amount, i.e., by ΔI. In other words, the total increase in income in time period one $\Delta Y_1 = \Delta I$. The additional income generated in each succeeding time period

goes on diminishing under the petering multiplying impact of the marginal propensity to consume which is positive but less than one. Eventually, in time period *t* (when *t* approaches infinity) the increase in income comes to a grinding halt approaching zero. In other words, the figure shows that the increase in aggregate income (shaded rectangle) in each succeeding time period diminishes and eventually disappears. The total accumulated increase in income from time period 0 to time period *t* (ΔY) which results from the initial given increase of ΔI amount in the autonomous investment is the entire shaded area extending upto time period *t* representing the sum of the rectangles upto time period *t* and is equal to the multiplier times the initial increase of ΔI amount in the autonomous investment. In other words, the total increase in income $\Delta Y = \Delta I.K$.

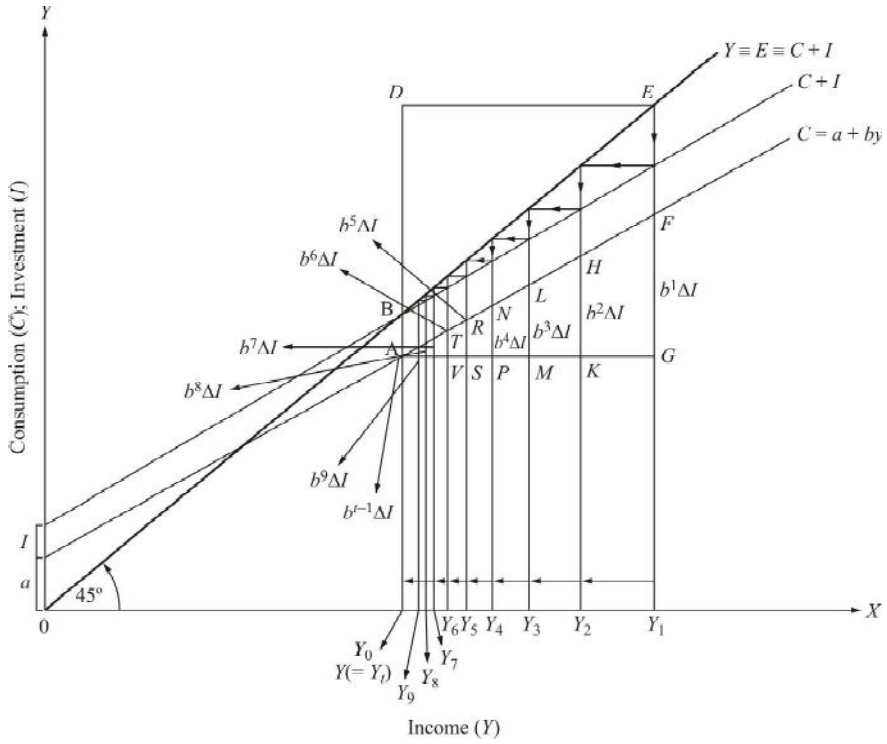


Fig. 1.15

The entire process of cumulative income propagation through the multiplier action can also be displayed by means of Figure 1.15. Consequent upon the one time given increase of ΔI (=*BD*) amount in the autonomous investment in time period 0, the aggregate income in the immediately succeeding time period 1 increases by full amount of the increase of ΔI (=*BD*) amount in the autonomous investment becoming EY_1 from BY_0.

Corresponding to the increase in the aggregate income of ΔI amount in time period 1, people will spend *b* (marginal propensity to consume) time of this additional income amount *FG* ($= b\Delta I$). Consequently, income in the succeeding time period 2 will be higher than the initial pre-increase in the autonomous investment income BY_0 by this amount, i.e., by *FG* amount. Again, *b* time of this additional income $b^2\Delta I$ (= *HK*) will be spent on consumption in time period 2. This will become the additional income available for consumption spending in time period 3. Again, *b* time of this smaller additional income $b^3\Delta I$ (= *LM*) will be spent on consumption. Consequently, the aggregate income in the succeeding time period 4 will increase by this amount.

The entire process of gradually petering out income propagation in succeeding time periods continues *ad infinitum*. In this manner, in each succeeding time period the propagation of additional income goes on diminishing and eventually in time period *t*, when *t* approaches infinity ($t \to \infty$), the increase in income approaches zero with the income in time period *t* (Y_t) becoming equal to the initial pre-increase in the autonomous investment income BY_0 in time period 0. The total cumulative increase in the aggregate income starting from time period 1 and ending in time period *t* (ΔY) consequent upon the given increase of ΔI amount in the autonomous investment in time period 0 will be:

$$\Delta Y = \Delta Y_1 + \Delta Y_2 + \Delta Y_3 + \Delta Y_4 + \Delta Y_5 + \Delta Y_6 + \Delta Y_7 + \ldots\ldots\ldots\ldots + \Delta Y_t$$

$$\Delta Y = \Delta I + b\Delta I + b^2\Delta I + b^3\Delta I + b^4\Delta I + b^5\Delta I + b^6\Delta I + \ldots\ldots\ldots\ldots + b^{t-1}\Delta I$$

$$\Delta Y = BD + FG + HK + LM + NP + RS + TV + \ldots\ldots\ldots\ldots + \ldots\ldots\ldots 0$$

In short, the total cumulative increase in the aggregate income (ΔY) beginning from time period 1 to time period *t* will be equal to the investment multiplier (*K*) times the initial increase of ΔI amount in the autonomous investment outlay, i.e., $\Delta Y = \Delta I.K$.

Criticism

The multiplier analysis suffers from several weaknesses. *First*, the multiplier analysis derives the multiplier on the assumption of constant marginal propensity to consume. Over the short period of a trade cycle, the marginal propensity to consume is not constant. In such a situation, we cannot think of any constant value of the multiplier. The multiplier $K = 1/1 - b$ will vary in value from one equilibrium income to another equilibrium income and we will have the marginal multiplier whose value changes from one equilibrium income to another equilibrium income instead of the simple investment multiplier with constant value.

Second, the simple multiplier analysis is faulty because it neglects the role of induced investment resulting from induced consumption in the determination of the equilibrium income. Cumulative changes in national income also result from factors other than the successive rounds of induced consumption spending. Changes in investment are also linked to changes in the national income. Consequently, the cumulative change in income will usually be greater than what is predicted by the simple investment multiplier.

Criticizing the concept of investment multiplier, Professor A G Hart has called it a useless 'fifth wheel' which adds nothing to the ideas already implied in the use of the consumption function. In his well-known article published in 1936, Haberler criticized Keynes' multiplier as a mere tautology. According to Haberler, Keynes has defined his multiplier in such a way that it is simply another name for the marginal propensity to consume. 'The multiplier is defined in terms of marginal propensity to consume. Instead of the multiplier we can always say $1/1 - \Delta C/\Delta Y$ and for marginal propensity to consume we can always substitute $1 - 1/K$. One and the same thing has got two names. Mr Keynes has fallen into the trap of treating such a relationship by definition as a causal or empirical relationship between investment and income and that, thereby, a large part of what he says about the multiplier and its probable magnitude is initiated.'

Third, the multiplier analysis has ordinarily been developed, under the assumption that labour and other fixed resources are idle or underutilized in the economy, i.e., surplus production capacity exists in the economy. In other words, the assumption made is that there is idle labour seeking employment and there is present idle plant capacity in the economy making it possible to expand output readily in response to a rise in demand.

NOTES

This assumption is, however, unrealistic because in any substantial expansion of output, or in any expansion which is different from the previous expansions, shortages of particular kinds of labour or of fixed resources are likely to occur at various points in the economy. Moreover, if the expansion takes place after a plateau of low output for a period long enough for the idle fixed assets to have become worn-out and not replaced, bottlenecks of fixed assets capacity may also be frequently encountered, thwarting the working of the multiplier action.

Fourth, the bottlenecks of particular kinds of labour or at particular places may block the expansion of employment and output in the economy. A theory of income flexibility in a broad sense cannot, therefore, be legitimately constructed without making a careful analysis of these bottlenecks and how these can be removed. Even in a less-than-full employment economy, the multiplier theory, particularly any dynamic discussion of the multiplier time periods, is seriously incomplete and misleading if it ignores the bottleneck problems. 'Even though the multiplier theory may be correctly regarded as a bold and challenging piece of analysis, it is true only as a first approximation under ideal conditions. Not only does income change only by a process that takes time, but to determine the speed and extent of that process involves an analysis of business practices, attitudes, responses; of technical conditions of production and supply; of consumer behaviour, income distribution, layoff and hiring procedures; of indirect impacts on government budgets through effects on tax collections, transfer payments, social security contributions; of indirect impacts on the money market, indeed of every aspect of the economic process. To understand them we need a vast fund of institutional knowledge about the business system, and, perhaps even more important, an understanding of the psychological frames of reference of business firms and individuals, which determine the way and the speed with which they respond to changes in objective facts.'

Notwithstanding its several weaknesses, the multiplier analysis, however, states the important truth that an initial increase in the autonomous investment, unless offset by a corresponding fall in the consumption expenditure, has the secondary, tertiary, quarternary and further gradually petering out effects which leads to a large increase in the total income depending on the value of the investment multiplier. This is a very important conclusion for economic policy decisions. It tells us that in depression, even if the increase in government spending is incurred on utterly unproductive and useless purposes, like leaf-raking, digging holes in the ground and filling them up again, it is worthwhile to spend because the induced increases in aggregate consumption spending and income will expand income and employment in the economy. The multiplier analysis shows that the social cost of eliminating unemployment for the community is much less than what the orthodox economic theory makes us believe.

1.7.1 Multiplier in an Open Economy

A full 'open' economy has all sectors, and therefore, three withdrawals – savings, taxation and imports.

This is indicated by the marginal propensity to save (mps) plus the extra income going to the government - the marginal tax rate (mtr) plus the amount going abroad – the marginal propensity to import (mpm).

By adding up all the withdrawals we get the marginal propensity to withdraw (mpw).

1.7.2 Balanced Budget Multiplier

The question is: what happens to the equilibrium income if the government budget is balanced, i.e., if the given increase in government expenditure is financed by an equal increase in taxes? In other words, will there be any impact on the economy if the increase in government purchases and taxes is balanced? According to the classical view, a balanced budget was neutral in its effect on the national income and employment since the increase in government spending was offset by the tax increase of an equal amount.

According to the modern view, however, a balanced budget is not neutral in its effect on income. It exerts a net expansionary effect on the national income because the decrease in the aggregate spending resulting from additional taxes will be less than the increase in the aggregate spending resulting from additional government expenditure. As a result of the additional tax levy, the aggregate consumption will not be reduced by the full amount of the tax because a part of the tax is paid out of saving. Substituting the value of the MPC into the government purchases and government tax multipliers, it can be seen that the difference between these two multipliers is one. With an MPC of 0.8, the government purchases multiplier is 1/1 – 0.8 while the government tax multiplier is 0.8/1 – 0.8. Since the tax multiplier is negative, the difference between the value of these two terms is one. This means that in the event of the government budget being balanced, the aggregate income will increase by full amount of the increase in the government expenditure. The balanced budget multiplier may be derived in the following manner.

$$Y = C + I + G \qquad ...(1.35)$$

Substituting in equation (1.35) for C we get–

$$Y = a + bY_d + \bar{I}_A + G \qquad ...(1.36)$$

Since Y_d equals $Y - T + R$, equation (1.36) can be rewritten as–

$$Y = a + b\,(Y - T + R) + \bar{I}_A + \bar{G} \qquad ...(1.37)$$

$$Y = a + bY - bT + bR + \bar{I}_A + \bar{G} \qquad ...(1.38)$$

$$Y - bY = a - bT + bR + \bar{I}_A + \bar{G} \qquad ...(1.39)$$

$$Y\,(1 - b) = a - bT + bR + \bar{I}_A + \bar{G} \qquad ...(1.40)$$

Dividing through by $1 - b$ we get–

$$Y = \frac{a - bT + bR + \bar{I}_A + \bar{G}}{1 - b} \qquad ...(1.41)$$

Consequent upon a given increase of ΔG amount in the government expenditure and an equal increase of ΔT amount in government taxes ($\Delta T = \Delta \bar{G}$ because the budget is balanced) the new aggregate equilibrium income will be–

$$Y + \Delta Y = a + b\,(Y + \Delta Y - T - \Delta T + R) + \bar{I}_A + \bar{G} + \Delta \bar{G} \qquad ...(1.42)$$

$$Y + \Delta Y = a + b\,Y + b\Delta Y - bT - b\Delta T + bR + \bar{I}_A + \bar{G} + \Delta \bar{G} \qquad ...(1.43)$$

$$Y - bY + \Delta Y - b\Delta Y = (a - bT - b\Delta T + bR) + \bar{I}_A + \bar{G} + \Delta \bar{G} \qquad ...(1.44)$$

$$Y\,(1 - b) + \Delta Y(1 - b) = a - bT - b\Delta T + bR + \bar{I}_A + \bar{G} + \Delta \bar{G} \qquad ...(1.45)$$

NOTES

Dividing both sides of equation (1.45) by $1 - b$ we get–

$$Y + \Delta Y = \frac{a - bT - b\Delta T + bR + \bar{I}_A + G + \Delta\bar{G}}{1 - b} \quad ...(1.46)$$

Subtracting equation (1.41) from equation (1.46) gives us–

$$\Delta Y = \frac{\Delta\bar{G} - b\Delta T}{1 - b} \quad ...(1.47)$$

$$\Delta Y = \frac{\Delta\bar{G} - b\Delta\bar{G}}{1 - b} \text{ (since } \Delta T = \Delta\bar{G}) \quad ...(1.48)$$

$$\Delta Y = \frac{\Delta\bar{G}(1 - b)}{1 - b} \quad ...(1.49)$$

$$\Delta Y = \Delta\bar{G} \quad ...(1.50)$$

Dividing both sides of the equation by $\Delta\bar{G}$ gives us the balanced budget multiplier K_b–

$$\frac{\Delta Y}{\Delta\bar{G}} = \frac{\Delta\bar{G}}{\Delta\bar{G}} = 1 \quad ...(1.51)$$

This states that when the government budget is balanced such that any given increase in the government expenditure incurred on the purchase of goods and services is financed through additional tax levies of an equal amount, the equilibrium aggregate income in the economy will increase by full amount of the additional government expenditure.

The balanced budget multiplier with the numerical value of one is, however, valid under the following assumptions:

1. The increase in government spending relates entirely to the purchases of goods and services by the government and not to government transfer payments. The balanced budget multiplier for government transfer expenditure will be zero because the positive transfer payments multiplier and the negative tax multiplier being equal, will offset one another thus having a neutral impact on the economy. This conclusion is, of course, valid on the assumption that the marginal propensities to consume of those who receive the benefit of additional government transfer payments and of those who bear the burden of additional tax levied to finance the additional transfer payments are equal.
2. It is assumed that the marginal propensity to consume of the tax payers equals the marginal propensity to consume of the sellers of goods and services to the government.
3. It is assumed that there is no impact of both the government purchases and government taxes on investment activity in the economy.
4. It is also assumed that taxes are autonomously determined and are not related to the level of income. In other words, the discussion considers only the lump-sum

If the increase of a given amount of $\Delta\bar{G}$ in government spending in the budget was not balanced by additional tax levies of ΔT (= $\Delta\bar{G}$), the increase in income would have been Y_1Y_3. However, since the additional spending is financed through additional taxes of an equal amount, the negative tax multiplier operates simultaneously partly offsetting the increase in income. The net increase in income will be equal to the difference between the total increase in income due to the increase of ΔG amount in government spending and the total decrease in income due to the levying of additional taxes (ΔT) of the same amount. The negative change in the aggregate income due to the additional tax levy of ΔT (= $\Delta\bar{G}$) amount is Y_2Y_3. Consequently, the net increase in the equilibrium aggregate income is Y_1Y_2 which equals the given increase in government spending of $\Delta\bar{G}$ amount.

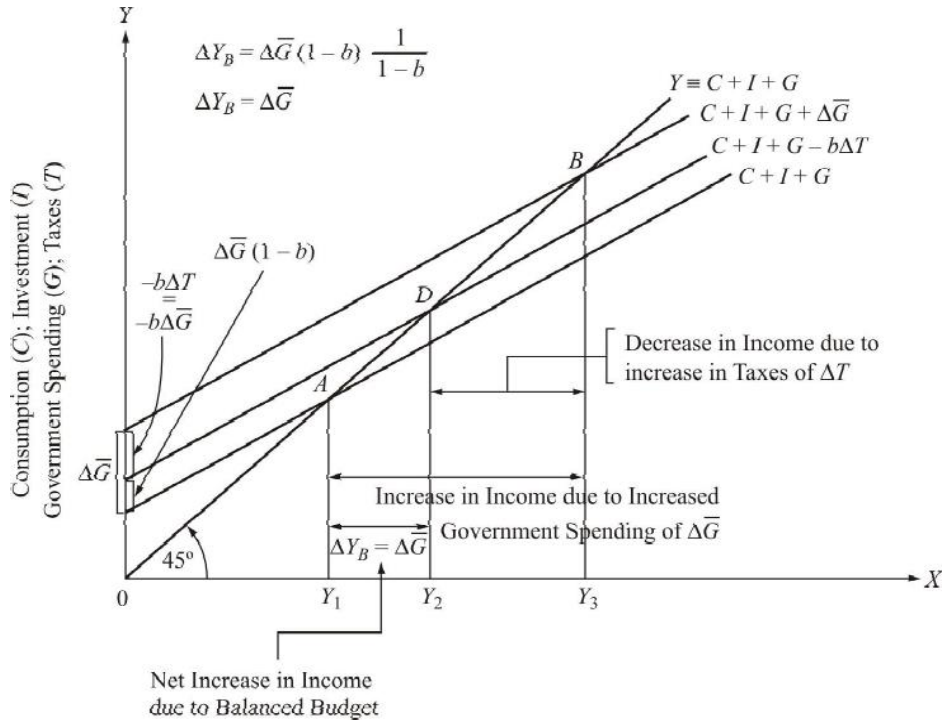


Fig. 1.16

Induced Taxes and Balanced Budget Multiplier

So far the balanced budget multiplier has been discussed under the assumption that taxes are wholly autonomously determined—these are a constant amount regardless of the level of income. In reality, however, not all government tax receipts are autonomous and the government's total tax income is also influenced by the level of people's income. If the tax payers' income increases, government's tax revenue goes up and *vice versa.* The problem is further complicated because not only does the total tax collection increase when tax payers' income increases but the tax rate also increases as income increases under the impact of the progressive income tax system. In a situation where the total government tax revenue comprises partly the autonomous tax and partly the induced taxes, the total tax function may be written as–

$$Y = d + tY; \text{ and } 0 < t < 1 \qquad ...(1.52)$$

where d is the constant autonomous component of total tax revenue and t is the proportion of tax related to the income level so that tY is the induced component of the total tax revenue. It is the marginal propensity to tax. Let us also assume that t is constant so that,

NOTES

in a way, we assume the proportional income tax rate system. In fact, *t* may be regarded as the marginal propensity to tax (MPT).

The balanced budget multiplier with the proportional income tax included, can be derived in the following manner–

$$Y = C + I + G \quad ...(1.53)$$

$$Y = a + bY_d + \bar{I}_A + \bar{G} \quad ...(1.54)$$

$$Y = a + b[Y - (d + tY) + R] + \bar{I}_A + \bar{G} \quad ...(1.55)$$

$$Y = a + bY - bd - btY + bR + \bar{I}_A + \bar{G} \quad ...(1.56)$$

Transferring the terms bY and btY to the left-hand side we get–

$$Y - bY + btY = a - bd + bR + \bar{I}_A + \bar{G} \quad ...(1.57)$$

$$Y(1 - b + bt) = a - bd + bR + \bar{I}_A + \bar{G} \quad ...(1.58)$$

Dividing through by $1 - b + bt$ we get–

$$Y = \frac{a - bd + bR + \bar{I}_A + \bar{G}}{1 - b + bt} \quad ...(1.59)$$

The new equilibrium income after the given increase of $\Delta\bar{G}$ amount in government expenditure which is wholly financed by an equal increase of ΔT amount in government taxes will be:

$$Y + \Delta Y = a + b[Y + \Delta Y - (d + tY + t\Delta Y + \Delta T) + R] + \bar{I}_A + \bar{G} + \Delta\bar{G} \quad ...(1.60)$$

$$Y + \Delta Y = a + bY + b\Delta Y - bd - btY - bt\Delta Y - b\Delta T + bR + \bar{I}_A + \bar{G} + \Delta\bar{G} \quad ...(1.61)$$

Transferring the terms with Y and AK to the left-hand side, we get–

$$Y - bY + btY + \Delta Y - b\Delta Y + bt\Delta Y = a - bd - b\Delta T + bR + \bar{I}_A + \bar{G} + \Delta\bar{G} \quad ...(1.62)$$

$$Y(1 - b + bt) + \Delta Y(1 - b + bt) = a - bd - b\Delta T + bR + \bar{I}_A + \bar{G} + \Delta\bar{G} \quad ...(1.63)$$

Dividing through by $1 - b + bt$, we get–

$$Y + \Delta Y = \frac{a - bd - b\Delta T + bR + \bar{I}_A + G + \Delta\bar{G}}{1 - b + bt} \quad ...(1.64)$$

By deducting equation (1.59) from equation (1.64), we get–

$$\Delta Y = \frac{\Delta\bar{G} - b\Delta T}{1 - b + bt} \quad ...(1.65)$$

$$\Delta Y = \frac{\Delta\bar{G} - b\Delta T}{1 - b + bt} \text{ (since } \Delta T = \Delta\bar{G}) \quad ...(1.66)$$

$$\Delta Y = \frac{\Delta\bar{G}(1 - b)}{1 - b + bt} \quad ...(1.67)$$

Dividing both sides by ΔG gives the modified balanced budget multiplier

$$\frac{\Delta T}{\Delta\bar{G}} = \frac{1 - b}{1 - b + bt} \quad ...(1.68)$$

The term $1 - b/1 - b + bt$ is, therefore, the balanced budget multiplier when the

Since both *b* and *t* are constant positive fractions, the value of the term $1 – b + bt$ in the denominator is higher than the value of the term $1 – b$ in the numerator. Consequently, the value of their coefficient $1 – b/1 – b + bt$, which is the balanced budget multiplier, must be positive but less than one. It is, therefore, obvious that the balanced budget multiplier is positive, i.e., greater than zero in both the cases. The balanced budget multiplier is, however, smaller when we consider the induced taxes along with the autonomous taxes compared with the balanced budget multiplier when only the autonomous taxes are considered. The balanced budget multiplier with a progressive income tax will be even smaller because the value of the term—the marginal propensity to tax will increase as a result of which the value of the term *bt*, which is a positive term in the denominator, will also increase. Consequently, the value of the balanced budget multiplier $1 – b/1 – b + bt$ will be smaller.

The use of the balanced budget to raise the level of aggregate income significantly in the economy is inefficient because a very large increase in government purchases will be necessary involving a sizeable shift in the economy's resource allocation and unnecessary increases in taxes which could cause further difficulties. Nevertheless, contrary to the old orthodox belief, the balanced budget multiplier theory stresses that a balanced budget is not neutral in its impact on the economy. It exerts its positive impact on the economy and makes the presence of the government felt in influencing the total size of economic activities in the economy even in the absence of resorting to a policy of deficit financing.

1.8 DIFFERENCE BETWEEN KEYNESIAN AND CLASSICAL MACRO-EQUILIBRIUM

The concept of the consumption function marks an important development in modern macroeconomic theory. Its discovery, inclusion and treatment as an important edifice of macroeconomic theory, and also its clear formulation owe much to the intelligence and clear insight of Keynes. Although Keynes recognized the importance of other factors in determining consumption, he, however, argued that income was the single most important determinant of consumption. According to Alvin H Hansen, Keynes' great contribution is the clear and specific formulation of the consumption function. Hansen regards the clear formulation of the consumption function as stated in *The General Theory* as Keynes' 'epoch-making contribution to the tools of economic analysis, analogous to, but even more important than Marshall's discovery of the demand function.'The concept of the consumption function is Keynes' monumental discovery and is the kingpin of the elegant edifice of the Keynesian theory of income and employment. It is the 'heart of the Keynesian analysis' which supports the extensive and continuing efforts to define, refine and to measure statistically the nature and stability of the relationship between income and consumption.

There was no classical consumption function relating the aggregate consumption to aggregate income because the classical economic analysis was developed in terms of the full employment economy. At any given time, income was a given constant. According to the classicists, out of the constant full employment real income, the decision of the people to consume more and save less and *vice versa* was influenced by changes in the interest rate. A high rate of interest induced the consumers to save more by postponing consumption while a low rate of interest discouraged saving. In the classical economic theory, saving was assumed to be a positive function of the rate of interest. Since higher

Check Your Progress

10. What causes shifts in the aggregate demand function?
11. What are the two fundamental features of investment multiplier principle?
12. State one ground upon which the logical multiplier theory has been criticised.

NOTES

saving out of any given income corresponding to a higher rate of interest meant lower aggregate consumption, consumption was regarded as a negative function of the rate of interest. Since saving was regarded as interest-elastic, consumption also was interest-elastic although this elasticity was negative. In the classical economic analysis, consumption was not a function of income because income was not a variable. In a fully employed economy, income was a constant. The assumption of full employment ruled out any possibility of an increase in the equilibrium aggregate income in the short period in the economy. This approach clashes with Keynes' approach according to which consumption is a stable function of income.

John Maynard Keynes' view about the income-consumption relationship can be summarized as follows.

1. According to Keynes, real consumption was a 'fairly stable function of real income'. He regarded this income-consumption relationship as a fundamental psychological law.
2. Keynes stated this 'fundamental psychological law' by observing that 'men are disposed, as a rule and on the average, to increase their consumption as their income increases, but not by as much as the increase in their income.' This means that the marginal propensity to consume (MPC) is positive and less than one.
3. According to Keynes, the short period MPC was less than the long period MPC because 'a man's habitual standard of life usually has the first claim on his income, and he is apt to save the difference which discovers itself between his actual income and the expense of his habitual standard; or, if he does adjust his expenditure to changes in his income, he will, over short periods, do so imperfectly. Thus a rising income will often be accompanied by increased saving, and falling income by decreased saving, on a greater scale at first than subsequently.' In other words, in the short period the average propensity to consume falls as income increases and *vice versa.*
4. According to Keynes, even in the long run, as a rule, a greater proportion of income will be saved as real income increases. This means that even in the long run the average propensity to consume will fall as income increases and *vice versa*. It is likely to be so because although 'the satisfaction of the immediate primary needs of a man and his family is usually a stronger motive than the motives toward accumulation', these latter motives 'acquire effective sway when a margin of comfort has been attained'. Keynes does not, however, seem to be certain about this tendency because in the subsequent lines of his book he has written: 'But whether or not a greater proportion is saved, we take it as a fundamental psychological rule of any modern community that, when its real income is increased, it will not increase its consumption by an equal *absolute* amount, so that a greater absolute amount must be saved, unless a large and unusual change is occurring at the same time in other factors.'

Empirical studies made about the nature of the consumption function have not verified all of Keynes' statements. Some of his arguments hold in the short period but not in the long period and *vice versa.* For example, his statement that consumption is a stable function of income is true in the long-run and the time-series data analysis shows a proportionality relationship between income and consumption pointing out that over a long period of time the average propensity to consume is constant. However, over a very short period of a quarter of a year, the association between consumption and income

may be quite erratic. Consequently, consumption and income may move in the opposite direction. Keynes' statement that the marginal propensity to consume is positive and less than one is true in the short, intermediate and long periods. Keynes' statement that the average propensity to consume falls as income increases in the long period has been proved wrong. In fact, Klein and Kosobud have argued that the average propensity to consume (income-consumption ratio) had actually recorded a slight increase over the long run, the increase being 0.129 per cent each year. Again, there is no evidence to support the view that the marginal propensity to consume falls as income increases, especially in the short-run cycle. Keynes' statement that the short-run MPC is less than the long-run MPC is now generally accepted. Keynes' statement that even in the long-run the APC (*C*/*Y*) will fall as income increases is wrong since it has been proved by several empirical studies that over the past hundred years the income-consumption ratio has remained constant.

1.9 THE PIGOU EFFECT: PIGOU'S CRITIQUE OF KEYNES' UNDEREMPLOYMENT EQUILIBRIUM

The *Pigou effect*, also called the *real-balance effect*, is named after the well-known Cambridge School economist, Arthur Cecil Pigou, who first formulated the relationship between the aggregate consumption, the real cash balances and the general price level.This particular effect was advanced to counter the Keynesian argument that a fall in wages and prices exerts its influence only through the interest rate changes which becomes inflexible in the downward direction at the liquidity trap interest rate where the aggregate effective demand was less than necessary to ensure full employment in the economy and to defend the classical position relating to the effect of the general wage-cut in achieving full employment in the economy during the course of a serious controversy which ensued in the early 40s between Arthur Cecil Pigou and John Maynard Keynes. Keynes had strongly refuted the classical argument that a general wage-cut could remove unemployment in the economy.

John Maynard Keynes and his followers had demonstrated the failure of a perfectly competitive free market economy to achieve a stable equilibrium at full employment. The Keynesians argument opened the floodgates of government intervention expressed in the economic articles. It was at this time that other economists, particularly Gottfried Von Haberler and Arthur Cecil Pigou took position to challenge this conclusion suggesting that the Keynesians had ignored the importance of the real-balance effect on an individual's behaviour. The arguments of both Pigou and Haberler were based on the assumption of the important role of wealth in the determination of the consumption function.

The Pigou effect or the real-balance effect measures, *ceteris paribus*, the influence of the change in an individual wealth-holder's real balances on the aggregate effective demand. Pigou had argued that a general price fall which was associated with a general wage-cut would, by increasing the real value of the cash balances of individuals, raise the level of aggregate demand in the economy by shifting the aggregate consumption function upward. If, in fact, an increase in the real value of wealth stimulates consumption, it could then be conceived that there would always be some amount of fall in the wages and prices which would be sufficient to increase the aggregate consumption enough to eliminate any deficiency in the aggregate effective demand at the full employment level in the economy. Pigou's following statement bears repetition here because different interpretations have been ascribed to it.

Check Your Progress

13. Why is the Keynesian statement that 'even in the long run the APC (C/Y) will fall as income increases' considered wrong?
14. What discovery by Keynes is considered as the heart of the Keynesian analysis?

NOTES

According to Pigou, 'as money wage rates fall, money income must fall also and go on falling. Employment, and so real income, being maintained, this entails that prices fall and go on falling, which is another way of saying that the stock of money, as valued in terms of real income, correspondingly rises. But the extent to which the representative man desires to make savings otherwise than for the sake of their future income yield depends in part on the size, in terms of real income, of his existing possessions. As this increases, the amount that he so desires to save out of any assigned real income diminishes and ultimately vanishes, so that we are back in the situation...where a negative rate of interest is impossible.'

But how does the Pigou effect operate? Assume that investment falls so that income and employment also fall in the economy. As a consequence of general unemployment, money wages in the economy fall which causes the production costs and prices to fall. In short, an all-around wage-price deflation grips the economy. Now what happens to the wealth portfolios of the wealth-holders? The total wealth of the wealth-holders comprises of the different forms of real and nominal assets. So far as the real assets—land, buildings, and common-stock shares—are concerned their prices will also fall with the general fall in prices in the economy. Consequently, the real value of these assets will not change. However, the general fall in prices will increase the real value of the fixed money or nominal assets such as money, savings bank deposits, government bonds, etc.

The increase in the real wealth of their fixed money assets will induce the wealth-holders to save a smaller fraction of their income and to spend a larger fraction of their income on consumption. Consequently, the aggregate consumption function will shift upward showing a higher aggregate consumption at each different level of aggregate income. This will raise the level of aggregate effective demand, output and employment in the economy. Thus, there would be a certain fall in the general price level which will raise the real value of a given stock of fixed money assets. This will further shift the aggregate consumption function upward by the amount necessary to shift the aggregate demand function upward to that position which yields the stable, full employment, equilibrium, aggregate, real income in the economy.

To the extent that the equilibrium aggregate real income and output in the economy can be raised through the operation of the Pigou effect, Pigou can be said to have succeeded in defending the classical position that in a perfectly competitive free market economy, full employment equilibrium was possible through the wage-price flexibility. It states that so long as the commodity prices, wages and interest rate are perfectly flexible, the system is capable of moving to the full employment level. Permanent unemployment in macrostatic analysis is possible if and only if one or more of these price variables are rigid. A rigid money wage can result in permanent unemployment. So also a rigid price level and a rigid interest rate can lead to permanent unemployment. In other words, Pigou won 'triumph' for the classical economic theory by showing that so long as wages and prices were flexible in the economy, full employment equilibrium output could be attained.

Basically, Pigou argued that if consumption depended on both the disposable personal income and wealth, the aggregate consumption function could be written as:

$C = f(Y_d', W)$

Real cash balances comprise a part of individuals' total stock of wealth. Unlike the physical wealth whose nominal value rises or falls with a rise or fall in the general price level, the nominal value of money remains constant and its real value moves inversely with changes in the general price level. For example, if prices double, the real value or purchasing power of `100 is halved and *vice versa.* If these real cash balances represent

net financial assets—financial claims against which there are no offsetting liabilities—changes in the general price level will inversely affect the net wealth of the entire economy. If the general price level falls when there is depression in a perfectly competitive economy, the net wealth would rise. Consequently, consumption spending out of any given income would also rise.

As consumption rises, the *IS* curve shifts rightward raising the aggregate demand in the process. In other words, a real-balance effect occurs in the household sector raising the aggregate demand in the economy as the general price level falls and *vice versa.* This is the Pigou effect.

Figure 1.17 shows the influence of the Pigou effect on the consumption-saving relationship. As the general price level falls from P_3 to P_2 to P_1 and eventually to P_0, the real cash balances and net wealth increase. Consequently, the short period consumption function shifts upward and concomittentlythe short period saving function shifts downward as shown in Figure 1.17(A). As the saving function shifts downward with each fall in the general price level, a new *IS* curve corresponding to each new general price level is generated as shown in Figure 1.17(B). With each change (fall) in the general price level, a new *LM* curve is also generated as shown in Figure 1.17(B) showing the money market equilibrium. The intersection points *A, B, C* and *D* in Figure 1.17(B) furnish us with the level of aggregate demand at each different general price level. The respective points on the aggregate demand curve which is generated by changes (falls) in the general price level have been labelled as *A′*, *B′*, *C′* and *D′* in Figure 1.17(C). Due to the presence of the Pigou effect in the commodity market, a price-elastic aggregate demand curve is generated which makes it possible to achieve a stable full employment equilibrium in a perfectly competitive economy notwithstanding the presence of the Keynesian liquidity trap. Even if instead of the presence of the Keynesian liquidity trap, the situation represented insufficiency of investment, the consequences of the presence of the Pigou effect would have remained unchanged.

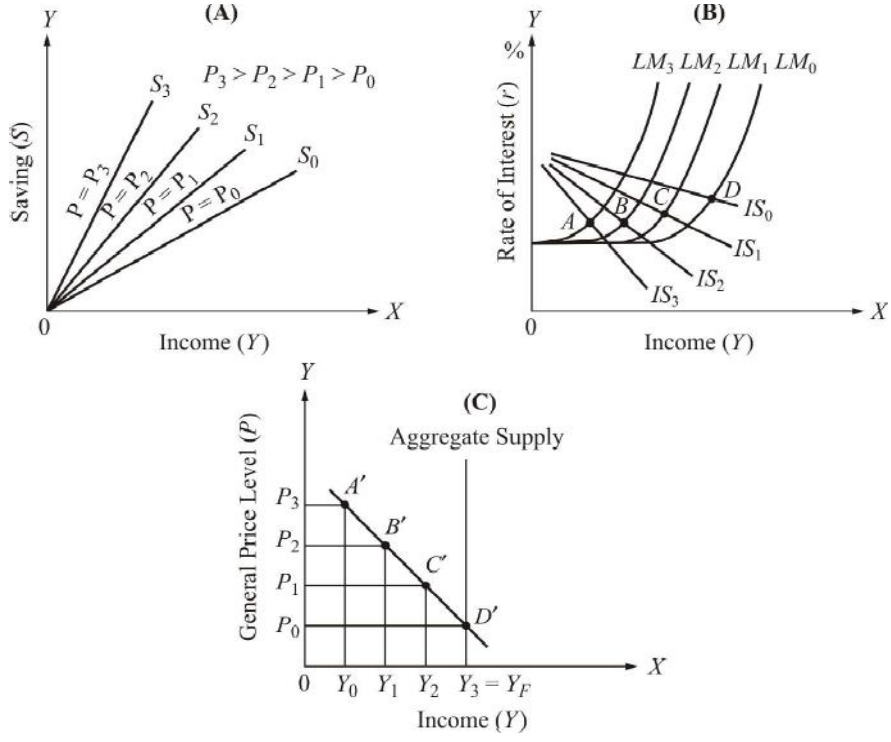


Fig. 1.17

In short, the Pigou effect is of great theoretical importance because it shows that a perfectly competitive economy is capable of achieving a stable full employment output equilibrium. Its presence ensures that the monetary policy will always be effective as it can directly increase the net wealth of the households.

NOTES

Criticism

Today, however, economists concede no such victory to Pigou and the classical macroeconomic theory. The Pigou effect has been criticized on several grounds and Pigou's efforts to salvage the classical theory from Keynes' attack has not proved successful. Pigou himself admitted that his analysis was simply an academic exercise 'of some slight use perhaps for clarifying thought, but with very little chance of ever being posed on the chequer board of actual life.' He felt that at least on social ground, no government would allow the extreme drastic cuts in money wage which were needed for the real-balance effect to be effective in eliminating the deficiency of the full employment aggregate effective demand in the economy.

1. The Pigouvian argument cannot apply to all fixed money assets. Although the real value of these assets held by the creditors increases as the general price level falls, however, this also increases the real burden of debt for the debtors. Consequently, the increase in the creditors' average propensity to consume may be offset by the decrease in debtors' average propensity to consume, leaving the average aggregate propensity to consume unchanged. This criticism could, however, be countered by arguing that the real-balance effect should be confined to only government bonds and money obligations because the government is a debtor whose spending will not be adversely affected by an increase in the government debt burden consequent upon the fall in the general price level. In short, deflation increases the real value of wealth-holders' government debt holdings—currency and bonds—and consequently stimulates consumption spending of the government debt holders without at the same time causing decrease in the aggregate government spending. The net effect, as was argued by Pigou, of an increase in the real cash-balances of the asset-holders was an increase in the aggregate spending.

   Assuming that the Pigouvian argument holds, the crucial question, however, is: how much rise in the aggregate consumption spending will any given fall in the general price level bring about? In the face of a certain amount of unemployment, it is one thing to say that a 5 per cent fall in the general price level and the accompanying increase in the real value of the cash balances is sufficient to raise the level of aggregate effective demand (by raising consumption) to rid the economy of unemployment and quite another to say that a 50 per cent fall in the general price level was necessary before unemployment could be removed through the remedy of Pigou effect. Unfortunately, there is nothing in the Pigou effect which could tell us about the magnitude of deflation needed to eliminate any given size of unemployment. At best, it speaks only about the direction of the effect of any given increase in the real value (worth) of liquid assets.

   If a severe deflation is advocated as a remedy for unemployment, it would be foolish to rely on the Pigou effect as a practical measure to restore full employment in the economy. In fact, a severe deflation which was characteristic of the great depression witnessed in the 30s of the 20th century and mass unemployment could coexist as economic phenomena although the Pigou effect makes us believe

to the contrary. Moreover, a once and for all deflation, even if it could be produced, would not work. When prices fall perceptibly, people begin to entertain the expectation that these will soon rise on the onset of recovery. In this situation, wealth-holders will treat the increase in the real value of their fixed rupee assets as purely temporary and they may not increase their consumption spending. Mayer's study suggests that the Pigou effect is too weak to be of any practical significance. As against this, however, the study made by Ta-Chung Liu suggests that the Pigou effect is of considerable strength and should not to be ignored.

2. The Pigou effect assumes that wealth-holders' taste for wealth does not increase with an increase in their wealth stock, i.e., their propensity to accumulate wealth remains constant. George Katona has argued that one's taste for wealth should increase with the increase in one's wealth. If this was so, the whole a *priori* basis of the Pigou effect is challenged.
3. A fall in the general price level may create expectations of a further fall in the general price level. Indeed, people may entertain the strong expectation that the fall in prices will continue unabated. In the face of such expectations, consumers will postpone their purchases because by doing so they hope to get more for their money. Consequently, the consumption function will not shift upward, rather it might shift downward contrary to the Pigouvian expectation. In the face of this possibility, deflation might make the situation worse by increasing the rot of unemployment further rather than restoring stability in the economy by reducing it.
4. The Pigou effect is a comparative static theoretical proposition. It says nothing about the dynamics of a slow adjustment to gradual deflation causing undesirable redistribution of income and wealth in favour of the 'rentier' class and against the 'active' entrepreneurial class reducing the employment opportunities in the process. The Pigou effect may be regarded as an argument favouring a long-run downward trend in the prices which would adversely affect the employment position of the mass of consumers by adversely affecting the business profits.
5. The Pigou effect is inconsistent with the neo-classical dichotomy between the real and the monetary sectors of the economy. If it is assumed (though wrongly), as the classicists did, that since in an economy the real markets are separate from the financial markets then how can changes in the general price level affect the level of aggregate real demand by affecting the real consumption spending. The real-balance effect integrates the monetary and value theories (the monetary and real sectors of the economy). It is, therefore, obvious that the inclusion of the real-balance effect in the classical model violates the classical assumption of the dichotomy between the real and the monetary sectors of the economy and the exclusion of the real-balance effect from the model involves it in a contradiction. Suffice it to say that generally a macroeconomic model is not dichotomizable since it is usually impossible to separate the system into a self-contained subset of markets and to determine the equilibrium values of the subset of variables. Normally, in the economy everything depends on everything else suggesting interdependence between the different real and monetary variables.
6. The Pigou effect ignores the adverse effect of deflation on the non-cash or real component of the total wealth portfolios of individual wealth-holders. The fall in prices, particularly if it is perceptible, will cause a substantial fall in the real value of property and other non-cash assets of wealth-holders. Consequently, the

favourable effect of a general price fall on the cash assets will be offset or may even be more than offset by the adverse effects of such a price fall on individual wealth-holders' non-cash assets. It is for these reasons that the economists do not consider the Pigou effect as meriting any serious attention. Even the neo-classical economists did not necessarily recommend that the government should wait passively for the deflationary disturbance to occur in order to bring about the required increase in the real balances to eventually re-establish the state of full employment in the economy. All they wanted to do was to demonstrate the theoretical rectitude and consistency of the classical macroeconomic assertion that in a world characterized by downward wage-price flexibility full employment was guaranteed. On the basis of the Pigou effect alone, it cannot be legitimately argued that social policy should be directed towards achieving the maximum degree of price flexibility in the pious hope of removing unemployment in the economy. In short, the Pigou effect has little to commend as a practical economic policy approach to the solution of the unemployment problem.

Although the Pigou effect is of little practical importance, yet it is of extreme theoretical importance in as much as it produces a stable full employment equilibrium in a perfectly competitive free enterprise market economy. In an economy characterized by downwardly rigid money wages, the invocation of the Pigou effect will not ensure full employment equilibrium and such an economy is destined to remain in a state of stable equilibrium at less than full employment.

In a regime of downwordly rigid wages and prices, the Pigou effect contributes virtually nothing to an automatic return to full employment equilibrium. Even if prices and wages are somewhat flexible, the magnitude of the Pigou effect would probably be too small to be of practical significance. The Pigou effect, however, asserts that monetary policy can restore full employment even if the special Keynesian conditions—insufficiency of investment and the liquidity trap—prevail in the economy. Thus, the Pigou effect demolishes the Keynesian argument against the effectiveness of monetary policy. Since recent research has shown that neither of these two conditions have actually occured and it seems unlikely that these will be encountered in the foreseeable future, there appears no basis to uphold the views of those who reject the monetary policy completely and place an exclusive reliance on the fiscal policy as an instrument of economic stabilization policy.

1.10 SUMMARY

In this unit, you have learnt that,

- The classical economists assumed that full employment was a normal feature in the economy.
- In a *laissez-faire* economy market forces operated in the system which maintained full employment and consequently kept the aggregate output at the level producible under conditions of full employment.
- The essential feature of classical macroeconomic analysis is that it presents a model of full employment in the economy in the long period.
- Underlying the analysis, are the assumptions of perfect competition in the factor and product markets and profit-maximization on the part of firms.

Check Your Progress

15. By what other name is the Pigou effect also known as?
16. Why is the Pigou effect considered of little practical importance but of extreme theoretical importance?

- In the classical economic theory, money does not matter and its function in the economy is merely to facilitate the real transactions by serving as a medium of exchange.
- The classical macroeconomic theory explains the determination of the equilibrium level of aggregate employment and output, real wage, saving and investment, rate of interest, general price level and money wage.
- The classical supply function of labour is positively sloping in relation to real wage showing that the amount of labour (total number of man-hours) offered by the workers for work increases as real wage increases.
- In the classical theory, a change in the aggregate money supply will not affect the real wage, employment and output in the economy.
- Voluntary unemployment is when an individual does not accept a suitable job at the current wage rate.
- Involuntary employment is when people are ready to accept suitable work at the current wage rate but they are not able to find work.
- The concept of liquidity trap has often been regarded as an article of faith and a bitter controversy has arisen about whether it constitutes the fundamental difference between the classical and the Keynesian economic analysis.
- Unlike the classical macroeconomic theory in which the real and monetary sectors of the economy remain dichotomized, in the Keynesian theory these are integrated forming parts of a compact whole.
- The simple Keynesian theory of income, output and employment determination can be studied either through the aggregate income-expenditure approach in the form of $Y = C + I$ or through the aggregate saving-investment approach in the form of $S = I$.
- The Keynesian theory has been criticized as being 'too static' in the sense of being concerned with equilibrium conditions during the short period in which technology and capital stock are given and are not likely to change.
- Consumption plays a crucial role in the determination of income and employment generation. If the consumption increases income and employment will also increase.
- Before John Maynard Keynes, the importance of increase in the new investment for causing the increase in aggregate income had been stressed by the business cycle theorists, more prominently by Knut Wicksell, Michael Tugan Baranowsky and Arthur Spiethoff.
- The concept of simple investment multiplier constitutes an important pillar of the whole edifice of the Keynesian theory of income and employment.
- The concept of investment multiplier, in the context of Keynes' theory of income and employment, refers to the increase (decrease) in the equilibrium level of national income consequent upon a given increase (decrease) in the autonomous investment spending.
- The Pigou effect, also called the real-balance effect, is named after the well-known Cambridge School economist, Arthur Cecil Pigou, who first formulated the relationship between the aggregate consumption, the real cash balances and the general price level.

NOTES

- According to Pigou, 'as money wage rates fall, money income must fall also and go on falling. Employment, and so real income, being maintained, this entails that prices fall and go on falling, which is another way of saying that the stock of money, as valued in terms of real income, correspondingly rises.

1.11 KEY TERMS

- **Consumption function:** The consumption function is a mathematical formula laid out by famed economist John Maynard Keynes.
- **Balanced budget multiplier:** Balanced budget multiplier indicates that an equal change in government spending (G) and taxes (T), which leaves the budget unchanged, still have an impact on the level of output and income (Y).
- **Pigou effect:** Pigou effect is the stimulation of output and employment caused by increasing consumption due to a rise in real balances of wealth, particularly during deflation.
- **Macroeconomic equilibrium**: Macroeconomic equilibrium is an economic state in an economy where the quantity of aggregate demand equals the quantity of aggregate supply.

1.12 ANSWERS TO 'CHECK YOUR PROGRESS'

1. Classical economic theory dealt with the following factors:
   - What goods and services would be produced in the economy with its given resources;
   - The allocation of the economy's given resources between their different rival uses;
   - The relative prices of different goods and services and of the factors of production; and
   - The distribution of income earned from production between the different cooperating factors of production.
2. The classical theory of employment, output and price was attacked by Thomas Robert Malthus, Jean Charles Leonard de Sismondi, Karl Marx, J A Hobson and Silvio Gesell.
3. The seven categories of unemployment are:
   - Structural unemployment
   - Cyclical unemployment
   - Marxian unemployment
   - Frictional unemployment
   - Seasonal unemployment
   - Hidden unemployment
   - Classical unemployment
4. Cyclical unemployment is usually influenced by the state of the economy. It may result in lowering of demand for products. Demand for products will go down because there is a reduction in the ability of consumers to spend. Lower demand will lead to cutting down of production. Low production may lead to forcible shrinking of the workforce which in turn leads to more cyclical unemployment.

5. The crucial argument in the invalidation of the classical theoretical economic system is that the ex ante investment is not always equal to the ex ante saving at some positive rate of interest. Once this follows, there will be idle cash balances, the velocity of money will fluctuate cyclically and the money stock will no longer be proportional to the money output even though the prices are flexible.
6. The most distinguished feature of the Keynesian approach to the demand for money is the speculative demand for money M_2 which is inversely related to the rate of interest.
7. The simple Keynesian theory of income, output and employment determination can be studied either through the aggregate income-expenditure approach in the form of $Y = C + I$ or through the aggregate saving-investment approach in the form of $S = I$.
8. The Keynesian theory is criticized as the Keynesian model is 'too aggregative'. It has fewer than necessary variables and relationships. It does not segregate the variables involved into their various component parts. For example, investment has been discussed as a single entity without breaking it into fixed plant investment, inventory investment, residential investment, etc.
9. Consumption function is also known as propensity to consume. Consumption plays a crucial role in the determination of income and employment generation. If the consumption increases income and employment will also increase.
10. The shifts in the aggregate demand function can be caused either by the shifts in the aggregate investment demand function or by the shifts in the aggregate consumption demand function or by the simultaneous shifting of both these demand functions.
11. Two fundamental features of the investment multiplier principle are:
    - The incurring of new expenditure will have an expansionary effect on the economy with unemployed resources which will be larger in magnitude than the size of the new expenditure itself; and
    - The expansionary process is necessarily limited and loses force on account of the leakages from the expenditure flow stream.
12. The logical multiplier theory has been criticized on the ground that it does not explain the path which is followed by income as it moves away from the original equilibrium to the new equilibrium position.
13. Keynes' statement that even in the long-run the APC (C/Y) will fall as income increases is wrong since it has been proved by several empirical studies that over the past hundred years the income-consumption ratio has remained constant.
14. The concept of the consumption function is Keynes' monumental discovery and is the kingpin of the elegant edifice of the Keynesian theory of income and employment. It is the 'heart of the Keynesian analysis' which supports the extensive and continuing efforts to define, refine and to measure statistically the nature and stability of the relationship between income and consumption.
15. The Pigou effect is also called the real-balance effect.
16. Pigou effect is of little practical importance, yet it is of extreme theoretical importance in as much as it produces a stable full employment equilibrium in a perfectly competitive free enterprise market economy. In an economy

characterized by downwardly rigid money wages, the invocation of the Pigou effect will not ensure full employment equilibrium and such an economy is destined to remain in a state of stable equilibrium at less than full employment.

NOTES

1.13 QUESTIONS AND EXERCISES

Short-Answer Questions

1. State in brief the evaluation of the Classical model.
2. What are the subjective factors affecting consumption and multiplier analysis?
3. Write a short note on pump-priming.
4. State the difference between Keynesian and Classical macro-equilibrium.
5. State the impact of structural unemployment.

Long-Answer Questions

1. Explain the Keynesian theory of income, output and employment. Why is it regarded as 'too static' and 'too aggressive'?
2. Discuss the different causes of unemployment.
3. Explain the full Keynesian model of income determination with all the important macroeconomic variables using suitable diagrams.
4. Analyse the meaning and importance of Pigou effect.
5. Explain how wage price flexibility can help in achieving full employment in a perfectly competitive economy.

1.14 FURTHER READING

Mankiw, N Gregory. 2010. *Macroeconomics.* Worth Publishers, New York.

Shapiro, Edward. 1996. *Macroeconomic Analysis.* Galgotia Publication, New Delhi.

Branson, WH. 1979. *Macroeconomic Theory and Policy*. Universal Book Stall, New Delhi.

Jha, R. 1999. *Contemporary Macroeconomic Theory and Policy*. New Age International, New Delhi.

Gupta, SB. 1995. *Monetary Economics: Instruments and Policy*. S Chand & Co. New Delhi.

UNIT 2 NEO-CLASSICAL AND KEYNESIAN SYNTHESIS

Structure

2.0 INTRODUCTION

The Neoclassical-Keynesian Synthesis refers to the Keynesian Revolution as interpreted and formalized by a largely American group of economists in the early post-war period. The centrepiece of the Neoclassical-Keynesian Synthesis (or the Neo-Keynesian system) was the infamous IS-LM Model first introduced by John Hicks (1937) and then expanded upon by Franco Modigliani (1944). The IS-LM model purported to represent the gist of John Maynard Keynes's General Theory (1936) in the form of a system of simultaneous equations.

One of the startling results of the IS-LM model was that it was unable to obtain the Keynesian result of an unemployment equilibrium. The model tended to yield the Neoclassical result of full employment. As a result, in order to generate an unemployment equilibrium as a solution to this system of equations, the Neo-Keynesians appealed to rigid money wages, interest-inelastic investment demand, income-inelastic money demand or some other imperfection to this system. Thus it is referred to as a synthesis of Neoclassical and Keynesian theory in that the conclusions of the model in the long run or in a perfectly working IS-LM system were Neoclassical, but in the short-run or imperfectly working IS-LM system, Keynesian conclusions held.

2.1 UNIT OBJECTIVES

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

- Analyse the classical and Keynesian theories of interest
- Differentiate between liquidity preference theory and modern theory

NOTES

- Discuss the Investment Saving-Liquidity Preference Money Supply model
- Describe the relative effectiveness of monetary and fiscal policies
- Analyse the extension of IS-LM model to labour market and flexible price
- Discuss the Mundell-Fleming model of an open economy

2.2 CLASSICAL AND KEYNESIAN THEORIES OF INTEREST

As a form of income accruing to its owner for the use of capital, interest has been a subject of a deep and fierce controversy from very early times. Savants, philosophers, religious leaders, social reformers, statesmen have all expressed their views on the desirability or otherwise of accepting the payment of interest on the part of the lender. In ancient times, interest was likened to usury and people were enjoined from accepting interest on ethical and religious grounds. Apart from these aspects, however, a plethora of analytical writings dwelling on the discussion of why interest is paid and what determines its equilibrium rate have been developed.

Early Views

As a payment made by a borrower to a lender of funds, interest was unknown in the ancient savage communities. It was, however, common in the early societies living in the eastern Mediterranean Persian Gulf region of Babylon, Egypt and Greece. The ancient Greek philosophers were generally hostile to the payment of interest by the borrower to the lender. Plato enjoined the lending of money on interest while Aristotle opposed the charging of interest on loans given for unproductive purposes. The position in Rome was not much different from that in Greece where interest was either abolished by law or was regulated through the royal edicts although evasion of the law was not unknown.

During the Middle Ages despite the ban on usury, it rather grew. The religious ban on accepting interest on loans drew its strength from the medieval church whose hold on the laity was quite firm and strong. It was a sin to lend on interest and the *Old Testament* enjoined the Christians from lending on usury. In 1311, Pope Clement V had declared all secular legislation permitting usury as null and void and declared as heretical those statements which did not declare interest taking as sinful. *The Gospel of Luke* prohibited usury stating that those who lend should not hope for anything again. Father Tertullian declared the taking of interest as immoral. The ecclesiastical prohibition on accepting interest by the lenders continued throughout the Middle Ages.

The impact of this thinking was evident till the end of the Middle Ages because even in the 13th century St. Thomas Aquinas had declared that the taking of usury (interest) for money lent was unjust in itself because this was to sell what did not exist. Obviously, in saying so Aquinas was using the ancient Greek philosopher Aristotle's argument about the barrenness of money. However, the ban on usury notwithstanding, governments and individuals regularly borrowed funds on interest. In short, while the medieval churchmen and monarchs laid down a standard of conduct free from the taint of usury, its observance in life could not be effectively enforced. Later, during the Protestant Reformation period, John Calvin, while disapproving of money lending as a business, did not emphasize on a general condemnation of usury. According to Calvin, interest was not to be demanded from the needy. He, however, said that interest could be taken on the productive loans. The French jurist Charles Dumsulin declared that usury was not forbidden by divine law

and that interest at moderate rate should be allowed. On the whole, the early discussion on the subject of interest (usury) lacked in the analytical rigour and was cast essentially in the ethical and religious mould.

NOTES

2.2.1 Pre-classical Approach[1]

The pre-classical approach to interest was largely concerned with the ideas on the subject contained in the mercantilist and physiocratic writings published largely during the 17th and 18th centuries. The mercantilist approach regarded interest as a payment for money rather than for capital. The mercantilists, who were men of affairs immersed in the daily affairs of the markets and political happenings, regarded interest as the price of money determined by the supply of currency in relation to the demand for money loans. Thus writing towards the close of the 17th century, the mercantilist writers emphasized the crucial role of the monetary factors–changes in the supply of and the demand for money–in explaining the levels of and changes in the interest rates.

About half a century later, the Physiocrats criticized the mercantilist view that the rate of interest was governed by the quantity of money. In his essay *Of Interest* which was included in his *Political Discourses* published in 1752, David Hume developed the idea that interest rate was determined by the supply of and the demand for capital. As against the mercantilist view that plenty of money lowered the rate of interest, Hume stated that an increase in the quantity of money had no effect other than to raise the prices in the economy. The supply of capital depended on savings which depended on the people's habits.

The analytical explanation of interest was chiefly developed in the 19th century and the 20th century at the hands of a long line of economists. Central to the theoretical literature on the subject are the four principal theories known as the classical theory, the loanable funds or the neo-classical theory, the Keynesian theory and the neo-Keynesian or modern theory of interest which have been discussed here.

2.2.2 Classical Theory[2]

The classical theory of interest rate cannot be ascribed to any one single writer belonging to the classical school. It has, in fact, to be distilled from the scattered writings of many writers. According to John Maynard Keynes, it is 'difficult to state it precisely or to discover an explicit account of it in the leading treatises of the modern classical school.'[3] Following Adam Smith, the classical writers being interested in those fundamental forces which determined the long term interest rate, disregarded those factors of temporary and 'secondary' nature which characterized the short-run disequilibrium situations. Consequently, the monetary factors, although these were important in influencing the short-term or market rate of interest, were ignored by these writers as unimportant trivial elements in the determination of the long-run level of the rate of interest. For example, Henry Thornton while emphasizing the importance of bank credit in relation to short-term interest rate stated that creation or destruction of the bank credit did not influence the long-term interest rate. Similarly, John Stuart Mill writing in 1848 observed that although an increase in currency tends to lower the market rate of interest, in the long period 'the greater or less quantity of money makes in itself no difference to the rate of interest.' In short, treating the monetary factors as an unimportant causative element in the determination of the long-term interest rate, the classical writers propounded the non-monetary theory of interest rate which involved the real flow variables of the saving supply and investment demand.

NOTES

The main propelling force behind the demand for investible funds was the physical productivity of present capital goods which provided incentive for the firms to undertake investment and add to their capital stock. Since, *ceteris paribus,* the productivity of capital diminishes with each addition to the capital stock, firms will increase their capital stock–make investment–when the rate of interest at the margin equals the productivity of capital that is diminishing. In other words, the investment demand function was negatively related to the rate of interest.

As regards the saving supply function, it was positively related to the rate of interest. Generally speaking, everybody prefers the present satisfaction of a given amount over the future satisfaction of an equal amount even when the uncertainty about the occurrence of the latter is zero and the present satisfaction of any given magnitude is no greater than future satisfaction of the same magnitude. This preference for the present pleasures over future pleasures is explained by the fact that on account of their 'defective telescopic vision of the future'[4], individuals preferred a given amount of the present goods over the same amount of future goods. Expressed differently, individuals preferred a certain given sum in the present to a similar sum in the future. The difference between these two sums represented their degree or extent of time preference. To overcome this difference in valuation, the future payment sum must exceed the present sum sacrificed by some given amount called the rate of interest. It was also argued that even if the individual savers' time preference was zero, since saving involved waiting, some inducement in the form of interest payment was necessary as a compensation for the sacrifice involved in the waiting for the enjoyment of material possessions.

Taking the position that interest was a payment for overcoming the sacrifice involved in waiting that accompanied saving Alfred Marshall observed: ' human nature being what it is, we are justified in speaking of the interest on capital as the reward of the sacrifice involved in the waiting for the enjoyment of material resources, because few people would save much without reward; just as we speak of wages as the reward of labour, because few people would work hard without reward.'[5] As the reward made for the sacrifice involved in waiting increased, people would be induced to increase their saving. Highlighting this view, Alfred Marshall has stated that 'a rise in the rate of interest offered for capital, i.e., in the demand price for saving, tends to increase the volume of saving. For in spite of the fact that a few people who have determined to secure an income of a certain fixed amount for themselves or their family will save less with a high rate of interest than with a low rate, it is a nearly universal rule that a rise in the rate increases the *desire* to save; and it often increases the *power* to save, or rather it is often an indication of an increased efficiency of our productive resources; but the older economists went too far in suggesting that a rise of interest (or of profits) at the expense of wages always increased the power of saving '[6]. In other words, the saving supply function was positively related to the rate of interest.

According to the classical approach, the rate of interest was a factor which brought into equilibrium the demand for investment and the supply of savings. Investment represented the demand for investible funds and savings represented the supply of these funds while the rate of interest was that 'price' of the investible funds at which the supply of and the demand for investible funds were equated. Under the play of free market forces, the long-run equilibrium rate of interest rested at that point where the total amount of investment at that rate was equal to the total amount of saving at that rate. The classical interest rate theory which was developed under the assumption of full employment of labour and capital is a *flow* analysis in which both investment and

saving are *flow* variables directing attention to a *period of time* rather than to a *point of time.* Since both investment and saving are flow variables, these can only be expressed as quantities per time unit. In short, the equilibrium relates to the capital market which must be continuously cleared of saving which represents flow on to the market and investment which represents flow off the market.

The classical theory of interest rate can be diagrammatically explained as shown in Figure 2.1 where the negatively sloping linear investment demand function *II* is a negative function of the rate of interest while the positively sloping linear saving supply function *SS* is a positive function of the rate of interest. In other words,

$$I = f(r); \text{ and } \frac{dI}{dr} < 0$$

$$S = g(r); \text{ and } \frac{dS}{dr} > 0$$

In Figure 2.1, the positively sloping saving supply curve *SS* and the negatively sloping investment demand curve *II* intersect each other at point *A* corresponding to which the long-run equilibrium rate of interest is $0r_e$ which will come to stay in the capital market since at this interest rate, the demand for and the supply of investible funds are equal leaving no scope for deviation of the rate of interest from the 'norm' as long as the underlying supply and demand schedules do not alter their position. The market rate of interest was regarded as fluctuating around this long-run equilibrium rate of interest.

The classical theory of the rate of interest has three important features. In the *first* place, it is a purely flow theory, i.e., the saving supply and investment demand variables which determine the rate of interest are the *flow* quantities as distinct from the *stock* quantities.

Second, both saving and investment are the *real* variables as distinct from the monetary variables. Savings represent the real resources which become available as a result of voluntary saving on the part of people. These are governed by the distribution of income and wealth, the nature of expected future income streams and wealth-owners' time preferences between the present and future goods. Investment represents net addition to economy's total capital stock involving an increase in the economy's total productive capacity. Although both saving and investment are reckoned in money units, money does not matter and it is *neutral* in the economy. Expressed differently, operations in the capital market relate to the supply of and the demand for real securities. Those who borrow capital funds are the suppliers or sellers of real securities while the suppliers of capital funds are the purchasers of these securities. Money acts only as a mediating agent—as a mere veil. In the classical theory the rate of interest is not determined by the quantity of money in circulation. Consequently, it is invariant with respect to changes in the money supply because any change in the money supply is neutralized by equi-proportionate change in the prices leaving the quantity of real money in the economy unchanged. Consequently, the demand and supply curves of money intersect at the same rate of interest.

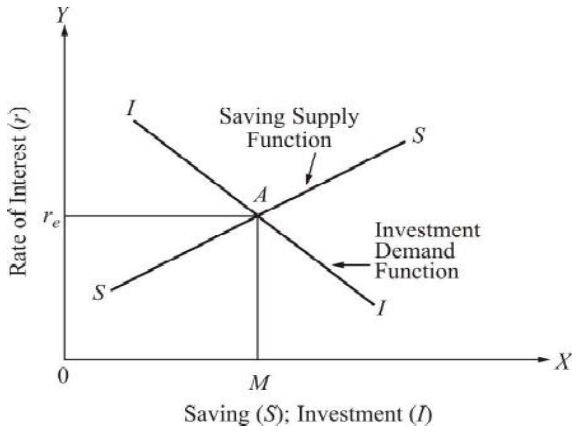


Fig. 2.1

Third, if the equilibrium between saving and investment is disturbed due to the shift in any one of the functions, it is reestablished exclusively through changes in the rate of interest without in any way affecting any other variable. In other words, the theory is self-contained and is relevant only to the equilibrium of economy's real sector while the monetary sector of the economy is completely left out of the purview of the theory.

Criticism

The classical theory of interest rate has been criticized on several grounds. Besides others, it has been severely criticized by Keynes. In the first place, the classical theory of interest rate is incomplete because it considers only the real as distinct from the monetary and only the flow as distinct from the stock variables. The result is that both the stock as well as the monetary variables which affect the rate of interest are completely left out from the discussion of interest rate determination. Any complete theory worth its name should be broad enough to include both the stock and the money variables. This weakness of the theory which flows from its excluding the consideration of monetary factors was recognized by the neo-classical economists, including the Swedish economist Knut Wicksell and the noted Cambridge economist Dennis Holme Robertson. These economists attempted to synthesize the monetary and non-monetary or real variables by developing the loanable funds theory according to which the equilibrium rate of interest is that rate which equates the supply of and the demand for the loanable funds. It was recognized that the classical theory was faulty because it considered the current voluntary savings as the only source of supply and investment as the only source of demand for the loanable funds.

Starting from a more realistic understanding of the operations taking place in the capital market, the exponents of the loanable funds theory recognized that the flow of money on to the market could be increased or decreased by the activities of the monetary authorities—central bank and the commercial banks—through the credit creation or credit squeeze and through the hoarding or dishoarding on the part of wealth-owners. Similarly, the flow of securities on to the market did not exclusively represent the issue of new securities to borrow the funds for investment; it was also fed by hoarding (borrowing in order to accumulate the surplus cash balances) on the part of wealth-owners.

Second, the classical theory of interest rate has been criticized by Keynes on several grounds. Keynes has questioned the nature of the classical saving supply and investment demand functions which are regarded as interest-elastic. According to Keynes, saving

is more a function of income rather than of the rate of interest for at very low levels of income, people will not save at all (they will rather dissave) even if they are offered inducement in the form of high rate of interest. Similarly, according to Keynes, investment is largely autonomous and at any rate the investment and interest rate relationship is a very weak and unreliable relationship. Had this relationship been strong, the monetary policy alone would have sufficed to ensure full employment in the economy and there would have been no necessity to resort to fiscal policy measures.

Besides questioning the form of the classical saving supply and investment demand schedules, Keynes has criticized the classical view that rate of interest is the 'price' which equates the demand for and the supply of investible resources. According to Keynes, 'the rate of interest is not the 'price' which brings into equilibrium the demand for resources to invest with the readiness to abstain from present consumption.'[7] Asserting that the rate of interest is a *purely monetary phenomenon* as distinct from the classical real phenomenon Keynes has stated that 'it is the 'price' which equilibrates the desire to hold wealth in the form of cash with the available quantity of cash...'[8] He has criticized the classical approach of regarding the rate interest as a return on saving on the ground that a man who hoards his savings in the from of cash earns no interest although he saves. In Keynes' view 'the mere definition of the rate of interest tells us in so many words that the rate of interest is the reward for parting with liquidity for a specified period.'[9]

According to Keynes, the classical theory of interest rate was faulty and misleading in another respect also. According to the theory, an upward or a downward shift in the investment demand schedule would result in the new equilibrium rate of interest determined by the point of intersection between the new investment demand curve and the given saving supply curve. In other words, when the investment demand curve shifts the saving supply curve does not shift. Keynes' criticism of this approach was that the assumption of a given saving supply curve corresponding to a shifting investment demand curve was faulty and untenable. Consequently, the conclusions that followed were also faulty. Keynes argued that when the aggregate investment outlay changed, the aggregate income also changed. Consequently, the amount saved at different interest rates also changed resulting in the shift of the saving supply curve simultaneously. If the aggregate investment outlay increases (decreases), the aggregate income must also increase (decrease) the multiplier times the increase in investment outlay.[10] In other words, it was faulty to assume the aggregate income as given when the aggregate investment was changing. And if the aggregate income changed when the aggregate investment changed (represented by shift in the investment demand curve), the saving supply curve would also appropriately shift because people would save different amounts out of the changed income at the different rates of interest. When both the saving supply and the investment demand curves shift simultaneously, the whole position becomes indeterminate.

Keynes' criticism of the classical theory of the rate of interest may be stated in his own words: 'The independent variables of the classical theory of the rate of interest are the demand curve for capital and the influence of the rate of interest on the amount saved out of a given income; and when, (*e.g.*,) the demand curve for capital shifts, the new rate of interest, according to this theory, is given by the point of intersection between the new demand curve for capital and the curve relating the rate of interest to the amounts which will be saved out of given income. The classical theory of the rate of interest seems to suppose that if the demand curve of capital shifts or if the curve relating the rate of interest to the amounts saved out of a given income shifts or if both these curves shift, the new rate of interest will be given by the point of intersection of the

NOTES

new positions of the two curves. But this is a nonsense theory. For the assumption that income is constant is inconsistent with the assumption that these two curves can shift independently of one another. If either of them shifts, then, in general, income will change; with the result that the whole schematism based on the assumption of given income breaks down In truth, the classical theory has not been alive to the relevance of changes in the level of income or the possibility of the level of income being actually a function of the rate of the investment.'[11]

Illustrating his above views diagrammatically, Keynes further stated 'the functions used by the classical theory, namely, the response of investment and the response of amount saved out of a given income to a change in the rate of interest, do not furnish material for a theory of the rate of interest. '[12]

Third, the classical theory of rate of interest is faulty because it ignores the influence which the bank created money (credit) exerts on the rate of interest.

The classical theory is also faulty since it completely ignores the consideration of the asset demand for money and regards money as being demanded exclusively for the transactions purpose in order to remove the difficulties of barter. In short, the classical theory of the rate of interest is at best a half-baked explanation of determination of the long-run equilibrium rate of interest. It cannot be accepted as a complete explanation of the complicated process through which the interest rate is determined in an economy where, far from being neutral, money plays an important and active role in shaping the entire processes of production and distribution by acting not only as the medium of exchange but also as the store of value.

Loanable Funds Theory

The loanable funds theory of the rate of interest, also known as the neoclassical theory of rate of interest, which represents an improvement over the classical theory of rate of interest was developed, among others, by the well-known Swedish economist Knut Wicksell and the British economist Dennis Holme Robertson.[13] The theory attempted to remove the basic weakness of the classical theory of interest rate by including the role of the monetary factors in determining the rate of interest. According to the loanable funds theory, the long-run equilibrium rate of interest is determined at the point of intersection between the demand curve for and the supply curve of loanable funds.

According to this theory, the supply of loanable funds comprises the current savings, dishoarding of the existing cash balances and the newly created money. Similarly, borrowing for investment, hoarding (to accumulate cash balances) and reduction in the money supply by the banking system constituted the sources of the demand for the loanable funds in the economy. The loanable funds theorists realized that the creation and destruction of credit by the banking system should be included in the flow of money into and off the capital market. The monetary authorities by their action could augment or diminish the total flow of money into the market. Furthermore, the flow of securities into the market depended, in addition to borrowing for the current investment activities, on the behaviour or inclination of investors towards the holding of existing securities. The loanable funds theorists recognized that quite independently of the real voluntary savings and/or investment outlay, the capital market could be swamped either by a selling pressure or buying pressure as a consequence of the investors desiring to become either more liquid (hoarding) or less liquid (dishoarding).

The loanable funds theorists took cognisance of the hoarding and dishoarding on the part of wealth-owners and the credit creating and destroying activities of the monetary authorities as a result of which the total supply of loanable funds in the market could deviate (increase or decrease) from the supply of voluntary savings. Since hoarding is a source of demand for loanable funds and dishoarding (negative hoarding) is a source of supply of these funds, we can obtain *net hoarding* by adding dishoarding to hoarding. Similarly, we can obtain *net new money* by adding the credit created and credit reduced. In this way, the sum of the voluntary savings and net new money constitutes the source of supply of the loanable funds while the sum of the investment outlay and net hoarding constitutes the demand for the loanable funds respectively.

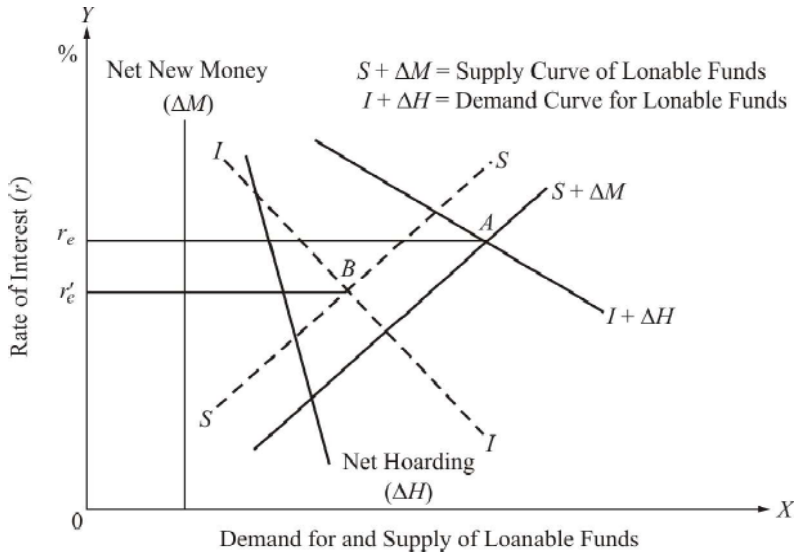


Fig. 2.2

The loanable funds theory of rate of interest can be diagrammatically explained as shown in Figure 2.2. It shows the equilibrium rate of interest $0r_e$ determined by the intersection of the loanable funds supply curve $S + \Delta M$ and the loanable funds demand curve $I + \Delta H$. The two curves intersect at point *A* corresponding to which the equilibrium rate of the interest is $0r_e$. The figure also shows that according to the classical version of the theory, the equilibrium rate of interest is $0r'_e$ determined by the intersection of the dotted saving supply and the investment demand curves. It is also obvious from the figure that at the equilibrium rate of interest $0r_e$ given by the loanable funds approach the aggregate saving exceeds the aggregate investment, i.e., $S > I$ and the amount of excess of saving over investment equals the algebraic sum of net new money (ΔM) and net hoarding (ΔH). In other words, the demand for funds to finance the investment outlay falls short of the total supply of current savings and the excess supply is absorbed in net hoarding.

Criticism

Like the classical theory, the loanable funds theory of interest rate has also been criticized on various grounds. In the first place, the theory is neither a purely flow theory nor a stock theory. It is a curious mixture of stocks and flows which hardly makes any sense. A misleading attempt has been made by the loanable funds theorists to add the stock quantities that are relevant to a *point of time* by showing these as flow schedules to the flow quantities that are relevant to a *period of time.* The successful effort to synthesize the *flows* and *stocks* was made by John R Hicks in 1937 in his general equilibrium

formulation developed in his now well-known article published in *Econometrica*[14] and which was subsequently pushed forward by Alvin H Hansen so as to become now the well-known Hicks-Hansen general equilibrium synthesis of the classical flow theory and the Keynesian stock theory of interest rate.

NOTES

Second, like the classical theory, the loanable funds theory is also indeterminate since the loanable funds' total supply schedule (since it includes voluntary savings which vary with income) varies with income. Not only does the saving portion of the supply schedule of loanable funds vary with income but even the 'new money and activated balances' which form part of the loanable funds rise and fall with the increase or decrease in the current income.[15]

Third, as is evident from Figure 2.2, according to the loanable funds theory, corresponding to the equilibrium rate of interest there is no equilibrium between current voluntary savings and the investment outlay and the disequilibrium between the two flow quantities is bridged by net hoarding. No difficulty arises so far. However, when this conclusion is presented in the form of a diagram through the aid of a set of interrelated schedules, a formidable problem of reconciling the stock and flow variables confronts us which cannot be got over without assuming a *period* too short to approach to a *point.* And the moment it is so, the distinction between a *period* and a *point* disappears and with it also disappear the concepts of *flow* and *stock.* In a way, then the entire loanable funds approach crashes like the frail house of cards and the theory is reduced to a mere pedagogic exercise with no bearing on the practical affairs.

The *fourth* criticism of the theory is that it is wrong to show the net new money and net hoarding as schedules in the diagram because (i) these changes in stocks are not functions of the rate of interest, and (ii) their interpretation in the schedules' sense clearly implies that at a certain rate of interest the monetary authorities go on adding continuously to the money supply in the economy at a constant or given rate per time period and similarly the people go on adding to their cash balances continuously at constant rate year after year. Such a conclusion hardly makes any sense without incorporating income growth in the analysis. Looking from the practical point of view, it has never been the experience of any real world economy that the central banking authority there ever indulged in such a regular credit creation or destruction nor does the behaviour of the wealth-owners support the argument.

The fact is that like those in the classical theory, in the loanable funds theory both saving and investment are purely flow concepts while hoarding and new money are, in reality, the changes in stocks. Consequently, these represent 'additions to or deductions from' the market flows within the period during which the change is taking place. But a situation in which such a change is taking place cannot be regarded as an equilibrium nor can a change in stocks from one level to a new desired level be commensurate with continuous flows. Voluntary saving and investment expenditure are continuous flows which, at any given rate of interest, will continue throughout the Marshallian 'short-run period' at least. In actual life, we may assume these rates of flows to persist over a year or more; it is otherwise with hoarding and the new money.

'It is one thing to say that during a very short interval market conditions will be affected by the creation of a certain additional sum of new money; it is clearly a very different thing to imply that at a certain rate of interest the monetary authorities will go on creating new money continuously at a constant rate throughout a period of years.'[16] The diagram gives this impression because by implication, the flow of the new money is commensurate with savings. Similarly, hoarding is commensurate with investment.

Criticizing the theory Hansen has stated that 'according to the loanable funds analysis the rate of interest is determined by the intersection of the demand schedule for loanable funds with the supply schedule. Now the supply schedule of loanable funds is composed of saving (in the Robertsonian sense) plus net additions to loanable funds from new money and the dishoarding of idle balances. But since the 'savings' portion of the schedule varies with the level of 'disposable' income, it follows that the total supply schedule of loanable funds also varies with income. Thus this theory is also indeterminate.'[17]

Liquidity Preference Theory

Keynes had criticized the classical theory and he was not satisfied with the neo-classical or loanable funds theory of rate of interest. According to him, the rate of interest was not a return on saving or waiting. 'It is the "price" which equilibrates the desire to hold wealth in the form of cash with the available quantity of cash.'[18] In other words, the rate of interest is purely a monetary phenomenon and its determination does not have anything to do with saving and investment.

In its simple form, the liquidity preference theory of interest rate states that the equilibrium rate of interest is determined at that point where the liquidity preference or the demand schedule for money intersects the supply schedule of money. The liquidity preference arises due to the transactions motive, the precautionary motive and the speculative motive. The transactions motive relates to the demand for money for transactions purpose, which in the case of business firms depends upon the size or volume of total turnover and in the case of individuals depends upon their income, the frequency with which income is received or the time interval between pay periods and the general practice of making payments. The precautionary motive gives rise to the demand for money for facing unforeseen contingencies such as sudden sickness, arrival of guests, accidental loss of life and property, etc.

Keynes lumps together these two demands and relates these to the level of income. The transactions demand and the precautionary demand for money are interest-inelastic. The speculative motive which is defined as 'the object of securing profit from knowing better than the market what the future will bring forth'[19] gives rise to the speculative demand for money which is interest-elastic with the elasticity increasing as the rate of interest falls until at some low enough rate of interest (around two per cent) the demand becomes perfectly interest-elastic. In other words, while the transactions and precautionary demand for money (M_1) are a positive function of the level of income (Y), the speculative demand for money (M_2) is a negative function of the rate of interest (r). Thus the total demand for money (M_d)is composed of M_1 which is a positive function of income and M_2 which is a negative function of the rate of interest. Thus,

$$M_d = M_1 = L_1(Y) + M_2 = L_2(r)$$

The supply of money in the economy is autonomously determined by the monetary authority. It comprises the currency and commercial banks' demand deposits, i.e., $M_s = C + D$. It is taken as fixed and is nonresponsive to changes in the interest rate.

Consequently, the money supply curve $M\overline{M_S}$ is a vertical straight line as shown in Fig. 2.3. The equilibrium rate of interest is determined corresponding to the point of intersection of the negatively sloping demand curve for money $M_d = M_1 + M_2 = L_1(Y) + L_2(r)$ and the vertical supply curve of money $M\overline{M_S}$ as has been shown in Figure 2.3.

Figure 2.3 shows the determination of the equilibrium rate of interest according to the liquidity preference theory of interest rate. In the figure, the demand curve for money $M_d = M_1 + M_2 = L_1(Y) + L_2(r)$ is negatively sloping, showing that the liquidity preference or the demand for money is a negative function of the rate of interest. The interest-elasticity of the curve increases as the rate of interest falls until at some critically low enough rate of interest (in the figure it is $0r_c$) it becomes perfectly horizontal, showing that the demand for money has become perfectly interest-elastic. In the figure at point *C* and beyond, the demand curve for money is perfectly interest-elastic. This situation in literature is known as the *liquidity trap*[20] which has given rise to fierce controversy among the economists.

In Figure 2.3, $M\overline{M}_S$ is the autonomously determined supply curve of money. Its interest elasticity is zero.[21] The two curves intersect at point *A* corresponding to which the rate of interest is $0r_e$. This rate of interest is the equilibrium rate of interest because the total amount of money demanded by the public at this rate of interest is just equal to the total supply of money made available by the monetary authority to satisfy the demand for money. At any rate of interest other than this particular rate, there will be disequilibrium between the demand for and the supply of money—either the total demand for money will exceed or fall short of the total supply of money available in the system. It is clear that the liquidity preference function remaining unchanged, any increase (decrease) in the money supply represented by the rightward (leftward) shift in the money supply function will cause a fall (rise) in the equilibrium rate of interest unless the equilibrium rate of interest has already touched the *liquidity trap* rate of interest. In Figure 2.3, when the money supply function shifts rightward from $M\overline{M}_S$, to $M'\overline{M}'_s$ the equilibrium rate of interest falls from $0r_e$ to $0r_e'$ showing that an expansionary monetary policy is effective in raising the level of aggregate employment and output in the economy by causing an increase in the aggregate investment depending upon the interest elasticity of the investment demand function.

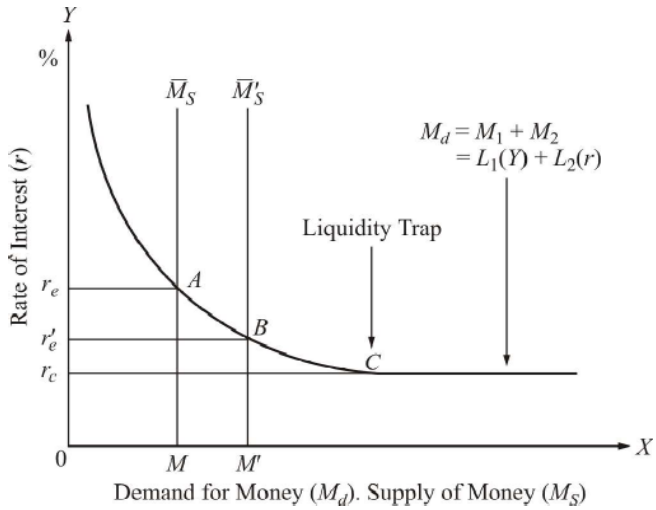


Fig. 2.3

Criticism

The liquidity preference theory of interest rate has not escaped criticisms. *First*, being completely a stock theory, the rate of interest in this theory is as much indeterminate as it was in the classical theory. According to Alvin H Hansen, 'the Keynesian theory, like the classical, is indeterminate. In the Keynesian case the money supply and demand

NOTES

schedules cannot give the rate of interest unless we already know the income level; in the classical case the demand and supply schedules for saving offer no solution until income is known. Keynes' criticism of the classical theory applies to his own theory.'[22]

Second, the theory cannot explain the determination of the rate of interest in the long period because it focusses attention only on those factors which are relevant only in the short period.

Third, the theory cannot explain the coexistence of the different interest rates on the basis of liquidity preference because interest rates will have to be perfectly uniform due to the perfect uniformity of money or cash balances.

The *Fourth* criticism of this theory is that it is wrong to say that rate of interest is not the reward for 'saving or waiting as such.' Keynes blissfully forgets that without saving or waiting, investment funds cannot be obtained. Jacob Viner stated the correct position when he asserted that 'without saving there cannot be liquidity to surrender... the rate of interest is the return for saving without liquidity.'

The *Fifth* criticism of the theory is that although in the short period at any given point of time the rate of interest will be such that the community's total holding of cash must be equal to the total existing stock of money but it is equally true that over a long period the rate of interest must tend to that level where the saving and investment flows are in equilibrium.

Last, Keynes' basic proposition in his theory is that the rate of interest and the demand for money, more particularly the speculative demand for money, are inversely related. According to Don Patinkin, 'Keynes'analysis of the implication of this dependence is repeatedly marred by a confusion (which characterizes the later Keynesian literature as well) between his basic proposition that the amount of money demanded is inversely dependent upon the rate of interest and the completely different proposition that the equilibrium rate of interest is inversely dependent on the amount of money. His discussion of the liquidity preference in *The General Theory* shifts uninhibitedly from one proposition to the other with never an indication that they are in any way not identical. More specifically, there is never a recognition that, in our terminology, the first of these propositions describes an individual experiment and the second a market experiment, and that the truth of the first does not imply the truth of the second.'[23]

Modern Theory

The modern or the neo-Keynesian theory of interest rate has been developed by John R Hicks by synthesizing the classical and the Keynesian theories. Its merit lies in successfully integrating together saving, investment, liquidity preference and the money supply. The theory shows that productivity (investment), thrift (saving), liquidity preference (demand for money) and the money supply are all needed in order to formulate a comprehensive and determinate interest rate theory.

In the modern theory, also known as the Hicks-Hansen synthesis of the classical and Keynesian theories, from the classical version has been derived the *IS* curve[24] which is the locus of those combinations of interest rate and income at which investment and saving are equal and consequently the real sector of the economy is in equilibrium. From the liquidity preference approach has been derived the *LM* curve[25] which is the locus of all those combinations of interest rate and income corresponding to which the demand for money equals the supply of money and consequently the monetary sector of the economy is in equilibrium. By successfully employing the tools of negatively sloping *IS* curve and the positively sloping *LM* curve, the modern theory of interest rate determines

simultaneously the equilibrium rate of interest (r) and income (Y) such that both the real and monetary sectors of the economy are simultaneously in equilibrium.

NOTES

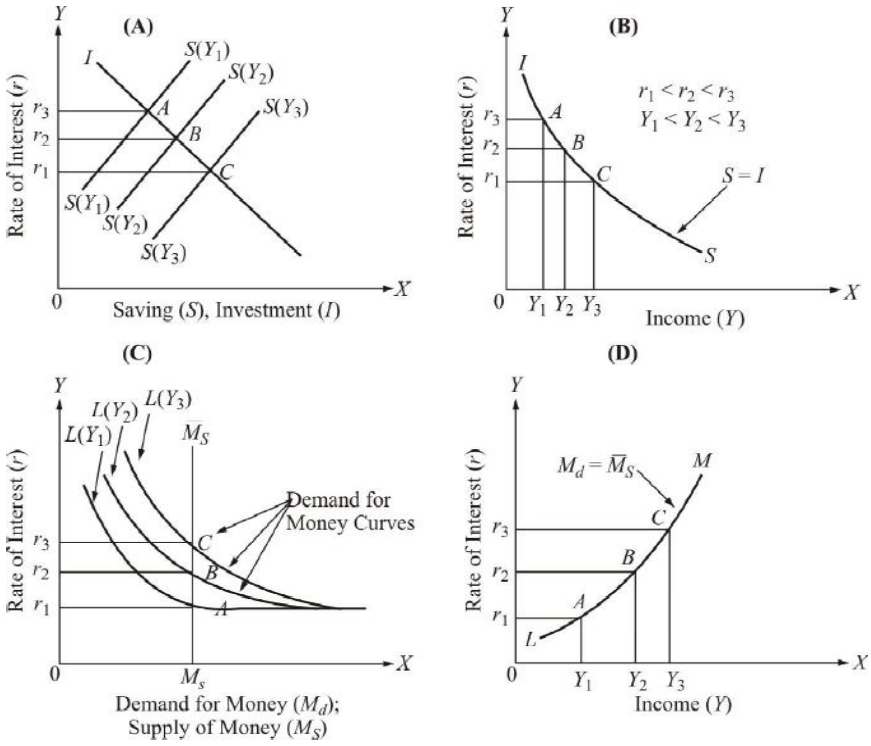


Fig. 2.4

In the theory, following the classical approach, investment has been treated as a negative function of the rate of interest while saving has been regarded as a positive function of the interest rate. Similarly, following the Keynesian approach, the liquidity preference or the demand for money has been treated as a function of the level of income and the rate of interest[26] while the supply of money has been treated as being autonomously determined by the monetary authority and it is treated as autonomously given.Thus,

$$I = f(r)$$

$$S = g(r)$$

$$M_d = h(r, Y)$$

$$M_S = \overline{M}_S$$

The equilibrium conditions are:

$$S = I$$

$$M_d = \overline{M}_S$$

The set of inter-relationships involved in the above equations are combined into a single composite diagram so as to derive the *IS* and *LM* curves as shown in Fig. 2.4.

Figure 2.4 has four parts. Parts *A* and *B* of the figure show the classical approach in which the rate of interest is indeterminate because mere saving and investment flows

cannot determine it without the level of income being known. As the *IS* curve in Part *B* shows, every pair of interest rate and income along the *IS* curve is an equilibrium combination. Parts *C* and *D* of the figure show the indeterminate character of the naive Keynesian or liquidity preference approach. The *LM* curve in part *D* does not alone determine the rate of interest because all along this curve we have those several combinations of income and interest rate corresponding to which the demand for and the supply of money in the economy are in equilibrium.

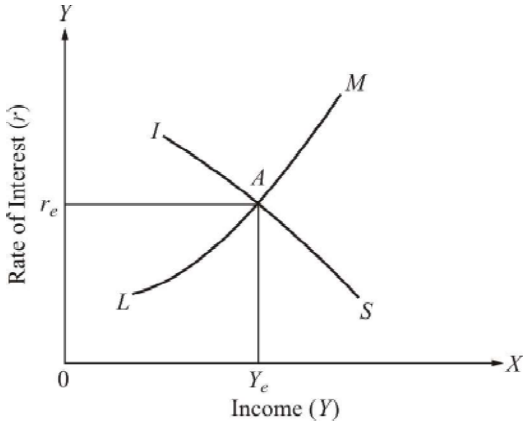


Fig. 2.5

In order to obtain the determinate solution, the classical and the Keynesian theories should be integrated by combining the *IS* curve of Part *B* and the *LM* curve of Part *D* in Figure 2.4 as shown in Figure 2.5 which shows that the equilibrium rate of interest determined by the intersection of the *LM* and *IS* curves is $0r_e$ corresponding to which the aggregate equilibrium income is $0Y_e$.

The chief merit of the modern theory of rate of interest is that it is free from all the criticisms which were valid in the case of the classical and the liquidity preference theories. It is also free from the crudities of the loanable funds theory in which the stock and flow quantities were unscientifically intermingled. The theory shows that the rate of interest is determined both by the stocks and flows and their mutual interaction in a system of general equilibrium. The equilibrium rate of interest $0r_e$ is a fully determinate rate because corresponding to it the double condition of equilibrium involving the simultaneous equality between investment, saving, demand for money and supply of money is satisfied, i.e., $I = S$ and $M_d = \overline{Ms}$. Consequently, the rate of interest so established has no scope for deviation. Corresponding to the $0r_e$ rate of interest, the economy's real and the monetary sectors are simultaneously in equilibrium. At any rate of interest other than this particular rate, economy's one or the other sector will be in disequilibrium. Consequently, that rate of interest will not be a stable equilibrium interest rate.

Check Your Progress

1. Why is the classical theory of rate of interest considered faulty?
2. What is the other name for loanable funds theory?
3. Who developed the neo-Keynesian theory of interest?

2.3 IS–LM MODEL

The IS–LM model (Investment Saving–Liquidity Preference Money Supply) is a macroeconomic tool that demonstrates the relationship between interest rates and real output, in the goods and services market and the money market. The intersection of the IS and LM curves is the 'general equilibrium' where there is simultaneous equilibrium in

both markets. This modern or the neo-Keynesian theory of interest rate was developed by John R. Hicks by synthesizing the Classical and the Keynesian liquidity preference theory. Hicks-Hansen's, IS-LM curves model seeks to explain a case of joint determination of equilibrium rate of interest and equilibrium level of income. IS curve is that curve which shows equilibrium in the commodity market corresponding to different pairs of level of income and rate of interest. The LM curve is obtained by determining equilibrium in the money market in terms of equilibrium between demand for money and supply of money corresponding to different pairs of interest rate and the level of income. Since IS curve and LM curve indicate equilibrium in the commodity market and equilibrium in the money market, respectively, the intersection of IS curve and LM curve shows the simultaneous equilibrium in both the commodity market and money market with equilibrium rate of interest and equilibrium level of national income.

2.3.1 IS Curve or Investment Curve

'IS curve shows the combination of level of output and interest rate in such a manner so that the desired expenditure equals to income.'

There is an inverse relationship between rate of interest and the investment expenditure. Therefore, slope of investment curve is negative. The higher the rate of interest the lower will be the investment expenditure and vice-versa.

The position of investment demand and schedule depends on the slope and the level of autonomous investment expenditure.

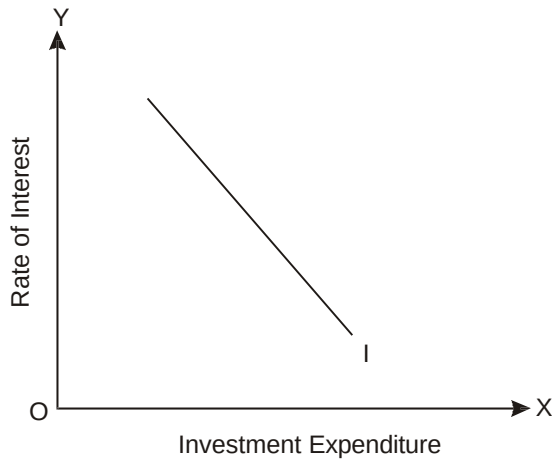


Fig. 2.6 *Investment Curve*

Any change in the autonomous investment shifts the investment schedule rightwards or leftwards. The increase in the investment expenditure shifts the schedule rightwards (see Figure 2.7A).

It means firms are planning to spend more on different levels of income whereas any decrease in the investment expenditure schedule will shift the schedule leftwards.

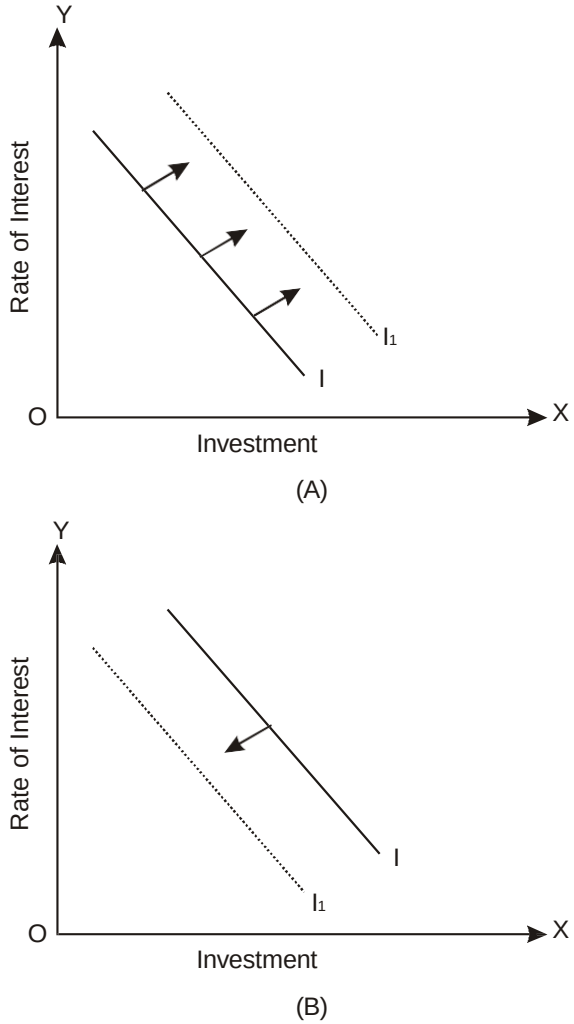


Fig. 2.7 *Changes in Investments*

As you know the IS curve shows the equilibrium in the goods market whereas the LM curve shows the equilibrium in the money market. Thus, the equilibrium in both goods and money market meets at that level where IS and LM curve interest each other.

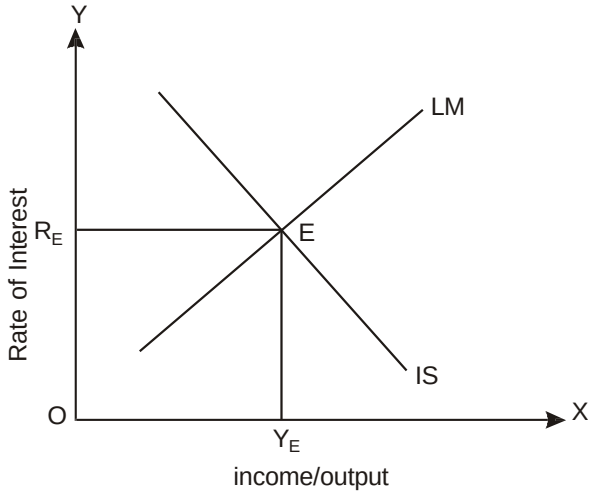


Fig. 2.8 *Equilibrium between Goods and Money Market*

According to Figure 2.8, the equilibrium between goods and money market is attained at

E point where LM = IS and at this point the equilibrium rate of interest is R_E and equilibrium income and output is Y_E.

Change in Equilibrium

The equilibrium level may change with change or shift in IS and LM curves. For example, with the increase in investment IS curve will shift to the right and LM curve may shift to the right with the increase in supply of money.

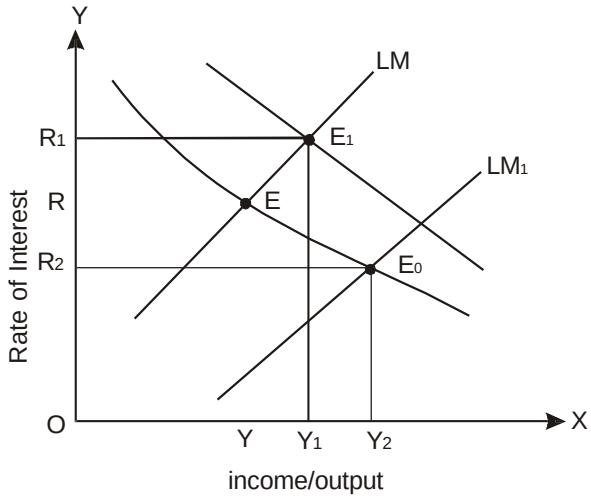


Fig. 2.9 *Change in Equilibrium*

According to Figure 2.9, due to increase in investment the IS curve shifts rightward to IS_1 as a result the level of interest shifts to R_1 and level of income shifts to Y_1, on the other hand due to decrease in money supply the LM curve shifts rightwards to LM_1 and new equilibrium shifts to the EO at which rate of interest falls to R_2 and income increase to Y_2.

Derivation of IS Curve

There are different ways of deriving the *IS* curve. We will use a shortcut method which is more intuitive than explanatory. This method derives directly from the theory of income determination, by assuming implicitly the effect of change in interest rate on the aggregate demand.

An open economy is in equilibrium where aggregate demand (AD) is equal to the aggregate supply (AS). That is, an economy is in equilibrium where

$$AD = AS$$

We know that at product market equilibrium, $AD = C + I + G + NX$, where $NX = X - M$. By substitution, the equilibrium condition may be expressed as:

$$AD = C + I + G + (NX) \quad ...(i)$$

Recall that

$C = a + b\,(Y - T)$

$I = -hi$ (where i is the interest rate)

$G =$ (Constant)

$T = t\,Y$ (t tax rate)

$NX =$ Net export $= X - M$ (constant)

By substituting these values in Equation (i), the product market equilibrium condition can be expressed as:

$$AD = a + b[Y - tY] + \quad ...(ii)$$
$$= a + bY - btY +$$

In order to shorten the Equation (ii), let us sum up the constant and denote the sum as *A*. Thus,

$$A = a +$$

By substituting *A* for constants in (ii), we get

$$AD = A + bY - btY - hi$$

or
$$AD = A + b(1 - t)Y + hi \quad ...(iii)$$

Having derived the aggregate demand curve, we can now use this equation to show the determination of the equilibrium of product market. In Figure 2.10(a), the AD given in Equation (iii), is shown by the curve marked as $AD_1 = A + b(1 - t)Y + hi_2$ where i_2 denotes the rate of interest. Given the rate of interest, as shown in panel (*b*) of Figure 2.10, the equilibrium level of income is determined at point E_1 determining the equilibrium level of income at Y_1.

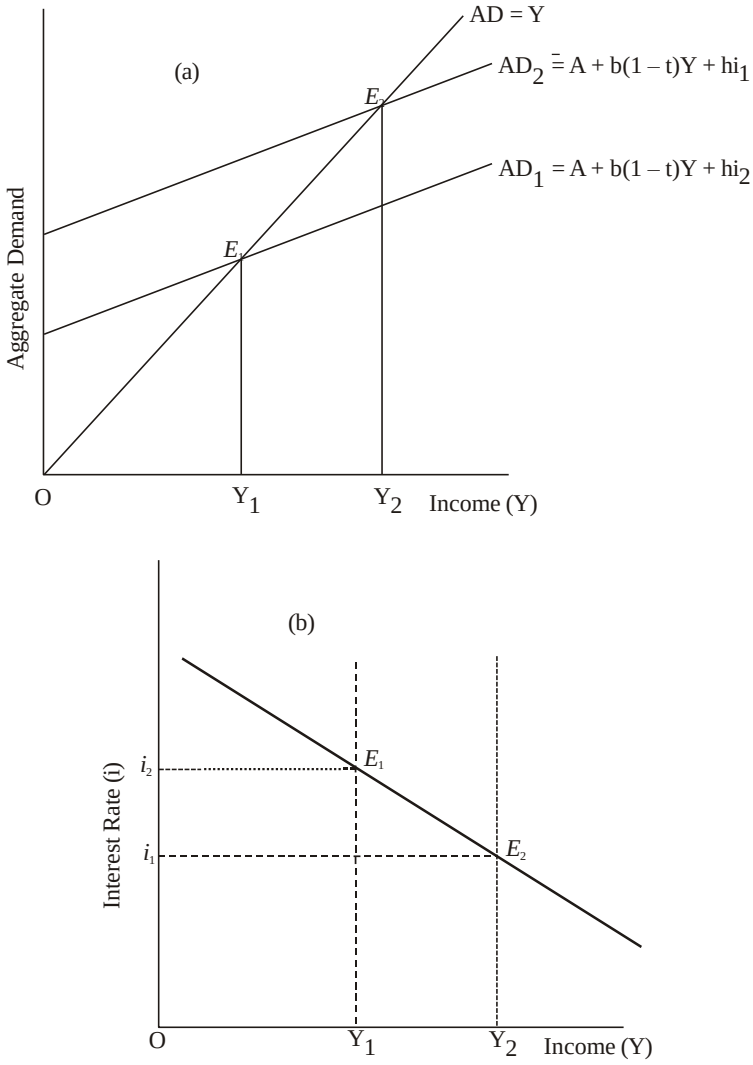


Fig. 2.10 *Derivation of the IS Curve*

NOTES

Note that at equilibrium income *Y*1, AD = AS and *I* = *S*, all other factors given. This point marks one point on the *IS* curve, at interest rate i_2. Now let the interest rate fall to *i*1. Recall the investment function,

$I =$ (where $h = \Delta I/\Delta i$)

Given the investment function, investment will increase with decrease in the interest rate from i_2 to i_1. With increase in investment (*I*), aggregate demand (AD) increases from AD_1 to a higher level, say, AD_2 all other factors remaining constant. Consequently, the aggregate demand curve AD_1 shifts upward to the position of AD_2. As a result, the equilibrium point shifts from E_1 to E_2 and income level increases to Y_2. Here again $S = I$.

The main conclusion that emerges from these changes is that when interest rate decreases, both savings and investment increase causing increase in income and $I = S$ at all the levels of income. This conclusion is plotted in panel (b) of Figure 2.10. Panel (b) shows the relationship between the interest rate and income with $I = S$. At interest rate i_2, given the savings and investment, equilibrium level of income is *Y*1. This is shown by point E_1 where $I = S$. When the interest rate falls from i_2 to i_1, investment increases causing increase in income from Y_1 to Y_2. When interest rate i_1 and income Y_2 are linked, it gives point E_2. Here again I = *S*. By joining points E_1 and E_2, we get the *IS* curve. This is how *IS* curve is derived for an open economy model. It is important to remember that all along the IS curve savings are equal to investment and the product market is in equilibrium.

2.3.2 Derivation of LM Curve

The *LM* curve is derived from the Keynesian theory of interest rate determination by linking the interest rate to income level. Therefore, before we proceed to show the derivation of the *LM* curve, let us have a glance at the Keynesian theory of money market equilibrium. According to Keynes, the equilibrium rate of interest is determined where

$$L = M \quad \text{...(iv)}$$

where *L* = liquidity preference, *Ms* = supply of money (prices remaining constant).

However, in modern interpretation of the Keynesian theory, *M* is taken to be the real balance of money expressed as *M*/*P* (where *P* = price level). Given this modification, Equation (iv) can be written as:

$$L = M/P \quad \text{...(v)}$$

The terms *L* and *M*/*P* in Equation (ii) need some elaboration. According to Keynes, the liquidity preference (*L*), the demand for money, has two components: (i) transaction demand for money (including precautionary motive), and (ii) speculative demand for money, i.e., money demanded for speculative purpose. The transaction demand for money (M_t) is the function of income: $M_t = f(Y)$. The transaction demand for money (M_t) is positively related to income, i.e., *Mt* increases with increase in income. The speculative demand for money (M_s) is the function of interest rate: $M_s = F(i)$ and *Ms* is negatively related to the interest rate, i.e., speculative demand for money decreases with increase in the interest rate. However, the aggregate demand for money ($M_t + M_s$) is negatively relative to the interest rate, according to the Keynesian theory.

Given these points of elaboration, the Keynesian aggregate demand for money can be expressed as:

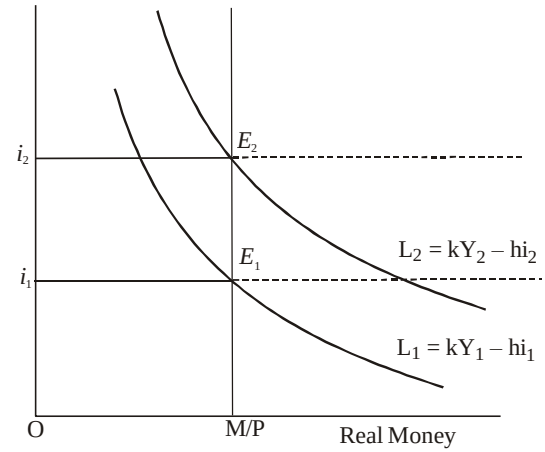
$$L = M_t + M_s = f(Y, i) \quad \text{...(vi)}$$

As regards the supply of money (M/P), according to Keynes, the total supply of money is determined by the central bank of the country (e.g., RBI in India) and it remains constant over a period of time.

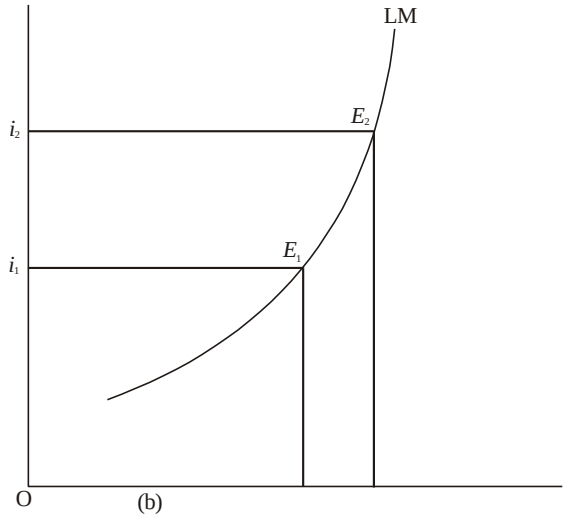
Having explained the Keynesian postulates regarding the demand for and supply of money, we can now proceed to derive the *LM* curve. Figure 2.11 presents both the Keynesian theory of money market equilibrium and the interest rate determination and also the derivation of the *LM* curve.

Panel (a) of Figure 2.11 shows the determination of money market equilibrium and interest rate. The real money balance is shown by the straight vertical line marked M/P, i.e., total money supply remains constant irrespective of the interest rate. The aggregate demand for money has been shown by the curves marked L_1 and L_2 which are interest elastic.

To show the derivation of the *LM* curve, let us suppose that at a given level of income, Y_1, the aggregate demand for money is shown by the curve $L_1 = Y + h_i$. The money demand curve L_1 intersects the M_s line at point E_1 determining the money market equilibrium, and equilibrium interest rate is determined at i_1. Note that at income level Y_1, the interest rate is i_1. This relationship between Y_1 and i_1 is shown by point E_1 in panel (b) of the figure.



(a)



(b)

NOTES

Now let the income increase from Y_1 to Y_2, interest rate remaining the same. Since M is positively related to income, demand for money demand curve L shift upwards to $L_2 = kY + h_i$. With the increase in money demand, money supply remaining the same, the equilibrium point shifts upward from E_1 to E_2 showing a rise in the rate of interest from i_1 to i_2. Note that with increase in money demand, money market attains its equilibrium at a higher level of income, Y_2, and at a higher rate of interest, i_2. Again, point E_2 shows a relationship between the level of income and the interest rate. This relationship is shown by point E_2 in panel (b) of Figure 2.11. The curve drawn through point E_1 and E_2 in panel (b) represents the *LM* curve. This is how *LM* curve is derived. It is important to remember that *all along the LM curve, money market is in equilibrium, i.e., L = M/P, at different levels of interest rate and income.*

Synthesis

Recall that the main objective of this section is to present a synthesis of the monetary and the real sectors as developed by Hicks. Having derived the *IS* and *LM* curves, we can now show the synthesis of the two sectors. The synthesis of the monetary and real sectors can be shown by combining the *IS* and *LM* curves. This task is accomplished in Figure 2.12.

In Figure 2.12, the *IS* curve represents the product sector. All the points on the *IS* curve show the equilibrium status of the product/real sector at different levels of interest rate. Similarly, the *LM* curve represents the money sector and all the points on the *LM* curve represent the equilibrium of the money sector.

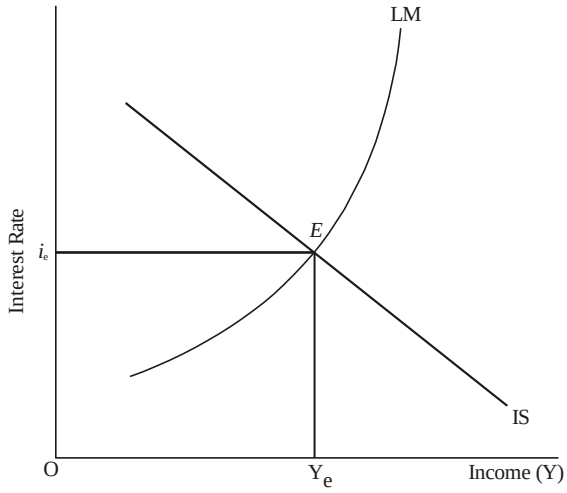


Fig. 2.12 *Synthesis of Monetary and Real Sectors*

Check Your Progress

4. How is the LM curve obtained?
5. What is general equilibrium?

As Figure 2.12 shows, the *IS* and *LM* curve intersect at point *E*. Note that at point *E* being on the *IS* curve establishes that $I = S$, i.e., savings are equal to investment, at interest rate i_e and income level Y_e. There is no excess saving or investment in excess of savings. This means that the product market is in equilibrium, all other factors given. Similarly, point *E* on the *LM* curve signifies that $L = M$, i.e., demand for money is equal to the supply of money at the interest rate *ie* and there is neither excess money demand nor excess money supply. This means that point E marks the simultaneous equilibrium of both the monetary and product markets at the same level of the interest rate (i_e) and at the same level of national income (Y_e). This is how Hicks has synthesized the monetary and the product sector. This is considered to be a great contribution by an all-time great economist, J.R. Hicks.

NOTES

2.4 RELATIVE EFFECTIVENESS OF MONETARY AND FISCAL POLICIES

In this section, you will learn about the effectiveness of monetary and fiscal policies.

2.4.1 Monetary Policy and its Effectiveness

Different monetary economists have defined monetary policy in their own ways. For example, Harry Johnson defines it as a 'policy employing central bank's control of the supply of money as an instrument of achieving the objectives of general economic policy'. Shaw defines monetary policy as 'any conscious action undertaken by the monetary authorities to change the quantity, availability or cost… of money'. Note that Johnson emphasizes only the control of money supply as the objective of monetary policy, Shaw emphasizes both supply of and demand for money. *Monetary policy is essentially a programme of action undertaken by the monetary authorities, generally the central bank, to control and regulate the demand for and supply of money with the public and the flow of credit with a view to achieving predetermined macroeconomic goals*. The objectives of monetary policy are the same as the objectives of fiscal policy, viz., *growth, employment, stability of price and also foreign exchange and balance-of-payment equilibrium*.

Scope of Monetary Policy The scope of monetary policy spans the area of economic transactions and the macroeconomic variables that monetary authorities can influence and alter through the monetary policy. From monetary instruments point of view, the scope of monetary policy includes (*i*) bank rate, (*ii*) cash reserve ratio, (*iii*) open market operations by the central bank and (*iv*) also the direct control measures as deemed fit by the central bank. From the view point of effectiveness, the scope of monetary policy depends, by and large, on two factors.

(i) The level of monetized economy and

(ii) The level of development of the capital market.

In a fully *monetized economy*, the scope of monetary policy encompasses the entire gamut of economic activities. For, in such an economy, all economic trans-actions are carried out with money as a medium of exchange. In that case, monetary policy works by changing the general price level. It is, therefore capable of affecting all economic activities—production, consumption, savings, investment and foreign trade. The monetary policy can influence all major macroeconomic variables—*GDP*, savings and investment, employment, the general price level and the foreign exchange.

The other contributory factor is the level of capital market development. While the change in the supply of money affects the level of economic activities through the price level, the other instruments of monetary control (bank rate and cash reserve ratio) work through the capital market. Where the capital market is fairly developed, monetary policy affects the level of economic activities through the changes in the capital market. It works faster and more effectively. Incidentally, a developed capital market is one which has the following features: (*i*) a large number of financially strong commercial banks, financial institutions, credit organizations and short-term bill market, (*ii*) a major part of financial transactions are routed through the capital markets, (*iii*) the working of the various capital sub-markets is inter-linked and inter-dependent and (*iv*) the commodity sector is highly sensitive to the changes in the capital market. It is important to note that the changes in the bank rate and cash reserve ratio work through the commercial banks.

NOTES

Therefore, for the monetary policy to have a widespread impact on the economy, it is necessary that the capital sub-markets have strong financial links with the commercial banks.

Instruments of Monetary Policy

The instruments of monetary policy refer to the monetary variables that the central bank can change at its discretion with a view to controlling and regulating the money supply and the availability of credit. The instruments are also called 'weapons of monetary control'. Samuelson and Nordhaus call these factors as "The Nuts and Bolts of Monetary Policy". The measures of monetary policy are generally classified under two categories:

(i) quantitative measures and

(ii) selective credit controls.

Quantitative Measures of Monetary Control

The *quantitative measures* or the *traditional measures* of monetary control are following:

(i) Open Market Operations,

(ii) Discount Rate or Bank rate and

(iii) Cash Reserve Ratio (CRR).

In addition to these measures, in a country like India with all major banks nationalized, the central bank uses direct measures to control supply of and demand for money. For example, RBI enhanced interest rates for the banks straightway to control inflation in April 2007.

Let us now briefly discuss the meaning and working of these measures.

(i) *Open Market Operations (OMO)*

The 'open market operation' comprises sale and purchases of government securities and treasury bills by the central bank of the country. When the central bank decides to increase the supply of money with the public, it purchases the government securities, i.e., bills and bonds, and when it decides to reduce money in circulation, it sells the government bonds and securities. The open market operation is the most powerful and widely used tool of monetary control.

Let us look into the working of open market operations. The central bank carries out its open market operations through the commercial banks—it does not deal directly with the public. The buyers of the government bonds include commercial banks, financial corporations, big business corporations and individuals with high savings. These customers of government bonds hold their accounts with the banks. When they buy government bonds, money is transferred from their account to RBI account. Therefore, when the central bank carries out its open market operations, it affects bank deposits and reserves and thereby, their capacity to create credit. For instance, suppose the central bank decides to reduce money supply with the public and the availability of credit with the objective of preventing inflation. To this end, the central bank will offer the government bonds and treasury bills for sale through the commercial banks. The task becomes easier when the government owns the commercial bank as is the case in India. The sale of government bonds and securities affects both the *supply of and demand for credit.*

As regards the supply of credit, it is affected adversely in the following ways.

(i) The people buy the government bonds and securities generally through cheques drawn on the commercial banks in favour of the central bank. In this process of sale of bonds, the money is transferred form the buyers' account to the central bank account. This reduces the total deposits with the commercial banks and also their cash reserves. As a result, credit creation capacity of commercial banks decreases. Consequently, the flow of credit from the commercial bank to the society decreases.

(ii) When the commercial banks decide to buy the government bonds and securities themselves, their cash reserves go down. This further reduces their credit creation capacity. The ultimate result is a fall in the flow of credit to the public.

As regards the *demand for credit*, when the central bank sells government bonds, their prices go down and, therefore, the rate of interest goes up. This causes an upward push in the interest rate structure. The rise in the rate of interest reduces the demand for credit. Thus, not only the supply of credit but also the demand for credit is affected by the open market operations.

On the contrary, when the central bank decides to increase money supply, it buys back government bonds and securities. Then the money flows out from the central bank account to the people's accounts with the commercial banks. As a result, deposits with the commercial banks and their cash reserves increase. This enhances their capacity to create credit. Other things given, the flow of money from the banks to the public increases.

Effectiveness of Open Market Operations The effectiveness of open market operation as a weapon of monetary control depends on the following factors.

(i) When commercial banks possess excess liquidity, the open market operation does not work effectively.

(ii) In a very buoyant market situation, the effective control of demand for credit through the open market operation is doubtful. And, during the period of depression, open market operations are not effective for lack of demand for credit.

(iii) In underdeveloped countries in which the banking system is not mature and security and capital markets are not interdependent, open market operations have a limited effectiveness.

(iv) The popularity of government bonds and securities with the public also matters a lot. The government debt instruments are generally not popular due to low rate of return. The central bank then has to use coercive measures and force the commercial banks to buy the government bonds, as is the case in India.

(ii) *Discount Rate or Bank Rate Policy*

Discount rate or Bank rate is the rate at which the central bank rediscounts the bills of exchange presented by the commercial banks. The RBI Act, 1935 defines 'bank rate' as the *"standard rate at which (the bank) is prepared to buy or rediscount bills of exchange or other commercial papers eligible for purchase under this Act"*. It rediscounts only approved bills and the 'first class bills of exchange'. Why do commercial banks get their bills of exchange rediscounted? What happens, in fact, when commercial banks are faced with a shortage of cash reserves, they approach the central bank to get

NOTES

their bills of exchange rediscounted. It is a common method of borrowing by the commercial banks from the central bank. The central bank rediscounts the bills presented by the commercial banks because it is a part of its functions—it is the *lender of the last resort.* For rediscounting the bills of exchange, the central bank charges a rate. This rate is traditionally called *bank rate.* A more appropriate name in usage now is the *discount rate*. However, *for all practical purposes, bank rate is the rate which the central bank charges on the loans and advances to the commercial banks.*

The central bank can change this rate—increase or decrease—depending on whether it wants to expand or reduce the flow of credit from the commercial bank. When it wants to increase the credit creation capacity of the commercial banks, it reduces the discount rate and *vice versa*. This action by the central bank is called the bank rate policy or more appropriately, the *discount rate* policy. The *bank rate policy* was first adopted by the Bank of England in 1839. It was the only and the most widely used weapon of credit control until the open market operation, first used in 1922, emerged as a more powerful instrument of monetary control. It still works as the main indicator of the central bank's monetary policy.

The working of the discount rate policy is simple. When the central bank changes its discount rate, commercial banks change their own discount rate. That is, when the central bank raises its discount rate, commercial banks raise their discount rates too. Generally, the central bank rate is 1 percentage point higher than the discount rate charged by the commercial banks. In order to look at how a change in discount rate affects the availability of credit, let us suppose that the central bank wants, as a matter of policy, to control the flow of bank credit. To achieve this objective, it will raise the discount rate. This action of the central bank reduces the flow of credit in three ways.

One, a rise in the discount rate (virtually the interest rate) reduces the net worth of the government bonds (the treasury bills and promissory notes) against which commercial banks borrow funds from the central bank. This reduces the commercial banks' capacity to borrow from the central bank. As a result, commercial banks find it difficult to maintain a high cash reserve. This reduces the credit creation capacity of the commercial banks. So the flow of credit to the market is reduced.

Two, when the central bank raises its discount rate, commercial banks raise their discount rate. This raises the *cost of credit* which discourages the business sector to get their bills of exchange discounted. Besides, a rise in the bank rate pushes the interest rate structure up. The rise in the interest rate reduces the demand for funds too. Such a policy is, therefore, called a *'dear money policy'*. A reverse process is used when central bank adopts a *cheap money policy*.

Three, bankers' lending rate is quickly adjusted to deposit rates. Therefore, a rise in the bank rate causes a rise in the deposit rate. This turns borrowers into depositors. Therefore, savings flow to the banks in the form of deposits. This is called *the deposit mobilization effect.*

Limitations of Discount Rate Policy The discount rate policy has lost its effectiveness as a weapon of monetary control over time for the following reasons.

1. The variation in the discount rate works effectively only when commercial banks approach the central bank for borrowing. In modern times, the commercial banks have built their financial resources. They are not dependent on the central bank for financial support. Therefore, their discount rate is not affected when central bank raises the bank rate.

2. With the growth of credit institutions and financial intermediaries, the capital market has widened extensively. The share of banking credit has declined. Therefore, variations made by the central bank in the discount rate, especially when it raises the rate, have only a limited impact on the credit market—it is limited to only bank credit.
3. Looking from the credit demand angle, variations in the discount rate become effective only where demand for credit is interest-elastic. The structure of the credit market in the less developed countries is such that the interest rates are sticky. Hence change in the discount rate has not been found to be very effective.

(iii) *The Cash Reserve Ratio or Statutory Reserve Ratio*

The cash reserve ratio (*CRR*) is the percentage of total deposits which commercial banks are required to maintain in the form of cash reserve with the central bank. The objective of the cash reserve ratio is to prevent shortage of cash in meeting the demand for cash by the depositors. The cash reserve ratio depends, normally, on the banks' experience regarding the cash demand by the depositors. But, "If there were no government rules, banks would probably keep only a very small fraction of their deposits in the form of reserves". Since cash reserve is *non-interest bearing*, commercial banks often keep their cash reserves below the safe limits. This situation might lead to financial crisis in the banking sector. So, in order to prevent this eventuality, the central bank imposes a *CRR* on the banks. This has become a handy tool for the central bank to control money supply. The central bank enjoys the legal powers to change the cash reserve ratio of the banks at its own discretion. The cash reserve ratio is a *legal requirement*. Therefore, it is also called *statutory reserve ratio (SRR)*.

By changing the *CRR*, the central bank can change the money supply overnight. When economic conditions demand a contractionary monetary policy, the central bank raises the *CRR*, and when economic conditions demand monetary expansion, the central bank cuts down the *CRR*. The effect of change in the *CRR* on the supply of money and credit can now be briefly explained. Suppose commercial banks possess a total deposit of Rs.100 million and *CRR* is 20 per cent. It means (*a*) that the banks can loan ` 80 million and (*b*) that the credit or deposit multiplier equals five. It also means that the banks can create, through the process of credit multiplier, a total credit of ` 500 million or an additional credit of ` 80 million × 5 = ` 400 million.

Now let the central bank decide to reduce the money supply with the public and raise the *CRR* to 25 per cent. Then the credit multiplier will go down to 4. With this provision, the banks can provide loan only to the extent of ` 75 million (= ` 100 million – ` 25 million). Thus, the total credit created by the banks goes down to ` 100 × 4 = ` 400 million and additional credit goes down to ` 75 × 4 = ` 300 million. A fall of ` 100 million in the bank credit is supposed to have considerable impact on the money market. The effect will be reversed when the central bank cuts back the *CRR* to 20 per cent.

Limitations of CRR as an Instrument of Monetary Control This method alone is effective where other measures fail. It proves more handy where open market operation and bank rate policy prove less effective. However, its effectiveness in terms of impact on the capital market depends on the share of the banking credit in the credit market. It is relatively more effective in the advanced countries with advanced banking system accounting for a major share in the capital market.

Selective Credit Controls

The *quantitative methods* of monetary control affect, (when they are effective), the entire credit market in the same direction. In other words, their impact on all the sectors of the economy is uniform. This may not be always desirable or intended by the policy-makers. The monetary authorities are often faced with the problems of (*a*) rationing the credit, (*b*) diverting the flow of credit from the non-priority sectors to the priority sectors and (*c*) curbing speculating tendency based on the availability of bank credit. These objectives of credit control are not well served by the quantitative measures of credit control. The monetary authorities then resort to *selective credit controls*.

Selective credit controls include the following credit control measures.

(i) ***Credit rationing.*** When there is shortage of institutional credit available for the business sector, the large and financially strong sectors or industries tend to capture the lion's share in the total institutional credit. As a result, the priority sectors and weaker but essential industries are starved of necessary funds, mainly because bank credit goes to the non-priority sectors. In order to curb this tendency, the central bank resorts to credit rationing measures. Generally, three measures are adopted: (*a*) imposition of upper limits on the credit available to large industries and firms, (*b*) charging a higher or progressive interest rate on bank loans beyond a certain limit. This is done with a view to making banking credit available to relatively weaker sectors and (*c*) providing credit to weaker sectors at lower internal rates.

(ii) ***Change in lending margins.*** The banks advance money more often than not against a mortgage of property – land, building, jewellery, shares, stock of goods, etc. The banks provide loans only upto a certain percentage of the value of the mortgaged property. The gap between the value of the mortgaged property and amount advanced is called 'lending margin'. For example, if the value of stock is ` 10 million and the amount advanced is only ` 6 million, the lending margin is 40 per cent. The central bank is empowered to increase or decrease the lending margin with a view to decreasing and increasing the bank credit. This method was used for the first time by the RBI in 1949 with the objective of controlling speculative activity in the stock market. Since 1956, the RBI has made an extensive use of this method with a view to preventing speculation in scarce agricultural products, namely, foodgrains, cotton, oilseeds, vegetable oils, sugar, *khandsari* and *gur*, and cotton textiles and yarns. However, it is important to note here that selective credit controls were eliminated in October 1996 (*Economic Survey*—1997–98, p. 43.)

(iii) ***Moral suasion.*** The moral suasion is a method of persuading and convincing the commercial banks to advance credit in accordance with the directive of the central bank in the economic interest of the country. This method is adopted in addition to quantitative and other selective methods, particularly when effectiveness of these methods is doubtful. Under this method, the central bank writes letters to and holds meetings with banks on money and credit matters with the objective of persuading banks to act according to the instructions and advise of the central bank in the interest of the economy as a whole.

(iv) ***Direct controls.*** Where all other methods prove ineffective, the monetary authorities resort to direct central measures with clear directive to the banks to carry out their lending activity in a specified manner. There are however rare instances of direct control measures.

In addition to traditional monetary control measures, RBI uses repo rate (repurchase operation rate) and reverse repo rate under its Liquidity Adjustment Facility (LAF) programme. Repo rate is the rate that RBI charges the banks when they borrow from the RBI. Reverse repo rate is the rate that it offers the banks willing to keep their money with it. Depending on the need of the country, the RBI keeps changing these rates. Repo operation increases liquidity and reverse repo rate reduces the liquidity (or money supply) in the country.

2.4.2 Fiscal Policy and its Effectiveness

As an instrument of macroeconomic policy, fiscal policy has been very popular with the modern governments to influence the size and composition of the national product, employment, industrial production, prices, etc., in the economy. The deliberate use of fiscal policy as a means to achieve and maintain full employment and price stability in the economy has been a characteristic feature of the past seven decades after the publication of John Maynard Keynes' well-known book titled *The General Theory of Employment, Interest and Money* in 1936. The post-Keynesian popularity of fiscal policy has been largely due to the following three factors:

1. Ineffectiveness of the monetary policy as a means of removing mass unemployment in the great depression of the 30s;
2. The development of 'new economics' by John Maynard Keynes with its stress on the role of aggregate effective demand; and
3. The growing importance of government spending and taxation in relation to the national income and output.

From its modest beginnings in the 40s, fiscal policy today has become a major macroeconomic policy instrument employed by the governments to achieve full employment, to prevent inflation and to promote rapid economic growth.

Following Keynes, economists have argued that substantial amount of spending and fund raising in the form of taxation by government are capable of changing the size of national product and the tempo of aggregate economic activity in the system. By determining what goods and services will be produced, the fiscal operations of the government affect significantly the direction of employment of the economy's resources.

Government expenditure and tax revenue are not, however, closely related to one another. In any given year, government's total expenditure and total tax receipts may be unequal in which case the budget will be either a deficit or a surplus budget. When the expenditure and income of the government are equal, the budget is said be a balanced budget. The use of budget deficit and surplus in order to affect the level of the aggregate economic activity or to maintain economic stability or to promote economic growth in the economy is the essence of fiscal policy. Both the Keynesian and the neo-Keynesian economists rely primarily on the fiscal policy to stabilize the economy. During a major recession, such as the one which occurred in the 1930s, even the monetarists believed that fiscal policy could be used more effectively to increase the level of aggregate demand in the economy.

Meaning of Fiscal Policy

In his epoch-making book *The General Theory of Employment, Interest and Money,* Keynes used fiscal policy when referring to the influence of taxation on savings and government investment spending financed through loans raised from the public. Keynes looked at it as a state policy which used public finance as a balancing factor in the economy's development. Ordinarily, by fiscal policy is meant a policy which affects the important macroeconomic variables—aggregate output, employment, saving, investment, etc., through the budgetary manipulation. Fiscal policy refers to the regulation of the level of government spending, taxation and public debt. According to Arthur Smithies, the term fiscal policy refers to 'a policy under which a government uses its expenditure and revenue programmes to produce desirable effects and avoid undesirable effects on the national income, production and employment.'According to Buehler, 'by fiscal policy is meant the use of public finance or expenditure, taxes, borrowing and financial administration to further our national economic objective.'According to Fred R Glahe, by fiscal policy is meant the regulation of the level of government expenditure and taxation to achieve full employment in the economy. While referring to fiscal policy here we mean *pure* fiscal policy. A fiscal policy affects the level of government spending or taxation while the nominal money supply remains constant.

Fiscal Policy and Economic Activity

Government expenditure, tax income and public debt act as important levers to influence aggregate outlay, employment and prices in the economy. A given change—increase or decrease—in aggregate government expenditure causes a change—increase or decrease—in the aggregate demand thereby increasing or decreasing the factor incomes. Government expenditure incurred on wages and salaries of its employees, interest paid on government debt, social security and old age pension payments, all tend to increase the disposable personal income of people as a consequence of which the aggregate demand for consumer goods increases. Thus an increase in the total expenditure of government tends to expand the aggregate economic activity in the economy. On the other hand, taxes levied on the people to finance government expenditure tend to reduce disposable personal and corporate incomes which could have been either spent on consumption or devoted to capital formation through saving. Thus taxes tend to reduce the aggregate demand and income in the economy. These effects of government budget are equally valid for the central, state and local government budgets although the budget of the central government is much more powerful in affecting the level of aggregate economic activity in the economy than are the combined budgets of all the states and local bodies like the municipal and district boards.

Government expenditure and revenue can be combined in several ways in order to stimulate or depress the aggregate effective demand and economic activity in the economy. A surplus in the budget will exert a deflationary effect on national income because the inflow of aggregate government expenditure into the circular income flow will be less than the tax leakage from the circular income flow. Conversely, a deficit in the budget expands the net national product since the leakage from the aggregate income flow due to taxes is less than the additional inflow into the circular flow in the form of government expenditure. It follows, therefore, that in slump when there is need for expanding the aggregate demand deficit budget while in inflation when the problem is of preventing the aggregate demand from exceeding the aggregate supply, surplus budget should be prepared. This generalization should not, however, lead us to conclude that a

balanced budget is neutral in its effects on the national income and economic activity in the system. Depending upon the particular circumstances, a balanced budget may be no less important than an unbalanced—deficit or suplus—budget.

For a correct appraisal of the effects of government's fiscal policy on the level of aggregate economic activity, apart from the magnitude of government expenditure and revenue, their composition or structure is also equally significant. A given amount of revenue can be realized by the government in several ways—by levying taxes, by increasing the area of and profits from commercial activities and by borrowing from the public. However, even though the revenue raised through these several alternative methods may be the same, each method of raising revenue will affect the economy differently. For example, the same amount of revenue may be raised either through taxing the people or through floating bonds in the market but the effect of each one of these two methods of raising the government revenue will be different. Even in the case of taxes the effects will be different in the case of different tax levies like the income-tax and excise duty.

Similarly, the government can incur a given expenditure in several ways. It might, for example, spend upon building hospital or slum clearance or on the construction of a sugar mill or on unemployment doles. The effect on the level of aggregate economic activity will be different although the total expenditure is the same in each case. An expenditure of ` 5 crore incurred on constructing a new national highway or on slum clearance will not affect the aggregate investment activity in the private sector adversely; if anything, it will affect private investment favourably by causing an increase in the demand for raw materials and equipment needed for road construction or for housing the slum dwellers. But if the same amount is spent for starting a new sugar factory, it might cause an offsetting fall in the aggregate private investment by depressing the marginal efficiency of capital in the private sector. Consequently, the beneficial effects of public expenditure on the level of aggregate economic activity will be partially lost. Thus a balanced budget is not neutral in its effects on national income and economic activity unless it is assumed that the composition of expenditure and income remains unchanged from year to year. Although the level of aggregate economic activity in the economy can be affected by varying the size of a balanced budget, the stabilizing effect of the fiscal policy depends largely on the size of the surplus or deficit in the budget. The extent to which fiscal policy can prove effective as an instrument of economic stability depends on the extent to which the government can vary the difference between the income and expenditure rather than upon the balanced budget and the change in its size.

Objectives of Fiscal Policy

As an instrument of macroeconomic policy, the goals of fiscal policy are likely to be different in different countries and in the same country in different situations. For example, while in a developed economy operating either at the full or at near-full employment level the goal of fiscal policy should be the maintenance of full employment while in a developing economy the main concern of fiscal policy has to be the promotion of economic growth with stability and reduction in the economic inequalities.

Broadly speaking, overall fiscal policy involves two types of important decisions. While one of these two decisions is related to the goal of full employment, the other is concerned with determining the social priorities. The second policy decision is concerned with the issue of allocation of economy's productive resources as between their different rival uses—should more resources be allocated for education, health care, public housing,

NOTES

slum clearance, transport, etc. The government expenditure on different items in any society will be determined by the prevailing social values.

Economists generally agree that fiscal policy should be employed to achieve full employment and economic stability in the economy. Before the great depression of the 30s, by economic stability was largely understood the stability of the general price level. The severity of the depression focussed attention on the need to remove unemployment and to employ fiscal policy for this purpose. The Employment Act of 1946 in the USA stated that it was the responsibility of the federal government to use all possible means, including fiscal policy, to promote maximum employment, production and purchasing power in the economy.

After the Second World War, inflation has become a worldwide problem. Consequently, economic stabilization has come to be widely defined so as to include the elimination of inflationary pressures in the economy. This means that the achievement of full employment and price stability should be simultaneously attained through the instrument of fiscal policy. At times, however, both these goals may be difficult to achieve as these might be mutually inconsistent. An economy which wants to achieve full employment must accept moderate price rise unless it resorts to price control, rationing and wage freeze policies.

2.5 EXTENSION OF IS-LM MODEL TO LABOUR MARKET AND FLEXIBLE PRICE

A complete Keynesian model should, include consideration of the aggregate production function and of the demand for and the supply of labour. As in the classical theory, here also the aggregate production is a function of employment with the marginal product of labour diminishing as employment increases. The demand for labour is determined, as in the classical model, under the assumption of (a) diminishing marginal returns in production; (b) profit-maximizing behaviour of the employers; and (c) perfect competition in the product and factor markets. The demand curve for labour is the marginal product of labour curve and is derived from the aggregate production function, In short, the Keynesian demand function for labour is similar to the classical demand function for labour. There is, however, a difference in the two approaches and this relates to the supply curve of labour. While according to the classicists the supply of labour was a function of the real wage which was flexible, according to Keynes the supply of labour depended on the money wage which he considered downwardly rigid because workers were subject to *money illusion* in the sense that the workers would not work at the reduced money wage although they would be willing to work at lower real wage caused by a rise in prices. Thus, the supply of labour in effect depends on the money wage and not on the real wage. Due to the presence of powerful trade unions and minimum wage legislation, money wage would be downwardly rigid below the minimum money wage rate which was fixed either by legislation or by social convention.

The importance of the minimum wage legislation follows from the fact that the money wage sought to be fixed as the 'minimum' has to be higher than the wage which workers can obtain in the perfectly competitive labour market because if it was lower than the equilibrium wage determined in the free labour market, it would become a mere farce. If money wage is downwardly rigid below the W_1 minimum money wage, the classical supply curve below this wage is suppressed while above this wage the normal

supply curve is still valid. Thus, Keynes partially suppressed the classical supply curve of labour. At this statutorily fixed minimum wage, employment is determined by the demand for labour for all levels of employment upto the maximum supply of labour available at the minimum wage. Above this minimum wage, labour supply is a function of money wage. In the Keynesian system, the level of money wage is determined by the historical and institutional factors such as the minimum wage laws, trade union pressures, public opinion and other such factors.

Assuming a rigid money wage fixed at W_1 with the prices free to vary, the analysis can be graphically illustrated as shown in Figure 2.13. A rigid money wage with flexible prices would mean a flexible real wage system. Figure 2.13 has seven parts *A, B, C, D, E, F* and *G*. Part *E* shows the short period aggregate production function. The aggregate output is increasing at the diminishing rate. Part *F* of the figure shows that the demand for labour is inversely related to real wage such that as the real wage of labour falls, the demand for labour by the profit-maximizing employers in the economy increases and *vice versa.* The demand curve for labour is the marginal physical product of labour curve. In otherwords, it is the slope of the aggregate production function at the different levels of employment as drawn in Part *E* of the figure. Part *G* of the figure shows the different combinations of money wage and the general price level which are consistent with any given real wage.

We start with part *A* of the figure. Corresponding to r_1 rate of interest, the equilibrium values of the other variables in the different parts of the figure have been marked with subscript 1. Corresponding to $0r_1$ rate of interest, the aggregate investment is $0I_1$. With the given *SS* saving supply function in Part *B*, this gives the equilibrium aggregate real income $0y_1$ corresponding to which $S_1 = I_1$. Part E shows that $0y_1$ aggregate real income (output) is produced through the employment of $0N_1$ workers. Since prices are free to vary, the transactions demand for money in Part *C* cannot be known by knowing $0y_1$ alone. For this to be known we also need to know the general price level *P,* Part *F* shows that employers will hire $0N_1$ workers only if the real wage is W/P_1. Given this real wage and the rigidly fixed money wage W_1, the appropriate general price level will be $0P_1$ corresponding to which W/P_1 real wage can prevail. This has been shown in Part *G*. Having found out the relevant general price level $0P_1$ from Part *G* and the aggregate real income $0y_1$ from Part *B*, the aggregate money income $0Y_1 = 0P_1.0y_1$ can be found out and once this is done the transactions demand for money M_1 in Part *C* can be easily found out. Given the total money supply $\bar{M}_s$ and the speculative demand for money schedule as shown in Part *D* we find that the equilibrium rate of interest in Part *D* is $0r_1$ which is the same with which we had started our analysis in Part *A*. If either the rate of interest or the value of any other variable was different from the one given by subscript 1 as we traced the corresponding values in all the diagrams it would not be possible to come back to the position from where we had started. For example, if the rate of interest was $0r_2$, the equilibrium values for all the other variables will necessarily change and these have been shown by subscript 2 in different parts of the figure. In other words, corresponding to the lower rate of interest $0r_2$, the aggregate real income will increase from $0y_1$ to $0y_2$, total employment will increase from $0N_1$ to $0N_2$, real wage will fall from W/P_1 to W/P_2, general price level will rise from $0P_1$ to $0P_2$, total money income will rise from $0Y_1$ to $0Y_{12}$ and the transactions demand for money will increase from M_1 to M'_1. With the given money supply $\bar{M}_s$ it will result in the higher rate of interest $0r_2$, which is clearly inconsistent with our assumed lower rate of interest.

NOTES

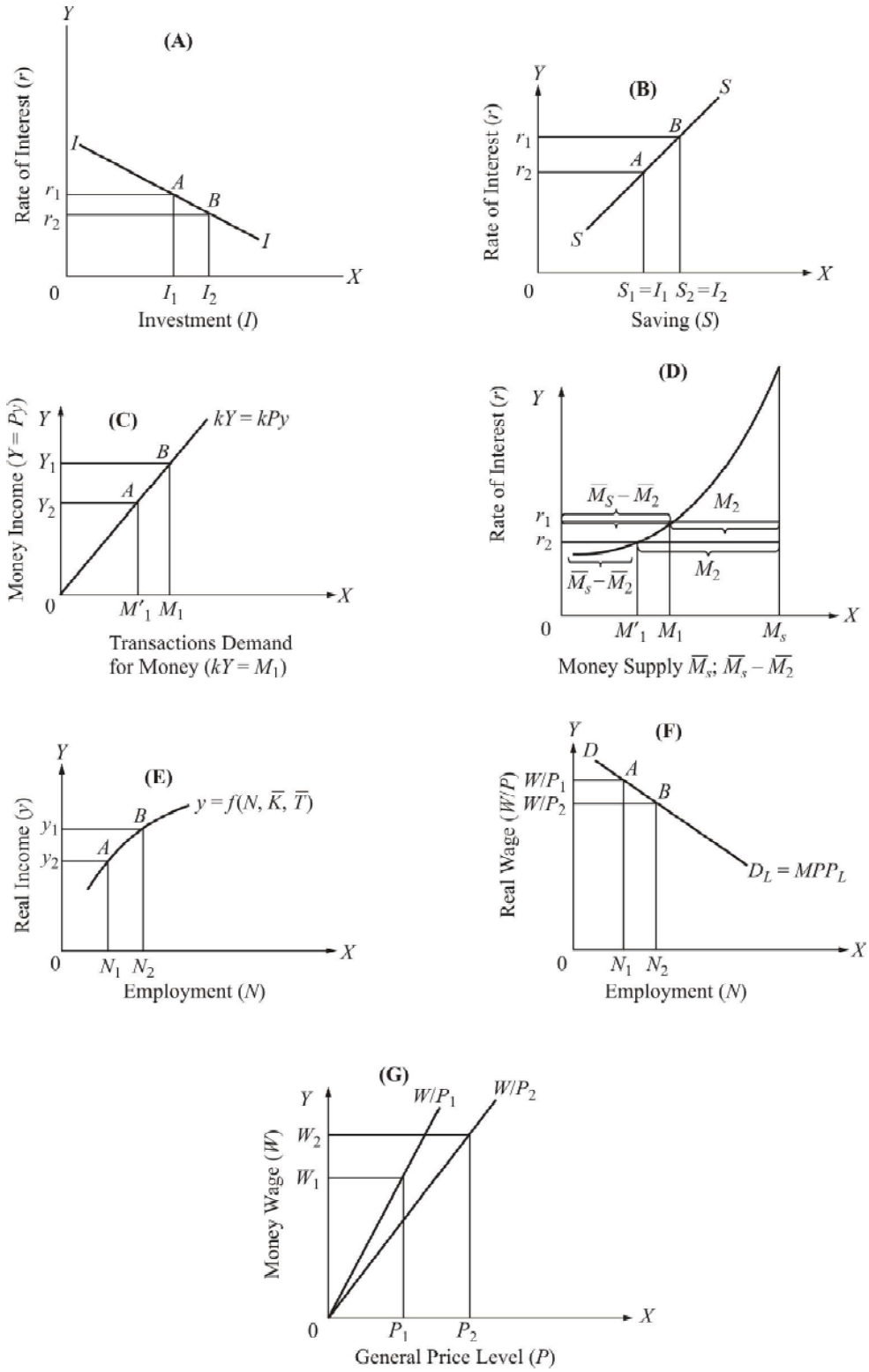


Fig. 2.13

Check Your Progress

6. State the factors on which the scope of monetary policy depends.
7. How is the level of money wage determined in the Keynesian system?

2.6 MUNDELL-FLEMING MODEL OF AN OPEN ECONOMY

NOTES

The Mundell–Fleming model, also known as the IS-LM-BoP model (or IS-LM-BP model), is an economic model first set forth (independently) by Robert Mundell and Marcus Fleming. The model is an extension of the IS-LM Model. Whereas the traditional IS-LM Model deals with economy under autarky (or a closed economy), the Mundell–Fleming model describes a small open economy. Mundell's paper suggests that the model can be applied to Zurich, Brussels and so on.

The Mundell–Fleming model portrays the short-run relationship between an economy's nominal exchange rate, interest rate, and output (in contrast to the closed-economy IS-LM model, which focuses only on the relationship between the interest rate and output). The Mundell–Fleming model has been used to argue that an economy cannot simultaneously maintain a fixed exchange rate, free capital movement, and an independent monetary policy. This principle is frequently called the 'impossible trinity,' 'unholy trinity,' 'irreconcilable trinity,' 'inconsistent trinity' or the 'Mundell–Fleming trilemma.'

Mundell developed a principle of effective market classification and suggested a rule for efficacy and stability of policy measure along the Tinbergen's rule. According to his rule, a policy instrument should be assigned a target which it can hit most effectively. Going by this rule, monetary policy or fiscal policy should be assigned the task which it can perform most successfully to achieve internal and external balance. Since monetary and fiscal policies affect both internal and external balance and both have their relative advantages and disadvantages, these policies need to be so combined that their positive effects are maximized and negative effects minimized.

Mundell's rule of policy assignment for the four different kinds of economic problems in the four zones is summarized in Table 2.1.

Table 2.1 *Mundell's Rule of Policy Assignment for Four Different Kinds of Economic Problems in Four Zones*

Zone	Nature of imbalance	Monetary policy	Fiscal policy
I	Unemployment and BOP surplus	Expansionary	Expansionary
II	Inflation and BOP surplus	Expansionary	Contractionary
III	Inflation and BOP deficit	Contractionary	Contractionary
IV	Unemployment and BOP deficit	Contractionary	Expansionary

Source: M. Chacholiades, *International Economics* (New York: McGraw-Hill, 1990), p. 422.

These rules offer a stable solution to the problem of internal and external balance provided (*i*) policies are chosen judiciously and implemented smoothly without discretionary changes, and (*ii*) there is no time lag in the working of monetary and fiscal policies. These are big conditions particularly the one to do with time lag. Therefore, Mundell's solution may be unstable.

NOTES

Shortcomings of the Monetary-Fiscal Policy Mix

The monetary-fiscal policy mix as a means of attaining internal and external balance has a strong theoretical underpinning. At a practical level, however, this approach has several shortcomings:

One, the monetary-fiscal mix approach assumes that policy-makers are fully aware of (*i*) the internal balance path (i.e., *IB* schedule), (*ii*) the external balance path (i.e., the *EB* schedule), (*iii*) the zone in which the economy is, and (*iv*) how far away the economy is from internal balance. In reality, however, the authorities lack data on these parameters.

Two, owing to the lack of necessary data, the determination of an exact combination of monetary and fiscal measures compatible with one another for achieving internal and external balance is an extremely difficult task. Therefore, some element of arbitrariness is bound to go into the policy formulation. Besides, political considerations affect the decision-making. Any mismatch in the monetary-fiscal mix, on account of these, affects the efficacy of the policy.

Three, the monetary-fiscal mix is based on some relationship between the interest rate and capital flows. This relationship may be disturbed by the implementation of policy because of some noneconomic factors or factors not accounted for in policy formulation. For example, the recent inflow of foreign capital to the Indian economy may be affected by the quick revival of growth in the US and the Euro zone.

Four, Mundell's approach does not provide a 'true adjustment mechanism.'The reason is Mundell's approach considers capital flows as autonomous whereas a considerable part of capital flows is accommodating, not autonomous. Accommodating capital flows are not affected by a change in the interest rate. This may seriously affect the efficacy of Mundell's solution.

Finally, the Mundellian approach assumes (implicitly) that other countries are not affected by the monetary and fiscal policies adopted by a country, and even if they are, they do not react. In reality, however, a great deal of conflict arises between the nations. Finding an appropriate monetary-fiscal mix compatible with that of other countries is a rather impossible task. Even if a combination of fiscal and monetary policies is somehow worked out by trial and error, such a policy may push the economy away from the equilibrium point rather than bringing it closer to it.

2.7 SUMMARY

In this unit, you have learnt that,

- A plethora of analytical writings dwelling on the discussion of why interest is paid and what determines its equilibrium rate developed in the recent years.
- As a payment made by a borrower to a lender of funds, interest was unknown in the ancient savage communities. It was, however, common in the early societies living in the eastern Mediterranean Persian Gulf region of Babylon, Egypt and Greece.
- The pre-classical approach to interest was largely concerned with the ideas on the subject contained in the mercantilist and physiocratic writings published largely during the 17th and 18th centuries.

Check Your Progress

8. Mundell-Fleming principle is also known by some other names. Name them.
9. State one shortcoming of the Monetary-Fiscal policy mix.

- The main propelling force behind the demand for investible funds was the physical productivity of present capital goods which provided incentive for the firms to undertake investment and add to their capital stock.
- According to the classical approach, the rate of interest was a factor which brought into equilibrium the demand for investment and the supply of savings.
- Investment represented the demand for investible funds and savings represented the supply of these funds while the rate of interest was that 'price' of the investible funds at which the supply of and the demand for investible funds were equated.
- The classical theory is faulty since it completely ignores the consideration of the asset demand for money and regards money as being demanded exclusively for the transactions purpose in order to remove the difficulties of barter.
- The loanable funds theory of the rate of interest, also known as the neoclassical theory of rate of interest, which represents an improvement over the classical theory of rate of interest was developed, among others, by the well-known Swedish economist Knut Wicksell and the British economist Dennis Holme Robertson.
- Like the classical theory, the loanable funds theory of interest rate has also been criticized on various grounds.
- Keynes had criticized the classical theory and he was not satisfied with the neo-classical or loanable funds theory of rate of interest.
- The modern or the neo-Keynesian theory of interest rate has been developed by John R Hicks by synthesizing the classical and the Keynesian liquidity preference theories.
- The IS–LM model (Investment Saving–Liquidity Preference Money Supply) is a macroeconomic tool that demonstrates the relationship between interest rates and real output, in the goods and services market and the money market.
- IS curve shows the combination of level of output and interest rate in such a manner so that the desired expenditure equals to income.
- The monetary policy has to be so devised that prices may fall without causing unemployment in the economy.
- The Mundell–Fleming model, also known as the IS-LM-BoP model (or IS-LM-BP model), is an economic model first set forth (independently) by Robert Mundell and Marcus Fleming.
- The Mundell–Fleming model portrays the short-run relationship between an economy's nominal exchange rate, interest rate, and output (in contrast to the closed-economy IS-LM model, which focuses only on the relationship between the interest rate and output).
- Mundell developed a principle of effective market classification and suggested a rule for efficacy and stability of policy measure along the Tinbergen's rule. According to his rule, a policy instrument should be assigned a target which it can hit most effectively.

2.8 KEY TERMS

- **Liquidity preference**: In macroeconomic theory, liquidity preference refers to the demand for money, considered as liquidity.

NOTES

- **New neoclassical synthesis:** New neoclassical synthesis or new synthesis is the fusion of the major, modern macroeconomic schools of thought, new classical and new Keynesian, into a consensus on the best way to explain short-run fluctuations in the economy.
- **IS-LM model:** The IS–LM model, or Hicks–Hansen model, is a macroeconomic tool that shows the relationship between interest rates and real output, in the goods and services market and the money market (also known as the assets market).

2.9 ANSWERS TO 'CHECK YOUR PROGRESS'

1. The classical theory of rate of interest is considered faulty because it ignores the influence which the bank created money (credit) exerts on the rate of interest.
2. The loanable funds theory of the rate of interest is also known as the neoclassical theory of rate of interest.
3. The modern or the neo-Keynesian theory of interest rate has been developed by John R Hicks by synthesizing the classical and the Keynesian liquidity preference theories.
4. The LM curve is obtained by determining equilibrium in the money market in terms of equilibrium between demand for money and supply of money corresponding to different pairs of interest rate and the level of income.
5. The intersection of the IS and LM curves is the 'general equilibrium' where there is simultaneous equilibrium in both markets.
6. The scope of monetary policy depends on two factors.
   (i) The level of monetized economy and
   (ii) The level of development of the capital market.
7. In the Keynesian system, the level of money wage is determined by the historical and institutional factors such as the minimum wage laws, trade union pressures, public opinion and other such factors.
8. Mundell-Fleming principle is frequently called the 'impossible trinity,''unholy trinity,' 'irreconcilable trinity,' 'inconsistent trinity' or the 'Mundell–Fleming trilemma.'
9. One shortcoming of the monetary-fiscal policy mix is that owing to the lack of necessary data, the determination of an exact combination of monetary and fiscal measures compatible with one another for achieving internal and external balance is an extremely difficult task. Therefore, some element of arbitrariness is bound to go into the policy formulation.

2.10 QUESTIONS AND EXERCISES

Short-Answer Questions

1. State in brief the relative effectiveness of monetary and fiscal policies.
2. Differentiate between the pre-classical and the classical approach of rate of interest.

3. What is the chief merit of the modern theory of rate of interest?
4. State in brief the basic principle of Mundell-Fleming model

Long-Answer Questions

1. Discuss the classical theory of rate of interest. On what grounds did Keynes criticize it?
2. Explain critically the liquidity preference theory of interest rate.
3. Discuss the Classical and Keynesian theories of interest.
4. Analyse the IS-LM model with the help of graphs showing the rate of interest.

2.11 FURTHER READING

Mankiw, N Gregory. 2010. *Macroeconomics.* Worth Publishers, New York.

Shapiro, Edward. 1996. *Macroeconomic Analysis.* Galgotia Publication, New Delhi.

Branson, WH. 1979. *Macroeconomic Theory and Policy*. Universal Book Stall, New Delhi.

Jha, R. 1999. *Contemporary Macroeconomic Theory and Policy*. New Age International, New Delhi.

Gupta, SB. 1995. *Monetary Economics: Instruments and Policy*. S Chand & Co. New Delhi.

Endnotes

[1] For a brief excellent review of interest rate theories refer to J Conard, *An Introduction to the Theory of Interest,* 1959.

[2] The term 'classical economics' of which the classical theory of interest rate is an inseparable part is difficult to define precisely including as it does the economic thought of a long line of writers. Karl Marx, who first used the term *classical economics* included in it the economic ideas of the writers starting from William Petty and ending with David Ricardo in England and from Pierre Boisguilbert to Jean Charles Leonard de Sismondi in France. Classical economics, according to Karl Marx, 'investigated the real relations of production in bour- geois society.' John Maynard Keynes, however, meant by classical economics the economic thought of David Ricardo and his followers, i.e., those who adopted and perfected the theory of Ricardian economics, including, for example, John Stuart Mill, Alfred Marshall, Francis Ysidro Edgeworth and Arthur Cecil Pigou.

[3] J M Keynes, *The General Theory of Employment, Interest and Money,* 1936, p. 175.

[4] A C Pigou, *The Economics of Welfare,* Fourth Edition, 1932, Reprinted 1950, p. 25.

[5] Alfred Marshall, *Principles of Economics,* Eighth Edition, 1920, Reset and Reprinted 1949, p. 193.

[6] Alfred Marshall, *op. cit.,* p. 196.

[7] J M Keynes, *op. cit,* p. 167.

[8] *Ibid.*

[9] *Ibid.*

[10] This argument implicitly assumes that there is no offsetting change in the aggregate consumption outlay.

[11] J M Keynes. *op. cit.,* p. 179–80.

[12] J M Keynes. *op. cit.,* p. 181.

[13] The theory was subsequently refined by the other Swedish economists comprising of Bertil Ohlin, Erik R Lindahl, Gunnar Myrdal and Bent Hansen.

[14] John R Hicks, 'Mr. Keynes and the 'Classics': A Suggested Interpretation,' *Econometrica,* Volume 5, 1937, p. 147–59.

[15] Alvin H Hansen, *A Guide to Keynes,* 1953, p. 141.

[16] W T Newlyn and R P Bootle, *Theory of Money,* Third Edition, 1978, p. 91.

[17] Alvin H Hansen, *op. cit.,* p. 141.

[18] J M Keynes, *op. cit.,* p. 167.

[19] J M Keynes, *op. cit.,* p. 170.

[20] The Keynesian concept of the liquidity trap severely limits the scope of monetary policy as an economic policy instrument to remove unemployment in the economy by raising the level of invest- ment through the interest rate effect. The phenomenon of liquidity trap shows that no amount of increase in the money supply on the part of monetary authority will push the rate of interest below the liquidity trap rate of interest. Thus an expansionary monetary policy is not of much help in depression.

UNIT 3 THEORIES OF CONSUMPTION

Structure

NOTES

3.0 INTRODUCTION

This unit discusses different theories of consumption. Keynes mentioned several subjective and objective factors which determine consumption of a society. However, according to Keynes, of all the factors it is the current level of income that determines the consumption of an individual and also of society. Since Keynes lays stress on the absolute size of income as a determinant of consumption, his theory of consumption is also known as absolute income theory.The three most important theories of consumption, namely, Relative Income Theory of Consumption, Life Cycle Theory of Consumption and Permanent Income Theory of Consumption have been discussed in detail in this unit.

3.1 UNIT OBJECTIVES

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

- Discuss the post-Keynesian developments made on the income-consumption relationship
- Analyse Duesenberry's theory of relative income hypothesis
- Explain life cycle hypothesis with the help of graphs
- Analyse the concept of permanent income hypothesis by the help of equations

3.2 EXTENSION OF KEYNESIAN CONSUMPTION FUNCTION TO LONG RUN

An enquiry about the determinants of consumers' demand may be restated in terms of the concept of the 'consumption function' which denotes the general income-consumption relationship. Like the familiar microeconomic market demand curve of a commodity which shows the different quantities of a particular good or service which will be demanded at the different prices, *ceteris paribus*, per any given time unit, the consumption function shows the total expenditure which, *ceteris paribus*, consumers will make on the purchase of consumer goods and services at different levels of income.

NOTES

The most widely accepted hypothesis about consumption is that its major determinant is the disposable personal income. The income-consumption relationship, for which the economists have coined the term 'consumption function', was originally called 'the propensity to consume' by John Maynard Keynes. According to Keynes, the propensity to consume is the functional relationship X between Y (a given level of income expressed in terms of wage-units) and C_w (the expenditure on consumption out of that level of income) so that–

$$C_w = X(Y_w)$$

Although Keynes has listed several objective and subjective factors in his book *The General Theory of Employment, Interest and Money* as co-determinants of consumption spending, he has singled out income as the main determinant of consumption. Following the notation used earlier, aggregate consumption as a function of aggregate income may be expressed as $C = f(Y)$. The short-period consumption function takes the general form of the straight-line CC in Figure 3.1 and is expressed in the form of equation $C = a + bY$, where a is the constant positive amount of autonomous consumption which is unrelated to income and bY is the induced consumption which is determined by the level of income Y and the constant proportion of income b spent on consumption. Since the linear consumption function $C = a + bY$ intersects the vertical axis above zero, consumption is positive at zero income and the slope of the consumption function is constant. This means that the consumption-income ratio—the average propensity to consume—is falling as income is increasing while the marginal propensity to consume—the marginal consumption-income ratio—is constant.

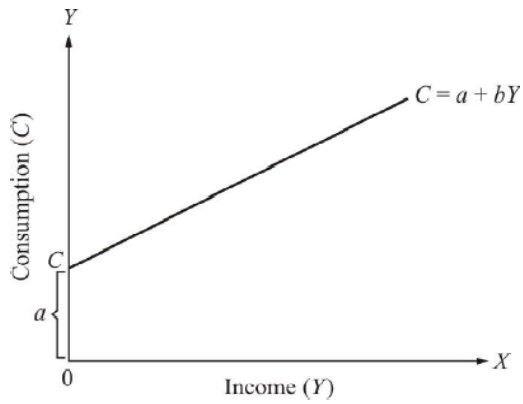


Fig. 3.1

While the short-run consumption function intersects the vertical axis above the point of origin of the axes showing a non-proportional income-consumption relationship, the long-run consumption function passes through the point of origin of the axes as shown in Figure 3.2. This shows that in the long-run total consumption bears a proportionality relationship to total income. Consequently, both the long-run average propensity to consume and the marginal propensity to consume are constant and equal. The general form of the long-run consumption function may be expressed by the equation $C = bY$, where b is the constant slope of the consumption function. The only restriction is that b is constant, less than one and more than zero. In Figure 3.2, the average propensity to consume (APC) equals the marginal propensity to consume (MPC) corresponding to all the different levels of income since the consumption function $C = bY$ passes through the point of origin of the axes.

NOTES

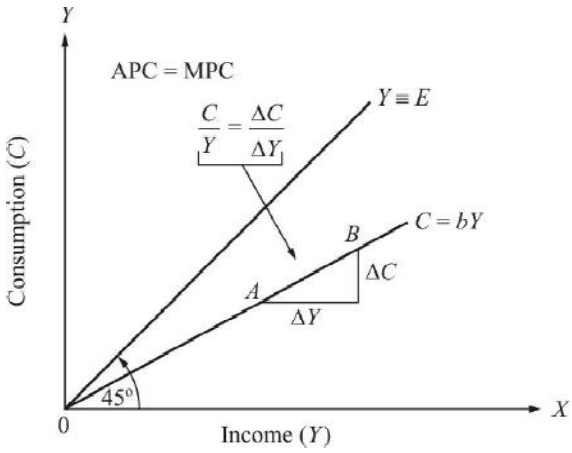


Fig. 3.2

The Classical and the Keynesian Views

The classical and the keynesian views have been discussed under section 1.8 of Unit 1.

The Post-Keynesian Developments

The post-Keynesian empirical studies made on the income-consumption relationship show that the short-run consumption function has a smaller slope relatively to the slope of the long-run consumption function. During the inter-war period, economists who estimated the consumption function for the US economy accepted Keynes' statement that even in the long run, a smaller proportion of income is consumed as income increases. Consequently, the APC falls as income increases. This was represented by a linear consumption function of the form $C = a + bY$, where $a > 0$ and $1 > b > 0$ = constant. In this consumption function, b is the marginal propensity to consume and no difference is made between the short-run and the long-run marginal propensities to consume. Several functions estimated during the inter-war period showed results clearly fitting this pattern and explained the data very well. Arthur Smithies and others used such a function to predict the post-war demand for the American economy. This consumption function shows that the basic long-run income-consumption relationship was one of non-proportionality.

This type of the Keynesian consumption function was, however, subsequently discarded as not being very useful in explaining the actual consumer behaviour because it was found that in the post-war period—1946 and 1947—the level of consumption was far above that predicted by the simplified consumption function of the form $C = a + bY$. This simplified consumption function wasalsodiscarded on additional grounds. The consumption function of the form $C – a + bY$ implies that the income-consumption ratio (C/Y) falls as income increases, i.e., the percentage of total income spent on consumption falls as income increases. On this basis, the saving-income ratio must increase as income increases. On the basis of a non-proportional income-consumption relationship, a sufficiently high level of investment was deemed necessary for attaining full employment in the post-war period. It was, however, experienced that government expenditure of even a smaller magnitude that was predicated by such a consumption function resulted in inflation.

The non-proportional income-consumption relationship was controverted by Simon Kuznet's now well-known study of the national income and consumption expenditure for the American economy during the period 1869 to 1929 which showed that the ratio

NOTES

of total consumption to national income (*C/Y*) had remained constant while income had quadrupled.[1] Simon Kuznet's finding was supported by Raymond Goldsmith's study relating to consumption and personal income. According to Goldsmith, 'a main enduring characteristic' of saving was the long-term stability of the aggregate personal saving at approximately one-eighth of income.[2] This means that the long-run income-consumption ratio (*C/Y*) was stable at seven-eighth of income.

While the income-consumption ratio (*C/Y*) has been constant over the long period of time, the cross-section data shows that the income-consumption ratio (*C/Y*) decreases as income increases. Moreover, studies also show that the *C/Y* ratio fluctuates cyclically—consumption decreases much less than proportionately during minor recessions and even increases sometimes in the face of falling income. In short, the empirical findings show a proportional relationship between income and consumption during the long period and a non-proportionality relationship during the short period. Figure 3.3 explains the short-run and long-run consumption functions. In the figure *aa*, *bb* and *dd* are the short-run consumption functions showing that the income-consumption ratio (*C/Y*) decreases as income increases. The long-run proportionality consumption function 0C = *bY* shows, however, that the income-consumption ratio *C/Y* remains constant regardless of the level of national income.

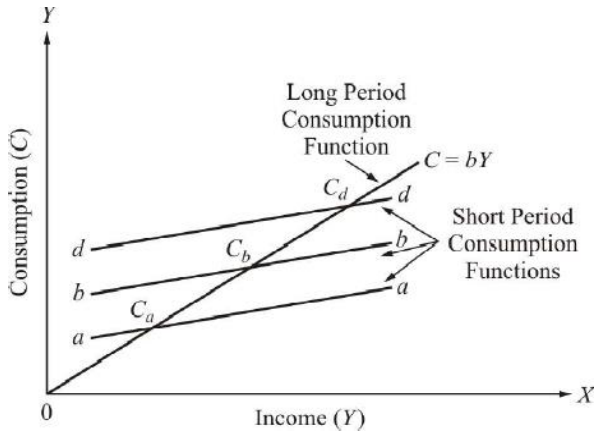


Fig. 3.3

Reconciliation of the Short Period and Long Period Consumption Functions

Different hypotheses have been developed by the economists in order to explain the apparent contradiction between the short-run non-proportional and the long-run proportional income-consumption relationship. According to the various studies made to study this relationship, while the long-run income-consumption ratio (*C/Y* = APC) has been constant, the short-run income-consumption ratio (*C/Y*) decreases as income increases. Thus, the consumption function takes two different forms on the basis of the short or long period. The short-run consumption function takes the form of equation $C = a + bY$ while the long-run consumption function takes the form of equation $C = bY$.

Check Your Progress

1. What do the post Keynesian empirical studies show?
2. Who controverted the non-proportional income-consumption relationship?

3.3 RELATIVE INCOME HYPOTHESIS

Reconciliation of the short-run and long-run consumption functions was first attempted by James S Duesenberry in his work entitled *Income, Saving and the Theory of Consumer Behaviour* published in 1949. His theory, known as the *relative income*

NOTES

hypothesis, is based on the assumptions that (1) the consumption behaviour of individuals is interdependent and not independent; and (2) consumption relations are irreversible over time.[3] On the basis of the first assumption, Duesenberry developed the proposition that the ratio of income consumed by an individual does not depend on his absolute income; instead, it depends on his relative income—on his percentile position in the total income distribution. In any given year, an individual will consume a smaller percentage of his income as his absolute income increases if his percentile position in the income distribution improves and *vice versa*. This is supported by evidence from the cross-section data. If, however, an individual's percentile position in the income distribution remains unchanged over time, he will continue to spend the same percentage of his income on consumption as his absolute income increases. In other words, if a family's relative position on the income scale remains unchanged when its absolute income has increased, the fraction of income spent on consumption and the fraction of income saved will remain unchanged—the division of income between consumption and saving will not change. For all individuals remaining in the same percentile position over time, the rise in their disposable personal incomes over the long period will cause an equal percentage increase in their consumption spending. Although the relative income position of some individuals may change, these changes will balance in the aggregate with the result that the long-run aggregate *C/Y* ratio will remain constant. This statement agrees with the evidence collected from the study of time-series data. Thus the relative income hypothesis explains the apparent paradox between the cross-section and the time-series data evidence.

The second assumption of this hypothesis explains the fluctuations in the aggregate income-consumption ratio *C*/*Y* during a trade cycle. Given the irreversible consumer standards, a fall in income during the cyclical downswing will cause a less than proportionate fall in the consumption because individuals base their current consumption partially on the previous higher levels of income and consumption represented by the previous peak income and consumption. This means that during a recession, the ratio of consumption to income (APC) increases and the saving-income ratio (APS) falls because consumers try to maintain their previous high level of consumption.

Starting from the peak income $0Y_2$ in Figure 3.4, let us first suppose that income falls. Consumers move back down on the short-run consumption function *bb* and their average propensity to consume (APC) rises as they try to maintain their previous high standard of living. With a successive fall in income, consumption will fall much less relative to the fall in income, with the consumers moving back down along the short-run consumption curve *aa* and not along the long-run consumption function $0C = bY$. With each successive fall in income and consumption along the *bb* consumption curve, the income-consumption ratio (*C/Y*) rises and the saving-income ratio (*S/Y*) falls.

When income increases, the income-consumption ratio (*C/Y*) falls while the saving-income ratio (*S/Y*) rises as consumers move back upward along the short-run consumption function *bb.* Once again there is a time-lag on the part of the consumers in adjusting their current consumption to their current rising income. But when income rises beyond the previous peak income $0Y_2$, consumption becomes proportional to income, i.e., the income-consumption ratio (*C/Y*) remains constant as the level of income continues to rise and consumers stop moving up along the short-run consumption curve *bb* any more. Instead, they move up along the long-run consumption function $0C = bY$ on which the APC is constant. This is the famous 'ratchet effect' which tells us that during a boom, consumption increases but as income falls, consumption does not fall to the previous level distorting the constancy of income-consumption ratio. According to Duesenberry,

'the ratchet keeps the economy from slipping back all the way and losing all the gains in income acquired during the preceding boom.'[4]

NOTES

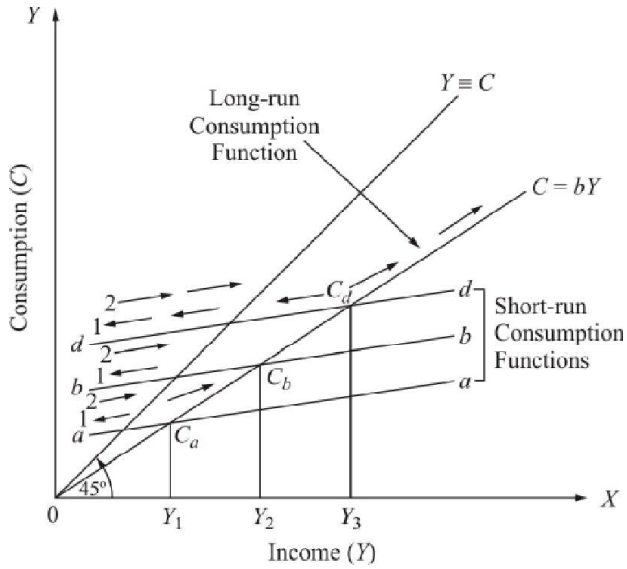


Fig. 3.4

According to Duesenberry, it is the rate of income growth which is important in determining the movement of consumers in the economy on the short-run or the long-run aggregate consumption function. After the peak income is achieved, assume a fall in income. Consumption does not fall proportionately along the long-run consumption function; rather it falls less than proportionately to the fall in income along the short-run consumption function. As the recovery begins, the aggregate consumption rises along the short-run consumption function until the previous peak income is experienced.

After this level of income is attained, the aggregate consumption no longer follows the path of the short-run consumption function. Abandoning the short-run consumption function, it takes a detour along the long-run consumption function and continues to increase proportionately to the increase in income until another depression starts when it again adopts one of the short-run consumption functions to retreat. From this pattern of consumer behaviour, it follows that so long as income grows steadily, the APC and the MPC will be equal. Consequently, increases in consumption will be equi-proportional to the increases in income as consumers stayon the long-run consumption function throughout the entire period of steady growth in national income. However, if recession occurs and income falls, the APC rises as consumers move backward along the short-period consumption function *bb* to repeat the foregoing phenomenon.

Duesenberry's hypothesis asserts that the consumption-income ratio depends on the ratio of the *current income* to the previous *peak income*. Thus, current consumption depends on the current income *relative* to the peak income previously attained. Duesenberry argues that income does not grow evenly over the long period, instead it displays cyclical spurts and dips which produce the non-proportional consumption-income relationship in the short period. In the absence of the cyclical short-run income fluctuations, consumption would follow the path shown by the long-run consumption function 0C = *bY* in Figure 3.4. However, there are cyclical fluctuations in income which push the economy off the long-run consumption function and on to the short-run consumption function in an unbroken series over time.

The relative income hypothesis can be functionally stated as:

$$\frac{C_1}{Y_1} = a + b\left(\frac{Y_1}{Y_0}\right); \text{ where } b < 0 \qquad ...(3.1)$$

where Y is the previous peak income and C and Y are the consumption and income respectively in current time period t. The value of the income-consumption ratio C/Y (APC) predicted by this function will be higher in recessions when the current income is less than the peak previous income, i.e., when $Y_t < Y_0$, compared with expansions when the current income is higher than the peak previous income, i.e., when $Y_t > Y_0$. Considering only the long-run patterns so that in general $Y_0 = Y_{t-1}$ (income of the previous year), Y_t/Y_0 will be equal to some constant value $1 + a$, where a is the annual growth rate of national income. In the long-run, the average propensity to consume, i.e., income-consumption ratio (C/Y) is itself equal to a constant value as suggested by this hypothesis.

Although the relative income hypothesis is ingenious in many ways, it suffers from many limitations. *First*, the hypothesis states that changes in consumption and income will always take place in the same direction. Yet at times, mild decreases in income and increases in consumption might often occur concomitantly. *Second*, the hypothesis says that increases in consumption are proportional to increases in income regardless of the magnitude of income increases. It seems reasonable to argue that unforeseen large increases in income will cause, at any rate initially, less than proportionate increases in consumption. *Third*, consumer behaviour is gradually reversible over time rather than being truly irreversible as assumed by Duesenberry. Consequently, the previous peak income exerts a smaller influence on current consumption, longer the time distance between the current and the previous peak income period. The *fourth* limitation is that Duesenberry stresses the significant part played by interdependence in consumer preferences. According to him, consumers in any given income group emulate or imitate the consumption behaviour of consumers in the higher income groups. Conclusive proof in this matter is still lacking. Although we have sufficient information on consumer behaviour, we have none or very little available on its causes. The *fifth* problem is that the relative income hypothesis fails to predict the post-World War II income-consumption relationship. *Lastly*, the relative income hypothesis is faulty because it represents the theory of consumer behaviour which relates a consumer's spending to his income, assuming all the other factors that influence consumption as given.

3.4 LIFE CYCLE HYPOTHESIS

The economists Franco Modigliani, Richard E Brumberg and Albert Ando also agree with Duesenberry. Like Duesenberry, they too have developed a similar form of the consumption function suggesting that the basic income-consumption relationship is one of proportionality. Their explanation is, however, somewhat different from Duesenberry's explanation.[5] Ando and Modigliani posit a consumption function in which individual consumption depends on the existing net worth (wealth) plus the present value of all the current and future labour earnings, rate of return on capital and the age of the individual. The Modigliani-Brumberg-Ando hypothesis is known as the *life-cycle hypothesis* of consumption. Essentially, the idea is that long-period consumption is related to lifetime average income and is, therefore, non-responsive to changes in the current income. This hypothesis begins with the empirical fact that the average individual's income is lower than his spending or consumption both at the beginning of his career and towards the end

NOTES

NOTES

of his life. In the middle years of his life-career, his total earnings or income is higher than his consumption resulting in positive savings. The income and spending (consumption) streams of the average individual appear as shown in Figure 3.5 where $t = t_0$ is the time of the start of his career and $t = T$ is the time of his death.

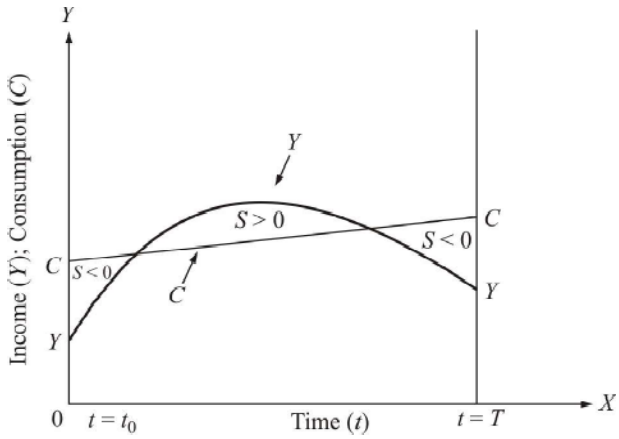


Fig. 3.5

According to Modigliani and Ando, age is a crucial variable in determining the relationship of consumption to the measured income and perhaps the relationship between consumption and wealth. Over their lifetime, individuals experience low income in the initial earning years, income rises to a peak in the middle or late working years and falls off sharply with old age and retirement. On the other hand, the typical consumption function is much 'flatter'—the typical individual's consumption remains fairly stable or rises gently over his lifetime. Consequently, in his youth and old age he dissaves while he saves in his middle years to repay the debts incurred earlier and to provide for his retirement in old age.

The concept of life-cycle consumption explains the non-proportional relationship between consumption and income which has been observed over the short period in the cross-sectional consumption studies of the *type* represented by Figure 3.1. If the profile of the lifecycle income and consumption streams is as shown in Figure 3.5, then the lifecycle hypothesis states that the average propensity to consume (APC) will be falling with higher levels of family income and *vice versa.* The APC is highest at the beginning of the lifecycle.

3.5 PERMANENT INCOME HYPOTHESIS

Yet another attempt that has been made to reconcile the short-run non-proportional income-consumption relationship with the long-run proportional income-consumption relationship is found in Milton Friedman's famous study entitled *A Theory of the Consumption Function* published in 1957. Friedman distinguishes between the current or observed or *measured income* of any given time period and the *permanent income* on which the consumers base their behaviour. Asimilar distinction has been made between the current or observed or *measured consumption and permanent consumption.* According to Friedman, permanent income is 'the amount a consumer unit could consume (or believes that it could) while maintaining its wealth intact' while permanent consumption is 'the value of the services that it is planned to consume during the period in question.'[6] It is the mean income which the family unit regards as permanent and would depend

upon the family unit's time horizon and farsightedness. The time span relevant to permanent income is the minimum period of time over which income influences must be maintained in order to make the income receiver treat these influences as permanent.

NOTES

The permanent income hypothesis states that the ratio of permanent consumption to permanent income is constant regardless of the level of permanent income. Since permanent consumption is proportional to permanent income, the long-run aggregate APC equals the long-run aggregate MPC. This is not to suggest, however, that the APC of every individual is equal. In particular, the average propensities to consume may depend upon such factors as the rate of interest, the ratio of non-human wealth to permanent income, the ages of members and the number of members in the consumer unit, the extent of income variability, etc. Notwithstanding the influence which these factors may exert on the individual consumer unit's consumption, the value of the income-consumption ratio is independent of the level of permanent income. Rich people do not consume a smaller percentage of their permanent income than do the poor people. Expressed differently, the 'rich' and the 'poor' devote the same fraction of their incomes to consumption. In other words, expressed as a ratio of their permanent consumption to their permanent income, the APC of the families at all levels of their income is constant.

According to Milton Friedman, the explanation of the short-run non-proportionality relationship between consumption and income despite the proportionality relationship between consumption[7] and income in the long-run lies in the fact that the observed or *measured* income and the observed or *measured* consumption of any short-period for the economy or the individual is composed of the permanent and transitory components. While the permanent components of the measured income and measured consumption are perfectly correlated, the transitory components of the measured consumption and measured income are neither correlated to their corresponding permanent components nor are they correlated to each other. The permanent income hypothesis can be set forth in terms of the following equations:

$$C_p = K(i, w, \mu)\, Y_p \qquad \text{...(3.2)}$$

$$Y = Y_p + Y_T \qquad b(Y_p\, Y_T) = 0 \qquad \text{...(3.3)}$$

$$C = C_p + C_T \qquad b(C_p\, C_T) = 0 \qquad \text{...(3.4)}$$

$$b(Y_T C_T) = 0 \qquad \text{...(3.5)}$$

where,

Y = measured or observed disposable personal income

C = measured or observed consumption

Y_P = permanent income

Y_T = transitory income

C_P = permanent consumption

C_T = transitory consumption

K = proportionality constant between permanent consumption and permanent income

i = rate of interest

w = ratio of non-human wealth to permanent income

μ = propensity of the consumer unit to add to consumption rather than to wealth. The most important factors which determine the value of μ are the number and ages of family members in the consumer unit and the importance of transitory factors which affect income and consumption, i.e., the extent of income variability, and b is the correlation coefficient term.

NOTES

Equation (3.2) states the relationship between permanent income and permanent consumption. The equation shows that the long-run average propensity to consume expressed as a ratio of permanent consumption to permanent income is constant, i.e., $C_P/Y_P = K$ = constant. In other words, the proportionality factor K in the equation is the permanent average propensity to consume whose value is independent of the level of income. It is, however, influenced by the terms i, w and μ which have been put inside the parenthesis. Equations (3.3) and (3.4) define the relationship between the permanent, transitory and measured magnitudes of the aggregate income and aggregate consumption. Both these equations show that the *measured* income and *measured* consumption in any given time period can be divided into permanent and transitory components and that these components are not correlated.

According to Milton Friedman, a family unit's measured or observed income for any given year will be larger or smaller than its permanent income depending on the presence of the positive or negative transitory income component. For example, if the family wage earner receives an extra unexpected bonus or overtime allowance during the year which he does not expect to receive in the following years, this is a positive transitory income component of the current or measured income. On the other hand, if due to a sudden unexpected shutdown of the plant, the wage earner suffers a loss in his measured income during the year but has no reason to expect this loss to continue in the future years, this is a negative transitory income component. In the first case, the measured income of the family unit rises above while in the second case it falls below its permanent income. Equation (3.5) shows that the transitory income and transitory consumption are independent. Consequently, a transitory increase in the measured income during the year, for instance, a windfall gain, will be entirely saved and none of it will be spent by the family unit.

The permanent income hypothesis can be explained diagrammatically as shown in Figure 3.6. The figure relates to a cross-section of population. The permanent or long-run consumption function $0C = bY$ passes through the point of origin of the axes. It shows that the average propensity to consume equals the marginal propensity to consume, i.e., the APC = MPC which is constant. Consequently, there is a proportionality relationship between the permanent income and permanent consumption. The $BB = a + bY$ line is the short-run or measured consumption function. It shows the non-proportionality relationship between the measured income and the measured consumption in any given time period. It shows that the short-run APC > MPC and it falls as income rises and *vice versa*. $0A$ is the average measured income for the whole community. The particular families having this ($0A$) average measured income have zero transitory component in it. Some families with higher permanent income have negative transitory income component while some others with lower permanent income have positive transitory income component in their measured income. The average permanent income of these families is nevertheless $0A$ which equals their average measured income. Corresponding to their $0A$ permanent income, these families have AK (= $0L$) permanent consumption which also equals their average measured consumption against their measured average income $0A$. The reason for this equality between permanent consumption and measured consumption resides in the absence of any negative or positive transitory income component in the average measured income.

NOTES

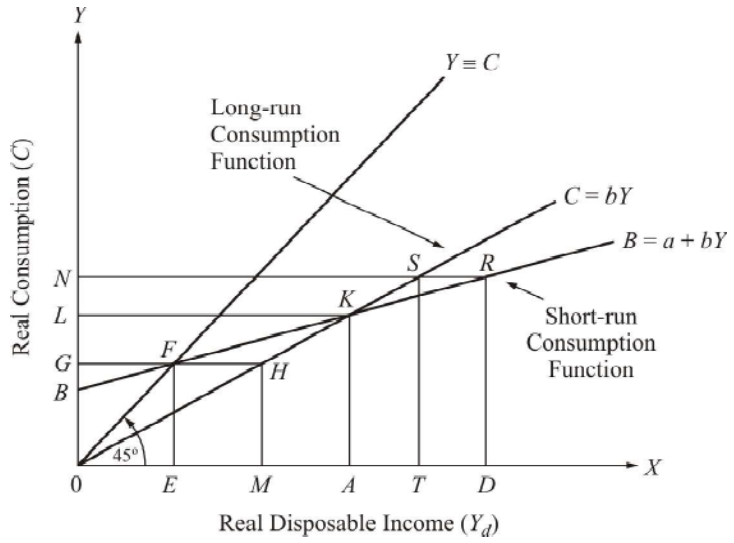


Fig. 3.6

However, those families whose measured income is less than the average—for instance, families with 0*E* measured income—include more than an average proportion of the families with the negative transitory income component. Their permanent income is, however, 0*M* because only against this permanent income can these families sustain the consumption of 0*G* (= *MH*) amount permanently. 0*G* is the measured consumption against the measured income 0*E*. Similarly, for those families whose measured income is 0*D*, their measured consumption is only 0*N* (= *DR*). This amount of consumption on a permanent basis can, however, be afforded only by 0*T* permanent income which is less than the 0*D* measured income showing that there is a positive transitory income component in their measured income and this transitory component equals *TD* amount.

To prove the consistency of his hypothesis, Milton Friedman has collected data from time-series data beginning from 1897. The permanent income hypothesis is not, however, unassailable. Agood deal of highly technical critical literature already exists.[8] Worthy of particular criticism are the assumptions that the 'permanent' and 'transitory' components of income and consumption are not correlated and that transitory consumption and the transitory income are uncorrelated, being independent of each other. In other words, it is asserted that a household unit is unresponsive to either a positive or negative transitory component in its measured income with regard to its consumption spending decisions. The household neither reduces its consumption when its measured income in any given year falls nor does it increase at all its total spending on consumption when its measured income in any given year increases unexpectedly. In short, unexpected increases or decreases in the observed or measured income are wholly saved or dissaved— the marginal propensity to consume out of the transitory income is zero. Friedman's assumption that an increase in the transitory component of measured income will cause no increase in consumption, i.e., it, will all be saved, runs counter to the accepted rational consumer behaviour. Critics have questioned the empirical validity of such an assertion. Criticizing the permanent income hypothesis, Houthakker has stated that according to the hypothesis 'the man who has a lucky day at the races does not buy his friends a drink and the poor fellow whose wallet is stolen does not postpone the purchase of a new overcoat.' Submitting evidence in support of his argument, Houthakker has stated that 'the lucky winner does not run to the savings bank but to the tavern and the victim of theft does cut his coat according to cloth.'[9]

NOTES

Moreover, Milton Friedman has lumped together the income received from human and non-human forms of wealth in the empirical form used by him to test his hypothesis. The conclusion that the 'rich'and 'poor'spend equal fractions of their incomes on consumption is in conflict with the observed behaviour of ordinary family units. This conclusion is based on the argument that the sole purpose of saving is to provide for future consumption for the family. The family saves primarily to even out consumption over a long period of time. Notwithstanding that this behaviour is necessary if the family units at all the income levels want to even out their consumption, the economists have questioned the empirical content of this assumed behaviour. Granting that families make some attempt to even out consumption in this manner, still it cannot be accepted that the preference for the present goods over future goods for the low income families is the same as it is for the high income families. Notwithstanding the strong urge among the low income families to provide for future consumption, it is unthinkable that these low income families would be able to save the same fraction of their low income as is saved by the high income families. It is far more difficult to resist the pressure to increase the present consumption at the low levels of current income than it is at the higher levels of current income.

The concept of permanent income also involves expected or anticipated income and this cannot be estimated in any direct manner. In his time-series data analysis, Milton Friedman has taken permanent income of any one year as a weighted average of the actual or measured income over a 17-year period. Measured income of the current year is assigned a weight of 33 per cent, that of the preceding year of 22 per cent and so on for the other 15 preceding years with the weights falling rapidly In this manner.[10] Friedman finds that, excluding the war-years, for the period 1905-51 the permanent consumption-income ratio was almost constant at around 0.88. There is, however, little rationale behind the weights chosen to be given to different years' measured incomes for estimating the permanent income of any given year. Albert Ando and Franco Modigliani have discarded Milton Friedman's proposition that the expected or permanent income can be measured as an exponentially weighted average of the past years' incomes.

The income-consumption relationship asserted by the permanent income hypothesis could be subjected to quite precise tests only if it was possible to ascertain what part of each year's aggregate measured income was permanent and what part was transitory. In this way, we would have in our possession the time-series data of permanent income and permanent consumption making it possible for us to test the income-consumption relationship set forth in the hypothesis. Such a separation of each year's aggregate measured income into permanent and temporary components requires this separation to be made on individual basis for each year's measured income. Unfortunately, no such data on an individual year basis is readily available at hand.

Eulogizing Milton Friedman's work, Michael K Evans has stated: 'Without making a final judgement on whether the strict terms of the permanent income hypothesis all hold, it can be fairly said that the weight of the evidence supports this theory. Even if parts of the hypothesis are ultimately shown to be incorrect, Friedman's formulation has reshaped and redirected much of the research on the consumption function. It is indeed unusual to discuss the consumption function today without referring to Friedman's terms of reference,'[11]

Check Your Progress

5. What is life-cycle hypothesis also known as?
6. What is permanent income hypothesis?

3.6 SUMMARY

NOTES

In this unit, you have learnt that,

- An enquiry about the determinants of consumers' demand may be restated in terms of the concept of the 'consumption function' which denotes the general income-consumption relationship.
- The most widely accepted hypothesis about consumption is that its major determinant is the disposable personal income.
- The income-consumption relationship, for which the economists have coined the term 'consumption function', was originally called 'the propensity to consume' by John Maynard Keynes.
- The concept of the consumption function marks an important development in modern macroeconomic theory.
- Its discovery, inclusion and treatment as an important edifice of macroeconomic theory, and also its clear formulation owe much to the intelligence and clear insight of Keynes.
- A high rate of interest induced the consumers to save more by postponing consumption while a low rate of interest discouraged saving.
- In the classical economic theory, saving was assumed to be a positive function of the rate of interest.
- In the classical economic analysis, consumption was not a function of income because income was not a variable. In a fully employed economy, income was a constant.
- The post-Keynesian empirical studies made on the income-consumption relationship show that the short-run consumption function has a smaller slope relatively to the slope of the long-run consumption function.
- Different hypothesis have been developed by the economists in order to explain the apparent contradiction between the short-run non-proportional and the long-run proportional income-consumption relationship.
- Reconciliation of the short-run and long-run consumption functions was first attempted by James S Duesenberry in his work entitled *Income, Saving and the Theory of Consumer Behaviour* published in 1949.
- Duesenberry's hypothesis asserts that the consumption-income ratio depends on the ratio of the current income to the previous peak income.
- Ando and Modigliani posit a consumption function in which individual consumption depends on the existing net worth (wealth) plus the present value of all the current and future labour earnings, rate of return on capital and the age of the individual.
- The Modigliani-Brumberg-Ando hypothesis is known as the life-cycle hypothesis of consumption.
- According to Modigliani and Ando, age is a crucial variable in determining the relationship of consumption to the measured income and perhaps the relationship between consumption and wealth.

NOTES

- According to Friedman, permanent income is 'the amount a consumer unit could consume (or believes that it could) while maintaining its wealth intact' while permanent consumption is 'the value of the services that it is planned to consume during the period in question.'
- The permanent income hypothesis states that the ratio of permanent consumption to permanent income is constant regardless of the level of permanent income.
- According to Milton Friedman, a family unit's measured or observed income for any given year will be larger or smaller than its permanent income depending on the presence of the positive or negative transitory income component.
- The income-consumption relationship asserted by the permanent income hypothesis could be subjected to quite precise tests only if it was possible to ascertain what part of each year's aggregate measured income was permanent and what part was transitory.

3.7 KEY TERMS

- **Relative income hypothesis**: Relative income hypothesis states that the satisfaction (or utility) an individual derives from a given consumption level depends on its relative magnitude in the society (e.g. *relative* to the average consumption) rather than its absolute level.
- **Life cycle hypothesis:** In economics, the life-cycle hypothesis (LCH) is a model that strives to explain the consumption patterns of individuals.
- **Permanent income hypothesis:** The permanent income hypothesis (PIH) is an economic theory attempting to describe how agents spread consumption over their lifetimes.

3.8 ANSWERS TO 'CHECK YOUR PROGRESS'

1. The post-Keynesian empirical studies made on the income-consumption relationship show that the short-run consumption function has a smaller slope relatively to the slope of the long-run consumption function.
2. The non-proportional income-consumption relationship was controverted by Simon Kuznet's now well-known study of the national income and consumption expenditure for the American economy during the period 1869 to 1929 which showed that the ratio of total consumption to national income (*C/Y*) had remained constant while income had quadrupled.
3. The relative income hypothesis, is based on the assumptions that (1) the consumption behaviour of individuals is interdependent and not independent; and (2) consumption relations are irreversible over time.
4. One drawback of the relative income hypothesis is that changes in consumption and income will always take place in the same direction. Yet at times, mild decreases in income and increases in consumption might often occur concomitantly.
5. Life-cycle hypothesis is also known as Modigliani-Brumberg-Ando hypothesis.
6. The permanent income hypothesis states that the ratio of permanent consumption to permanent income is constant regardless of the level of permanent income.

NOTES

3.9 QUESTIONS AND EXERCISES

Short-Answer Questions

1. State in brief the method of reconciliation of the short period and long period consumption functions.
2. What are the drawbacks of relative income hypothesis?
3. Why is age considered a crucial variable in determining the relationship of consumption to the measured income?
4. State the different equations of permanent income hypothesis.

Long-Answer Questions

1. Discuss the long-run and the short-run income-consumption relationship.
2. Although the short period consumption function is of the form $C = a + bY$ showing a non-proportional relationship between consumption and income, yet the long period consumption function is of the form C.= bY, giving a proportional relationship between the consumption and income. Discuss the various hypotheses which have been advanced to explain this paradox.
3. Discuss the absolute income, the permanent income and the relative income hypotheses about the consumption function.
4. What do you understand by the term 'consumption function'? Explain briefly the permanent income hypothesis propounded by Milton Friedman to reconcile the short-run non-proportional income-consumption relationship with the long-run proportional income-consumption relationship.

3.10 FURTHER READING

Mankiw, N Gregory. 2010. *Macroeconomics.* Worth Publishers, New York.

Shapiro, Edward. 1996. *Macroeconomic Analysis.* Galgotia Publication, New Delhi.

Branson, WH. 1979. *Macroeconomic Theory and Policy*. Universal Book Stall, New Delhi.

Jha, R. 1999. *Contemporary Macroeconomic Theory and Policy*. New Age International, New Delhi.

Gupta, SB. 1995. *Monetary Economics: Instruments and Policy*. S Chand & Co. New Delhi.

Endnotes

[1] Simon Kuznets, *Uses of National Income in Peace and War,* NBER, 1942, p. 32.

[2] Raymond W Goldsmith, *A Study of Saving in the United States,* Volume. 1, 1955, p. 22.

[3] In the beginning, Duesenberry had referred to the *S/Y* ratio rather than to the *C/Y* ratio. He, however, subsequently referred to the *C/Y* ratio.

[5] Franco Modigliani, 'Fluctuations in the Saving-Income Ratio: A Problem in Economic Forecasting,' published in *Studies in Income and Wealth,* Volume II, NBER, 1949; Franco Modigliani and Richard E Brumberg, 'Utility Analysis and the Consumption Function: An Interpretation of Cross-section Data,' published in Kenneth K Kurihara (ed.), *Post-Keynesian Economics,* Rutgers, 1954; Albert Ando and Franco Modigliani, 'The Life-Cycle Hypothesis of Saving: Aggregate Implications and Tests,' *The American Economic Review,* Volume 53, No. 1, March 1963, p. 55–84.

[6] Milton Friedman, *A Theory of the Consumption Function,* 1957, p. 10–11.

NOTES

[7] According to Milton Friedman, consumption is the spending on services and non-durable goods plus the depreciation of consumer durable goods. A net addition to the family unit's stock of durable goods is treated as saving.

[8] H S Houthakker, 'The Permanent Income Hypothesis,' *The American Economic Review,* June 1958: R Eisner, 'The Permanent Income Hypothesis: Comment,' *The American Economic Review,* December 1958; I Friend and I B Kravis, 'Consumption Patterns and Permanent Income,' *The American Economic- Review,* May 1957; R Bodkin, 'Windfall Income and Consumption,' *The American Economic Review.* September 1959; M E Krienen, 'Windfall Income and Consumption,' *The American Economic Review*. June 1961.

[9] H S, Houthakker, *op. cit.,* p. 398 and 404.

[10] It does not amount to saying that the current year's permanent income is 33 per cent of the current year's measured income since it also includes 33 per cent of the previous year's income and the declining percentages of the 15 preceding years' measured income, When these amounts are added, the permanent income of the current year may exceed the current year's measured income. It will be so if the current year's measured income has substantially declined in relation to the preceding year's measured income.

[11] Michael K Evans, *Macroeconomic Activity,* 1969, p. 34.

UNIT 4 THEORIES OF INVESTMENT

Structure

NOTES

4.0 INTRODUCTION

Most financial decisions, such as the purchase of assets or procurement of funds, affect the firm's cash flows in different time periods. The firm's cash balance will increase at the time shares are issued, but, as the firm pays dividends in future, the outflow of cash will occur. Sound decision making requires that the cash flows, which differ in timing and risk, are not directly comparable.

Cash flows become logically comparable when they are appropriately adjusted for their differences in timing and risk. In this unit, the different theories of investment and risk are discussed in detail.

4.1 UNIT OBJECTIVES

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

- Analyse the present value criterion for investment
- Discuss the payback period and the internal rate of return method of evaluating investment proposals
- Discuss the concept of marginal efficiency of investment (MEI)
- Describe the principle of acceleration and its derivation from profit maximizing behaviour

4.2 INVESTMENT DEMAND: PRESENT VALUE CRITERION FOR INVESTMENT

The Net Present Value (NPV) is the difference between the current value of the cash that is flowing in (inflow) and the current value of cash that is flowing out (outflows). NPV is used for the purpose of budgeting of capital to analyze how profitable a project or a projected investment will be.

NOTES

Given below is the formula used to calculate NPV.

$$NPV = \sum_{t-1}^{T} \frac{C_t}{(1+r)^t} - C_o$$

In the formula:

C_t denotes net cash inflow for the period *t*

C_o denotes the cost of the initial investment

r denotes the discount rate

t denotes the number of time periods

If there is a positive net present value, it implies that the projected earnings that will accrue from an investment or project (in current currency value) will be in excess of the cost that has been anticipated for the investment or project (in current currency value). Mostly, it is expected that an investment or project that shows a positive NPV will be one that is profitable while on the other hand that which shows a negative NPV will cause a net loss. The concept given above is derived from the Net Present Value Rule, according to which one should make only those investments that provide a positive NPV value.

In case the question of investment pertains to a merger or acquisition, it is recommended that the Discounted Cash Flow (DCF) metric should be used.

Besides the formula itself, many times it is possible to calculate net present value by using tables, spreadsheets and other tools available in the market or on the Internet.

The task of determining the value of a project is extremely challenging. This is due to the fact that there exist varied ways of measuring the value of future cash flows. Due to time value of money (TVM), a specific sum of money at the current time has much more value than it will have at a point in time in future. This is due both to the earnings that could potentially be made using the money during the intervening period and due to inflation. That is to say, a thousand rupees earned at a point in time in future will hold a lower value than it does at the present time. To account for this difference, the NPV formula's discount rate element is used. It is possible that companies have different means to identify discount rate. Some common methods used to determine discount rate are: using the expected return of choices of investment that have a similar level of risk.

A major problem with using NPV in finding out the profitability of an investment is that there are a number of estimates and assumptions that NPV relies on. This can easily cause errors to creep in. Some of the factors that need to be estimated for NPV are: projected return, discount rate and cost or investment. Oftentimes, one might need to make unforeseen investments in a project to get it started or even during the actual running of the project or its culmination. This cannot be built into the calculation at the stage of calculating project profitability since it is unknown.

Furthermore, it is not essential that discount rates and cash inflow estimates are able to inherently account for the risk associated with the project and these might assume the maximum possible cash inflows over an investment period. This may occur as a means of artificially increasing investor confidence. So, there might be a need to adjust these factors for accounting for unexpected losses or even costs or losses or for projections of cash inflow that are majorly optimistic.

Check Your Progress

1. State the uses of Net Present Value (NPV).
2. State the formula that is used to calculate Net Present Value.

4.3 PAYBACK PERIOD

NOTES

The payback (PB) is one of the most popular and widely-recognized traditional methods of evaluating investment proposals. Payback is the number of years required to recover the original cash outlay invested in a project. If the project generates constant annual cash inflows, the payback period can be computed by dividing cash outlay by the annual cash inflow. That is:

$$\text{Payback} = \frac{\text{Initial investment}}{\text{Annual cash inflow}} = \frac{C_0}{C} \quad \ldots(1)$$

Illustration 4.1: Payback (constant cash flows)

Assume that a project requires an outlay of `50,000 and yields annual cash inflow of `12,500 for 7 years. The payback period for the project is:

$$\text{PB} = \frac{\text{`}\ 50,000}{\text{`}\ 12,000} = 4 \text{ years}$$

Unequal cash flows: In case of unequal cash inflows, the payback period can be found out by adding up the cash inflows until the total is equal to the initial cash outlay. Consider the following example.

Illustration 4.2: Payback (uneven cash flows)

Suppose that a project requires a cash outlay of `20,000, and generates cash inflows of ` 8,000; `7,000; `4,000; and `3,000 during the next 4 years. What is the project's payback? When we add up the cash inflows, we find that in the first three years `19,000 of the original outlay is recovered. In the fourth year cash inflow generated is `3,000 and only `1,000 of the original outlay remains to be recovered. Assuming that the cash inflows occur evenly during the year, the time required to recover `1,000 will be (`1,000/`3,000) × 12 months = 4 months. Thus, the payback period is 3 years and 4 months.

Acceptance Rule

Many firms use the payback period as an investment evaluation criterion and a method of ranking projects. They compare the project's payback with a predetermined and standard payback. The project would be accepted if its payback period is less than the maximum or standard payback period set by management. As a ranking method, it gives highest ranking to the project, which has the shortest payback period and lowest ranking to the project with highest payback period. Thus, if the firm has to choose between two mutually exclusive projects, the project with shorter payback period will be selected.

Evaluation of Payback

Payback is a popular investment criterion in practice. It is considered to have certain benefits.

- **Simplicity:** The most significant merit of payback is that it is simple to understand and easy to calculate. The business executives consider the simplicity of method as a virtue. This is evident from their heavy reliance on it for appraising investment proposals in practice.

NOTES

- **Short-term effects:** A company can have more favourable short-run effects on earnings per share by setting up a shorter standard payback period. It should, however, be remembered that this may not be a wise long-term policy as the company may have to sacrifice its future growth for current earnings.
- **Risk shield:** The risk of the project can be tackled by having a shorter standard payback period as it may ensure guarantee against loss. A company has to invest in many projects where the cash inflows and life expectancies are highly uncertain. Under such circumstances, payback may become important, not so much as a measure of profitability but as a means of establishing an upper bound on the acceptable degree of risk.
- **Liquidity:** The emphasis in payback is on the early recovery of the investment. Thus, it gives an insight into the liquidity of the project. The funds so released can be put to other uses. In spite of its simplicity and the so-called virtues, the payback may not be a desirable investment criterion since it suffers from a number of serious limitations:
- **Cash flows after payback:** Payback fails to take account of the cash inflows earned after the payback period. For example, consider the following projects *X* and *Y*:

	Cash Flows (`)					
Project	C_0	C_1	C_2	C_3	*Payback*	*NPV*
X	– 4,000	0	4,000	2,000	2 years	+ 806
Y	– 4,000	2,000	2,000	0	3 years	– 530

As per the payback rule, both the projects are equally desirable since both return the investment outlay in two years. If we assume an opportunity cost of 10 per cent, Project *X* yields a positive net present value of `806 and Project *Y* yields a negative net present value of `530. As per the NPV rule, Project *X* should be accepted and Project *Y* rejected. Payback rule gave wrong results, because it failed to consider `2,000 cash flow in third year for Project *X*.

- **Cash flows ignored:** Payback is not an appropriate method of measuring the profitability of an investment project as it does not consider all cash inflows yielded by the project. Considering Project *X* again, payback rule did not take into account its entire series of cash flows.
- **Cash-flow patterns:** Payback fails to consider the pattern of cash inflows, i.e., magnitude and timing of cash inflows. In other words, it gives equal weights to returns of equal amounts even though they occur in different time periods. For example, compare the following projects *C* and *D,* where they involve equal cash outlay and yield equal total cash inflows over equal time periods:

	Cash Flows (`)					
Project	C_0	C_1	C_2	C_3	*Payback*	*NPV*
C	– 5,000	3,000	2,000	2,000	2 years	+ 881
D	– 5,000	2,000	3,000	2,000	2 years	+ 798

Using payback period, both projects are equally desirable. But Project *C should be* preferable as larger cash inflows' come earlier in its life. This is indicated by the NPV

rule; project *C* has higher NPV (`881) than Project *D* (`798) at 10 per cent opportunity cost. It should be thus clear that payback is not a measure of profitability. As such, it is dangerous to use it as a decision criterion.

NOTES

- **Administrative difficulties:** A firm may face difficulties in determining the maximum acceptable payback period. There is no rational basis for setting a maximum payback period. It is generally a subjective decision.
- **Inconsistent with shareholder value:** Payback is not consistent with the objective of maximizing the market value of the firm's shares. Share values do not depend on payback periods of investment projects.

Let us re-emphasize that the payback is not a valid method for evaluating the acceptability of the investment projects. It can, however, be used along with the NPV rule as a first step in roughly screening the projects. In practice, the use of DCF techniques has been increasing, but payback continues to remain a popular and primary method of investment evaluation (Exhibit 4.1).

Exhibit 4.1 Capital Budgeting Methods in Practice

- In a study of the capital budgeting practices of 14 medium to large size companies in India, it was found that all companies, except one, used payback. With payback and/or other techniques, about two thirds of companies used IRR and about two-fifths NPV. IRR was found to be the second most popular method.
- The reasons for the popularity of payback in order of significance were stated to be its *simplicity to use and understand, its emphasis on the early recovery of investment* and *focus on risk.*
- It was also found that one third of companies always insisted on the computation of payback for all projects, one-third for majority of projects and remaining for some of the projects. For about two thirds of companies standard payback ranged between 3 and 5 years.
- Reasons for the secondary role of DCF techniques in India included *difficulty in understanding and using these techniques, lack of qualified professionals* and *unwillingness of top management* to use DCF techniques. One large manufacturing and marketing organization mentioned that conditions of its business were such that DCF techniques were not needed. Yet another company stated that replacement projects were very frequent in the company, and it was not considered necessary to use DCF techniques for evaluating such projects.

Source: Pandey, I M, "Capital Budgeting Practices of Indian Companies", *MDI Management Journal*, Vol. 2, No. 1 (Jan. 1989).

Payback Reciprocal and the Rate of Return

Payback is considered theoretically useful in a few situations. One significant argument in favour of payback is that its reciprocal is a good approximation of the rate of return under certain conditions.

The payback period is defined as follows:

$$\text{Payback} = \frac{\text{Initial investment}}{\text{Annual cash inflow (annuity)}} = \frac{C_0}{C} \qquad \ldots(2)$$

The formula for the present value of an annuity is given by the following equation:

$$C_0 = C\left[\frac{1-\frac{1}{(1+r)^n}}{r}\right] = \frac{C}{r} - \frac{C}{r}\left[\frac{1}{(1+r)^n}\right]$$

NOTES

Multiplying both sides by r, we get

$$rC_0 = C - C\left[\frac{1}{(1+r)^n}\right]$$

Solving for r, we find

$$r = \frac{C}{C_0} - \frac{C}{C_0}\left[\frac{1}{(1+r)^n}\right] \quad \ldots(3)$$

where C_0 is the initial investment, C is annual cash inflow, r is rate of return and n is the life of investment.

In Equation (3), the first right-hand term is the reciprocal of the payback period. The second right-hand term is payback reciprocal multiplied by $1/(1 + r)^n$. If n is very large or extends to infinity, the second term becomes insignificant (almost equal to zero), and we are left with the term C/C_0. Thus, IRR is equal to the reciprocal of payback.

The reciprocal of payback will be a close approximation of the internal rate of return if the following two conditions are satisfied:

- The life of the project is large or at least twice the payback period.
- The project generates equal annual cash inflows.

The payback reciprocal is a useful technique to quickly estimate the true rate of return. But its major limitation is that every investment project does not satisfy the conditions on which this method is based. When the useful life of the project is not at least twice the payback period, the payback reciprocal will always exceed the rate of return. Similarly, it cannot be used as an approximation of the rate of return if the project yields uneven cash inflows.

Discounted Payback Period

One of the serious objections to the payback method is that it does not discount the cash flows for calculating the payback period. We can discount cash flows and then calculate the payback. The discounted payback period is the number of periods taken in recovering the investment outlay on the present value basis. The discounted payback period still fails to consider the cash flows occurring after the payback period.

Let us consider an example. Projects *P* and *Q* involve the same outlay of `4,000 each. The opportunity cost of capital may be assumed as 10 per cent. The cash flows of the projects and their discounted payback periods are shown in Table 4.1.

Table 4.1 *Discounted Payback Illustrated*

	Cash Flows (`)					*Simple*	*Discounted*	*NPV at*
	C_0	C_1	C_2	C_3	C_4	*PB*	*PB*	*10%*
P	– 4,000	3,000	1,000	1,000	1,000	2 yrs	—	—
PV of cash flows	– 4,000	2,727	826	751	683		2.6 yrs	987
Q	– 4,000	0	4,000	1,000	2,000	2 yrs	—	—
PV of cash flows	– 4,000	0	3,304	751	1,366		2.9 yrs	1,421

The projects are indicated of same desirability by the simple payback period. When cash flows are discounted to calculate the discounted payback period, Project *P* recovers the investment outlay faster than Project *Q*, and therefore, it would be preferred over Project *Q*. Discounted payback period for a project will be always higher than simple payback period, because its calculation is based on the discounted cash flows.

Discounted payback rule is better as it discounts the cash flows until the outlay is recovered. But, it does not help much. It does not take into consideration the entire series of cash flows. It can be seen in our example that if we use the NPV rule, Project *Q* (with higher discounted payback period) is better.

4.3.1 Internal Rate of Return

The internal rate of return (IRR) method is another discounted cash flow technique, which takes account of the magnitude and timing of cash flows. Other terms used to describe the IRR method are yield on an investment, marginal efficiency of capital, rate of return over cost, time-adjusted rate of internal return and so on. The concept of internal rate of return is quite simple to understand in the case of a one-period project. Assume that you deposit `10,000 with a bank and would get back `10,800 after one year. The true rate of return on your investment would be:

$$\text{Rate of return} = \frac{10{,}800 - 10{,}000}{10{,}000} = \frac{10{,}800}{10{,}000} - 10{,}000 = 1.08 - 1 = 0.08 \text{ or, } \quad 8\%$$

The amount that you would obtain in the future (`10,800) would consist of your investment (`10,000) plus return on your investment (0.08 × `10,000):

$$10{,}000\,(1.08) = 10{,}800$$

$$10{,}000 = \frac{10{,}800}{(1.08)}$$

You may observe that the rate of return of your investment (8 %) makes the discounted (present) value of your cash inflow (`10,800) equal to your investment (`10,000).

We can now develop a formula for the rate of return (r) on an investment (C_0) that generates a single cash flow after one period (C_1) as follows:

$$r = \frac{C_1 - C_0}{C_0} = \frac{C_1}{C_0} - 1 \qquad \ldots(4)$$

Equation (4) can be rewritten as follows:

$$\frac{C_1}{C_0} = 1 + r$$

$$C_0 = \frac{C_1}{(1+r)} \qquad \ldots(5)$$

From Equation (5), you may notice that the rate of return, r, depends on the project's cash flows, rather than any outside factor. Therefore, it is referred to as the internal rate of return. The internal rate of return (IRR) is the rate that equates the investment outlay with the present value of cash inflow received after one period. This also implies that the rate of return is the discount rate which makes NPV = 0. There is no satisfactory way of defining the true rate of return of a long-term asset. IRR is the best available concept. We shall see that although it is a very frequently used concept in finance, yet at times, it can be a misleading measure of investment worth. IRR can be determined by solving the following equation for r:

$$C_0 = \frac{C_1}{(1+r)} + \frac{C_2}{(1+r)^2} + \frac{C_3}{(1+r)^3} + \square + \frac{C_n}{(1+r)^n}$$

$$C_0 = \sum_{t=1}^{n} \frac{C_t}{(1+r)^t}$$

NOTES

It can be noticed that the IRR equation is the same as the one used for the NPV method. In the NPV method, the required rate of return, *k*, is known and the net present value is found, while in the IRR method, the value of *r* has to be determined at which the net present value becomes zero.

Uneven Cash Flows: Calculating IRR by Trial and Error

The value of *r* in Equation (6) can be found out by trial and error. The approach is to select any discount rate to compute the present value of cash inflows. If the calculated present value of the expected cash inflow is lower than the present value of cash outflows, a lower rate should be tried. On the other hand, a higher value should be tried if the present value of inflows is higher than the present value of outflows. This process will be repeated unless the NPV becomes zero. The following illustration explains the procedure of calculating IRR.

Illustration 4.3: Trial and error method for calculating IRR

A project costs `16,000 and is expected to generate cash inflows of `8,000, `7,000 and `6,000 at the end of each year for next 3 years. We know that IRR is the rate at which project will have a zero NPV. As a first step, we try (arbitrarily) a 20 per cent discount rate. The project's NPV at 20 per cent is:

$$\begin{aligned}\text{NPV} &= -\text{`}16{,}000 + \text{`}8{,}000(\text{PVF}_{1,0.20}) + \text{`}7{,}000(\text{PVF}_{2,0.20}) + \text{`}6{,}000(\text{PVF}_{3,0.20}) \\ &= -\text{`}16{,}000 + \text{`}8{,}000 \times 0.833 + \text{`}7{,}000 \times 0.694 + \text{`}6{,}000 \times 0.579 \\ &= -\text{`}16{,}000 + \text{`}14{,}996 = -\text{`}1{,}004\end{aligned}$$

A negative NPV of `1,004 at 20 per cent indicates that the project's true rate of return is lower than 20 per cent. Let us try 16 per cent as the discount rate. At 16 per cent, the project's NPV is:

$$\begin{aligned}\text{NPV} &= -\text{`}16{,}000 + \text{`}8{,}000(\text{PVF}_{1,0.16}) + \text{`}7{,}000(\text{PVF}_{2,0.16}) + \text{`}6{,}000(\text{PVF}_{3,0.16}) \\ &= -\text{`}16{,}000 + \text{`}8{,}000 \times 0.862 + \text{`}7{,}000 \times 0.743 + \text{`}6{,}000 \times 0.641 \\ &= -\text{`}16{,}000 + \text{`}15{,}943 = -\text{`}57\end{aligned}$$

Since the project's NPV is still negative at 16 per cent, a rate lower than 16 per cent should be tried. When we select 15 per cent as the trial rate, we find that the project's NPV is `200:

$$\begin{aligned}\text{NPV} &= -\text{` }16{,}000 + \text{` }8{,}000(\text{PVF}_{1,0.15}) + \text{` }7{,}000(\text{PVF}_{2,0.15}) + \text{` }6{,}000(\text{PVF}_{3,0.15}) \\ &= -\text{` }16{,}000 + \text{` }8{,}000 \times 0.870 + \text{` }7{,}000 \times 0.756 + \text{` }6{,}000 \times 0.658 \\ &= -\text{` }16{,}000 + \text{` }16{,}200 = \text{` }200\end{aligned}$$

The true rate of return should lie between 15–16 per cent. We can find out a close approximation of the rate of return by the method of linear interpolation as follows:

		Difference
PV required	`16,000	
		200
PV at lower rate, 15%	16,200	
		257
PV at higher rate, 16%	15,943	

$r = 15\% + (16\% - 15\%)200/257$
$= 15\% + 0.80\% = 15.8\%$

Level Cash Flows

An easy procedure can be followed to calculate the IRR for a project that produces level or equal cash flows each period. To illustrate, let us assume that an investment would cost `20,000 and provide annual cash inflow of `5,430 for 6 years. If the opportunity cost of capital is 10 per cent, what is the investment's NPV? The `5,430 is an annuity for 6 years. The NPV can be found as follows:

$$NPV = -\text{`}\ 20{,}000 + \text{`}\ 5{,}430(PVFA_{6,0.10}) = -\text{`}\ 20{,}000 + \text{`}\ 5{,}430 \times 4.355 = \text{`}\ 3{,}648$$

How much is the project's IRR? The IRR of the investment can be found out as follows:

$$NPV = -\text{`}\ 20{,}000 + \text{`}\ 5{,}430(PVFA_{6,r}) = 0$$

$$\text{`}\ 20{,}000 = \text{`}\ 5{,}430(PVFA_{6,r})$$

$$PVFA_{6,r} = \frac{\text{`}\ 20{,}000}{\text{`}\ 5{,}430} = 3.683$$

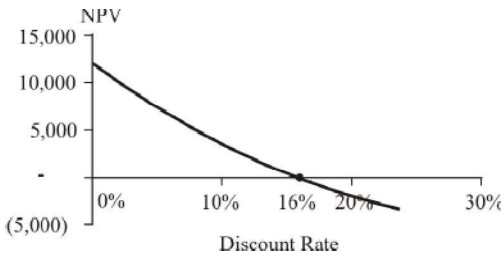
The rate, which gives a PVFA of 3.683 for 6 years, is the project's internal rate of return.

NPV Profile and IRR

We repeat to emphasize that NPV of a project declines as the discount rate increases, and for discount rates higher than the project's IRR, NPV will be negative. NPV profile of the project at various discount rates is shown in Table 4.2. At 16 per cent, the NPV is zero; therefore, it is the IRR of the project. As you may notice, we have used the Excel spreadsheet to make the computations and create the chart using the Excel chart wizard.

Table 4.2 *NPV Profile*

1	*NPV Profile*		
2	*Cash Flow*	*Discount rate*	*NPV*
3	– 200,00	0%	12,580
4	5,430	5%	7,561
5	5,430	10%	3,649
6	5,430	15%	550
7	5,430	16%	0
8	5,430	20%	(1,942)
9	5,430	25%	(3,974)



Acceptance Rule

The accept-or-reject rule, using the IRR method, is to accept the project if its internal rate of return is higher than the opportunity cost of capital ($r > k$). Note that k is also known as the required rate of return, or the cut-off, or hurdle rate. The project shall be rejected if its internal rate of return is lower than the opportunity cost of capital ($r < k$). The decision maker may remain indifferent if the internal rate of return is equal to the opportunity cost of capital. Thus, the IRR acceptance rules are:

- Accept the project when $r > k$
- Reject the project when $r < k$
- May accept the project when $r = k$

The reasoning for the acceptance rule becomes clear if we plot NPVs and discount rates for the project given in Table 4.2. It can be seen that if the discount rate is less than 16 per cent IRR, then the project has positive NPV; if it is equal to IRR, the project has

NOTES

a zero NPV; and if it is greater than IRR, the project has negative NPV. Thus, when we compare IRR of the project with the opportunity cost of capital, we are in fact trying to ascertain whether the project's NPV is positive or not. In case of independent projects, IRR and NPV rules will give the same results if the firm has no shortage of funds.

Exhibit 4.2:
CALCULATION OF IRR

We can easily calculate IRR using the Excel function for IRR. The spreadsheet below gives the cash flows of the project. We write the IRR formula in column C7: =IRR(C3:C6). The project cash flows, including the cash outlay in the beginning (C_0 in year 0) are contained in column C3 through column C6. It is optional to include the "guess" rate in the formula.

	A	B	C	D
1	**IRR of An Investment Project**			
2	Year	Cash Flow (Rs)		
3	0	C_0	−16000	
4	1	C_1	8000	
5	2	C_2	7000	
6	3	C_3	6000	
7		IRR	15.8%	IRR(C3:C6)

Evaluation of IRR Method

IRR method is like the NPV method. It is a popular investment criterion since it measures profitability as a percentage and can be easily compared with the opportunity cost of capital. IRR method has following merits:

- **Time value:** The IRR method recognizes the time value of money.
- **Profitability measure:** It considers all cash flows occurring over the entire life of the project to calculate its rate of return.
- **Acceptance rule:** It generally gives the same acceptance rule as the NPV method.
- **Shareholder value:** It is consistent with the shareholders' wealth-maximization objective. Whenever a project's IRR is greater than the opportunity cost of capital, the shareholders' wealth will be enhanced.

Like the NPV method, the IRR method is also theoretically a sound investment evaluation criterion. However, IRR rule can give misleading and inconsistent results under certain circumstances. Here we briefly mention the problems that IRR method may suffer from.

- **Multiple rates:** A project may have multiple rates, or it may not have a unique rate of return. As we explain later on, these problems arise because of the mathematics of IRR computation.
- **Mutually exclusive projects:** It may also fail to indicate a correct choice between mutually exclusive projects under certain situations. This pitfall of the IRR method is elaborated later on in this chapter.
- **Value additivity:** Unlike in the case of the NPV method, the value-additivity principle does not hold when the IRR method is used—IRRs of projects do not add. Thus, for Projects *A* and *B*, IRR(*A*) + IRR(*B*) need not be equal to IRR (*A* + *B*). Consider the following example.

The NPV and IRR of Projects *A* and *B* are given as follows:

Project	C_0 (`)	C_1 (`)	*NPV @ 10%* (`)	*IRR* (%)
A	– 100	+ 120	+ 9.1	20.0
B	– 150	+ 168	+ 2.7	12.0
A + *B*	– 250	+ 288	+ 11.8	15.2

NOTES

It can be seen from the example that NPVs of projects add:

$$\text{NPV}(A) + \text{NPV}(B) = \text{NPV}(A + B) = 9.1 + 2.7 = 11.8, \text{ while}$$
$$\text{IRR}(A) + \text{IRR}(B) \neq \text{IRR}(A + B) = 20\% + 12\% \neq 15.2\%$$

4.4 MARGINAL EFFICIENCY OF INVESTMENT (MEI)

In economics, expected rates of return on investment as additional units of investment are made under specified conditions and over a stated period of time. A comparison of these rates with the going rate of interest may be used to indicate the profitability of investment. The rate of return is computed as the rate at which the expected stream of future earnings from an investment project must be discounted to make their present value equal to the cost of the project.

As the quantity of investment increases, the rates of return from it may be expected to decrease because the most profitable projects are undertaken first. Additions to investment will consist of projects with progressively lower rates of return. Logically, investment would be undertaken as long as the marginal efficiency of each additional investment exceeded the interest rate. If the interest rate were higher, investment would be unprofitable because the cost of borrowing the necessary funds would exceed the returns on the investment. Even if it were unnecessary to borrow funds for the investment, more profit could be made by lending out the available funds at the going rate of interest.

The British economist John Maynard Keynes used this concept but coined a slightly different term, the marginal efficiency of capital, in arguing for the importance of profit expectations rather than interest rates as determinants of the level of investment.

4.5 ACCELERATION PRINCIPLE AND ITS DERIVATION FROM PROFIT MAXIMIZING BEHAVIOUR

Although the discussion of the acceleration principle owes much to the pioneering efforts of Erik Lundberg, Paul A Samuelson, Roy F Harrod, John R Hicks, William J Baumol, Richard M Goodwin and others, it is also found in the pre-Keynesian literature. Its origin is traceable in the writings of Albert Aftalion, CF Bickerdike, Ralph George Hawtrey and Ragnar Frisch. The best-known early study of the acceleration principle is found in John Maurice Clark's article titled 'Business Acceleration and the Law of Demand' published in *The Journal of Political Economy* in March 1917. Clark studied the fluctuations in the railroad traffic and the demand for railroad cars. His main conclusion was that the demand for railroad traffic fluctuated more closely with the change in the railroad traffic than with the level of the railroad traffic. This, in brief, is the naive acceleration principle which stresses a certain fixed relationship between changes in the capital stock and changes in the aggregate output.

NOTES

After Clark, not much headway was made until Ragnar Frisch's well-known article published in 1933 opened the vast possibilities of new development on the subject. Among the more recent contributions, the studies made by John R Hicks and Paul A Samuelson have ascribed trade cycle to the interaction of the multiplier and the accelerator. Great interest in the study of acceleration principle—*relation* as Harrod called it—was aroused by Roy F Harrod's excellent work published in 1936. According to Harrod, accelerator played a significant role in the generation of the trade cycle. Highlighting it, Harrod wrote: 'It is a relation which has, indeed, been noted by learned writers often enough Its simplicity, ineluctability, and independence of all special theories as to the workings of the cyclical process demand for it pride of place.'

The acceleration principle states that the demand for capital goods varies directly with the change in the level of aggregate output. The extent of change in the demand for capital goods depends on the capital-output ratio and the change in the level of output. Since the change in aggregate output depends on the change in aggregate expenditure or aggregate demand which itself equals the change in the level of equilibrium income, we might say that total investment in the economy in any given time period depends on the change in the aggregate demand which in equilibrium equals the increase in the national income plus the replacement investment which is assumed constant. Thus the gross investment in the economy during any given time period t will be equal to the increase in the national income during that time period times the capital-output ratio (K/Q) plus the replacement of capital consumed in the process of production. Designating the 'capital-output ratio'or 'capital coefficient' by v, the aggregate income of time periods t and $t-1$ by Y_t and Y_{t-1} respectively and the replacement investment by R, the gross investment (I_g) in any given time period t will be–

$$(I_g)t = v\,(Y_t - Y_{t-1}) + R \quad \text{.....(7)}$$

$$= v\,\Delta\,Y_t + R \quad \text{.....(8)}$$

The average capital-output ratio $v = (K/Q)$ is called the *accelerator* or *relation.*

The rigid acceleration principle in its naive form, however, assumes a given fixed relationship between capital and output. It states that the net induced investment in any given time period t is entirely a function of the growth of final output in that time period, i.e.,

$$(I_n)_t = v\,(Y_t - Y_{t-1}) \quad \text{...(9)}$$

$$(I_n)_t = v\Delta Y_t \quad \text{...(10)}$$

or $$(I_n)_t = \Delta K_t = v\Delta Y_t^{\,3} \quad \text{...(11)}$$

According to equations (10) and (11), the net investment in the economy in any given time period t is acceleration times the change in the aggregate output or income in that time period. Expressed differently, the accelerator states that if the aggregate output stayed at a high level but ceased growing, then the net investment would eventually become zero. This has been shown in Figure 4.1 where so long as the aggregate output shown by the aggregate output curve $0Q$ is increasing over time, the net investment shown by the curve I_nI_n is positive. It is, however, falling because the aggregate output per time period is growing at a diminishing rate. The negative slope of the net investment curve reflects the diminishing positive slope (growth) of the aggregate output curve. When the slope of the aggregate output curve becomes zero, the net investment also becomes zero, i.e., when the aggregate output attains the plateau, the net investment in the economy becomes zero.

NOTES

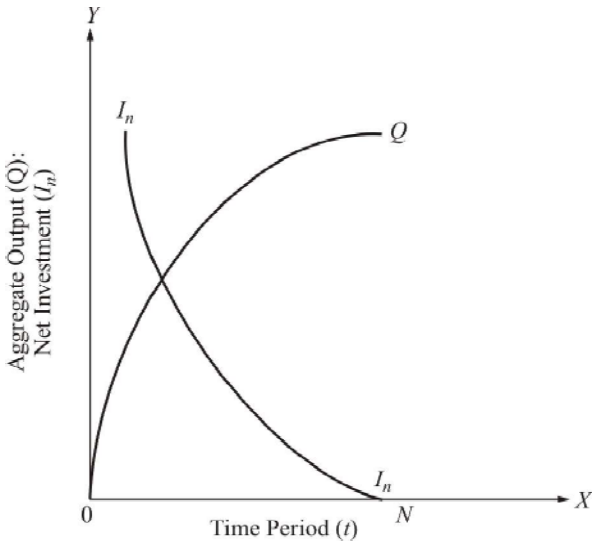


Fig. 4.1

The acceleration principle expressed in equation (11) states that if the existing capital stock of the economy is fully utilized, i.e., if there is no redundant capital in the economy and if the capital-output ratio (K/Y) is fixed, a given increase in the final output will require an increase in the economy's total capital stock which will be equal to the increase in the aggregate output times the accelerator. If the value of the accelerator exceeds one, the required increase in the aggregate capital stock will exceed the increase in the demand for final output. In short, with an accelerator of more than one, a given increase in the aggregate demand for the final output will magnify the derived demand for the capital stock needed to produce the additional output. The working of the acceleration principle can be understood by tracing the changes in the aggregate output and net and gross investment over a number of time periods as shown in Table 4.3 and Figure 4.2.

Table 4.3 *Acceleration Principle*

Time Period	***Aggregate Output (Q)***	***Required Capital (v = 3)***	***Replacement Investment (R)***	***Net Investment (I_n)***	***Gross Investment (I_g)***
1	***2***	***3***	***4***	***5***	***6***
t	200	600	60	0	60
$t + 1$	200	600	60	0	60
$t + 2$	220	660	60	60	120
$t + 3$	250	750	60	90	150
$t + 4$	270	810	60	60	120
$t + 5$	280	840	60	30	90
$t + 6$	270	810	60	−30	30
$t + 7$	250	850	60	−60	0
$t + 8$	240	720	60	−30	30

The table extends over nine time periods beginning from time period t and ending with time period $t + 8$. It shows the relationship between changes in the total output, changes in the capital stock, net and gross investment. The value of the accelerator v has been assumed 3 and the replacement investment demand has been assumed constant at 10 per cent of the initial capital stock on the assumption that the average age of a machine is ten years. The table shows that an increase of 10 per cent in the aggregate output in time period $t + 2$ raises the net investment from 0 in time period $t + 1$ to 60 in

NOTES

time period $t + 2$ while the gross investment increases from 60 to 120, i.e., the gross investment demand is doubled. A given change (increases or decrease) in the final output causes a magnified change in both the net and gross investment.

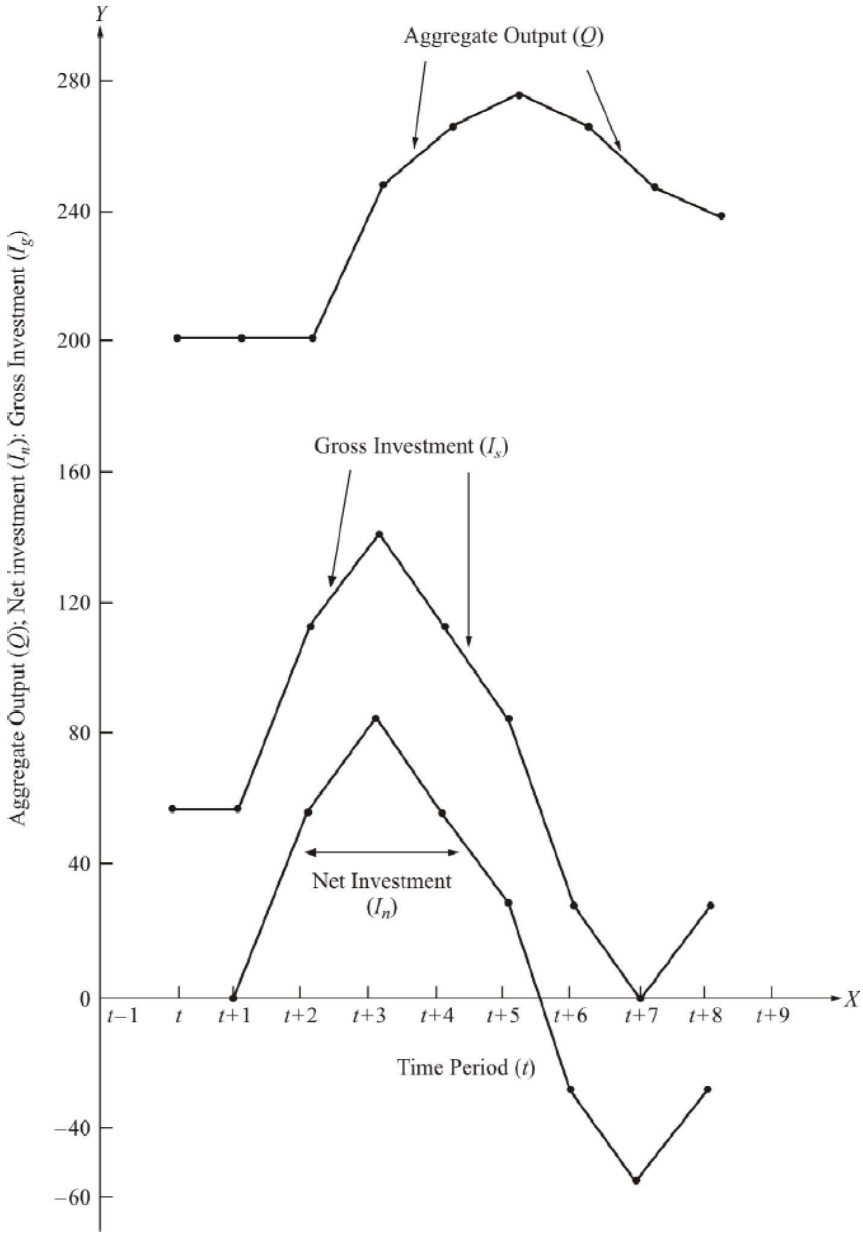


Fig. 4.2

The extent to which the demand for capital stock or investment will be magnified as a result of the given increase in the demand for final output will depend on the durability of the capital goods and the value of the accelerator. Given the value of the accelerator, greater the durability of the capital equipment, more violent will be the fluctuations in the investment demand consequent upon any given change in the demand for final output. Greater durability of the capital goods and a high accelerator reinforce each other causing explosive oscillations in the demand for capital or investment goods and *vice versa*.

Figure 4.2 illustrates the acceleration principles table graphically. At the top of the figure is the total output curve which shows changes in the total output per time period. The gross investment curve in the middle and the net investment curve at the bottom of

the figure show the impact of changes in the total output on the gross and net investment in the economy. A given change in the demand for total output causes a greater change in the net investment compared with the change in the gross investment.

The acceleration principle is useful for the analysis of the downturn and of the long-run growth. To an extent, the theory of pump-priming rests on the acceleration principle. Apart from helping in the formulation of the trade cycle models, it has also helped in the formulation of neat models of long-run economic growth path of the economy. The acceleration principle has also been usefully employed in explaining the changes in the investment in inventory stocks and in the durable consumer goods.

Assumptions

The simple acceleration principle is based on several assumptions. *First,* it is essential for the operation of the principle that the total capital stock of the economy must be fully utilized. In other words, there should be no idle or surplus plant capacity present in the economy. Consequently, the acceleration principle becomes non-operative in depression when due to the presence of idle or surplus plant capacity in the economy the output can be increased without requiring any increase in economy's capital stock. This essential condition for the operation of the acceleration principle was stated by J M Clark in his well-known article in these words: ' the first increase in demand for finished products can be taken care of by utilizing the excess producing capacity which an industry using much machinery habitually carries over a period of depression. Thus, they do not need to buy more equipment the instant the demand begins to increase.' In other words, the operation of the acceleration principle is asymmetrical.

Second, it is assumed that the firms increase their capital stock to meet the increase in the demand for their products promptly without considering the nature of the increase in the demand, i.e., the firms add to their plant capacity even if the increase in the demand for their products is short-lived and temporary.

Third, the naive acceleration principle assumes a fixed capital-output ratio. Consequently, changes in the capital-output ratio under the impact of technological improvements are ruled out.

The *fourth* assumption is that the acceleration principle assumes that there is no ceiling to investment, i.e., the capital goods' supply function is perfectly elastic placing no barrier on the expansion of capital stock when needed. The acceleration theory assumes that it does not matter how rapid is the increase in the demand for final products. The necessary capital goods can be immediately produced so that the optimum capital stock and the actual capital stock are always equal.

The *fifth* assumption is that the naive acceleration principle assumes that an increase in the rate of output growth is accompanied by an increase in the investment. The increase in investment follows rather than precede the increase in output.

The *sixth* assumption is that the increase in the aggregate output does not alter the structural composition of the aggregate output.

Lastly, the acceleration theory assumes that the capital goods are perfectly divisible in any required size.

Criticism

The naive acceleration principle has been criticized on several grounds. The assumption of constant capital-output ratio has been vehemently criticized. It has been argued that the empirical studies do not lend support to the assumption of fixed accelerator. In this

NOTES

age of rapid scientific advancement, technology has been fast changing over time. Even if the technology was assumed to remain static, the extent of the needed increase in the capital stock consequent upon a given increase in the demand for final output will depend on the distribution of this increase in the demand for total output between different goods and services produced by the different industries with different capital-output ratios. If the increase in the demand for output was concentrated on the demand for the products of high accelerator industries, a greater increase in the capital stock would be needed compared with the situation in which the demand was concentrated on the products of low accelerator industries.

Second, by disaggregating the investment by industries it can be shown that net investment in the economy might take place even if the total output has not increased. A redistribution of the existing aggregate demand for goods may cause, through the accelerator, more net investment in those industries which experience an increase in the demand for their products compared with the disinvestment in those industries which are faced with a falling demand for their goods because disinvestment in any industry cannot exceed the rate at which the capital stock is consumed.

Third, the acceleration theory ignores the role of expectations in the investment decisions of the firms. Entrepreneurs will not increase their plant capacity even if the demand for their products has increased unless they expect the increase in the demand to be permanent, According to Jan Tinbergen, the acceleration principle cannot accurately depict the formation of investment decisions. David McCord Wright has rejected the acceleration principle as an unimportant factor in the firms' investment decisions.

Entrepreneurs are influenced by many factors other than the current increase in the demand for their products while making investment decisions. Past memories affect reactions to present events. Regardless of the state of the present demand, firms will not increase their existing plant capacity unless their expectations about the future warrant it. According to Simon Kuznets, if the equipment is more durable, the entrepreneurs will be more wary before installing additional equipment in response to an increase in the demand for their products. The fact that in their investment decisions, firms are alive to distant possibilities divorced from the current market changes restricts the utility of the acceleration principle for purposes of analysis.

The *fourth* criticism is that full capacity which is a prerequisite for the operation of the acceleration principle, is absent in the early stages of the cyclical upswing. Consequently, the operation of the accelerator is asymmetrical as between the cyclical upswings. Stating this weakness of the acceleration principle, Tinbergen writes: 'Very strong decreases in consumers' goods production must not occur. If the principle were right, they would lead to a corresponding disinvestment and this can only take place to the extent of replacement. If annual replacement amounts 10 per cent of the stock of capital goods, then a larger decrease in this stock than 10 per cent per annum is impossible. A decrease in consumers' goods production of 15 per cent could not lead to a 15 per cent decrease in physical capital as the acceleration principle would require. It is interesting that this limit is sharper the greater the duration of life of the capital goods considered'. Tinbergen has also criticized the acceleration principle as being useless in practice. Since full capacity is a prerequisite for the operation of the acceleration principle, according to Tinbergen, statistical evidence shows that this condition is very rarely, if ever, met with in practice.

The *fifth* criticism is that the acceleration principle ignores the technical factors in investment. Capital equipment may be bulky and indivisible and the use of additional

plant would be justified only when the demand for output has increased considerably. This factor is all the more important because usually what is added is not a single machine but a complex of machines. For example, when the demand for railway traffic increases, additional capacity will not be employed unless the increase in the traffic demand justifies the running of an additional train and the construction of the additional rail track, etc. If net investment is to be strictly a function of the rate of growth of output, units into which the stock of capital equipment is divisible must be the same as those into which the output is divisible. In other words, corresponding to each output unit there must exist an appropriate capital equipment unit to produce that output unit. This is not, however, possible. Indivisibility of capital plant is at the root of the economies of scale in production.

Criticizing the acceleration principle, Hutt has stated that the 'sophisticated accelerationists contend that net accumulation (*S*) will tend to vary directly with the rate of growth of output, so that–

$$S = K \frac{dO}{dt}$$

I contend, on the contrary, (a) that the magnitude replacement + net accumulation tends to vary in direct proportion to output and with the rate of change in output; (b) that the magnitude net accumulation ('investment') is arbitrarily related to output because it is an overlap dependent upon the rate of decumulation (consumption); and (c) that fallacy lurks in the concept of the time increment, for when output is related to the provision of assets, the minimum comparable time unit is the economic life of the assets.'

The *sixth* point of criticism is due to the acceleration principle which ignores the role of movements in the rate of interest on causing fluctuations in the investment schedule due to reactions through the monetary sector. Moreover, changes in income are partly discounted by the business firms. Consequently, these changes do not cause equal changes in expected output. Further, it is criticized due to a serious drawback of the acceleration theory in that it assumes that all investment is a function of change in income ΔY while only a component of the net investment is a function of change in income ΔY.

Lastly, the naive acceleration principle does not take into account the limits on the rate of investment. The theory assumes that the supply function of capital goods (and that too in the short run) is perfectly elastic so that it is possible to increase the capital stock to any extent required by the increase in the demand for output. In reality, however, it is not so.

In the light of all these criticisms of the acceleration principle, one might ask: to what extent, if any, does the simple acceleration principle explain the changes in investment which we see around us? Although the acceleration principle does not operate as effectively in practice as the theory assumes, however, it works to some extent. For example, who can deny that entrepreneurs' decisions to add to their capital stock are partly influenced by the increase in the demand for their products? Although the acceleration principle cannot explain the lower turning point of the cycle, it is, however, useful in the analysis of the down-turn and of the long-run growth of the economy. In short, the acceleration principle, together with the other factors, helps in explaining the cyclical oscillations which are observed in the investment activity in the economy.

On account of these limitations, various attempts have been made to modify the simple acceleration theory described above in several ways. One modification is to assume that the firm's capital requirements are related to the previous period's output

NOTES

rather than to current output such that K_t is proportional to Y_{t-1} rather than to Y_t. Consequently, net investment in any given time period *t* is functionally related to an increase in the immediately preceding time period's income ΔY_{t-1} (= $Y_{t-1} - Y_{t-2}$). Although this version of the acceleration principle is an improvement over the naive version in certain respects yet the limitations mentioned above still apply. Goodwin and Chenery have suggested a stock-adjustment version of the acceleration principle.

Interaction between the Multiplier and Accelerator

The acceleration and the multiplier principles have been combined into a single model to show their interaction. This follows from the fact that an increase in the investment or consumption spending involving either the multiplier or the accelerator has an impact on the other depending on the presence or absence of the excess plant capacity in the economy. In other words, an increase in the autonomous investment increases output and income which in turn increases consumption. The induced consumption increases the induced investment (if there is no excess plant capacity in the economy) via the acceleration principle. The chain of causation between the multiplier and the accelerator is roughly of the following structure:

$$\Delta I_a \rightarrow \Delta Y \rightarrow \Delta C \rightarrow \Delta I_i \rightarrow \Delta Y \rightarrow \Delta C \rightarrow \Delta I_i$$

Different multiplier-accelerator models have been constructed by the economists to show how the fluctuations in the level of aggregate economic activity come about and how these can be self-generating. However, any desired result can be obtained by choosing the appropriate values for the accelerator (*v*) and the marginal propensity to consume (*b*). Explosive trade cycle will be produced if the sum of the values of *v* and *b* is high; if this sum is low, cyclical fluctuations in income will occur, the magnitude of the fluctuations depending on the chosen values of the accelerator and the marginal propensity to consume. If their sum is close to 1, the oscillations will be small. If it is more than 1, the oscillations will be large. A very large *v* combined with a high *b* leads to an increasing rate of growth of income.

It is the value of v which determines the degree and the kind of the trade cycle which is generated. If *v* is zero, the multiplier would simply work its way out without causing any cyclical fluctuations. When the value of *v* is high, income explodes following an increase in the autonomous component of the aggregate demand (investment or consumption) and the multiplier is overpowered. For a value of *v* between 0 and 5, high cyclical fluctuations will occur. For a high value of *v*, the explosive cycles with increasing size will occur. A smaller value of *v* produces the damped trade cycles. At some in-between moderate value of *v*, perfectly symmetrical trade cycles will be produced. The working of the multiplier and the accelerator can be understood with the help of the following equations and Table 4.4.

$$Y_t = C_t + I_t \qquad ...(12)$$

$$C_t = a + bY_{t-1} \qquad ...(13)$$

$$I_t = I_A + \boldsymbol{v}\,(Y_{t-1} - Y_{t-2}) \qquad ...(14)$$

In equation (13), the consumption function shows that the induced consumption has one time period lag in relation to income, i.e., the consumption of any given time period *t* depends on the income of the immediately preceding time period *t* – 1. In equation (14), the investment demand function shows that the induced part of total investment has a two time period lag in relation to income, i.e., net induced investment in time period *t* is a function of the increase in income in the preceding time period *t*–1 or of ΔY_{t-1}. Thus, both a second-order lag and an accelerator are required to generate a business cycle.

Taking the value of *v* as 3, the value of *b* as 0.5, the value of the autonomous consumption *a* as 10, the value of the autonomous investment I_A as 50 and the initial equilibrium income as 120 corresponding to which the induced consumption is 60, we assume that there is an increase of 10 units in the autonomous investment.

Under the impact of the interaction between the multiplier and the accelerator, the aggregate income will increase in the manner as shown in Table 4.4.

Table 4.4 *Interaction of the Multiplier and Accelerator*

Time Period	***Autonomous Investment*** I_A	***Induced Investment*** $I_i = v(Y_{t-1} - Y_{t-2})$ $(v = 3)$	***Consumption*** $C = a + bY_{t-1}$ $(b = 0.5)$	***Total Income*** $Y = C + I_A + I$
1	*2*	*3*	*4*	*5*
$t-2$	50	0	10 + 60	120
$t-1$	50	0	10 + 60	120
t	60	0	10 + 65	135
$t+1$	60	30	10 + 65	165
$t+2$	60	105	10 + 82.5	257.5
$t+3$	60	277.5	10 + 128.75	476.25
$t+4$	60	656.25	10 + 238.12	964.38
...	...	...	...	...
...	...	...	...	...
...	...	...	...	...

The table shows that consequent upon a given increase in the aggregate autonomous investment, the aggregate equilibrium income grows through an explosive path through time and there are no oscillations. This is so because the value of the accelerator v chosen is high (it is 3) while the value of the marginal propensity to consume *b* (= 0.5) is not low.

4.5.1 Derivation from Profit Maximizing Behaviour

The acceleration principle from profit maximizing behaviour can be derived using Jorgenson's neoclassical theory of investment.

The assumptions of the theory are as follows:

- There is perfect competition and no uncertainty and no adjustment costs.
- There is full employment in the economy where prices of labour and capital are perfectly flexible.
- The company operates in a perfect financial market where it can borrow or lend at a given rate of interest.
- There are diminishing returns to scale and the capital stock is fully used.
- The company maximises the present value of its current and future profits with perfect foresight in relation to all future values.

In the model, there is a production process with a single output (Q), a single variable input labour (L), and a single capital input (I-investment in durable goods), and p, w, and q representing their corresponding prices. The flow of net receipts (R) at time t is given by

$$R(t) = p(t)\,Q(t) - w(t)\,L(t) - q(t)\,I(t) \quad ...(15)$$

Where Q is output and p is its price; L is the flow of labour services and w the wage rate; I is investment and q is the price of capital goods.

NOTES

The present value is defined as the integral of discounted net receipts which is represented as

$$W = \int_0^{\infty} e^{-r t} R(t) dt \quad ...(16)$$

Where W is the present value (net worth); e is the exponential used for continuous discounting; and r is the constant rate of interest.

The present value is maximised subject to two limitations. These are:

(i) The rate of change of the flow of capital services is proportional to the flow of net investment. The constant of proportionality may be understood as the time rate of utilisation of capital stock, i.e., the number of units of capital service per unit of capital stock. Net investment is equal to total investment less replacement investment where replacement investment is proportional to capital stock. This limitation takes the form:

$$K(t) = I(t) - \delta K(t) \quad ...(17)$$

Where K (t) is the time rate of change of the flow of capital services at time (t) while δ is the rate of depreciation attached to capital stock. This limitation is application at each point of time so that K, K and I are functions of time.

(ii) The levels of output and the levels of labour and capital services are constrained by a production function:

$$F(Q, L, K) = 0 \quad ...(18)$$

The marginal productivity of labour is equal to the real wage:

$$\partial Q / \partial L = w/p \quad ...(19)$$

Likewise, the marginal productivity of capital is equal to its real user cost:

$$\partial K / \partial L = w/p \quad ...(20)$$

$$\text{Where } c = q(r + \delta) - q \quad ...(21)$$

In equation (21), q is the average price of capital assets, r is the rate of discount, δ is the rate of depreciation of capital goods and q is the rate of appreciation of capital assets or time derivative of q. Thus, the critical determinant of the optimal capital stock is c, the user cost of capital.

The accelerator is merely a special case of the neoclassical theory of investment where the price variables have been reduced to constants. If the price of output is assumed to be constant and the price variables in Jorgenson's user cost of capital (equation 21) are fixed, we get the accelerator principle where the desired capital stock is assumed to be proportional to output.

Check Your Progress

7. Define Marginal Efficiency of Investment (MEI).
8. What is the other name for flexible accelerator theory?
9. State the principle of acceleration.
10. State the third assumption of the acceleration principle.

4.6 FLEXIBLE ACCELERATOR

With the flexible accelerator theory it has become possible to do away with a key weakness associated with the simple acceleration principle which is that the capital stock is optimally adjusted without any time lag. With flexible accelerator, lags exist in the process of adjustment between the level of capital stock and the level of output.

The flexible accelerator theory is also called the capital stock adjustment model. This theory has been developed in different forms by Chenery, Goodwin, Koyck and Junankar with the theory propounded by Koyck being the most accepted.

Lags existing in the adjustment between output and capital stock have been discussed by Junankar. They have been explained by him at the firm level and extended to aggregate level. Assume that demand rises for an output. For meeting this demand, the inventories will be utilized by the firm as also the capital stock that it has will be made use of in a more intensive way.

In case a large and persistent demand arises for the output, the firm might go in for an increase its stock capital demand. We refer to this as the decision-making lag. Also, there might exist the administrative lag of capital ordering.

Since in the financial capital market, capital is neither in abundance nor easily available, a financial lag exists in obtaining finance to purchase capital. Also, delivery lag exists between placing order for capital and it being delivered.

Derivation

The flexible version of the acceleration principle removes some of the rigid assumptions of the original accelerator theory. That is why it is called the flexible version of the acceleration theory. One such assumption is the acquisition of desired additional capital in one period. The modified version assumes instead that the gap between the actual and desired capital stock is filled over a number of periods. There are at least two possible reasons for this: *first*, the production of additional capital equipment takes a longer time than implied in the simple version of the acceleration principle, and *second,* acquisition of desired capital stock is usually based on long-run considerations. This is obviously a more realistic approach than that of simple acceleration principle.

The flexible version of the acceleration principle allows a time lag in filling the gap between the desired capital stock (K^*_t) in period t and the actual capital stock (K_{t-1}) in period $t-1$. In period t, therefore, only a fraction (λ) of K^*_t is procured. This relationship is expressed as

$$K_t - K_{t-1} = \lambda(K^*_t - K_{t-1}),\ (0 < \lambda < 1) \quad \text{... (22)}$$

where, K_t is the actual capital stock in period t; K_{t-1} is the actual capital stock in period $t-1$; K^*_t is the desired capital stock in period t; and λ is a constant (proportion).

Since $K_t - K_{t-1}$ equals net investment (I_n) and $I_n = (I_t - R_t)$—where R_t is replacement capital in period t—Eq. (22) can be written as

$$I_t - R_t = I_n = \lambda(K^*_t - K_{t-1}) \quad \text{... (23)}$$

Equation (23) reads that net investment in period t *equals a fraction* of the difference between the desired capital stock in period t and actual capital stock in period $t-1$.

Since, given the technology, K^*_t equals capital-output ratio (k) times Y_t, the output in period t, Eq. (23) can be written as

$$I_n = \lambda(kY_t - K_{t-1}) \quad \text{... (24)}$$

and gross investment (I_g) can be expressed as

$$I_g = \lambda(kY_t - K_{t-1}) + R_t \quad \text{... (25)}$$

To conclude, the flexible version of the acceleration principle incorporates a partial adjustment mechanism between the desired and the actual capital stock rather than instantaneous adjustments. However, the basic principle remains the same.

4.7 SUMMARY

NOTES

In this unit, you have learnt that,

- The Net Present Value (NPV) is the difference between the current value of the cash that is flowing in (inflow) and the current value of cash that is flowing out (outflows).
- NPV is used for the purpose of budgeting of capital to analyse how profitable a project or a projected investment will be.
- If there is a positive net present value, it implies that the projected earnings that will accrue from an investment or project (in current currency value) will be in excess of the cost that has been anticipated for the investment or project (in current currency value).
- The payback (PB) is one of the most popular and widely-recognized traditional methods of evaluating investment proposals.
- Payback is the number of years required to recover the original cash outlay invested in a project.
- Many firms use the payback period as an investment evaluation criterion and a method of ranking projects. They compare the project's payback with a predetermined and standard payback.
- The most significant merit of payback is that it is simple to understand and easy to calculate. The business executives consider the simplicity of method as a virtue.
- Payback is considered theoretically useful in a few situations. One significant argument in favour of payback is that its reciprocal is a good approximation of the rate of return under certain conditions.
- One of the serious objections to the payback method is that it does not discount the cash flows for calculating the payback period. We can discount cash flows and then calculate the payback.
- The internal rate of return (IRR) method is another discounted cash flow technique, which takes account of the magnitude and timing of cash flows.
- The internal rate of return (IRR) is the rate that equates the investment outlay with the present value of cash inflow received after one period. This also implies that the rate of return is the discount rate which makes NPV = 0.
- The accept-or-reject rule, using the IRR method, is to accept the project if its internal rate of return is higher than the opportunity cost of capital.
- In economics, expected rates of return on investment as additional units of investment are made under specified conditions and over a stated period of time.
- As the quantity of investment increases, the rates of return from it may be expected to decrease because the most profitable projects are undertaken first.
- The acceleration principle states that the demand for capital goods varies directly with the change in the level of aggregate output. The extent of change in the demand for capital goods depends on the capital-output ratio and the change in the level of output.
- The acceleration and the multiplier principles have been combined into a single model to show their interaction.

- An increase in the investment or consumption spending involving either the multiplier or the accelerator has an impact on the other depending on the presence or absence of the excess plant capacity in the economy.

NOTES

4.8 KEY TERMS

- **Payback period:** Payback period in capital budgeting refers to the period of time required to recoup the funds expended in an investment, or to reach the break-even point.
- **Payback reciprocal**: The payback reciprocal is a crude estimate of the rate of return for a project or investment. The payback reciprocal is computed by dividing the digit "1" by a project's payback period expressed in years.
- **Uneven cash flows**: Any series of cash flows that doesn't conform to the definition of an annuity is considered to be an uneven cash flow stream.
- **Multiplier**: The multiplier refers to the phenomenon whereby a change in an injection of expenditure (either investment, government expenditure or exports) will lead to a proportionately larger change (or multiple change) in the level of national income.
- **Accelerator principle:** The accelerator principle indicates how changes in the level of current income will have an accelerated impact on the level of investment and is therefore one explanation of economic instability and the upward and downward swings of the trade cycle.

4.9 ANSWERS TO 'CHECK YOUR PROGRESS'

1. NPV is used for the purpose of budgeting of capital to analyse how profitable a project or a projected investment will be.
2. The formula that is used to calculate Net Present Value is:

$$NPV = \sum_{t-1}^{T} \frac{C_t}{(1+r)^t} - C_o$$

3. The risk of the project can be tackled by having a shorter standard payback period as it may ensure guarantee against loss. A company has to invest in many projects where the cash inflows and life expectancies are highly uncertain. Under such circumstances, payback may become important, not so much as a measure of profitability but as a means of establishing an upper bound on the acceptable degree of risk. This process is called risk shield.
4. One of the serious objections to the payback method is that it does not discount the cash flows for calculating the payback period. We can discount cash flows and then calculate the payback.
5. The internal rate of return (IRR) is the rate that equates the investment outlay with the present value of cash inflow received after one period. This also implies that the rate of return is the discount rate which makes NPV = 0. There is no satisfactory way of defining the true rate of return of a long-term asset.
6. IRR method has the following merits:

- **Time value:** The IRR method recognizes the time value of money.

NOTES

- **Profitability measure:** It considers all cash flows occurring over the entire life of the project to calculate its rate of return.
- **Acceptance rule:** It generally gives the same acceptance rule as the NPV method.
- **Shareholder value:** It is consistent with the shareholders' wealth-maximization objective. Whenever a project's IRR is greater than the opportunity cost of capital, the shareholders' wealth will be enhanced.

7. In economics, expected rates of return on investment as additional units of investment are made under specified conditions and over a stated period of time. This is called marginal efficiency of investment.
8. The flexible accelerator theory is also called the capital stock adjustment model.
9. The acceleration principle states that the demand for capital goods varies directly with the change in the level of aggregate output. The extent of change in the demand for capital goods depends on the capital-output ratio and the change in the level of output.
10. The third assumption of the acceleration principle is that the naive acceleration principle assumes a fixed capital-output ratio. Consequently, changes in the capital-output ratio under the impact of technological improvements are ruled out.

4.10 QUESTIONS AND EXERCISES

Short-Answer Questions

1. State the benefits of the payback period.
2. Mention the limitations of the Internal Rate of Return (IRR) method.
3. State the factors that make reciprocal of payback a close approximation of internal rate of return method.
4. What are the acceptance rules of Internal Rate of Return method?

Long-Answer Questions

1. Discuss the internal rate of return method as a cash flow technique.
2. Compute the method of calculating IRR by trial and error method.
3. Analyse the acceleration principle and its derivation from profit maximizing behaviour.
4. Discuss the concept of flexible accelerator.

4.11 FURTHER READING

Mankiw, N Gregory. 2010. *Macroeconomics.* Worth Publishers, New York.

Shapiro, Edward. 1996. *Macroeconomic Analysis.* Galgotia Publication, New Delhi.

Branson, WH. 1979. *Macroeconomic Theory and Policy*. Universal Book Stall, New Delhi.

Jha, R. 1999. *Contemporary Macroeconomic Theory and Policy*. New Age International, New Delhi.

Gupta, SB. 1995. *Monetary Economics: Instruments and Policy*. S Chand & Co. New Delhi.

UNIT 5 THEORIES OF TRADE CYCLE

NOTES

Structure

5.0 INTRODUCTION

Trade or business cycles are those fluctuations which recur in aggregate economic activities with a certain degree of regularity following the pendulum like oscillations. According to Wesley Clair Mitchell, who did some pioneering work in this field, 'business cycles are a type of fluctuations found in the aggregate economic activity of nations that organize their work mainlyin business enterprises.Acycle consists of expansions occurring at about the same time in many economic activities followed by similarly general recessions, contractions and revivals which merge with the expansion phase of the next cycle, this sequence of change is recurrent but not periodic.' This definition reveals that business cycles are fluctuations in the aggregate economic activity and, therefore, are concerned with the economy as a whole. Apart from this feature, business cycles are confined to only those fluctuations which recur with regularity.

In this unit, different trade cycle theories, particularly Schumpeterian theory, Multiplier-acceleration theory and Kaldor's theory have been explained in detail.

5.1 UNIT OBJECTIVES

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

- Discuss the innovations theory of trade cycle associated with Schumpeter
- Analyse the multiple-accelerator interaction and the principle behind it
- Describe the investment-consumption function according to Kaldor's theory
- Analyse the process of the re-emergence of monetary policy
- Discuss the various instruments of monetary policy

5.2 SCHUMPETERIAN THEORY

NOTES

The innovations theory of trade cycle, associated with the name of well-known economist Joseph A Schumpeter, makes innovations the central cause of the recurrence of business cycle in the modern industrial economies. According to Schumpeter, the key to the explanation of the trade cycle lies in the recurring bursts of innovational investment activity which dominates the capitalist economies. Schumpeter views the business cycle as a natural outgrowth of economic progress. As Schumpeter puts it:

> 'The booms consist in the carrying out of innovations in the industrial and commercial organism. By innovations I understand such changes of the combinations of the factors of production as cannot be affected by infinitesimal steps of variations on the margin. They consist primarily in changes in methods of production and transportation, or in changes in industrial organization, or in the production of a new article, or in the opening of new markets or of new sources of material. The recurring periods of prosperity of the cyclical movements are the form progress takes in capitalistic society.'[1]

According to Schumpeter, innovations are lumpy and bunched in time leading to the corresponding surge of investment outlay necessary for their commercial exploitation. The lumpy and surging investment activity is sufficient to create conditions of boom in the economic activity. According to Schumpeter, an innovation is different from an invention.

An innovation is the initial application of an invention to commercial production while an invention is the discovery or development of a new process, product or service by scientists or engineers. Despite the fact that inventions may appear more or less continuously through time, innovations—initial commercial exploitation of inventions—show marked discontinuities because a majority of the businessmen being risk avoiders are reluctant to innovate at all under ordinary circumstances. An innovator is one who is the first to make the commercial application of invention and consequently bears the risks involved in introducing a new product or service or process in the market. The success or failure of an innovation can only be proved by actually producing and marketing the new product. This virtually amounts to groping in the dark, inviting avoidable risks which most firms would choose to avoid preferring to produce and market only the old and tried goods and services produced through the traditional processes and methods. The majority of firms are simply the imitators.

According to Schumpeter, the innovators, formally designated as entrepreneurs, play a leading role in the entire process of capitalistic evolution. They comprise the microscopic group of dashing pioneering members of the business community who are constantly on the look-out for something new and different. They translate inventions into innovations relying largely on their business acumen and believing that success will be theirs. When the innovation proves successful extraordinary profits accrue to the innovators. But very soon the innovation is imitated by others by adopting new processes of production and producing and marketing the goods and services similar to those produced and marketed by the innovators. Others imitate the innovator because they cannot allow themselves to be browbeaten by an aggressive rival. As more firms imitate, the pressure on those not imitating increases. Moreover, the innovation, especially when it comes about at a time when the economy is operating at the near-full or full employment level, causes a rise in prices because entrepreneurs compete in bidding for the available supply of resources. Profit opportunities appear brighter and investment is stimulated. A

surge of investment activity grips the economy. However, as an increasing swarm of imitators expand their investment activity, cut-throat competition results in bringing down perceptibly the profits of the innovator. Although we started off here with one innovation and one industry, the fact is that several other innovations are being simultaneously exploited in the economy. For some time, the business activity continues to expand as the innovators reap the increasingly rich harvest of extraordinary profits under the impact of increasing commercial application of inventions. As each successful innovator creates a swarm of imitators and as the expanding business activity reinforces the prospects for further success of still other innovations, the economy moves into a period of expansion characterized by a bunching of innovations and rapid growth in investment spending which accompanies the innovation process.

NOTES

But what causes this expanding investment activity to come to an end? The expansion comes to an end as the process of introducing innovations in the economy comes to an end for the time being. The prosperity was spurred by investment activity caused by innovations. The innovation opportunities are not, however, limitless. Consequently, after some time the opportunities for investment in the new spheres decline. For each innovation taking the form of production of some new product or service it is just a matter of time until the new factories needed to produce these goods are constructed. When this has happened, the innovational net advantage disappears. In short, investment outlays in general whether incurred to produce the new goods or services, reduce costs, open new markets or develop new sources of supply, eventually will have been made and the actual capital stock will correspond to the desired capital stock in the economy. When this situation is reached, all those innovations deemed worthwhile by the entrepreneurs have been adopted. The boom comes to an end because the new investment outlays generated by the innovations slow down to a trickle.

The upswing turns into a downswing as the impact of the maladjustments which were built into the system during the process of innovational activity begin to manifest themselves. According to Schumpeter, innovational activity is 'lopsided, discontinuous, disharmonious by nature the disharmony is inherent in the very *modus operandi* of the factors of progress.'[2] Being unbalanced, the prosperity cannot be sustained and 'depression is nothing more than the economic system's reactions to the boom or the adaptation to the situation into which the boom brings the system.'[3]

The lower turning point of trade cycle, i.e., recovery occurs when following a period of readjustment to the changed economic circumstances created by the previous boom, prices have fallen to a point where the pioneering businessmen, entertaining the hope that any further fall is most unlikely to happen, once again start innovating. The introduction of innovation heralds the beginning of the recovery which soon leads to rapid expansion of the economy.

In Schumpeter's trade cycle theory, *first*, the innovators are assumed to appear in a cluster causing the process of economic development to proceed discontinuously rather than in a smooth continuous manner. *Second*, the assumption is that the majority, if not all, of the innovations are bank-financed by way of credit expansion.[4] *Third*, special significance attaches to equilibrium because different phases of trade cycle are viewed as deviations of the economy from the equilibrium path. According to Schumpeter, trade cycles of different lengths occur simultaneously reacting upon one another. Although these different length trade cycles are caused by one and the same cause, namely innovations, the difference in their lengths is due to the fact that 'the periods of gestation and of absorption of effects by the economic system will not, in general, be equal for all the innovations that are undertaken at any time.'[5]

NOTES

Criticism: The innovations theory of trade cycle suffers from several weaknesses. The depth of the depression is not entirely determined by the disturbances caused by the innovations; it is also attributable to other factors which vary in intensity from one trade cycle to another.

1. The extent to which innovations are introduced before the previous trade cycle has run out its course and before readjustments to the previous cycle have been made, significantly affect the depth and behaviour of trade cycle. Thus, if the innovations are made in the depression phase of the previous cycle, many firms that are doomed to liquidation get a fresh lease of life. Consequent upon the introduction of innovations, prices are prevented from falling. The shift of labour from one industry to another needed for the equilibrium adjustment remains incomplete. For these reasons, the early appearance of a new period of prosperity apart from involving its own maladjustments also carries some maladjustments to the next period of depression.
2. The amount of bank credit created during the prosperity also affects the nature and course of the business cycle. The greater the amount of bank credit absorbed in the system of financing innovations by the old industry, the higher will be the rise in prices. Consequently, the greater will be the readjustment needed to bring the prices into equilibrium. Further, greater the extent of speculative over-buying financed by the bank credit, greater will be the likelihood of the bank failures and financial crisis.
3. The nature of the competitive impact of innovations is also very significant in determining the magnitude of the trade cycle. If the impact of demand for the new product is reflected in a small decrease in the demand for all other products in general, the loss of demand felt by each one of the old industry will be comparatively mild and readjustment will be easily accomplished. On the contrary, if the competitive impact of the new product is limited to only a few old industries, a large number of firms in these industries will go out of production leading to mass unemployment and severe depression in the economy.
4. If the industries suffering from the competitive impact of the new product are highly localized, the unemployment of labour and other factors of production due to immobility will be far more severe than if the industries were widely dispersed in the country or the world.
5. The period required for the construction of factories for innovations to be carried through is also important in determining the duration and magnitude of the trade cycle. Longer the period required by the innovation to be carried through, more violent is the necessary readjustment. If the capital equipment for new innovations can be manufactured quickly, the new products by becoming available for sale in the market quickly will cause readjustments to take place before the development of extreme maladjustments. On the contrary, when the capital equipment involves a long gestation period for construction (e.g., railways), the entire economy becomes adjusted to the temporary demands of the construction period. At the end of the long period of construction when shifts in the demand take place, a far greater and far more severe readjustment becomes necessary.
6. The theory assumes that resources in the economy are fully employed and the innovations are entirely financed by means of the bank credit. In an economy with vast unemployed reserves of labour and other factors of production, the introduction of an innovation will not cause the withdrawal of labour and other

resources from the old industries. Consequently, the supply of old goods will not fall. Nor will the costs of old goods and services increase under the competitive impact of an innovation. Similarly, when the innovation is financed through real savings, there will be no inflationary price rise. Consequently, in an underemployed economy, innovations financed through real savings may not initiate or generate a trade cycle. Since full employment is an exception rather than a rule and since most innovations are self-financed by the firms by drawing on their development cash reserves, the innovations theory is at best an inadequate explanation of the trade cycle in a modern business economy. Friedrich A von Hayek is correct in holding that innovations alone cannot explain the phenomenon of business cycles without a substantive monetary explanation.

5.3 MULTIPLIER-ACCELERATOR INTERACTION

The acceleration principle explains that the net investment I_t during any given time period *t* depends upon the value of the accelerator v and the change in income in time period *t*, i.e.,

$$I_t = \nu \Delta Y_t = \nu (Y_t - Y_{t-1})$$

Assuming a change in the aggregate income *Y*, the net investment I_t in any given time period *t* will differ from investment I_{t-1} of the immediately preceding time period *t*–1 as indicated by the acceleration principle. However, an increase or decrease in investment will affect the aggregate income *Y* in the succeeding time periods as indicated by the multiplier principle. But any change in income from one time period to the next will in turn affect the net investment due to the operation of the accelerator which in turn will affect the income due to the action of the multiplier and so on. This process may continue endlessly due to the interaction between the multiplier and accelerator.

In Samuelson's model, interaction between the multiplier and the accelerator is an endogenous force which generates business cycle in the economy. Following Samuelson, national income Y_t in any given time period *t* can be written as the sum of the autonomously determined government expenditure G_t, consumption expenditure C_t, and private induced investment I_t. Thus,

$$Y_t = G_t + C_t + I_t$$

$$C_t = bY_{t-1}$$

$$I_t = \nu (C_t - C_{t-1})$$

On substituting for C_t, and I_t, the equilibrium national income can be rewritten as:

$$Y_t = G_t + bY_{t-1} + v (C_t - C_{t-1})$$

$$Y_t = G_t + bY_{t-1} + bv (Y_{t-1} - Y_{t-2})$$

This equation states that the aggregate income Y_t in time period *t* is the sum of the autonomous government expenditure G_t, consumption expenditure C_t which depends on the marginal propensity to consume *b* and the income of the preceding time period Y_{t-1} and the investment outlay which depends on the capital-output ratio or the accelerator *v* and the change in consumption ΔC ($= C_t - C_{t-1}$) in time period *t*. In other words, given

Check Your Progress

1. Why is the period required for the construction of factories for innovations considered important?
2. Who is an innovator?

NOTES

period t and the amount of consumption in time period t and the preceding time period $t - 1$ are known we can determine the equilibrium income Y_t for that time period by substituting the values of v and b in the above equilibrium equation.

Although in the original multiplier-accelerator model of the trade cycle developed by Samuelson, investment has been related to the change in consumption, however, Hicks' version of this model in which investment is related to the change in income, has become more common.[6] We shall discuss here Hicks' version in which–

$$C_t = bY_{t-1}$$

$$I_t = v(Y_{t-1} - Y_{t-2})$$

$$Y_t = C_t + I_t + G_t$$

Fluctuations in the aggregate income Y around its equilibrium value after an initial exogenous shock might be of the nature as shown in Figure 5.1. The shape of the time path of income Y will depend on the arithmetical values of b and v. In general, higher the value of v, greater is the possibility of an explosive time path of the aggregate income Y; higher the value of b, smaller is the possibility that oscillations will occur. The following analysis is based on the assumption that b is constant and less than one and v is greater than zero, i.e., $0 < \bar{b} < 1$ and $v > 0$.

In this model, no trade cycle can be generated without an accelerator relationship and with only the first order lag. In other words, an accelerator and a second-order lag are both needed to generate a trade cycle. The multiplier itself is not needed to cause the trade cycle. In fact, as the value of b increases, the possibility of generation of the trade cycle is diminished. The model can be expressed through the following equations:

$$I_t = v(Y_{t-1} - Y_{t-2})$$

$$Y_t = C_t + I_t + G_t$$

The above model will generate a trade cycle for any value of v ranging between 0 and 5. But with the inclusion of the multiplier this simple model becomes more realistic and causes the trade cycles to occur which are similar to those that occur in the economy.

With or without a consumption function, two time periods after an initial increase in the exogenous variable—increase in the autonomous government expenditure—has raised the level of income Y; the negative term $v\,Y_{t-2}$ will exert a negative effect on investment. Unless the accelerator is very strong or the marginal propensity to consume is very close to one, this will result in a slight decrease in investment and a slackening of the rate of increase in income. This will ordinarily be sufficient to cause a fall in investment (I) and in the rate of increase in income (ΔY). Soon a turning point is reached and ΔY becomes negative. When this happens, the forces of reversal start to operate. The low value of Y tends to reduce the negative effect of the term $v\,Y_{t-2}$ on investment with the result that investment ceases to fall rapidly. When this happens, ΔK increases and with it I also increases. Consequently, the economy begins to move upward. This process will be repeated until the economy either reaches a new equilibrium or continues to fluctuate so widely that the negative values of Y are reached. Such explosive cycles are not, however, likely to occur in the real world as the system always tends to return to equilibrium. Table 5.1 shows that with different combinations of the values chosen for b and v, the interaction of the multiplier and accelerator generates different patterns of income oscillations.

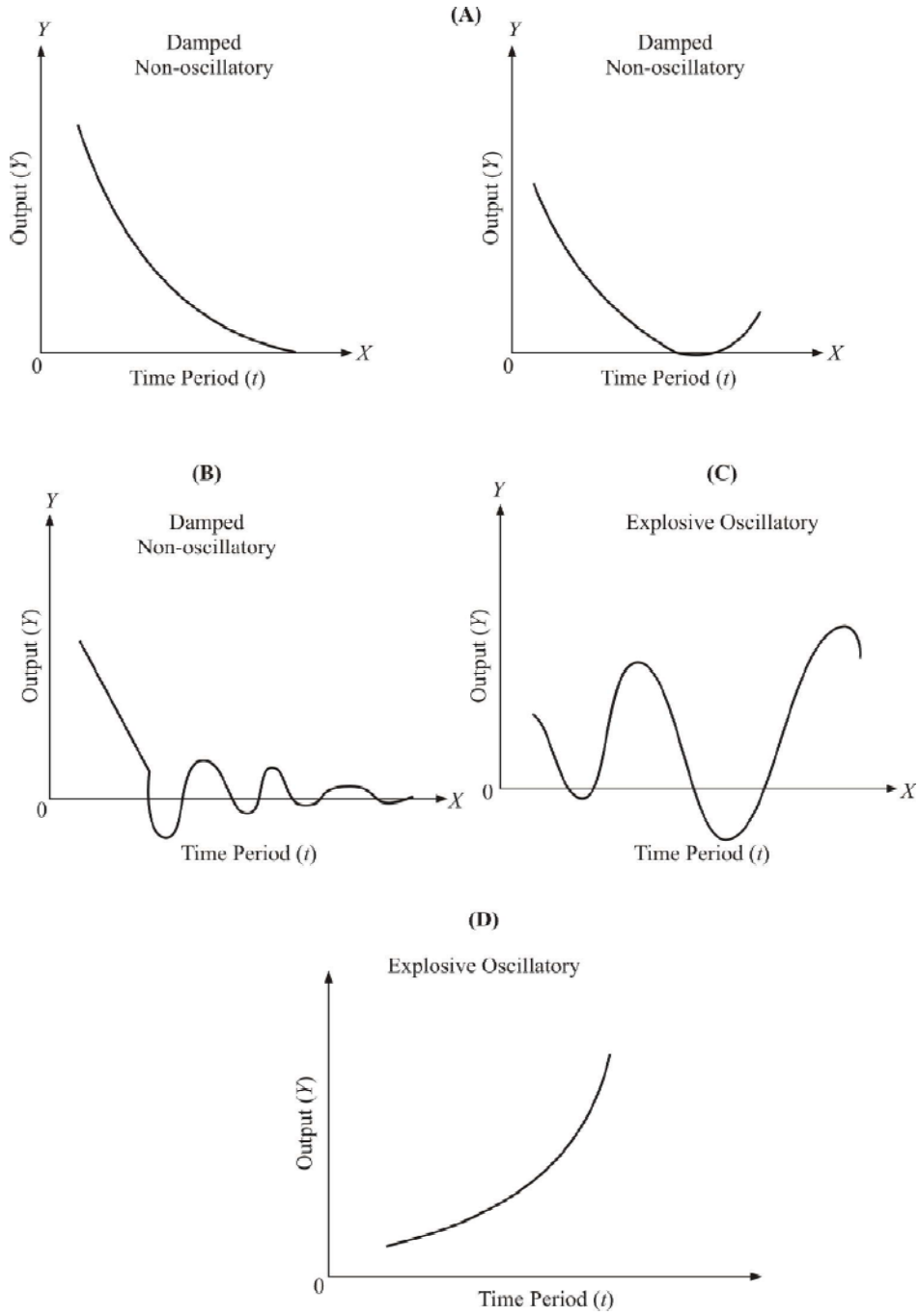


Fig. 5.1

Different possible combinations of *b* with values ranging between zero and 1 and of *v* with values ranging between zero and 5 will generate four different income patterns as shown in Figure 5.2. Those combinations of *b* and *v* which fall in the area labelled *A* in the figure will generate damped non-oscillatory income movements such as those shown in Figure 5.1(A). Similarly, region *B* shows those different possible combinations of the values of *b* and *v* which produce damped oscillatory movements in income as shown in Figure 5.1(B). Region *C* in Figure 5.2 shows those combinations of *b* and *v* which produce explosive oscillatory movements in income like the one shown in Figure 5.1(C). Region *D* shows those different combinations of *b* and *v* which will produce explosive non-oscillatory movements in income as shown in Figure 5.1(D).

Table 5.1 *Interaction of Multiplier and Accelerator*[1]

	A								***B***								***C***								***D***							
Time Period	***Damped; No Oscillations b = 0.5; v = 0.05***								***Damped Oscillations b = 0.5; v = 0.8***								***Explsive Oscillations b = 0.5; v = 1.5***								***Explosive; No Oscillations b = 0.5; v = 3.0***							
	C	**+**	***I***	**+**	***G***	**=**	***Y***		***C***	**+**	***I***	**+**	***G***	**=**	***Y***		***C***	**+**	***I***	**+**	***G***	**=**	***Y***		***C***	**+**	***I***	**+**	***G***	**=**	***Y***	
t–2	50		0		50		100		50		0		50		100		50		0		50		100		50		0		50		100.	
t–1	50		0		50		100		50		0		50		100		50		0		50		100		50		0		50		100	
t	0		60		110		50		0		60		110		50		0		60		110		50		0		0		60		110	
t+1	55		0.5		60		115.5		55		8.0		60		123		55		15		60		130		55		30		60		145	
t+2	57.7		0.3		60		118.0		61.5		10.4		60		132		65		30		60		155		72.5		105		60		237.7	
t+3	59.0		0.1		60		119.1		66		7.2		60		133.2		77.5		37.5		60		175		118.8		227.5		60		456.3	
t+4	59.5		0.1		60		119.6		66.6		1		60		127.6		87.5		30.0		60		177.5		228.2		656.4		60		944.6	
t+5	59.8		0		60		119.8		63.8		–4.5		60		119.3		88.8		3.7		60		152.5		472.3		1465		60		1,997.3	
t+6	59.9		0		60		119.9		59.7		–6.6		60		113		76.2		–37.5		60		98.7									
t+7	60.0		0		60		120.0		56.5		–5		60		111.5		49.4		–89.7		60		28.7									
t+8									55.8		–1.2		60		114.6		14.4		–100.0		60		–30.6									
t+9									57.3		–2.5		60		119.8		15.3		–89.0		60		–44.3									
t+10									59.9		4.2		60		124		22.2		–20.6		60		17.2									
t+11									62		3.4		60		125.4		8.6		92.2		60		160.8									
t-12									62.7		1.4		60		123.8		80.4		215.4		60		355.8									
t-13									61.9		–1.3		60		120.6		178		292.5		60		530.5									
Y_{eq}	60		0		60		120.0		60		0.0		60		120.0		—		—		—		—									

[1] Figures in the table have been reproduced from Michael K Evans, book entitted, *Macroeconomic Activity*, 1969, p. 365.

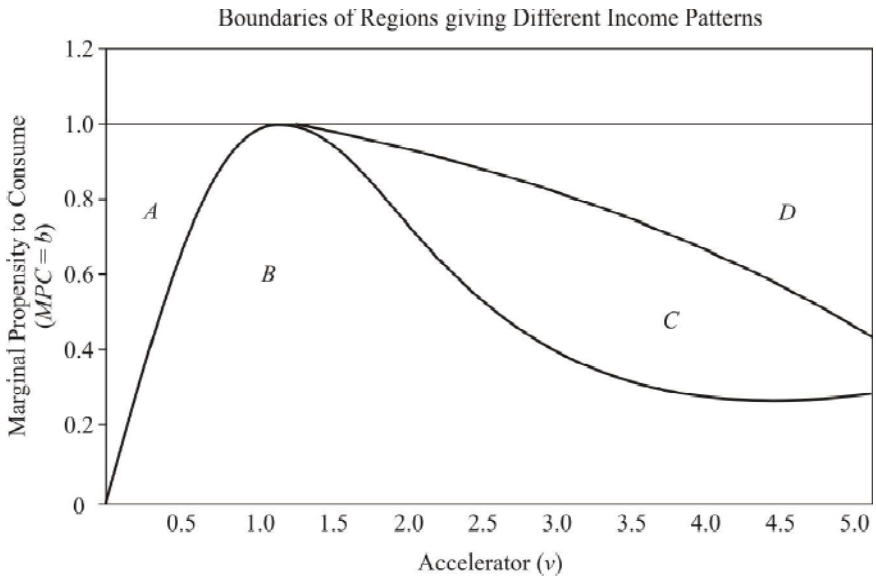


Fig. 5.2

The basic criticism of this model is the weaknesses of the rigid acceleration principle. Since the naive acceleration principle is unacceptable, the multiplier-accelerator interaction which incorporates the acceleration principle is also unacceptable as a perfect explanation of the occurrence of the trade cycle. James S Duesenberry has correctly stated that although 'the basic concept of multiplier-acceleration is an important one but we cannot really expect to explain observed cycles by a mechanical application of that concept.'[7]

5.4 KALDOR'S THEORY[8]

In the model of the trade cycle developed by Nicholas Kaldor, consumption (or saving) is a function of income while investment is directly related to income and inversely related to the stock of capital. In other words, the investment demand function is of the capital adjustment type. Kaldor's theory of trade cycle appeared in 1940, barely 4 years after the publication of Keynes' book *The General Theory of Employment, Interest and Money.* Although Keynes devoted Chapter 22 of the *General Theory* to 'Notes on the Trade Cycle,' he did not develop any full-fledged theory of the trade cycle because his chief concern was to develop an alternative theory of income and employment which could replace the classical theory. Kaldor's theory is a very simple and neat discussion of the trade cycle based on the Keynesian saving investment analysis.

Kaldor's trade cycle theory is simply an extension of the income determination model where the saving-supply function was of the form $S = -a + sY$ and the investment-demand function was of the form $I = I_A + eY$. The stability condition requires that the marginal propensity to invest (MP1) should be less than the marginal propensity to save (MPS), i.e., MPI < MPS. In other words, the slope of the investment demand function should be less than the slope of the saving supply function. This means that for the stability of equilibrium, the investment demand function must intersect the saving supply function from above. If the MPI > MPS, i.e., if the investment demand function intersects the saving supply function from below, the resulting equilibrium will be unstable.

In Figure 5.3, the saving supply and investment demand functions are linear. From

the point of view of trade cycle theory, these offer very little as in either case–stable or

Check Your Progress

3. State the principle of acceleration.
4. State one limitation of multiplier-accelerator interaction.

NOTES

unstable equilibrium–trade cycle will not occur while in the real world cyclical fluctuations have certainly occurred. For his business cycle analysis, Kaldor takes the nonlinear saving and investment functions. According to Kaldor, the saving supply and the investment demand functions cannot both be linear over the entire range of changes in the income which take place during the course of a business cycle.

Dividing the full trade cycle into relatively low, normal and relatively high income phases, the marginal propensity to invest will not be the same during all the three phases. The linear investment demand function, however, makes us believe that it will be uniform. During the course of business cycle, the nonlinear investment demand function will behave in such a way that the MPI or the slope of the investment demand function will be relatively low at both relatively low and relatively high levels of income. To express differently, the investment demand function is likely to be income-inelastic at low income levels due to the presence of excess plant capacity in the economy. It is also likely to be so at very high levels of income due to the high construction costs and the high cost and increasing difficulty of borrowing the funds.

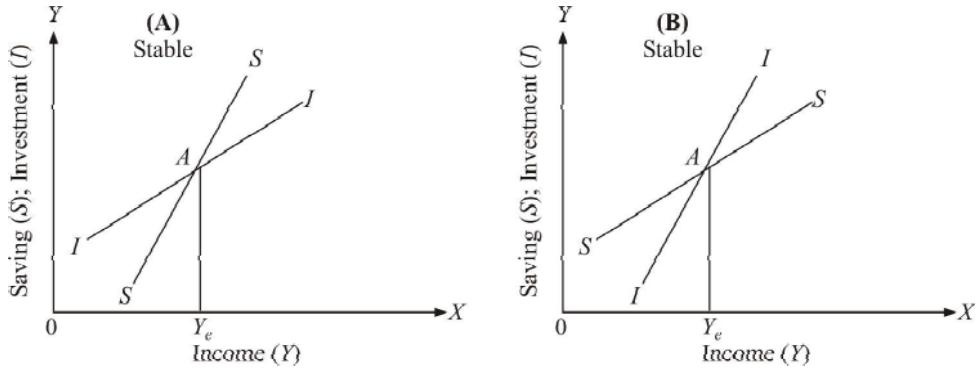


Fig. 5.3

According to Kaldor, the nonlinear investment demand function has the shape as shown in Figure 5.4(A). Like the investment demand function, the saving supply function is also nonlinear with the MPS varying corresponding to different income ranges. Thus, during the course of the trade cycle both for the relatively low and relatively high levels of income the MPS and therefore, the slope of the saving supply function will be high. Figure 5.4(B) shows such a saving supply function. In other words, the saving supply function is income-elastic at both very low and very high levels of income. The saving supply function behaves so because when income is very low, people try to maintain their former high standard of living to which they are accustomed with the result that a further fall in their income is accompanied by almost the full amount of fall in savings. When the income is very high, further increases in the income are accompanied by a very large proportional increase in savings because people expect these high levels of income to be transitory. Consequently, they do not increase their consumption at all or at any rate increase it very insignificantly. Thus, the saving supply function has a steep slope, both at relatively low and at relatively high levels of income.

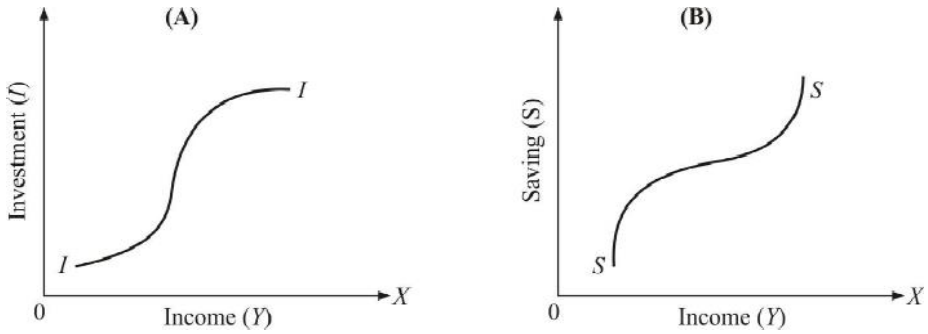


Fig. 5.4

By combining the investment demand and saving supply functions of Figure 5.4(A) and Figure 5.4(B) into a single diagram as shown in Figure 5.5, we obtain multiple equilibriums— both stable and unstable. At the equilibrium points *A* and *B*, the slope of the saving supply function is higher than the slope of the investment demand function. Consequently, the marginal propensity to save is higher than the marginal propensity to invest, i.e., MPS > MP1. Accordingly, the equilibrium points *A* and *B* which correspond to relatively low and relatively high levels of national income are the points of stable equilibrium and Y_A and Y_B are stable income levels. Below Y_A income and between the Y_B and Y_C levels of income, investment is more than saving, i.e., $I > S$. Consequently, income rises until it reaches the Y_A or Y_C level of income. At income levels ranging between Y_A and Y_B or above the Y_C levels of income, saving exceeds investment, i.e., $S > I$. Consequently, income falls until it reaches the Y_A or Y_B. level of income. *C* is the point of unstable equilibrium. Consequently, the corresponding income Y_B is the unstable income. If income is between Y_C and Y_B, it will rise to Y_C and if it is between Y_B and Y_A, it will fall to Y_A. In other words, if the level of income is disturbed even slightly from Y_B, it will never tend back to Y_B. Moving farther away from Y_B, it will move either toward Y_C or toward Y_A stabilizing itself at either of these two stable income levels depending upon the direction of disturbance. The economy will attain stability only either at some particular high level of income, such as Y_C or at some particular low level of income, such as Y_A.

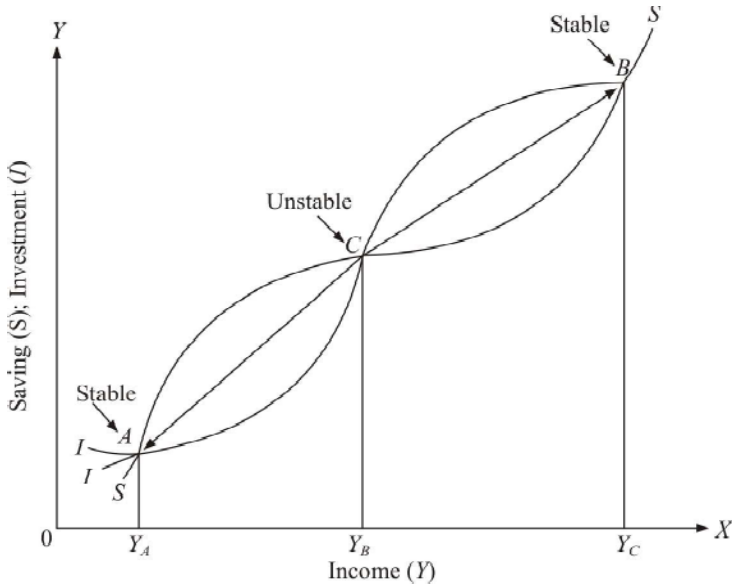


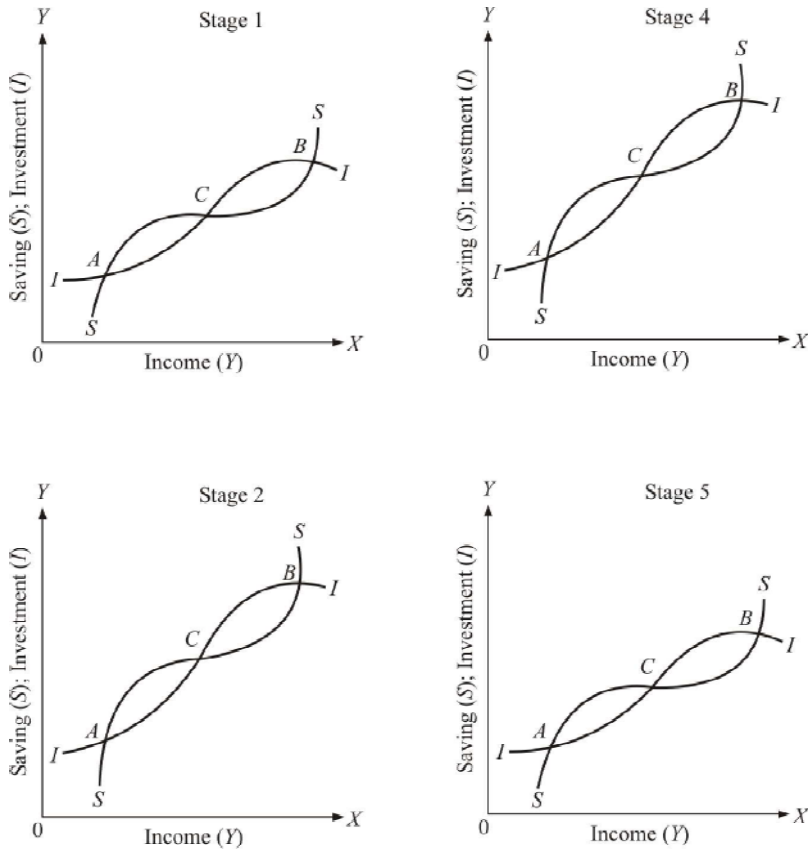
Fig. 5.5

NOTES

Figure 5.5 does not, however, explain anything about the occurrence of the trade cycle composed of the alternating contractions and expansions. It shows nothing more than the two positions of stable equilibrium towards either of which the income will tend to move. According to Kaldor, 'the key to the explanation of the trade cycle is to be found in the fact that each of these two positions is stable only *in the short period*, that as activity continues at either one of these levels, forces gradually accumulate which sooner or later will render that particular position unstable.'[9] If we can show that stable equilibrium at *B* becomes unstable over time forcing a movement towards *A* and *vice versa* we can move ahead to show the generation of the trade cycle.

Figure 5.5 shows that both saving and investment vary as income changes during the course of trade cycle. However, apart from income, investment and saving are also affected by another factor. According to Kaldor, the capital stock in the economy also causes cyclical changes in saving and investment. Saving is a direct function of the capital stock such that higher the capital stock, higher is the amount of saving and *vice versa*. Investment is an inverse function of the capital stock such that for any level of income higher is the capital stock, smaller is the amount of investment and *vice versa*.

According to Kaldor, the investment and saving curves drawn in Figure 5.5 are short-run curves which shift over time. At high levels of income, investment will increase more rapidly. Consequently, the economy's capital stock will continue to grow. After sometime, however, additions to capital stock will cause a downward shift in the investment demand curve. The saving supply curve will shift upward at high levels of income over time. This statement is similar to the view held by the under-consumptionists. This saving supply curve will shift upward at high levels of income because purchases of consumer durables will not increase as income increases.[10] Figure 5.5 shows changes in the equilibrium position as the capital stock of the economy changes over time.



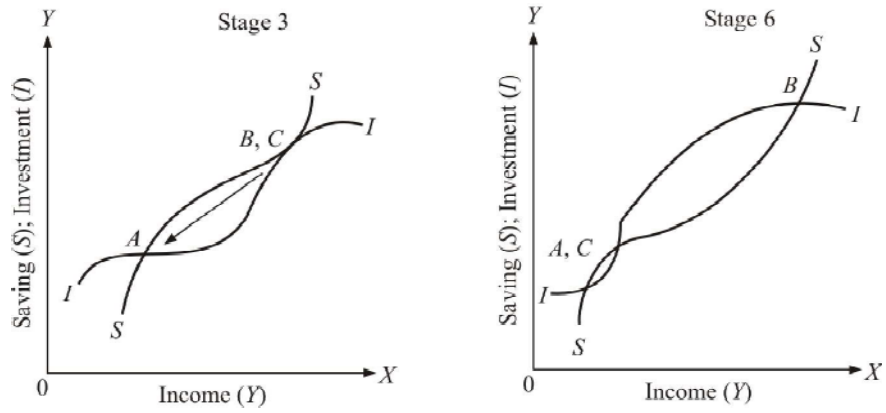


Fig. 5.6

In Figure 5.5, Stage 1 corresponds to Figure 5.5. We assume, to begin with, that the economy is initially in equilibrium at point *B* at relatively high or above normal level of income. Corresponding to this high level of equilibrium income, investment is correspondingly high. Higher the rate of investment, more rapid is the increase in an economy's total capital stock. As the capital stock grows, *ceteris paribus*, the marginal efficiency of capital falls causing a downward shift in the marginal efficiency of investment schedule which in terms of Figure 5.6 means a downward shift in the investment demand curve. At the same time, an increase in economy's capital stock, which is an increase in economy's wealth, shifts the saving supply curve upward. The upward shift in the saving supply curve and a downward shift in the investment demand curve gradually shift the position of *B* to the left and that of *C* to the right bringing the two points closer to one another as shown in Stage 2 in the Figure 5.6. Eventually, as a result of the gradual upward and downward shifts in the saving supply and investment demand curves respectively, these curves become tangential and points *B* and *C* coincide as shown in Stage 3 diagram of Figure 5.6.

Both to the right and left of the point of tangency between the investment demand and saving supply curves, saving is higher than investment, i.e., $S > I$. Consequently, the equilibrium at the *B, C* position in Stage 3 is unstable in the downward direction. Since deflationary pressures are at work, the economy will move downward from this unstable equilibrium position towards point *A* which shows the stable equilibrium position corresponding to a very low level of income. At this low level of income, there will be the problem of excess plant capacity. Consequently, the entire capital stock which has depreciated will not be replaced. Thus gross investment will be smaller than depreciation and, therefore, capital stock will decrease tending to shift the investment demand curve *II* upward. On the other hand, at low income the saving supply curve *SS* will either fall (as consumer durables wear out creating a demand for their replacement and as the demand increases for buying more new consumer durable goods) or flatten out. These shifts in the *II* and *SS* curves will tend to move point *C* closer to point *A* as shown in Stage 5 in Figure 5.6. Eventually, points *C* and *A* will coincide as shown in Stage 6 in Figure 5.6. Since both to the right and left of the point of tendency between the *II* and *SS* curves, shown by the position *A* and C, investment is more than saving, i.e., $I > S$, inflationary pressures are present in the economy. Consequently, the economy will move upward toward point *B*. This process in which the economy moves downward and upward between points *A* and *B* can continue indefinitely. Figure 5.6 shows one complete trade cycle. The cycles generated by this mechanism will not necessarily be of the same length nor will expansions and contractions be necessarily symmetrical. These

NOTES

characteristics will depend on the slopes of and the rate at which the investment demand curve *II* and the saving supply curve *SS* shift. It is also possible that no trade cycle will occur.

Kaldor's theory is simple to understand and is quite ingenious. The cyclical process explained by Kaldor is self-generating. The upward movement to high levels of income generates forces which produce a downward movement at low levels of income and *vice versa.* These cycle generating forces—changes in the average propensity to save and the increase and decrease in the capital stock—are inherent in the economic process. In other words, these forces are endogenous. In Kaldor's model of trade cycle, the non-linearities of the investment demand and saving supply functions which are due to excess capacity, the high costs of construction and the high cost and increased difficulty of borrowing the funds are crucial in explaining the occurrence of the trade cycle.

Kaldor's theory does not employ the acceleration principle in order to explain the trade cycle. In Kaldor's model, investment is related directly to the level of income and inversely to the amount of the economy's capital stock. This approach does away with the unrealistic and inflexible tying of investment to changes in income which is implied in the rigid acceleration principle, retaining, all the same, the basic idea of the acceleration principle. Kaldor's model, which employs the investment demand function of the type $I_t = I_a = eY_{t-1} - hK_t$ incorporating the capital stock adjustment principle, does not make any direct reference to the rate of change of income and output over successive periods. Although the model retains the link between changes in the aggregate output and investment, it has been done in such a way so as to avoid some of the weaknesses of the rigid acceleration principle.

5.5 MONETARY AND FISCAL POLICY FOR ECONOMIC STABILIZATION

Macroeconomic policy has become a very important economic policy instrument of modern welfare state to achieve the desired changes in the size and composition of national income and employment in the economy. After the publication of Keynes' well-known work *The General Theory of Employment, Interest and Money* in 1936, the achievement of price stability, guaranteeing employment to every able-bodied and willing-to-work citizen and raising the level of national product in order to raise the living standards of people in the country are now regarded as important social obligations of the state. To achieve these social goals, the government initiates an appropriate economic policy at the national level. The evolving of a proper economic policy for promoting economic growth with stability and ensuring economic justice for all the people involves the framing of an appropriate economic policy which, while aiming at promoting the rapid economic growth of the economy, reduces the glaring incomes and wealth inequalities in the system.

Check Your Progress

5. How is Kaldor's theory of trade cycle different from other theories?
6. How is the saving supply function similar to the investment demand function?

Indeed the goals of state's economic policy may be many and varied. At different times in history the state has employed economic policy for different purposes which always did not serve the interest of the community at large. For example, Queen Cleopatra of Greece had devalued the Greek drachma by a hefty 75 per cent to finance her luxurious living. Similarly, Charles the Bad, King of Navarre in 1383 had debased his country's coinage by one-third to secure funds to celebrate the release of the heir to the throne from French captivity. Again, one can find many instances of the French kings during the Middle Ages debasing their coins to finance their defensive wars against the English invasions. The goal of the Nazi government's economic policy in Germany was

to mobilize large surpluses to finance Hitler's massive rearmament programme of 1933–39 and his war of 1939–45. Today, however, the three major goals of government's economic policy are the achievement of full employment, stability of foreign exchange rate and price stability and promoting economic growth. All these goals are covered under the slogan 'rapid economic growth with stability'.

The two important subdivisions of macroeconomic policy are the monetary policy and the fiscal policy. These two policies are applied as mutually complementary policies to serve as instruments of the government's economic policy which is applied to achieve certain given social goals. Although there is often considerable overlapping between the monetary and fiscal policies because it is almost impossible to envisage any major fiscal or monetary policy measure which does not affect the other, still it is necessary to draw a meaningful distinction between the monetary policy on the one hand and fiscal policy on the other in order to circumscribe the scope of both these instruments of macroeconomic policy.

Demise and Re-emergence of Monetary Policy

Monetary policy had held the field for a considerable time and had occupied the pride of place in the classical and the neoclassical economic analysis. From Knut Wicksell to John Maynard Keynes (*Treatise on Money*) it was generally believed that by lowering the interest rate and concurrently the supply price or cost of capital goods, monetary policy was capable of raising the investment sufficiently to maintain constancy of the price level or sufficient to achieve utilization of all the factors of production. Conversely, it was believed that a rise in the interest rate would deter the marginal borrowers enough to contain total spending within the limits of total supply at the prevailing general price level. This view about the effectiveness of monetary policy assumed that the interest-elasticity of investment demand function was high enough to ensure absorption in the long period, if not in the short period of a business cycle, of full employment savings.

Soon, however, this high faith in the effectiveness of monetary policy was replaced by despair and the skeptics questioned whether small increases or decreases in the rate of interest, which was a small or rather minor element in total cost of production, could influence the spending decision of the borrowers. Consequently, monetary policy was unceremoniously dragged down from the position of honour. It suffered almost complete emasculation as a result of Keynes' attack. It lost all its prestige and had been given up as an economic policy instrument in the great depression of the 30s. We have seen earlier that in the 'Keynesian range', characterized by the liquidity trap, monetary policy is rendered completely helpless and no amount of increase in the money supply can raise income and employment since the rate of interest becomes 'sticky' at the liquidity trap rate of interest. This Keynesian theoretical innovation exercized a tremendous impact on the development of economics in the post-depression period. Prior to the Keynesian revolution, economists assumed, more or less unanimously, that fiscal policy—government's tax and spending policies—should be tailored to meet the community's public goods' requirement.

Following the Keynesian revolution, all this changed and it began to be increasingly argued that if monetary policy was useless in bringing about full employment, then government should pursue an active fiscal policy by spending more and taxing less. Several empirical studies[11] were mostly made during the 1930s and early 1940s which provided strength to the critics of monetary policy by concluding that firms' investment decisions were insensitive to changes in interest rate so that investment demand could be regarded as insensitive to changes in the interest rate. Based largely on the alleged interest inelasticity

NOTES

of the investment demand schedule, the critics of monetary policy received further support from the Keynesian view of liquidity preference which stated that monetary policy had little effect on interest rate which could not fall below some positive level, say below two per cent, due to the presence of liquidity trap. In due course of time, fiscal policy completely annihilated the monetary policy. Throughout the 1940s and even upto the mid-1950s, economists sang the unsavoury song of 'money does not matter'. Consequently, monetary policy was completely upstaged and fiscal policy was all the rage.

However, by the mid-1950s a reaction set in. It was being questioned by the skeptics that there may be a liquidity trap but who has ever seen it? The question asked was: does the mere theoretical possibility that in a very rare and limited set of circumstances, an expansionary monetary policy may fail to raise employment and output in the economy justify our discarding the monetary policy for all times and in all situations? The supporters of the monetary policy pleaded that it was unwise to regard monetary policy as generally ineffective on the basis either of the analysis or of the conditions of the 1930s. Those who still believed in the efficacy of monetary policy undertook serious researches to test the effectiveness of the monetary policy. These research studies clearly showed that 'money does matter'. Consequently, monetary policy staged a 'grand' comeback after remaining browbeaten for almost two decades. In the grand revival of monetary policy in the mid-1950s, the contributions of Milton Friedman, his collaborators, students and other followers of the Chicago School of economists have been very important. The Chicago University's Department of Economics has long been associated with the intellectual tradition of the quantity theory of money.

The brute force of events in the post-war period produced a counter-revolution. In many countries, including the United States of America, cheap money policies were pursued partly under the influence of Keynes' ideas. What was the result? Every such country either suffered from hectic open inflation or witnessed a network of partly effective and partly ineffective controls aimed at suppressing the inflationary pressure. In every case the stock of money increased as a consequence of pursuing the cheap money policy and so did the prices either openly or in the disguise of controls. No country that did not check the growth in money stock, succeeded in checking inflation while every country which could effectively stem the growth in the stock of money successfully checked the price rise.

Milton Friedman and some of his students published a collection of essays in 1956 which refocused attention on both the theoretical and empirical relevance of money.[12] Apart from this collection of essays, in 1963 was published Milton Friedman's and Anna Jacobson Schwartz's monumentally scholarly work entitled *A Monetary History of the United States—1867–1960*. The main thesis of this landmark study is that money does matter very much. In short, monetary policy has now regained its lost status and claims parity with fiscal policy as an area of concern, as a field of research and as a technique of economic stabilization.

Instruments of Monetary Policy

The various instruments of monetary policy which the central bank employs to achieve the goals of economic policy can be classified into the general[13] or quantitative instruments and the selective[14] or qualitative instruments. The general instruments employed by the central bank to carry out its monetary policy are the open market operations, changes in the minimum legal cash reserve ratio and changes in the bank or discount rate. All these three monetary policy instruments influence the credit-creating capacity of commercial banks in the economy by operating directly or indirectly on their excess cash reserves.

When the banks grant credit by making loans and purchasing securities their assets and liabilities portfolios are influenced. On the one hand, banks' assets increase but on the other hand there takes place a corresponding increase in their deposits, primarily the demand deposits. Since apart from keeping operating and safety margin cash reserves, banks are statutorily required to maintain a certain minimum percentage of their total deposit liabilities in the form of cash reserves with the central bank, their capacity to lend more is limited by their excess cash reserves position. If the banks have only sufficient cash reserves with them to meet their minimum legal reserves requirement against their total deposits, they cannot grant any more loans and purchase more securities. In such a situation, when the banks have the total cash reserves which are just adequate to cover their total deposits, further acquisition of earning assets by increasing their deposit liabilities would create a deficiency of cash reserves. Since the bulk of the deposits of banks comprise the money supply, its rate of growth is under the central bank's control by exercizing control over the volume of cash reserves held by the commercial banks by raising the percentage of the minimum legal cash reserve ratio requirement which the banks must maintain against their deposits.

The central bank can influence—increase or decrease—the commercial banks' cash reserves through its open market operations. These operations are broadly defined as the bulk purchase or sale of financial instruments by the central bank either in the primary market (open market type operations) or in the secondary market (full-fledged open market operations). The instruments which are commonly used for this purpose include treasury bills, central bank bills or prime commercial paper. The instrument of open market operations is the most effective instrument which is available to the central bank to carry out the objectives of its monetary policy. Being flexible, it enables the central bank to change the direction of its open market operations according to circumstances from a policy of increasing the cash reserves of the commercial banks to decreasing their cash reserves and *vice versa.*

Open market operations are either defensive or dynamic. Defensive operations are those which are taken to offset the other factors that change the volume of banks' cash reserves. For example, if gold outflows or increases in treasury deposits at the central bank are tending to reduce the commercial banks' cash reserves, the central bank may make offsetting government security purchases even though it is not trying to ease its credit policy. Conversely, it may buy securities during the tight money periods if other factors are tending to reduce the commercial banks' cash reserves too fast. Thus it is impossible to tell from a mere sale or purchase of securities whether the central bank is pursuing a defensive or a dynamic open market operations policy without knowing how the other factors are affecting the commercial banks' cash reserves.

In the developed countries with highly developed financial markets, open market operations are usually conducted with treasury bills and/or central bank bills. However, those countries where the financial markets are not fully developed[15] can conduct the open market type operations through the central bank intervention in the primary markets for securities. A common approach employed is the holding of regular auctions of the treasury or central bank bills and vary the net amount auctioned in order to influence the bank reserves. This instrument is often used in combination with other instruments—use of rediscount facilities and changes in the reserve requirements—to achieve the desired reserve impact and to smoothen the day-to-day fluctuations in liquidity.

The instrument of variable minimum legal cash reserve ratio requirement affects not only the total amount of commercial banks' reserves but also the amount of their excess cash reserves which in turn affects their total ability to lend. Thus, the central

NOTES

bank can carry out its expansionary monetary policy by increasing their total cash reserves through the open market purchases of securities or, their total cash reserves remaining unchanged, by decreasing the minimum legal cash reserve ratio requirement. As a result of decreasing the minimum legal cash reserves requirement, a part of the existing required cash reserves is reclassified as excess cash reserves and consequently becomes available for credit creation by the banks. In terms of the lending power of commercial banks, cash reserves so released are in effect similar to an addition in banks' excess reserves produced through the open market operations conducted by the central bank. However, due to practical difficulties inherent in making the frequent small changes in the minimum legal reserve ratio requirement, the central bank takes recourse to this instrument of monetary policy only in those situations in which relatively large increases or decreases in the excess cash reserves of banks are considered essential. In India, the Reserve Bank of India has raised the minimum legal cash reserve ratio several times during the 1970s and 1980s in order to exercize credit restraint on the commercial banks. In the 1990s, this ratio had, however, fallen substantially and at present stands at 6.5 per cent.

Changes in the discount or bank rate do not *per se* affect the cash reserves of the commercial banks. Such changes affect the cost at which financial accommodation in the form of borrowing can be made available to the banks from the central bank. From the point of view of controlling the lending or credit-creating capacity of the banks, the instrument of discount or bank rate is the least important of the three general instruments of credit control at the disposal of the central bank because banks generally borrow from the central bank not to expand their earning assets but to meet the shortfall in their cash reserves. And whenever the banks are forced to take recourse to borrowing from the central bank they repay such loans promptly.

Discount or bank rate and other forms of central bank credit control instruments are employed for three purposes—to relieve the shortage of liquidity with the banking system (lender of last resort); to control monetary and credit conditions; and to allocate credit in a selective manner. In operating the discount or bank rate facility, the central bank limits access in various ways. Some central banks rely on the market to limit the access. In such a situation, the discount rate has to be kept high enough so that as a first resort the banks are discouraged from approaching the central bank for financial accommodation and instead obtain funds from other financial sources such as the deposits and inter-bank markets. Other central banks, such as the US Federal Reserve and the German Bundsbank maintain the discount rate below the market levels and limit access to the facility administratively.

Some central banks use changes in the discount rate primarily to signal a change in the monetary policy. Still others use the discount rate window as their main instrument to influence the money market conditions.

The cash reserve ratio requirement instrument directly links the central bank and commercial bank liabilities by forcing the banks to hold a prescribed fraction of public's deposits in the form of currency or deposits with the central bank. It can be used as a means to sterilize excess liquidity with the banks. However, in so far as this instrument lacks flexibility, frequent and large changes in it would be disruptive and costly for the banks. Moreover, in so far as unremunerated reserves requirement is akin to a tax, it can lead to financial disintermediation—to a shift in the flow of funds away from the formal financial sector to the unregulated or informal financial markets. It happens to the extent that, the share of the financial holdings over which the authorities can exert monetary control decreases.

As the central bank indulges in the open market sales of government securities to restrain the lending or deposit-creating power of the banks, the move simultaneously exerts an upward pressure on the whole structure of interest rates because the bulk sale of securities, which has to be on a mass scale if the credit-creating power of the banks has to be curtailed, is possible only at the falling prices of the government securities sold by the central bank. A fall in the prices of securities raises the yields on these securities and tends to raise the yields on other securities among the higher interest rates. To the extent that demand for bank loans is interest-elastic, a rise in the interest rates cuts the aggregate demand for bank credit. The same result follows by increasing the minimum legal cash reserve ratio requirement for the banks. As a result of this action, banks' excess reserves on which they can raise the pyramid of credit are reduced, forcing them to raise the entire structure of their lending rates in order to discourage borrowers from borrowing. Generally, the central bank reinforces the action of one instrument by applying the other monetary policy instruments also. For example, the restrictive action through the open market sale of securities must be followed by raising the bank rate if the restrictive action has to be really effective. Although the banks typically borrow from the central bank to meet their cash reserves shortages but unless the central bank raises its bank rate as it reduces their total cash reserves to an extent which forces the banks to raise interest rates, there will be a danger of excessive borrowing on the part of banks from the central bank. Commercial banks might use the 'bank rate window' to replenish their cash reserves lost due to open market sales of securities by the central bank and relend these funds at higher interest rate. A rise in the bank rate can prevent the misuse of central bank's bank rate window. When the bank rate is raised under these circumstances, such rise in the bank rate is the effect of the credit tightening made possible through the open market sales of securities undertaken by the central bank.

Certain vital effects of changes in the central bank's bank or discount rate are psychological. Such effects are particularly important when observers feel that the bank rate is being used by the central bank to signal a shift in the direction of the monetary policy. In such cases, the financial markets react immediately—sometimes even in advance of central bank's action—when the move is anticipated. If the bank rate is raised, interest rates—particularly those on short-term securities—generally rise and credit markets tighten. Conversely, a cut in the bank rate, which clearly signals an easing of the central bank's monetary policy, is ordinarily followed by easy conditions in the money and capital markets in the economy.

Apart from these three general instruments of credit control there are the selective or qualitative credit control instruments which are employed by the central bank from time to time. Unlike the general instruments which affect the total volume of credit directly, the selective instruments of monetary policy affect the types of credit extended by the banks—these instruments affect the composition rather than the size of the loan portfolios of the commercial banks. The immediate object of imposing the selective credit controls is to regulate both the amount and the terms on which credit is extended by the banks for selective purposes. For example, since 1956 the Reserve Bank of India byemploying the instrument of selective credit control flexibly has successfully determined both the total amount of credit and the terms on which such amount of credit can be granted by the banks for certain specified purposes. In fact, ever since this instrument was first employed by the Reserve Bank of India in 1956, a spate of directives has been issued to the banks instructing them to change both the amount and the terms of credit supply for different purposes from time to time. There has hardly been any year, rather

any month, which did not witness the use of the selective credit control instruments being made by the Reserve Bank of India for some purpose or the other. This instrument has, however, now been discontinued.

NOTES

Unlike the general instruments of credit control, the selective credit control instruments enable the central bank to restrict unhealthy expansion of credit for specific purposes say, for financing the speculative purchase of sugar or rice and paddy which may be in short supply and whose stock-building on the part of speculators in likely to lead to price spurt without at the same time curbing credit expansion in general. In this respect, selective credit controls in certain such situations prove effective in meeting the goals of monetary policy where the general credit control instruments may fail.

Contra-cyclical Fiscal Policy

If fiscal policy has to be employed as an instrument of economic stability, it has to be contra-cyclical in nature. The government can contribute to raise the levels of employment, income and economic activity by spending more than its current income. Conversely, it will exert a contractionary effect on employment, income and economic activity by collecting more revenue from the people in the form of taxes than it spends. To use its fiscal policy as an instrument of economic stability, the government should carefully regulate both the time and size of its spending and tax revenue operations. A deficit in the budget in inflation will further aggravate inflation and will, therefore, act as a destabilizing factor rather than act as a stabilizing factor in the economy. But the same policy if enforced in recession will promote economic stability in initiating recovery. Similarly, surplus budgeting in recession by aggravating the fall in the level of aggregate demand will convert a mild recession into a great depression. The same policy, however, if pursued during boom will promote economic stability in the system.

If fiscal policy is to be used as an instrument of economic stability, it is essential to abandon the current practice of balancing the budget annually in the face of fluctuating employment and income. The spending and revenue programmes of the government, which constitute the budget, must be flexible. Rather than balance its budget annually, the government should balance the budget over the period of a trade cycle. A fiscal policy that would contribute most to the economic stability must be such as to produce a surplus of revenue over spending in prosperity with comparatively full employment and a surplus of spending over revenue in a period of depression with abnormally high unemployment. This means that the annual budget should be kept unbalanced. A balanced budget would only be desirable when the economy was operating at full employment level and showed no tendency either to expand or to contract. The fiscal policy of the government should have a feature of automatic stability so that needless delays pending the passage of new appropriation or tax laws may not hamper the smooth operation of fiscal policy. It should have built-in stabilizers which will function automatically and shall remove delays in the execution of the fiscal policy in the absence of built-in stabilizers. The Committee on Economic Development stated the principle of guidance for incorporating the built-in stabilizers in the fiscal policy in the following words:

> 'Set tax rates to balance the budget and provide a surplus for debt retirement at an agreed high level of employment and national income. Having set these rates, leave them alone unless there is some major change in national policy or condition of national life.'

The merit of this policy is not difficult to see. With the fall in national income, government revenue falls relatively to government outlays leading to deficit budget and

vice versa. As a built-in stabilizer, the fiscal policy cushion's the fluctuations bywithdrawing more purchasing power from the economy than it injects in the economy during a boom and *vise versa*.[16]

NOTES

Fiscal Policy in Inflation

When resources are fully employed and the economy is tormented by inflation, the appropriate fiscal remedy is to create a budget surplus in order to reduce the aggregate spending. If the total tax collections exceed the total government expenditure, the reduction in private spending caused by tax collections is not fully offset by government expenditure. Consequently, total spending will be less than what it would have been had the budget been balanced. This policy will directly attack the cause of inflation—the rate of increase in the aggregate spending which exceeds the rate of increase in the volume of goods and services which are available for making the purchases in the economy.

A budget surplus will have the largest impact on total spending and, therefore, in checking inflation if the surplus is impounded by the government. *First*, in and of itself the surplus reduces total spending. But if the surplus in impounded, the quantity of money in circulation will fall causing aggregate spending to be reduced still further. Since the taxes which give rise to the surplus in government budget are paid with cheques drawn by the public on commercial banks, the net effect is to reduce the demand deposits in banks by the amount of the budget surplus. Further, as the cheques are deposited by the government in its account with the central bank, the commercial banks' deposits at the central bank are transferred from commercial banks to the government account. This reduces the commercial banks' cash reserves and to the extent it reduces these reserves below the required or desired level, it forces the commercial banks to contract their loans. All this will hold good only if the surplus is impounded by the government.

It is, however, possible that the government may use this surplus to payoff its debt, i.e., to retire or purchase the outstanding government bonds. If the surplus is utilized for retiring the outstanding debt, total spending may or may not be reduced depending upon who owns the bonds which are retired. There are three possibilities. The bonds might be held by the (i) central bank; (ii) commercial banks; and (iii) public. If the budget surplus is used to retire the outstanding government bonds which are held by the individuals and business institutions who would hoard the money received for bonds, the reduction in the aggregate spending occasioned by the surplus in the budget would be the same as it would have been had the entire surplus been impounded by the government. Consequently, the total stock of money in the economy is not changed by the surplus accrual and the debt retirement; the velocity of money is, however, decreased. But the initial surplus and the fall in the velocity of circulation tend to reduce the aggregate spending. On the other hand, if the public spends the entire money received by it due to government retiring its debt the aggregate spending will not be reduced.

If the budget surplus is used to retire the outstanding government securities held by the central bank, the effect is precisely the same as it is when the government impounds the surplus. If the government employs the budget surplus for purposes of retiring the government securities held by the commercial banks, i.e., if it uses its deposits held at the central bank to buy bonds from the commercial banks, then as result of this operation the cash reserves of the commercial banks will increase enabling them to expand credit. Since the commercial banks' cash reserves are increased by the full amount of the budget surplus, their reserves are raised to the same level at which they had stood before the budget surplus had accrued. As a consequence, the banks are able

NOTES

to expand their loans and demand deposits to the same old level at which these stood before the budget surplus was built up. Although the primary reduction in aggregate spending caused by the surplus is still effective yet there is no net fall in the money supply and the secondary reductions which would have been caused in spending from this source are eliminated.

Fiscal Policy in Depression

In depression, the economy suffers from rising unemployment, falling income and shrinking economic activity. In slump, the private investment is very small. There is a large idle plant capacity awaiting utilization. Resources are there in the economy but there is no demand for them. The aggregate demand for current output falls very low. The economy faces the paradox of 'actual poverty amidst potential plenty'. In depression, when the existing aggregate private and government spending is too low to achieve full employment, the government must increase public spending by undertaking public works programmes on a massive scale and indirectly inducing people to spend more. The amount of government spending incurred on unemployment doles and payments made to veterans and the aged should be increased. The great merit of public works programmes is that they raise personal incomes and consumption by multiplier time of the original expenditure without depressing the marginal efficiency of investment in the private sector. Aggregate spending can be increased also by reducing the taxes. The effect of a tax-cut would be to increase the amount of disposable income of the individuals and business firms. Sales tax should be abolished and excise duties on consumer goods satisfying the community's basic needs must be reduced.

To relieve the economy of depression, it is not enough to increase the aggregate consumption; aggregate investment should also be simultaneously raised. Fiscal policy can induce changes in the aggregate investment demand by making appropriate changes in the tax structure. Since the marginal efficiency of capital of private investment should be raised, business and corporate taxes should be reduced. Firms engaged in the capital formation in depression should be allowed tax concessions. Government's debt policy should be so designed that public debt should be retired in depression so that the disposable income of the bond-holders may increase causing substantial increase in the aggregate spending in the economy. During depression, like the one of the 1930s, when the *LM* curve becomes almost flat at very low rate of interest, fiscal policy action in the form of increase in government expenditure is most effective in raising the level of aggregate effective demand and employment in the economy.

Fiscal Policy and Economic Growth

The use of fiscal policy for attaining full employment and stable price level in the economy is a development of the past six decades beginning during the 1930s. It was due to (i) the ineffectiveness of monetary policy as a means to remove unemployment during the great depression; (ii) the 'new economies' which was developed by Keynes; and (iii) the increasing importance of government spending and taxation in national income and output. As an instrument of growth with stability, fiscal policy should be so employed that while promoting consumption and investment to the level of optimum utilization of economy's resources it may check inflation. Accelerating the rate of growth requires the allocation of a higher proportion of the fully employed resources to those activities which increase the productive capacity of the economy. In other words, the fraction of the full employment real output devoted to consumption must decrease while that devoted to investment

should increase. Fiscal policy through its tax instruments should encourage investment and discourage consumption so that the production may increase. It is also necessary to increase the rate of capital formation in the economy by reducing the high income-tax rates on personal income.

NOTES

Fiscal Policy Lags

Like the lags in monetary policy, fiscal policy is also subject to inside and outside lags. So far as the inside recognition lag is concerned, it is more or less the same as in the case of monetary policy. So far as the inside action lag is concerned, this lag arises on account of delay on the part of the government to act in the matter. Due to the fact that all significant changes in tax and expenditure require the prior approval of parliament and state legislatures, the action lag for fiscal policy is long and variable. The actual legislative process surrounding the fiscal policy decisions is very cumbersome and time consuming and renders the fiscal policy a wholly inappropriate instrument of economic stabilization and growth. While the inside action lag for fiscal policy is longest, the outside lag in fiscal policy is shorter than the outside lag in monetary policy because the full effects of fiscal policy actions are felt by the economy significantly more quickly than the full effects of monetary policy actions are felt. According to Rasche and Shapiro, 75 per cent of the full effect of changes in federal defence expenditure is realized in 9 months and of changes in federal personal income tax in six months.[17]

Monetary and Fiscal Policies are Complementary

As instruments of government's economic policy, monetary and fiscal policies are complementary. While the monetary policy influences the level of aggregate income and spending in the economy by influencing the total money supply and the cost of borrowmg funds from the banks, fiscal policy affects income and spending through its effects on the size, composition and timing of the government spending and revenue. In inflation, economic stability can be achieved quickly and effectively by combining the policy of surplus budgeting with dear money policy. Conversely, in slump recovery can be started more quickly by reinforcing the policy of deficit budgeting with the cheap money policy. Thus, for achieving the economic stability quickly it is necessary to coordinate effectively the two macroeconomic stability instruments. The importance of the monetary and fiscal policies in achieving economic stability was stressed by Mr J Cameron Thomson of the Committee on Economic Development in his testimony before the Douglas Sub-Committee in the following words.

'Fiscal, monetary, and debt policies are appropriate means for attacking the problem of instability in a free society. The problem of instability is essentially a problem of broad forces affecting the overall magnitudes of the economy. The problem arises when millions of workers are simultaneously unemployed, or when there is a general, although probably uneven, rise of most prices. The advantage of fiscal, monetary and debt policies is that they allow the government to influence the overall forces—especially the level of aggregate demand—that determine the stability of the economy without necessarily involving the government in detailed control of the particulars of the economy. These overall measures will, of course, affect different individuals and businesses differently. But the differences are determined by the market process, not by government decisions...'

Check Your Progress

7. What are the two important subdivisions of macroeconomic policy?
8. What are defensive open market operations?
9. State the result of an increase or raise in the bank rate.
10. State one merit of public works programmes.

5.6 SUMMARY

NOTES

In this unit, you have learnt that,

- The innovations theory of trade cycle, associated with the name of well-known economist Joseph A Schumpeter, makes innovations the central cause of the recurrence of business cycle in the modern industrial economies.
- An innovation is the initial application of an invention to commercial production while an invention is the discovery or development of a new process, product or service by scientists or engineers.
- According to Schumpeter, the innovators, formally designated as entrepreneurs, play a leading role in the entire process of capitalistic evolution.
- The innovators comprise the microscopic group of dashing pioneering members of the business community who are constantly on the look-out for something new and different.
- The lower turning point of trade cycle, i.e., recovery occurs when following a period of readjustment to the changed economic circumstances created by the previous boom, prices have fallen to a point where the pioneering businessmen, entertaining the hope that any further fall is most unlikely to happen, once again start innovating.
- The innovations theory of trade cycle suffers from several weaknesses. The depth of the depression is not entirely determined by the disturbances caused by the innovations; it is also attributable to other factors which vary in intensity from one trade cycle to another.
- The nature of the competitive impact of innovations is also very significant in determining the magnitude of the trade cycle.
- If the impact of demand for the new product is reflected in a small decrease in the demand for all other products in general, the loss of demand felt by each one of the old industry will be comparatively mild and readjustment will be easily accomplished.
- In Samuelson's model, interaction between the multiplier and the accelerator is an endogenous force which generates business cycle in the economy.
- In the model of the trade cycle developed by Nicholas Kaldor, consumption (or saving) is a function of income while investment is directly related to income and inversely related to the stock of capital.
- Kaldor's trade cycle theory is simply an extension of the income determination model where the saving-supply function was of the form $S = -a + sY$ and the investment-demand function was of the form $I = + eY$.
- In Kaldor's model, investment is related directly to the level of income and inversely to the amount of the economy's capital stock.
- This approach does away with the unrealistic and inflexible tying of investment to changes in income which is implied in the rigid acceleration principle, retaining, all the same, the basic idea of the acceleration principle.
- Macroeconomic policy has become a very important economic policy instrument of modern welfare state to achieve the desired changes in the size and composition of national income and employment in the economy.

- The goals of state's economic policy may be many and varied. At different times in history the state has employed economic policy for different purposes which always did not serve the interest of the community at large.
- The two important subdivisions of macroeconomic policy are the monetary policy and the fiscal policy. These two policies are applied as mutually complementary policies to serve as instruments of the government's economic policy which is applied to achieve certain given social goals.
- The various instruments of monetary policy which the central bank employs to achieve the goals of economic policy can be classified into the general or quantitative instruments and the selective or qualitative instruments.
- Open market operations are either defensive or dynamic. Defensive operations are those which are taken to offset the other factors that change the volume of banks' cash reserves.
- Discount or bank rate and other forms of central bank credit control instruments are employed for three purposes—to relieve the shortage of liquidity with the banking system (lender of last resort); to control monetary and credit conditions; and to allocate credit in a selective manner.
- When resources are fully employed and the economy is tormented by inflation, the appropriate fiscal remedy is to create a budget surplus in order to reduce the aggregate spending.
- The use of fiscal policy for attaining full employment and stable price level in the economy is a development of the past six decades beginning during the 1930s.
- Fiscal, monetary, and debt policies are appropriate means for attacking the problem of instability in a free society. The problem of instability is essentially a problem of broad forces affecting the overall magnitudes of the economy.

5.7 KEY TERMS

- **Cash reserve ratio**: Cash Reserve Ratio (CRR) is a specified minimum fraction of the total deposits of customers, which commercial banks have to hold as reserves either in cash or as deposits with the central bank.
- **Contra-cyclical fiscal policy:** Government policy aimed at reducing or neutralizing anti-social effects of economic cycles is known as contra-cyclical fiscal policy.

5.8 ANSWERS TO 'CHECK YOUR PROGRESS'

1. The period required for the construction of factories for innovations to be carried through is important in determining the duration and magnitude of the trade cycle. Longer the period required by the innovation to be carried through, more violent is the necessary readjustment. If the capital equipment for new innovations can be manufactured quickly, the new products by becoming available for sale in the market quickly will cause readjustments to take place before the development of extreme maladjustments.

NOTES

2. An innovator is one who is the first to make the commercial application of invention and consequently bears the risks involved in introducing a new product or service or process in the market. The success or failure of an innovation can only be proved by actually producing and marketing the new product.
3. The acceleration principle explains that the net investment I_t during any given time period *t* depends upon the value of the accelerator v and the change in income in time period *t*, i.e.,

   $I_t = \nu \Delta Y_t = \nu (Y_t - Y_{t-1})$.
4. One limitation of multiplier-accelerator interaction is the weaknesses of the rigid acceleration principle. Since the naive acceleration principle is unacceptable, the multiplier-accelerator interaction which incorporates the acceleration principle is also unacceptable as a perfect explanation of the occurrence of the trade cycle.
5. In the model of the trade cycle developed by Nicholas Kaldor, consumption (or saving) is a function of income while investment is directly related to income and inversely related to the stock of capital. In other words, the investment demand function is of the capital adjustment type.
6. Like the investment demand function, the saving supply function is also nonlinear with the MPS varying corresponding to different income ranges.
7. The two important subdivisions of macroeconomic policy are the monetary policy and the fiscal policy. These two policies are applied as mutually complementary policies to serve as instruments of the government's economic policy which is applied to achieve certain given social goals.
8. Defensive open-market operations are those which are taken to offset the other factors that change the volume of banks' cash reserves. For example, if gold outflows or increases in treasury deposits at the central bank are tending to reduce the commercial banks' cash reserves, the central bank may make offsetting government security purchases even though it is not trying to ease its credit policy.
9. If the bank rate is raised, interest rates—particularly those on short-term securities—generally rise and credit markets tighten.
10. The great merit of public works programmes is that they raise personal incomes and consumption by multiplier time of the original expenditure without depressing the marginal efficiency of investment in the private sector. Aggregate spending can be increased also by reducing the taxes.

5.9 QUESTIONS AND EXERCISES

Short-Answer Questions

1. What are trade cycles? Is it correct to say that the trade cycle is generated by the interaction of the multiplier and the accelerator?
2. State five limitations of innovations theory of trade cycle.
3. Write a short note on the cyclical process in Kaldor's theory.
4. State the inside and outside lags of the fiscal policy.

Long-Answer Questions

1. Explain critically Hicks' theory of the trade cycle.
2. Discuss Schumpeter's theory of the business cycle.
3. Explain the concepts of multiplier and accelerator and the effect of their combined action on the generation of trade cycle.
4. Discuss the instruments of monetary policy.

5.10 FURTHER READING

Mankiw, N Gregory. 2010. *Macroeconomics.* Worth Publishers, New York.

Shapiro, Edward. 1996. *Macroeconomic Analysis.* Galgotia Publication, New Delhi.

Branson, WH. 1979. *Macroeconomic Theory and Policy*. Universal Book Stall, New Delhi.

Jha, R. 1999. *Contemporary Macroeconomic Theory and Policy*. New Age International, New Delhi.

Gupta, SB. 1995. *Monetary Economics: Instruments and Policy*. S Chand & Co. New Delhi.

Endnotes

[1] Joseph A Schumpeter, 'The Explanation of the Business *Cycle*,' *Economica*, December 1927, p. 295.

[2] Joseph A Schumpeter, *op. cit.,* p. 102.

[3] Joseph A Schumpeter, *The Theory of Economic Development,* 1934, p. 224.

[4] When the innovations are financed out of real savings, the rate of innovations will be slower. Furthermore, in the absence of inflation caused by the credit financing of the innovations, the disturbances in the price structure will be far less severe than those in the case of credit-financed innovations.

[5] Joseph A Schumpeter, *op. cit.,* p. 166–7.

[6] It can be easily shown that there is no fundamental difference between Samuelson's and Hicks' versions of the multiplier-accelerator model. It can be seen that if we convert, Hicks' version of the model into Samuelson's version of the model, the value of the accelerator v according to Hicks will be higher than the value of the accelerator bv in Samuelson's version since $b < 1$. It is evident from the following two equations:

$$Y_t = G_t + bY_t + bv\,(Y_{t-1} - Y_{t-2}) \text{ (According to Samuelson)}$$

$$Y_c = G_t + bY_t + v\,(Y_{t-1} - Y_{t-2}) \text{ (According to Hicks)}$$

Both these equations are, however, identical except the additional presence of b in the first equation. Since the value of b is less than one, the value of bv will be less than the value of v alone.

[7] James S Duesenberry, 'Hicks on the Trade Cycle,' *The Quarterly Journal of Economics*, August 1950, p. 464–76.

[8] This discussion is based on Kaldor's.article 'A Model of the Trade Cycle.' published in *The Economic Journal*, Volume 50, March 1940, p. 78-92, reprinted in *Essays on Economic Stability and Growth*, 1960, Free Press. New York, p. 177–2.

[9] Nicholas Kaldor, *op. cit.,* p. 182.

[10] It is more likely that the slope of the saving supply function will fall with the people becoming more accustomed to the high levels of income. They will, therefore, save a smaller proportion of their high income. However, this modification does not affect the general nature of the argument. The argument will hold even if the saving supply curve remains stahle and only the investment demand curve shifts downwards.

[11] See the articles by James E Meade and R W S Andrews reprinted in *Oxford Studies in the Price Mechanism,* 1951, p. 27–30 and p. 51–66.

[12] Milton Friedman (ed.), *Studies in the Quantity Theory of Money,* 1956.

[13]

UNIT 6 DEMAND FOR AND SUPPLY OF MONEY

Structure

6.0 INTRODUCTION

The demand for money is the desired holding of financial assets in the form of money: that is, cash or bank deposits. It can refer to the demand for money narrowly defined as M_1 (non-interest-bearing holdings), or for money in the broader sense of M_2 or M_3.

Money in the sense of M_1 is dominated as a store of value by interest-bearing assets. However, money is necessary to carry out transactions; in other words, it provides liquidity. This creates a trade-off between the liquidity advantage of holding money and the interest advantage of holding other assets. The demand for money is a result of this trade-off regarding the form in which a person's wealth should be held. In economics, the money supply or money stock, is the total amount of monetary assets available in an economy at a specific time. In this unit, the several theories and concepts regarding the demand and supply of money have been discussed in detail.

6.1 UNIT OBJECTIVES

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

- Discuss the concept of demand for and supply of money with the help of equations
- Differentiate between endogenous and exogenous money supply
- Analyse Tobin's portfolio selection and Baumol's transaction demand for money
- Discuss Friedman's restatement of quantity theory of money its criticism
- Differentiate between the concept of inside and outside money
- Explain the concept of financial intermediaries as explained by Gurley and Shaw

6.2 TOBIN'S PORTFOLIO SELECTION AND BAUMOL'S TRANSACTION DEMAND

NOTES

In his classic article, James Tobin has developed a sophisticated analysis of portfolio selection under uncertainty which meets the shortcomings present in Keynes' analysis of the speculative demand for money. According to Keynes, an individual wealth-holder discretely decides to hold either bonds or money and not both. According to Tobin, rational behaviour induces an individual to hold a wealth portfolio which is comprised of both bonds and money.

In order to demonstrate Tobin's analysis of rational portfolio selection under uncertainty, let us assume that at the beginning of some given time period, initially an individual wealth-holder possesses a wealth portfolio of certain given size W_0. We also make the most likely assumption—which is an essential part of rational behaviour—that the individual wealth-holder prefers more wealth to less wealth. Given these two assumptions, the individual wealth-holder is faced with the formidable problem of deciding as to what fraction of his total wealth portfolio he should hold in the form of consols (we assume that consol is the only form of bond available to him) and what fraction he should hold in money.

According to the Keynesian theory, the individual holds either all the money or all consols based on the expected future rate of interest—the *normal rate*. However, if the more realistic assumptions of uncertainty about the future rate of interest and of the equal probability on the part of average individual wealth-holder to underestimate and overestimate this interest rate are made, then after many years the average capital gain or loss will be zero although in any given year either a capital gain or a capital loss can occur. The extent to which a capital gain or loss will occur will depend on the amount of uncertainty about the future interest rate. Tobin demonstrated that for a given uncertainty about the future interest rate, the wealth-holder bears a greater or smaller risk as he holds a larger or smaller proportion of his total wealth portfolio in the form of consols.

A wealth-holder who operates under the above assumed condition will maximize the expected growth of wealth if he holds his total wealth in the form of consols. He will, however, simultaneously also bear the maximum risk of a possible capital gain or loss. If, on the other hand, the wealth-holder holds his entire wealth in the form of money, he will assume zero risk but his wealth will not grow at all. In what exact money-consol mix he will hold his total wealth will *ceteris paribus* depend on the temperament and attitude towards risk of the individual wealth-holder. If he is a risk averter–preferring less to more risk at a given rate of return–then the decision about the proportion in which to hold his wealth in the form of consols and money will depend upon his attitude towards the trade-off–substitution of certainty with zero growth in wealth and risk with growth in wealth. Figure 6.1 shows the individual wealth-holder's attitude towards risk and the growth in his total wealth.

In Figure 6.1, on the X-*axis* has been shown the portfolio risk while on the Y-axis has been shown the expected worth of the portfolio at the end of the given time period. Since the portfolio risk depends on the fraction of the total wealth portfolio held in the form of consols, an increasing portfolio risk implies an increasing fraction of the portfolio held in the form of consols. In the figure, I_{-1} , I_0, I_1 and I_2 are the individual wealth-holder's indifference curves. Indifference curve I_2 is preferred to indifference curve I_1. Indifference curve I_1 is preferred to indifference curve I_0 and difference curve I_0 is

preferred to indifference curve I_{-1}. The slope of the indifference curves is positive due to our assumption that increased risk is desirable only for a higher level of wealth at the end of the period and that higher growth of portfolio is preferred for a given level of portfolio risk. The indifference curves are convex towards the origin of axes showing that with the increase in wealth, the marginal utility of wealth for the individual wealth-holder falls and he is less willing (more reluctant) to bear the greater risk in order to increase the total stock of his wealth.

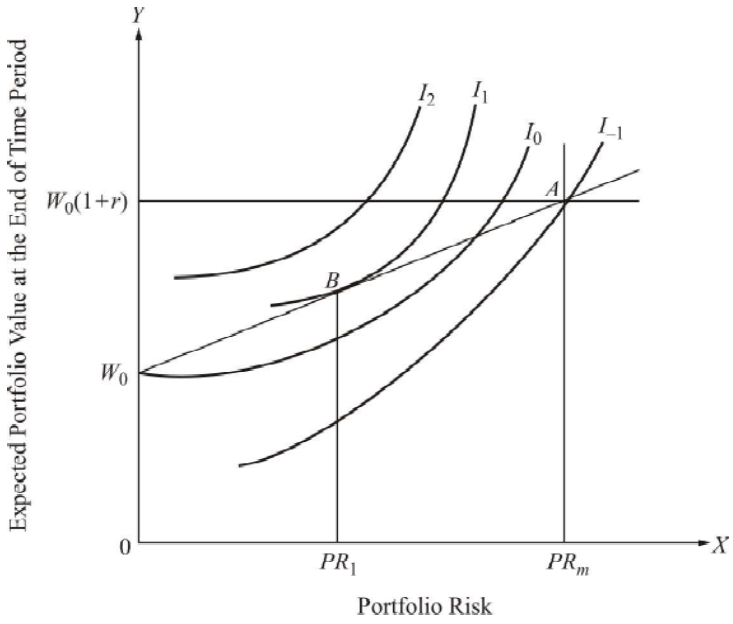


Fig. 6.1

If the individual wealth-holder prefers to bear no risk and keeps his entire initial stock of wealth W_0 in the form of money, then at the end of any given time period his wealth stock will not increase and it will remain fixed at W_0 as shown in Figure 6.1. Consequently, the wealth-holder's welfare does not increase but remains constant at indifference curve I_0. On the other hand, if the wealth-holder decides to keep his entire wealth in the form of consols, then his expected wealth at the end of the time period will become $W_0(1 + r)$, where r is the rate of return on consols. It should however, be stressed that the expected wealth is not the amount of wealth which the individual has acquired at the end of any single given time period; it is rather the average wealth at the end of many similar time periods. At the end of any given time period, the stock of wealth may be either below or above $W_0(1 + r)$. The attitude of the wealth-holder, who is a risk averter, will be asymmetrical towards capital gain or loss since he considers the marginal utility of a rupee worth of capital gain worth less than the marginal disutility (pain) suffered from a rupee worth of capital loss.

Given the amount of uncertainty associated with the future interest rate, the portfolio risk will be maximum when the entire stock of wealth W_0 is held in the form of consols. This maximum risk has been mentioned as PR_m in Figure 6.1. Corresponding to this maximum risk, the expected portfolio value is also maximised at $W_0(1 + r)$ indicated by point A in the figure. For a given interest rate, the expected gain in wealth by holding wealth in different money-consol combinations ranging from holding only money W_0 to holding only consols $W_0(1 + r)$ shown by point A has been shown by the straight line W_0A. This line is like the budget constraint line in the theory of consumer choice. The

individual wealth-holder will maximize his total welfare at point *B* where indifference curve I_1 is tangential to W_0A line showing that the wealth-holder diversifies his wealth portfolio between money and consols rather than hold his wealth either entirely in the form of money or entirely in the form of consols. If he does the former, he will only attain indifference curve I_1 while by holding entirely consols (point *A*) he will only worsen his position since he will come down on a still lower indifference curve I_{-1} representing smaller welfare.

The liquidity preference function–Keynes' speculative demand for money function–for an individual wealth-holder may now be derived by analysing the effect on the portfolio selection under uncertainty of varying the rate of interest, keeping the wealth-holder's total endowment of wealth constant at W_0. We also assume that there is no change in the amount of uncertainty regarding the future rate of interest. Under these two assumptions, the maximum portfolio risk remains unchanged at the PR_m level. Figure 6.3 shows that the vertical intercept W_0 of the budget constraint line W_0A is the initial given wealth stock and the slope of the line is–

$$\frac{W_0(1+r) - W_0}{PR_m} = \frac{rW_0}{PR_m}$$

Since both W_0 and PR_m are constants, the slope of the budget constraint line W_0A depends on the rate of interest. An increase in the rate of interest expected to be paid on the consols will mean an increase in the slope of the line and *vice versa*.

A change in the expected rate of interest on consols will affect the portfolio selection decision of the individual wealth-holder by causing a change in the proportion in which he will hold money and consols. In the normal course, as the rate of interest increases, the proportion in which the individual wealth-holder will keep his total wealth in the form of consols will increase. It is so because as the rate of interest increases, with the given uncertainty about the future rate of interest, the opportunity cost in the form of interest-yield forgone of holding money increases inducing the wealth-holder to decrease the proportion of money in his portfolio. Figure 6.2(A) shows the effect of changes in the interest rate on the portfolio selection under uncertainty while Figure 6.4(B) shows the negatively sloping liquidity preference (speculative demand) curve of the individual wealth-holder which has been derived from Figure 6.2(A).

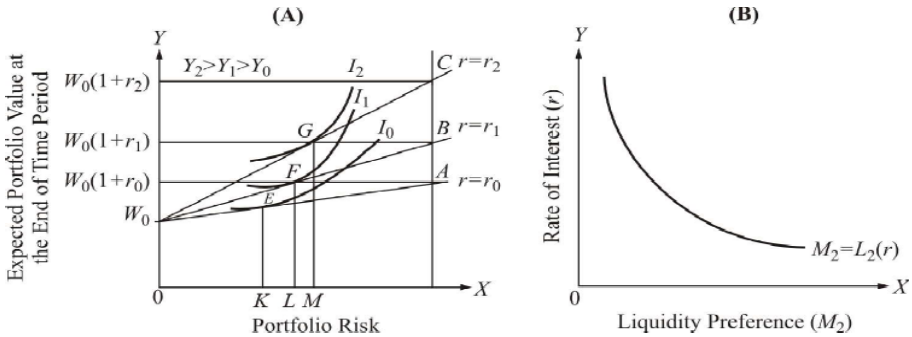


Fig. 6.2

In Figure 6.2(A), three budget-constraint lines have been drawn for the three different interest rates r_0, r_1, and r_2 such that $r_2 > r_1 > r_0$. As the rate of interest increases from r_0 to r_1 to r_2, the rational individual wealth-holder moves away from the initial

equilibrium portfolio-mix shown by the tangency point *E* to that shown by the tangency point *F* and finally to the next one shown by the tangency point *G*. As the interest rate increases, the increase in the wealth-holder's welfare which has been shown by the movement on to the higher indifference curves is accompanied by an increase in the amount of consols (or decrease in the amount of money held) and also by increase in the portfolio risk. The liquidity preference curve for the individual wealth-holder is a continuously downward sloping curve as shown in Figure 6.2(B). This kind of negatively sloping curve is supported empirically by the studies which have been made in order to measure the interest elasticity of the demand for money.

Although Tobin's theory of the demand for money yields results similar to Keynes' analysis, it is superior to Keynes' theory. *First*, it reveals rationality on the part of the individual wealth-holders who exhibit a continuous liquidity preference function. *Second*, the fact that they simultaneously hold cash and consols in their asset portfolio is not based on the belief that in future the interest rate will change in only one direction; it is rather based on the likely assumption of uncertainty entertained by the wealth-holders regarding the future movements in the rate of interest. *Third*, Tobin's analysis, in which the demand for money is determined by the individual wealth-holder's behaviour towards risk, can be extended to the problem of asset choice when the wealth-holder has more alternatives available to him than the only two alternatives—money and consols—of holding his wealth portfolio. William J Baumol has developed the theory of the transactions demand for money on the basis of business control. Business firms demand the inventory of money in order to facilitate their operations. However, by holding money the firm incurs cost in the form of interest income forgone. Consequently, the firm will strive to hold the optimal inventory of money as it will minimize its cost.

In order to find a solution to the problem of inventory management, it is assumed that the firm has knowledge about the size of its total future transactions. Baumol concentrates only on the transactions demand for money excluding the precautionary and speculative liquidity preference. It is also assumed that the payments are made in a steady flow and that during any given time period (say one year) *y* nominal rupees's worth of income transactions will take place. For the sake of simplicity, it is further assumed that the firm obtains money in order to carry out the transactions by selling the bonds. When the firm converts into cash a rupee's worth of bond, it incurs the opportunity cost (interest income forgone) of *r* rupees per rupee per year. If at the beginning of the year, the firm encashes bonds equivalent to an amount of cash M_1 which is enough to meet the total transactions during the year, then the interest opportunity cost O_c will be:

$$O_c = \frac{yr}{2} = \frac{M_1 r}{2}$$

where *r* = the annual interest rate.

The opportunity cost is $\frac{M_1 r}{2}$ rather than $M\ r$ because under the assumption made the transactions take place in a steady flow as shown in Figure 6.3 where at the beginning of the year the firm (or individual) has $M_1 = y$ rupees and at the end of the year it has zero rupees. Consequently, the firm possesses on an average of $\frac{M_1}{2}$ rupees at any given time. If the withdrawals are made twice in a year, the amount of money withdrawn will be:

$$M = \frac{y}{}$$

NOTES

and the average amount of rupees held by the firm will be:

$$\frac{M_1}{2} = \frac{y/2}{2}$$

The opportunity cost for the half-year will be:

$$\frac{(y/2)(y/2)}{2}$$

Consequently, the yearly opportunity cost O_c will be:

$$O_c = \left(\frac{y}{2}\right)\left(\frac{r}{2}\right)$$

where $(r/2)$ is the six-monthly interest rate. Since $y = 2M_1$ the annual interest opportunity cost O_c can also be expressed as:

$$O_c = \frac{M_1 r}{2}$$

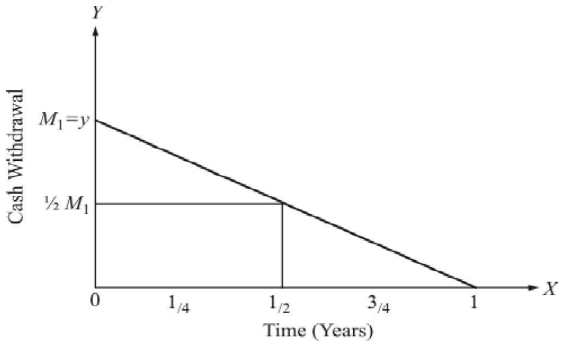


Fig. 6.3

The opportunity cost does not, however, comprise the interest cost alone. The firm also incurs some fixed cost, say of *b* rupees, for each withdrawal. This non-interest fixed cost consists of all non-interest expenses including the brokerage paid for converting bonds into cash, transport costs which are incurred in visiting the bank and so on. Since each year y/M_1 withdrawals are made, the yearly non-interest fixed cost incurred by the firm is by/M_1. This gives us the total (interest and non-interest) opportunity cost π as:

$$\pi = \frac{by}{M_1} + \frac{M_1 r}{2}$$

The problem involved here for the firm is of determining that amount of inventory cash M_1 for which the total opportunity cost π is minimized. According to Baumol, this will occur when–

$$M_1 = \sqrt{\frac{2by}{r}}$$

By dividing both sides of the equation by the general price level *P*, we get the transactions demand for real money balances which is:

$$M_1 = \sqrt{\frac{2(b/P)(y/P)}{r}} = \sqrt{\frac{2by}{r}}$$

where M_1 = transactions demand for money balances

b = real non-interest opportunity cost

y = real income.

The equation states that the loss-minimizing firm's demand for the real cash balances for transactions purpose is inversely proportional to the square-root of the real income and inversely proportional to the square-root of the interest rate. This proposition is frequently referred to as the square-root rule.

Baumol's analysis has two chief merits. *First,* it shows that the transactions demand for cash balances is interest-elastic—a conclusion which is insignificant in Keynes' approach. *Second,* it shows that the transactions demand for cash balances increases less than proportionately with the increase in income due to the economies of scale experienced in the use of money—a conclusion which is in sharp conflict with the cash balances quantity theory of money approach according to which the transactions demand for money bears a rigid proportionality relationship to the level or income.

Besides the approaches of James Tobin and William Baumol, there is Milton Friedman's approach to the demand for money. According to Friedman, the demand for money is simply an application of a more general theory of the demand for capital assets and it depends upon several factors. Friedman's theory of the demand for money is part of his modern quantity theory of money. To state very briefly, Milton Friedman argues that the demand for money by the wealth-holders depends on the following factors.

- Total real wealth
- Fraction or percentage of total wealth kept in the non-human form of wealth
- Opportunity cost of holding money
- Other variables which determine the utility of money

6.3 FRIEDMAN'S RESTATEMENT OF QUANTITY THEORY OF MONEY

John Maynard Keynes' monumentally scholarly work titled *The General Theory of Employment, Interest and Money* was highly critical of the old quantity theory of money. It was Keynes' contention that the quantity theory of money was wrong in singling out the general price level as the sole determinant of the demand for money and changes in the general price level as being principally determined by changes in the supply of money. As a result of Keynes' attack on it, the quantity theory of money was dismissed by most economists until it was revived in the late 1950s as result of the serious academic work done by the economists at the University of Chicago. The monetary theory had been discredited and the view that 'money does not matter' had assumed great importance in the academic circles. Consequently, monetary policy had suffered emasculation being largely replaced by the fiscal policy. The revival of economists' interest in the quantity theory of money is largely due to the sincere and pioneering efforts of the neo-quantity theorists led by Professor Milton Friedman and his worthy students. Milton Friedman's writings restating the quantity theory are a part of the oral tradition of Chicago of which the quantity theory was a central and vigorous part throughout the 1930s and 1940s. At the University of Chicago, Henry Simons and Lloyd Mints taught and developed a more subtle version of the quantity theory of money in which the theory 'was connected and integrated with general price theory and became a flexible and sensitive tool for interpreting movements in aggregate economic activity and for developing relevant policy

Check Your Progress

1. State two chief merits of Baumol's analysis.
2. On what factors does the demand for money by wealth-holders depend upon?

NOTES

prescriptions'. The quantity theory of money as developed at Chicago was a theoretical approach which insisted that money does matter. It asserted that any interpretation of short-term cyclical movements which take place in the economic activity which neglected the role of monetary changes and left unexplained the question as to why people were willing to hold a particular nominal quantity of money was seriously faulty and misleading.

Agreeing with the early quantity theory of the demand for money, Friedman holds that the quantity of money demanded by the public will vary directly and proportionately with changes in the level of prices. In other words, he accepts the view that the demand for money is unit-elastic with respect to the general price level. He also agrees that real income is a major determinant of the demand for money. But he rejects the early quantity theory position that the demand for money is unit-elastic with respect to income.According to him, the income elasticity of the demand for money is greater than unity, being in the neighbourhood of 1.8, i.e., the quantity of money which people want to hold increases or decreases more than proportionately to the increase or decrease in their incomes. Friedman explains this relationship by considering money as a luxury good similar to education and recreation. He also agrees with Pigou that money is only one among the many assets that are held for the sake of services rendered by these assets to the asset-holders.

Friedman treats the demand for money as nothing more than the application of a more general theory of demand for capital assets. He derives the demand function for money by specifying those variables which determine this demand.According to Friedman, the demand for money, besides being determined by the level of prices and income is also determined by the cost of holding money. The cost of holding money depends on (i) the rate of interest which could be earned if the wealth-holders lent money instead of holding it in the barren or unproductive form of cash; and (ii) the rate of change in the general price level. When an individual decides to hold cash balances, he forgoes income which he could have earned by holding fixed interest-yield giving asset such as bond. As the market rate of interest rises, the opportunity cost of holding cash balances in the form of interest income forgone increases. In short, the price of acquiring or holding money rises. Assuming the demand for money to behave akin to the demand for other assets, less money will be demanded as the cost (price) of holding it increases. Thus, the demand for money and the rate of interest are inversely related.

When the general price level rises, the real value of the nominal cash balances falls. As the rate of price change $\Delta P/P$ increases, the opportunity cost of holding money also increases. In this situation, we may consider money as being similar to a bond on which a bond-holder earns a negative interest rate. As the rate of inflation increases, the negative rate of interest becomes large inducing an individual to demand less money. Conversely, if the prices are falling, the opportunity cost of holding money decreases as the rate of deflation increases rendering money similar to a bond on which its owner earns a positive interest rate. An increasing rate of deflation amounts to an increasing positive rate of return on a bond inducing an individual to demand more money. In short, the opportunity cost of holding money is directly related to the rate of rise (fall) in the general price level causing the demand for money to decrease (increase) as the rate of rise (fall) in the general price level increases. An increase in either of these two or both the determinants of the opportunity cost of holding money will cause the amount of money which people would want to hold to decrease. At the higher cost of holding money, people will avoid being burdened with the higher cost by economizing on their cash balances. They will strive hard to pay their bills and meet their obligations with less cash in hand and with less money in the

bank. Conversely, a fall in the rate of interest or in the rate of increase in the general price level reduces the cost of holding money. Consequently, people will be induced to hold large amount of cash balances with them. In short, the demand for money and the opportunity cost of holding money are inversely related. Milton Friedman identifies the following four determinants of the demand for money.

- Level of prices
- Level of real income and output in the economy
- Rate of interest
- Rate of change (increase or decrease) in the general price level

Changes in the first two determinants of the demand for money cause changes in the demand for money in the same direction while changes in the last two determinants cause changes in the demand for money in the opposite direction. Furthermore, while changes in the demand for money caused in response to changes in the general price level are equi-proportional, changes in the demand for money are more than equi-proportional to changes in the real income.

Holding that 'the quantity theory of money is a term avocative of a general approach rather than a label for a well-defined theory,' Milton Friedman asserts that the empirical validity of the theory is not open to question. The strength of the quantity theory of money was derived from the frequently observed high correlation between the general price level and substantial changes in the supply of money over a short period of time. Accordingly Milton Friedman states:

'There is perhaps no other empirical relation in economics that has been observed to recur so uniformly under so wide a variety of circumstances as the relation between substantial changes over short periods in the stock of money and prices; the one is invariably linked with the other and is in the same direction; this uniformity is, I suspect, of the same order as many of the uniformities that form the basis of the physical sciences. And the uniformity is more than direction. There is an extraordinary empirical stability and regularity to such magnitudes as income velocity that cannot but impress anyone who works extensively with monetary data. This very stability and regularity contributed to the downfall of the quantity theory, for it was overstated and expressed in an unduly simply form; the numerical value of velocity itself, whether income or transactions, was treated as a natural constant. Now this is not; and the failure to be so, first during and after World War I and then, to a lesser extent, after the crash of 1929, helped greatly to foster the reaction against the quantity theory.'

Who is a Quantity Theorist?

What does it mean when it is said that a writer is a quantity theorist? What is the general approach of the quantity theory of money? According to Milton Friedman, a quantity theorist believes in three things. *First,* he believes that the demand for money is highly stable–he regards the demand function for money as being more stable than the Keynesian consumption function. In other words, according to a quantity theorist, an increase in the supply of money will by raising the level of aggregate spending, increase the general price level in the economy as against the Keynesian approach that it would disappear into hoarding or speculative cash balances with little or no effect on prices. But his belief in the high stability of the demand for money does not mean that the real quantity of money demanded per unit of output or the velocity of circulation of money(V) is constant over time. In other words, the stability of the demand function for money does in no way mean that the demand function for money is invariant. Consequently, a rapid increase in (V) during a hyperinflation is no contradiction of the stability of the demand function for

money if the function includes a variable referring to expected price changes. The stability which he expects lies in the functional relationship between the quantity of money demanded and the variables which determine it. A sharp increase in (V) during hyperinflation is entirely consistent with a stable functional relationship. Furthermore, the quantity theorist limits and specifies explicitly the variables which on account of their empirical importance have to be included in the function. Apart from regarding the demand function for money as stable, the quantity theorist also regards it as playing a vital role in determining those variables, such as the level of money income or of prices, which, according to him, are of great importance for the analysis of the economy as a whole.

Second, the quantity theorist believes that the important factors affecting the supply of money are independent of those factors which affect the demand for money. This is why the quantity theorists, ever since the days of Henry Thornton, have attacked the real-bills doctrine according to which changes in the demand for money caused corresponding changes in the supply of money which cannot change without a change in the demand for money because the banks extended credit on the basis of self-liquidatiing 'real' bills.

Third, the quantity theorist believes in the 'realness' of the rate of interest being determined by the forces of thrift and productivity. He criticizes the Keynesian approach which regards the interest rate as a 'purely monetary phenomenon'. He asserts that had the rate of interest been a purely monetary phenomenon, being *solely* determined in the money market (quite independent of the forces of thrift and productivity), the monetary authority could have pushed it to any chosen level. This might be regarded as increasing the prestige of monetary policy and consequently the appeal of the quantity theory. However, if the rate of interest was entirely determined in the money market by the forces of demand for and the supply of money, manipulations of the rate of interest would have no effect on the real economic activity. There are, however, certain limitations on the ability of the monetary authority to establish any given structure of interest rates for the debts of different maturities and risks. Of particular mention is the situation of the liquidity trap in which the monetary authority would have lost effective control over the rate of interest.

In his classic article in 1956, Milton Friedman had stated the essence of the Chicago School's approach to the demand for money in the following words.

> 'To the ultimate wealth-owning units in the economy money is one kind of asset, one way of holding wealth; to the productive enterprise, money is a capital good, a source of productive services that are combined with other productive services to yield the products that the enterprise sells. Thus, the theory of the demand for money is a special topic in the theory of capital,...'

According to Friedman, the quantity theory of money is a theory of the demand for money. He considers the analysis of the demand for money on the part of the ultimate wealth-owning units in society formally identical with the analysis of the demand for a consumer durable good or capital. Money is one of the several forms of assets in which wealth may be held. Like the theory of consumer choice, the demand for money (or any other particular asset) on the part of wealth-owners depends on (i) the total wealth to be held in the different forms (budget constraint); (ii) the price of and the return on money or any other particular asset and the alternative forms of holding wealth; and (iii) tastes and preferences of the wealth-owning units. Unlike the Keynesian approach to the demand for money, in Milton Friedman's approach there is absence of

the three separate 'transactions', 'precautionary' and 'asset' demands for money. All these three demands are treated as one asset and not as three separate assets and the demand for one asset reflects the wealth, relative prices and tastes and preferences of the wealth-owners.

According to Friedman, 'the substantive differences from the analysis of the demand for a consumption service are the necessity of taking account of inter-temporal rates of substitution in (ii) and (iii) and of casting of budget restraint in terms of wealth.' Since wealth can be held in several forms, the ultimate wealth-owning unit will keep his total wealth in different forms so as to maximize 'utility'. To make this possible, he must divide his wealth in different alternative forms such that the rate at which one form of wealth can be substituted for another form of wealth equals the rate at which he is willing to substitute one form of wealth for another form of wealth.

Friedman considers five different forms in which wealth can be held, namely, money (*M*); bonds (*B*); equities (*E*); physical non-human capital goods (*G*); and human capital or wealth (*H*). The composition of one's total wealth portfolio will be determined by the returns available on cash and the rival forms of assets in which wealth can be held. The demand for money will depend on the relative rates of returns obtainable on the different competing forms of assets in which wealth can be held. Friedman takes the nominal returns from each asset, except human capital (for which there is a limited market in the modern non-slave free societies). Consequently, the rate of substitution of human capital for other forms of capital cannot be expressed in terms of the market prices. Here, however, he sees some possibility of substitution of non-human capital for human capital in an individual's total wealth holdings. At any given point of time an individual wealth-holder's asset portfolio will be divided in some way between human and non-human wealth. Although this division of the asset portfolio between human and non-human wealth can undergo a change over time but it is assumed as *given* at any given point of time and is expressed by (*w*) which is defined as the ratio of non-human wealth to human wealth or equivalently, as the ratio of income from non-human wealth to income from human wealth. Given the tastes of the wealth-owners, Milton Friedman has given the following demand function for money:

$$M = f\left(P, r_b - \frac{1}{r_b}\frac{dr_b}{dt}, r_e + \frac{1}{P}\frac{dP}{dt} - \frac{1}{r_e}\frac{dr_e}{dt}, \frac{1}{P}\frac{dP}{dt}; w; \frac{Y}{r}; \mu\right) \quad ...(6.1)$$

where

M is the total demand for money;

P is the general price level;

P is a variable which affects the 'real' yield of every asset. Since it is assumed that money gives returns solely in kind in the usual forms of convenience, security, etc. the magnitude of this return in 'real' terms depends on the quantity of goods that a money-unit can buy, i.e., on the general price level;

r_b is the market bond interest rate;

r_e is the market interest rate on equities;

$\frac{1}{P}\frac{dP}{dt}$ is the size of nominal return per $1 of physical goods which together with goods; *P* defines the 'real' return obtainable from holding $1 in the form of physical

w is the ratio of physical non-human wealth (G) to human wealth (H) or the ratio of income from non-human wealth to income from human wealth; Y/r is the total wealth since Y represents the total return on all forms of wealth including money and physical capital goods owned and held directly by the ultimate wealth-owners. The assumption is that some imputed income from the stock of money is included in Y and directly owned physical capital goods; μ stands for the utility determining variables–variables affecting the tastes and preferences of the ultimate wealth-owning units. By assuming r_b and r_e to be stable over time and stating that the rate of change of prices is required separately, Milton Friedman replaces the unwieldy variables representing the nominal return $r_b - \frac{1}{r_b}\frac{dr_b}{dt}$ on bonds and on equities $\left(r_e + \frac{1}{P}\frac{dP}{dt} - \frac{1}{r_e}\frac{dr_e}{dt}\right)$ in equation (6.1) by simply r_b and r_e. As a result of this replacement, the demand function for money expressed in equation (6.1) can be written as:

$$M = f\left(P, r_b, r_e, \frac{1}{P}\frac{dP}{dt}; w; Y; \mu\right) \quad ...(6.2)$$

According to Friedman, this demand equation must be considered independent in an essential way of the nominal units to measure the money variables. If the unit in which prices and money income are expressed is changed, the amount of money demanded should change proportionately. This means that equation (6.2) pertaining to the demand function for money must be regarded as homogeneous of degree 1 in prices (P) and the money value of wealth (Y) so that–

$$\lambda M = f\left(\lambda P, r_b, r_e, \frac{1}{P}\frac{dP}{dt}; w; \lambda Y; \mu\right) \quad ...(6.3)$$

Putting $\lambda = \frac{1}{P}$ equation (6.3) can be rewritten as–

$$\frac{M}{P} = f\left(r_b, r_e, \frac{1}{P}\frac{dP}{dt}; w; \frac{Y}{P}; \mu\right) \quad ...(6.4)$$

In this form, the equation expresses the demand for real balances as a function of the 'real' variables, independent of the nominal monetary values.

Putting $\lambda = \frac{1}{Y}$ equation (6.3) can be rewritten as–

$$\frac{M}{Y} = f\left(r_b, r_e, \frac{1}{P}\frac{dP}{dt}; w; \frac{P}{Y}; \mu\right) \quad ...(6.5)$$

or $$M = f\left(r_b, r_e, \frac{1}{P}\frac{dP}{dt}; w; \frac{P}{Y}; \mu\right) Y \quad ...(6.6)$$

According to Friedman, although equation (6.6) represents the total demand for money on the part of wealth-holders who consider money as an asset in their asset portfolios, it can also be used to represent the demand for money on the part of business firms to whom money is similar to a factor of production. There is, however, one difference between these two types of demanders. In contrast to the individual wealth-holders,

because the business firms react in the same manner as do the individuals to changes in the rate of interest and prices. Consequently, the equation will represent the total demand for money for the economy.

NOTES

Criticism

Criticizing Friedman's approach, Professor M L Burstein has stated that although a strong positive correlation has been found between the money stock and the nominal GNP, this correspondence is nowhere near those forming the 'basis of the physical sciences'. Obviously, the relationship between the stock of money and the flows highly correlated with the GNP is subject to substantial variation in both the short and long run. "Unduly simple" characterizations of the association of monetary and price (and/or income) series are apt to be a crashing failures.' Burstein further states that 'more elaborate formulations can permit better statistical fits — perhaps at the expense of introducing a large number of variables as parameters that are not controllable or even forecastable. These estimating equations might be empirically useless; they might achieve retroactive accuracy, but of what use is it to know that variations in *M* will lead shortly to well-defined variations in GNP only if innumerable other variables are kept under control or can accurately be predicted? Indeed there is a more important consequence of "complicated" formulations, at least for our immediate purposes: they imply that there will not be proportional variation in the short run between the stock of money and, say nominal GNP even when the economy is at full employment; at least not unless an extraordinary concatenation of events occurs.'

According to Milton Friedman, his theory has rich empirical contents. In his article titled 'The Demand for Money: Some Theoretical and Empirical Results', which was published in *The Journal of Political Economy* in 1959, Friedman tested the proposition that the demand for money varies directly and proportionately to the change in the general price level and directly but more than proportionately to a change in the level of income More specifically, he tests the equation $M_d = aPy_b$, where M_d is the demand for money, P is the general price level, y is the aggregate real income, i.e., total output of goods and services, a is a positive constant indicating that the demand for money changes in the same direction in which the prices and income change and b is a constant with value greater than one signifying that the demand for money changes more than proportionately with changes in the level of income. Dividing both sides of the equation by P, Milton Friedman reduces the above equation to the following equation:

$$M_d/P = ay^b$$

In the above equation, the term M_d/P is treated as a single variable and b is simply the income elasticity of demand for money which according to Milton Friedman is 1.8. For the income variable, Milton Friedman uses 'permanent income' while for the money variable he uses the sum of currency, demand deposits and commercial bank time deposits.

A major criticism of Friedman's empirical work is that the results he derived depend upon the manner in which he has defined money. His definition of money is too broad. If we define money conveniently excluding the time deposits from it, the income elasticity of the demand for money will be closer to unity rather than being as high as 1.8. Moreover, Milton Friedman's suggestion that money is a luxury is misleading.

Although Milton Friedman has emphasized the relationship between monetary stocks and aggregate wealth, still he has not found the interest rate as being empirically significant as a determinant of the demand for money. Friedman's research suggests that the relationship

between the demand for money and interest rate is weak. This weak relationship between the demand for money and interest rate results from the broad definition of money adopted by Friedman. Currency and demand deposits are those assets on which there is no explicit return as these are non-interest bearing assets. For these two assets, the rate of interest paid on the alternative assets represents the opportunity cost of holding money and demand deposits. This is not, however, true of time deposits which earn an explicit return in the form of interest paid on such deposits. As a result, when interest rates, including those on the time deposits, rise, the demand for currency and demand deposits falls whereas the demand for time deposits ordinarily increases. If the increase in the demand for time deposits is combined with the decrease in the demand for currency and demand deposits, the decrease in the total will be smaller than the decrease in the demand for currency and demand deposits alone. This is why Friedman finds the interest rate as having an effect which is too small to be statistically significant on the secular demand for money. Needless to say that this conclusion, which is not accepted by every economist, depends on the use of the permanent income in the sweepingly broad definition of money which consists of the currency in circulation plus demand deposits and time deposits. The inclusion by Friedman of the time deposits in the definition of money blunts the interest elasticity of the demand for money since it conceals the shifts between the demand and time deposits occasioned by changes in the interest rates.

6.4 INSIDE AND OUTSIDE MONEY

Money as we all know is a social construct whose main purpose is of a medium of exchange. Man is a social species that exchanges services and goods with the help of various tools. The most prominent form of modern money is what we refer to as 'fiat money'. It is a legally mandated specific unit of account, but has no inherent value as it is not physically present. In India, this legally mandated form of money is the INR.

In the modern market economy, money is mainly distributed via the competitive banking process and with banks competing for demand of loans in the modern system which is market based. Money mainly is found as bank deposits in bank accounts, which are an outcome of loans. Deposits are created by loans and it is possible for banks to create these new loans without dipping into their reserve position. It is crucial to understand that the money multiplier is false. Banks do not multiply their reserve balances. Instead, banks lend and then seek out reserves, that too only if it is essential. This is a way of distributing money in the market economy. This brings us to the concept of inside and outside money.

Check Your Progress

3. State four determinants of the demand for money identified by Friedman.
4. On what factors does the demand for money on the part of the wealth holders depend upon?
5. What was the major criticism against Friedman's theory?

The concept was first introduced by John G. Gurley and Edward S. Shaw in 1960 in their book *Money in a Theory of Finance*. In the book, they attempted to develop a theory of finance that encompasses the theory of money and a theory of financial institutions that included banking theory.

In a market economy, the money people use every day is nearly completely in the control of private banks and is referred to as 'inside money'. This is that money which is created inside the private sector. In other words, 'inside money' is that which is issued as debt by private intermediaries. It includes bank deposits which exist as a result of the loan creation process. This type of money is increasingly becoming the foremost form of money with the economy becoming more and more electronic.

When there are interest rate differentials, issuers of debt earn profits by borrowing long term and lending out short term. This money or debt that is exchanged is the inside money and is a liability to the one who issues it. In any economy, net inside money is zero

and the majority of the money that is circulating within an economy is also the inside money. To put it another way, inside money is in zero net supply within the private sector. The qualifier inside is short for (backed by debt from) inside the private sector. Since inside money issuing entities are inherently unstable, this money too is inherently unstable. These entities are profit seeking and in every situation look to maximize their own profits. Inside money is stabilized by outside money.

Outside money plays a crucial role in facilitating inside money. Outside money is money that is either of a fiat nature (unbacked) or backed by some asset that is not in zero net supply within the private sector of the economy. Outside money then is a net asset for the private sector. The qualifier outside is short for (coming from) outside the private sector. Outside money does not become a liability to those within or 'inside' the economy. For example, foreign cash, stocks or bonds, gold or other assets that are foreign currency denominated.

Since in the modern economy private economy is looked upon as being the 'inside', all money issued by the government becomes 'outside money' since it is not created in the private sector. In an economy, outside money includes bank reserves, coins and cash notes. For a private entity, outside money includes gold or assets denominated in foreign currency or otherwise backed up by foreign debt, like foreign cash, stocks or bonds.

In a market economy, the most vital form of outside money is bank reserves or deposits held at a country's reserve banks. These deposits are held for two purposes:

- to settle payments in the interbank market
- to meet reserve requirements

Inside money is private debt that also circulates as a tangible medium of exchange. Thus, an economy with inside money must perform a delicate balancing act. Firstly, it must have enough commitment or enforcement for credit to be feasible, but at the same time, credit must not function too well, because otherwise a tangible medium of exchange would become unnecessary. For example, the American economist Kocherlakota shows that a tangible medium of exchange is not important if agents can commit to future actions or if their trading histories are public. Starting from this observation, Cavalcanti and Wallacein their paper *Inside and Outside Money as Alternative Media of Exchange* consider an environment where trading histories are public for a subset of agents but private for the rest, and show that a social optimum requires note issue by those agents with public trading histories. Further, those notes are in turn used in trade among the agents whose trading histories are private. Thus, in their environment, an optimum requires inside money.

The economists Kiyotaki and Moore instead study an environment where everyone is unspecified, and stress the significance of the agents' capability to make bilateral and multilateral commitments. These include:

- The degree of (bilateral) commitment a borrower can make to an initial lender when selling a paper claim places a bound on the entire stock of private debt.
- The degree of (multilateral) commitment a borrower can make to repay any bearer determines the extent to which the borrower's debt can circulate in equilibrium.

They find that only outside money circulates in economies with very low degrees of bilateral commitment. For higher but still low degrees of bilateral commitment, outside and inside money circulate alongside in equilibrium. For yet higher degrees, only inside money circulates, and when the agents' ability to make bilateral commitments is large enough, the economy can manage without any money, inside or outside.

NOTES

6.5 GURLEY-SHAW APPROACH

It is proposed by John G. Gurley and Edward S. Shaw that NBFCs are the providers of liquidity and safety to financial assets and aid in the transfer of funds to eventual lenders to decisive borrowers. Capital formation caused by this increases fiscal development. With buying securities from eventual borrowers and through sale to eventual lenders of indirect securities, accessibility to credit is overpowered by the mediators as also is the level and structure of interest rated. Credit is generated by them which is different from that of the commercial banks. Also, new liabilities and assets are generated by them that can overpower money supply and disrupt even an effective fiscal policy.

Gurley and Shaw state that both the NBFCs' savings deposits and commercial banks' demand deposits are similar and so it is uncomplicated for NBFCs to convert savings to cash. All savings deposits, be they of NBFCs or commercial banks, are considered liquid just like demand deposits, and if held NBFCs they are called 'near money'.

Since the central bank has no control over these demand deposits they hinder the effectiveness of the fiscal policy and do not allow the central bank to prevent surplus liquidity. Further, according to Gurley and Shaw in case the central bank is implementing an anti-inflationary fiscal policy, this difficulty increases. One reason for this is that in the event of the NBFCs raising their interest rates on savings deposits, there will be a fall in the public demand for money leading to fall in market rate of interest.

It is possible for NBFCs to decrease liquidity and render the expansionary fiscal strategy unproductive. It is opined by Gurley and Shaw that there must be a comprehensive control by the central bank on NBFCs when it comes to an effectual fiscal strategy since NBFCs create greater near fund assets that disrupt overall liquidity and cause an overpowering of the complete fiscal and demand performance.

Gurley and Shaw: Financial Intermediaries

The work of Gurley and Shaw (1955, 1956, 1960), gained its inspiration and motivation from the work done by Goldsmith. Gurley and Shaw opine that process of financial development parallels real economic growth.

According to them, it is an important reciprocal relationship that exists between real and financial development. To quote: 'Development involves finance as well as goods.' (1955).

They believe that the real growth process comprises external finance where the surplus spending units will transfer their saving to deficit units as bonds via financial intermediaries. The system only has commercial banks and in such a process 'diversification demand for money' is created with increased holding of bonds by the surplus spending units, with money being extremely liquid.

According to Gurley and Shaw, there is need for some appropriate adjustment in supply of money so that rates of interest can be eased.

According to them financial intermediation hold an important place in the process of economic growth because of the relative importance of non-banks over commercial banks.

It is argued by them in their financial theory of growth that debt accumulation and growth of non-monetary intermediaries as an instrument proves to be of greater use

than short period liquidity theory in analyzing economic development. It is believed by them that the financial part must get integrated with real development.

Non-banks' liabilities which are called 'indirect financial assets' prove to be a preferred substitute to meet diversified demand for interest, security and various other services. The duo are of the opinion that:

The necessary growth in the money supply may be high or low, positive or negative, depending on the growth of income, the share of spending that is externally financed (especially by long-term securities), the growth in demand by spending units for direct relative to indirect financial assets, and on the development of financial intermediaries whose indirect debt issues are competitive with money (1955).

From the pioneering work that has been done by Gurley and Shaw (1960) it can be gleaned that 'the principal function of financial intermediaries is to purchase primary securities from ultimate borrowers and to issue indirect debt for the portfolios of ultimate lenders.'

A key role is played by financial intermediaries of providing a market mechanism that will transfer claims on real resources from the savers within the surplus sectors to the deficit sector's efficient investors. The degree of perfectness of the financial market decides the degree of optimum allocation of investment. The path taken by the transmission is of credit supply through financial intermediaries. If the financial intermediary is efficient, it can transform the process of 'internal finance' into 'external finance' making the debt-asset system work as the prime technique for mobilizing savings.

In the writings of Gurley and Shaw, financial assets are categorized as direct and indirect securities. Primary or direct securities comprise debts, besides those of financial intermediaries.

Indirect debt/securities comprise obligations of all "Financial Intermediaries" (banks too), and in as much also comprise demand deposits. When primary/direct securities get accumulated, it leads to real financial growth. With the main function non-monetary and monetary financial intermediaries buying direct debt from those who issue indirect debt to non-financial spending units, both indirect and direct debt are useful concepts.

It is believed by Gurley and Shaw that both financial instruments and financial transactions instruments affect the economic behaviour that corresponds to flow and stock relationships in two distinct ways. One is referred to 'Intermediation Effect', which is due to the properties of the financial assets and not possessed by tangible assets. Technically speaking, the indirect exchange via intermediation of financial instruments proves more efficient for want satisfaction than does direct exchange.

Intermediation of money, that mostly is an accepted way to make payments, allows the sale and purchase of commodities decomposed into two acts that are special in time. So, money usage does away with the problems faced during exchange while using the barter system.

Furthermore, a key role is played by financial instruments in production, integration and wealth ownership as also for economic activity creation. These stock implications of financial assets, called 'Asset Transmutation Effect', follow likewise, from observation that financial goods possess characteristics, which is absent in tangible goods. So, real wealth's indirect ownership via financial assets proves to be technically more efficient manner of want satisfaction compared with having direct ownership of tangible assets.

It is stated by Gurley and Shaw (1960): "The design and performance of a financial system may stimulate saving and investment in efficient uses or it may retard saving and

divert it to inefficient uses." They opine that in an economy that is rudimentary though various restraints of a financial nature exist on economic growth, "An immature financial system is in itself an obstacle to economic progress". In an economy of this type, the sole financial asset is money and the sole financial institution is the Government. All money is 'outside money'. There is demand for real money balance since there is an implicit marginal deposit rate for money which is higher than the marginal returns compared with investment and consumption. There is not neutral monetary impact and this financial system does not support fast growth in real output, because there are no different types of markets and financial assets to stimulate savings to put in investments.

Gurley and Shaw state that on money entering complex financial structure, there is the rise of a new theory of money in which an important role is played by financial intermediaries. Where there is a financial structure which has extremely well developed non-bank financial intermediaries (NBFls), 'inside money' comes up. Then, a change in the price level will lead to wealth transfer between the two private sectors, with one losing and the other one gaining equally, ensuring that the money is not neutral. To support this argument, the duo introduced the concept of 'overall liquidity' where all assets have their own different liquidity characteristic. In the theory of finance proposed by Gurley and Shaw, they show that the combination of 'inside' and 'outside' money goes to show that changes in the quantity of money will create an up or down movement in the general price level and also cause changes to happen in relative prices and invalidating money neutrality.

A distinction is made by Gurley and Shaw between real savings and financial savings. While we can transform the latter into investments we cannot do the same to the former. The duo opines that there is a distinct relationship that exists between financial development and economic growth. They show that capital stock which has been accumulated via investments in economic growth plays a crucial role and it can be stimulated by financial intermediaries who lay the key role of diverting savings into productive investments. An economy that does not have financial intermediaries, all investment is dependent on own resources, which might not be able to meet the actual requirement for investment, and more so for such investment that is crucial for the process of development. So, when there is higher level of intermediation in the financial sector, more saving will get mobilized and there will be more investment, leading to raised economic growth levels.

6.6 EXOGENOUS AND ENDOGENOUS MONEY SUPPLY AND KALDOR'S THEORY

It has been a matter of debate whether the central banks or the banking sector determines the money supply.

The terms endogenous and exogenous have been coined in the context of a model. It is possible to design a model such that 'it treats something as exogenous, meaning that the determination of it is not explained in the model but taken for granted, or it can be designed in such a way that that thing is endogenous or determined in the model'. Neither of the two is either wrong or right.

Everything in the real world is endogenous (besides the laws of nature and some set of initial conditions).

According to Keen, the debate is exogenous to the economy. Yet there are several others who are of the opinion that how the central bank will behave depends on a major

Check Your Progress

6. What is outside money?
7. State the key role played by financial intermediaries.

extent on the conditions that are prevalent in the economy at that time, making it endogenous.

It is claimed by the believers of the exogenous theory that the central bank is the one to determine the quantity of base money and after which it is the market which determines the interest rate. According to the believers of the endogenous theory, the reverse happens and the sole role played by the central bank merely is to determine the interest rate, following which the economy determines how much base money is needed which the central bank then supplies.

According to Keen, in general, the monetary policy is not exogenous.

Central banks have no alternative but to accept this course of events, their only option being to vary the short-term rate of interest at which they supply liquidity to the banking system on demand.

Commercial banks are now in a position to supply whatever volume of credit to the economy that their borrowers demand.

In brief we can say:

- The supply of money and credit is determined by the demand for money and credit. There is no independent supply curve as in standard micro theory.
- All the state can do is affect the price of credit (the interest rate).

It is important to understand that for a monopolist there will not be supply curve. Since the price can be controlled by the monopolist, his quantity is determined by more than just the demand curve. He controls the quantity and the price is determined by the demand curve based upon that quantity.

So it is possible to say that it is the central bank that is the one to set the price, following which the demand curve and that price together determine the quantity. This is not different from Central bank setting the quantity and the demand curve in hand with this quantity, determining the price.

It is claimed by both Keen and Moore that it is not money supply that will "cause" price increases but it is increase in price that leads to increase in the money supply. There is, then, no possible way to determine a price level or change in it. So, everything cannot be endogenous in a model and every exogenous variable that needs to be made endogenous will need another equation to identify it.

According to the Business Dictionary:

> 'Theory that money exists just as it's needed by the economy, because bank system reserves are increased or decreased to accommodate for demand is endogenous money. Under the endogenous money theory, if banks can borrow money at the Federal Reserve discount rate and still lend money profitably, then the money available for banks to borrow will become available as necessary to support the level of consumer lending individual banks require.'

In the field of economics, endogenous money refers to that money supply of an economy whose quantity of money present is governed endogenously meaning that it depends on the interactions of other economic variables. This money supply is not determined exogenously (autonomously) through some authority, like central bank.

According to the theoretical argument it is the need of the real economy that leads to the existence of money and there is a contraction/expansion of the reserves of the banking system due to the needed for accommodating demand for loan at the interest rates that are prevailing.

There are three key claims that form the basis for this theory:

- **Loans create deposits:** For the complete banking system, when a non-bank borrower drawing down a bank loan, a new deposit is created while repayment of a bank loan destroys deposits. Though quantity of bank loans and deposits within an economy might not be the same, it is considered that a deposit is also a logical concomitant of a loan – there is no need for a bank to increase deposits before making out a loan. Even if a bank is capital-constrained, in majority of the nations across the world, no bank if it is solvent will be funding-constrained or reserve-constrained: funding and reserves can be obtained by it from either the central bank or the interbank market. Banks rationally follow all profitable lending opportunities which can be identified by them till the level that is consistent with their level of capital. At such times, they look at funding issues and reserve requirements as matters that will be addressed at a later point – more so, at an aggregate level. So, in an economy, the amount of broad money that exists is determined endogenously: this implies that, the quantity of deposits that the non-bank sector holds will 'flex' down or move up based on the non-banks' aggregate preferences. It is important to note that it is stated by the theory that if the deposits of the non-bank sector are augmented due to some policy-driven exogenous shock (for example, quantitative easing), it is to be expected of the sector that it find ways to 'shed' majority or all the excess deposit balances with making payments to banks (such as purchases of securities or bank loan repayments).
- **Central banks implement policy primarily through controlling short-term interest rates:** When the above happens, the money supply adjusts to the changes in demand for credits and reserves that have been created by the change in rate of interest. There is a rightward shift in the supply curve when new substitutes for money are issued by financial intermediaries, in reaction to opportunities for profit making during the cycle. While the monetaryauthority might decline from accommodating changes of this nature, banks may continue to increase reserves for loan demand through their own initiatives.
- **Given available credit, investment precedes and 'forces' the saving necessary to finance it:** It is investment that goes to determine saving and not savings determining investments. In the current times, there cannot be any saving without income distribution, and no income can be distributed without entrepreneurs getting into debt. So, for investment plans there is no need for considering savings. When there is a positive balance of payments, excess reserves are not maintained by banks, and they either lend their excess to other banks which are facing a deficit on their balance of payments, or purchase government debt for profit. It is precisely the credit-creation process dictated by profit seeking motives which causes instability. Central banks and governments, in the role of lender of last resort and of regulator of financial practices, help to monitor the quality and level of debt, and are capable of preventing a downward profit trend, which is the key variable for debt validation and for asset prices.

Keynes in the *General Theory of Employment, Interest, and Money* (1936) treated the money supply as exogenous, but in *A Treatise on Money* (1930) and his article 'Alternative Theories of the Rate of Interest' (Keynes 1937), he had recognised the concept (Arestis 1992: 180). In the latter work, Keynes had stressed the finance motive as a basis of endogenous money (Keynes 1937).

King (2002: 161) contends that Richard Kahn and Joan Robinson were the first to develop the Post Keynesian theory of endogenous money, even if in a somewhat limited and incomplete manner.

Nicholas Kaldor continued to develop the theory in his polemics against monetarism (King 2002: 166–167), and particularly in his classic book *The Scourge of Monetarism* (Oxford and New York, 1982).

The fierce debate with monetarists actually inspired Post Keynesians to clarify and formulate a more rigorous endogenous money theory (King 2002: 172).

6.7 SUMMARY

In this unit, you have learnt that,

- James Tobin has developed a sophisticated analysis of portfolio selection under uncertainty which meets the shortcomings present in Keynes' analysis of the speculative demand for money.
- According to Keynes, an individual wealth-holder discretely decides to hold either bonds or money and not both.
- According to Tobin, rational behaviour induces an individual to hold a wealth portfolio which is comprised of both bonds and money.
- According to the Keynesian theory, the individual holds either all the money or all consols based on the expected future rate of interest—the normal rate.
- A change in the expected rate of interest on consols will affect the portfolio selection decision of the individual wealth-holder by causing a change in the proportion in which he will hold money and consols.
- In the normal course, as the rate of interest increases, the proportion in which the individual wealth-holder will keep his total wealth in the form of consols will increase.
- The quantity theory of money as developed at Chicago was a theoretical approach which insisted that money does matter.
- The quality theory asserted that any interpretation of short-term cyclical movements which take place in the economic activity which neglected the role of monetary changes and left unexplained the question as to why people were willing to hold a particular nominal quantity of money was seriously faulty and misleading.
- Friedman treats the demand for money as nothing more than the application of a more general theory of demand for capital assets. He derives the demand function for money by specifying those variables which determine this demand.
- Milton Friedman asserts that the empirical validity of the theory is not open to question. The strength of the quantity theory of money was derived from the frequently observed high correlation between the general price level and substantial changes in the supply of money over a short period of time.
- A major criticism of Friedman's empirical work is that the results he derived depend upon the manner in which he has defined money. His definition of money is too broad.

Check Your Progress

8. How has the business dictionary defined endogenous money?
9. According to Keen and Moore, what leads to an increase in money supply?

- Although Milton Friedman has emphasized the relationship between monetary stocks and aggregate wealth, still he has not found the interest rate as being empirically significant as a determinant of the demand for money.
- In modern monetary systems, money is a social construct whose main purpose is of a medium of exchange. Man is a social species that exchanges services and goods with the help of various tools.
- In modern times money's most prominent form is what we refer to as 'fiat money' which is a legally mandated specific unit of account but has no inherent value as it is not physically present. In India, this legally mandated form of money is the INR.
- Outside money plays a crucial role in facilitating inside money. Outside money is that form of money which is in the economy as net profit amounts and does not become a liability to those within or 'inside' the economy.
- 'Outside money' is found in its most important form in deposits or reserves in banks, on reserve at Federal Reserve banks with a two-fold purpose: 1) payment settlement in interbank market; 2) meeting reserve requirements.
- It is proposed by John G. Gurley and Edward S. Shaw that NBFCs are the providers of liquidity and safety to financial assets and aid in the transfer of funds to eventual lenders to decisive borrowers.
- Gurley and Shaw state that both the NBFCs' savings deposits and commercial banks' demand deposits are similar and so it is uncomplicated for NBFCs to convert savings to cash.
- All savings deposits, be they of NBFCs or commercial banks, are considered liquid just like demand deposits, and if held NBFCs they are called 'near money'.
- A key role is played by financial intermediaries of providing a market mechanism that will transfer claims on real resources from the savers within the surplus sectors to the deficit sector's efficient investors.
- The terms endogenous and exogenous have been coined in the context of a model. It is possible to design a model such that 'it treats something as exogenous, meaning that the determination of it is not explained in the model but taken for granted, or it can be designed in such a way that that thing is endogenous or determined in the model'.
- It is claimed by the believers of the exogenous theory that the central bank is the one to determine the quantity of base money and after which it is the market which determines the interest rate.
- In the field of economics, endogenous money refers to that money supply of an economy whose quantity of money present is governed endogenously meaning that it depends on the interactions of other economic variables.

6.8 KEY TERMS

- **Quantity theory**: In monetary economics, the quantity theory of money (QTM) states that money supply has a direct, proportional relationship with the price level.
- **Inside money:** Money held in a form such as bank deposits which is an asset to the holder but also represents a liability for someone else.

- **Outside money:** Money held in a form such as gold which is an asset for the holder and does not represent a corresponding liability for someone else.
- **Endogenous money:** In economics, endogenous money is the money supply of an economy in which the quantity of money in existence is determined endogenously—that is, as a result of the interactions of other economic variables, rather than exogenously (autonomously) by an authority such as a central bank.

6.9 ANSWERS TO 'CHECK YOUR PROGRESS'

1. Baumol's analysis has two chief merits. First, it shows that the transactions demand for cash balances is interest-elastic—a conclusion which is insignificant in Keynes' approach. Second, it shows that the transactions demand for cash balances increases less than proportionately with the increase in income due to the economies of scale experienced in the use of money.
2. The demand for money by the wealth-holders depends on the following factors:
   - Total real wealth
   - Fraction or percentage of total wealth kept in the non-human form of wealth
   - Opportunity cost of holding money
   - Other variables which determine the utility of money.
3. Milton Friedman identified the following four determinants of the demand for money.
   - Level of prices
   - Level of real income and output in the economy
   - Rate of interest
   - Rate of change (increase or decrease) in the general price level.
4. Like the theory of consumer choice, the demand for money (or any other particular asset) on the part of wealth-owners depends on (i) the total wealth to be held in the different forms (budget constraint); (ii) the price of and the return on money or any other particular asset and the alternative forms of holding wealth; and (iii) tastes and preferences of the wealth-owning units.
5. A major criticism of Friedman's empirical work is that the results he derived depend upon the manner in which he has defined money. His definition of money is too broad.
6. Outside money is that form of money which is in the economy as net profit amounts and does not become a liability to those within or 'inside' the economy. For example, foreign cash, stocks or bonds, gold or other assets that are foreign currency denominated.
7. The key role played by financial intermediaries is providing a market mechanism that will transfer claims on real resources from the savers within the surplus sectors to the deficit sector's efficient investors. The degree of perfectness of the financial market decides the degree of optimum allocation of investment.
8. According to the Business Dictionary: 'Theory that money exists just as it's needed by the economy, because bank system reserves are increased or decreased to accommodate for demand is endogenous money.'
9. It is claimed by both Keen and Moore that it is not money supply that will "cause" price increases but it is increase in price that leads to increase in the money

supply.

6.10 QUESTIONS AND EXERCISES

Short-Answer Questions

1. State in brief how Keynes' liquidity preference function is derived.
2. Differentiate between inside and outside money in brief.
3. State the three key claims to the theory of endogenous money.
4. State the essential requirements of Baumol's transaction demand for money.

Long-Answer Questions

1. Discuss Tobin's portfolio selection theory and the risk associated with portfolio selection.
2. Analyse Friedman's restatement of quantity theory of money.
3. 'According to Gurley and Shaw, there is need for some appropriate adjustment in supply of money so that rates of interest can be eased.' Discuss.
4. Discuss the theory of endogenous and exogenous money supply.

6.11 FURTHER READING

Mankiw, N Gregory. 2010. *Macroeconomics.* Worth Publishers, New York.

Shapiro, Edward. 1996. *Macroeconomic Analysis.* Galgotia Publication, New Delhi.

Branson, WH. 1979. *Macroeconomic Theory and Policy*. Universal Book Stall, New Delhi.

Jha, R. 1999. *Contemporary Macroeconomic Theory and Policy*. New Age International, New Delhi.

Gupta, SB. 1995. *Monetary Economics: Instruments and Policy*. S Chand & Co. New Delhi.

UNIT 7 INFLATION, MONETARISM AND NEW-CLASSICAL MACROECONOMICS

Structure

7.0 INTRODUCTION

Macroeconomic theory has its origins in the study of business cycles and monetary theory. In general, early theorists believed monetary factors could not have an impact on real factors such as real output. John Maynard Keynes attacked some of these 'classical' theories and produced a general theory that described the whole economy in terms of aggregates rather than individual, microeconomic parts.

Attempting to explain unemployment and recessions, he noticed the tendency for people and businesses to hoard cash and avoid investment during a recession. He argued that this invalidated the assumptions of classical economists who thought that markets always clear, leaving no surplus of goods and no willing labour left idle. 'New Classical' economists are more likely to accept ideas of rigidities in prices and wages. In this unit, the theories of inflation, monetarism and new-classical macroeconomics have been discussed in detail.

NOTES

7.1 UNIT OBJECTIVES

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

- Analyse Keynesian theory of inflation and differentiate between demand-pull inflation and cost-push inflation
- Discuss the concept of inflationary gap
- Analyse Phillips curve and the causes of high unemployment
- Discuss monetarist theory of income determination and inflation
- Analyse Friedman-Phelp critique of Phillips curve and money supply rule
- Comment on the functioning of real business cycle and economic policy
- Differentiate between adaptive and rational expectations

7.2 KEYNESIAN THEORY OF INFLATION

Inflation, which today confronts the economic policy-makers throughout the world in the form of a most dominant economic problem, is not a new phenomenon because from the earliest days of recorded history, mankind has been puzzled and discomfited by rising prices. Throughout the ancient period, the Mediterranean civilizations frequently experienced higher prices in terms of metallic currency due to the discovery of new mines and improved methods of mining gold. The early Mediterranean history also contributed to inflation with the release of hoards of the metallic money accumulated through pillage. The capture of the Persian gold by Alexander the Great was perhaps the largest single inflationary act of the ancient period. The acquisition of substantial gold and silver by the conquering Roman emperors also had an inflationary impact in the kingdom. The frequent debasement of coins in ancient China, Greece and Rome had resulted in inflation. The metallic inflation which followed the discovery of America constituted one of the most important instances of inflation in history.

Meaning of Inflation

There has been a proliferation of definitions of inflation. Many of these definitions, however, embody the description of the processes by which the underlying causes of inflation–demand–pull, cost-push, etc.,–reveal themselves. Consequently, the fundamental connection between an increase in the total money supply and the rise in prices in the economy is obscured.

According to the public understanding, inflation means a condition which produces a rising trend in the general price level in the economy. Inflation may, however, be present in the economy if the sustained price rise, which would have otherwise occurred, is prevented from occurring by imposing the price and physical controls in the economy. Such a situation is called 'suppressed inflation'. Inflation is not amenable to any one definition. According to the *Chambers' Twentieth Century Dictionary*, inflation is an 'undue increase in quantity of money in proportion to buying power, as on an excessive issue of fiduciary money.' Gardner Ackley has defined inflation 'as a persistent and appreciable rise in the general level or average of prices.'According to this definition, a sporadic price spurt or an imperceptible rise in prices will not be inflation. Elaborating further, Ackley has stated: 'We define inflation as *rising prices*, not as "high" prices. In some sense, then inflation is a disequilibrium state; it must be analysed dynamically

rather than with the tools of statics.'According to Crowther, 'inflation is a state in which the value of money is falling, i.e., prices are rising.'According to Pigou, inflation exists 'when money income is expanding relatively to the output of work done by the productive agents for which it is the payment.' In general, inflation may, therefore, be defined as a sustained rise in the general price level brought about by high rates of expansion in the aggregate money supply although in the contemporary discussions on inflation it is defined as a sustained rise in the general price level, howsoever generated. All these definitions have a common feature of stressing the point that inflation is a process of rising prices and not a state of high prices, showing a state of disequilibrium between the aggregate supply and the aggregate demand at the existing or current prices necessitating a rise in the general price level in the economy.

According to the market laws of supply and demand, an increase in prices *per se* should not be inflationary. Indeed, if anything, it should be anti-inflationary because consequent upon a given price rise, the total amount of goods and services demanded should decrease while the amount supplied should increase. This must be so unless the aggregate demand and aggregate supply functions are perfectly inelastic. Inflation emerges in the economy on account of the increase in the money incomes of certain sections of the community without any corresponding increase in their productivity, giving rise to an increase in the aggregate demand for goods and services which cannot be met at the current prices by the total available supply of goods and services in the economy.

A sustained rise in prices of about 2 per cent per year may be called 'creeping' inflation to distinguish it from 'galloping' (or hyper) inflation, which occurs when monthly price rise of about 500 or 600 per cent or more occurs, and from 'trotting' inflation in which the price rise occurs at the intermediate rates. The basic characteristic of creeping inflation is that the annual price rise is almost imperceptible so as to be lost sight of by the casual observers. Any complacency in controlling creeping inflation is likely to prove disastrous for the economic and political stability of the economy because creeping inflation must eventually accelerate through the trotting stage until it is galloping at an even faster rate culminating in the complete collapse of the currency and the consequent disruption of the political and economic life of the community. Such hyperinflation in which due to the astronomical rise in the prices, money becomes almost worthless and causes unbelievable hardships to people was witnessed in Germany in 1923, in Hungary in 1947 and in China in 1949. In hyperinflation, it becomes senseless to hold money for the precautionary or speculative purposes since the real capital losses on cash balances become prohibitive.

For a milder sustained price rise, economists have used the terms walking inflation and running inflation. In walking inflation, a sustained price rise may be of about 8-10 per cent yearly. For the higher two-digit sustained annual price rise the term 'running inflation' has been used. For example, in India during 1973 and 1974 the inflationary situation could be characterized as being one of running inflation because prices rose by 26 per cent in 1973 and by 19 per cent in 1974. Thus, it is the rate of price rise that justifies our calling a particular situation as being one of creeping, walking, running or of hyperinflation. Sometimes there may be a good deal of overlapping between these terms depending upon the rate of price rise adopted for purposes of classification. The difference between creeping, walking, running and hyperinflation has been explained diagrammatically in Figure 7.1.

Figure 7.1 shows that in hyperinflation the price increase is so rapid that the slope of the curve ($dP/P/dt$) which shows the annual rate of price rise is almost infinite showing that there is almost no limit to price rise. In hyperinflation, when due to the almost

astronomical price increase, money becomes worthless and there is a virtual flight from currency, people revert back to barter or adopt some other country's currency whose value is relatively stable to express the deferred payment contracts. It happened in Germany when the deferred payments contracts were expressed in US dollar instead of in German mark which had become worthless. The climax of hyperinflation is reached when the flight from currency becomes so fantastically high that the velocity of money in circulation approaches infinity. In the case of running inflation, the increase in prices is relatively mild although it is quite high compared with that in walking and creeping inflation. The price rise is least in the creeping inflation as the slope of the price rise curve is gentle.

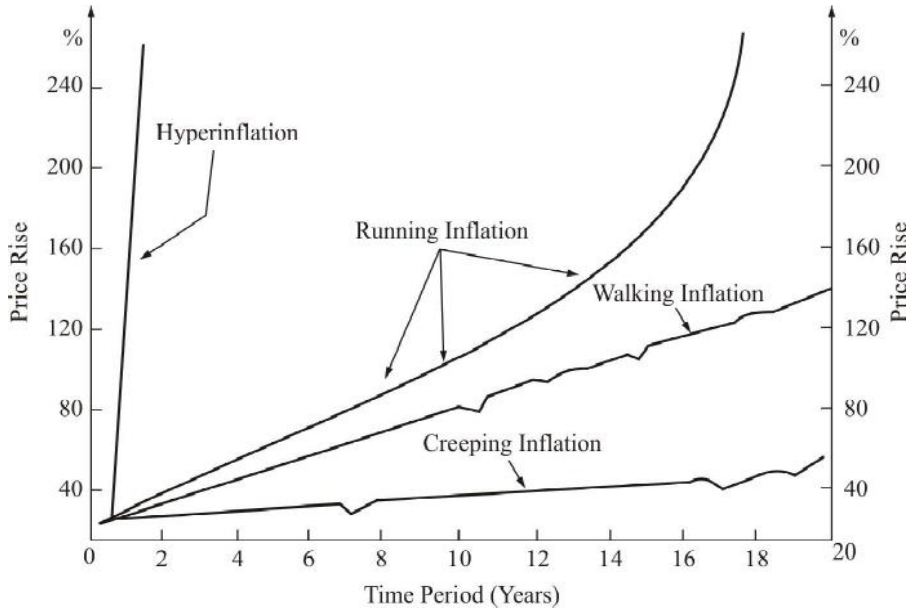


Fig. 7.1 *Different Kinds of Inflation*

Keynes has mentioned in his *A Tract on Monetary Reform* that during the post-war hyperinflation in Russia, Moscow grocers, upon selling their cheese, ran to the central market in order to replenish their stocks rather than hold steadily the depreciating money. He also mentions a Viennese witticism that 'a prudent man at a cafe ordering a bock of beer should order a second bock at the same time, even at the expense of drinking it tepid lest the price should rise in the meanwhile.' Similarly, James Hicks has written that in Chile, which experienced inflationary pressure for a long time, it had become customary for the patrons of Chilean sporting houses to place double orders, even at the expense of fatigue. There are, however, counter-examples of strong but non-accelerating inflations in Brazil, Argentina and certain other Latin American economies in the 1950s and 1960s which show that price acceleration and collapse are by no means inevitable.

We have associated inflation with a situation of sustained rising prices. It is not, however, the only meaning which has been given to inflation in the past. For the quantity theorists, for instance, inflation was synonymous with an increase in the quantity of money which, on the assumption of given velocity and transactions, caused a rise in the general price level. This means that a rise in prices was the effect of inflation and not inflation itself. Similarly, according to the Keynesians, inflation is an excess of the aggregate demand over the aggregate supply at full employment and a given price level.

In a closed economy with low inventories and absence of institutional barriers to price rise, the excess of the aggregate demand over aggregate supply would certainly cause a rise in prices. It is, however, the excess demand which would be inflation and the rise in prices would merely be the symptom showing its existence. However, excess demand need not necessarily lead to a rise in prices; in an open economy it can be satisfied by increasing the imports with the result that so long as it is possible to finance a balance-of-trade deficit, the price level will not rise even in the face of an excess demand phenomenon in the economy. Furthermore, a sustained price rise can be occasioned by factors other than excess demand at full employment and it would be wrong to refuse to call such a price rise inflation on the ground that this has not been caused by excess demand. For these reasons, it is proper to have a definition which refers directly to the phenomenon of sustained rise in prices rather than to any possible causes of this phenomenon which are undoubtedly quite significant in suggesting possible remedies against inflation. It must, however, be stressed that it is not every rise in prices but only a sustained rise in prices which deserves the label of inflation. A sporadic increase in prices in the economy cannot be called inflation.

7.2.1 Demand-pull or Excess Demand Inflation

According to the classical economic analysis, the general price level depends directly and proportionately on the supply of money. According to the classicists, inflation occurs when the quantity of money increases and comes to a halt when the quantity of money becomes stable. The rate of inflation will depend on the rate at which the new money is created, i.e., on $dM/M/dt$. Thus, if $dM/M/dt$ is 5 per cent per year, the general price level will also rise at 5 per cent per year. This is the quantity theorist's explanation of the inflationary process.

In its naive form, the quantity theory of money is faulty because it does not explain the process by which an increase in the quantity of money (M) causes an increase in the aggregate money spending which, assuming a given or constant output, raises the level of prices in the economy. It was the well-known Swedish economist Knut Wicksell who removed this defect by arguing that the creation of credit money by the banking system which flowed in the economy in the form of loans to businessmen to finance investment in excess of the current rate of saving represented a net excess of the aggregate demand (at the given general price level) over the constant aggregate supply (since the economy was operating at full employment) bidding up the prices of goods and factors of production in the economy. Consequent upon the creation of bank credit, the market rate of interest becomes lower than the 'natural rate' of interest—rate at which total real saving equals the total real investment—as a result of which investment (demand for investible funds) exceeds saving (supply of investible funds). Inflation would halt only if the banks ceased to expand the money supply raising the market rate of interest sufficiently so as to become equal to the natural rate of interest. According to Wicksell, so long as the banking system continued to finance investment that was not sustainable by the availability of savings by creating credit, inflation would continue to haunt the economy.

Keynes' analysis of the excess demand inflation assumes aggregate demand to exceed the aggregate supply at the full employment level. Starting with the situation of full employment equilibrium, if investment increases then the aggregate demand for goods and services will exceed their aggregate supply at full employment level assuming a given level of prices, i.e., $C_F + I_F > Y_F$ at constant prices. This is a situation of

disequilibrium which can be corrected only either through an increase in the prices or through an increase in the aggregate real output or through increase in both the prices and output. But since under our assumption, the economy is already operating at full employment level, no increase in aggregate output is possible. Consequently, prices rise sufficiently so as to bring about equilibrium between the aggregate demand and aggregate supply. Since consumption is a function of real income, the excess demand will persist because the rise in prices raises people's money income as a result of which the real income remains unchanged.

Keynes severed the close relationship between the quantity of money and the level of aggregate demand by showing that even with constant money supply, some inflation may be experienced. With the total quantity of money held constant, an increase in prices occasioned by increased aggregate demand would raise the transactions demand for the cash balances. The increased transactions demand for money, with the total money supply held constant ($\overline{M}_s$), can be satisfied by releasing the sufficient cash from wealth-holders' speculative cash balances. This is, however, possible only if the rate of interest increases, i.e., if the security prices fall sufficiently to induce the wealth-holders to make the required quantity of money available. The rise in the rate of interest would release some cash from the speculative cash balances to feed the additional transactions demand for cash. As the security prices fall, the wealth-holders would substitute securities for cash. This would, however, check the future fall in the security prices (rise in the rate of interest). Consequently, the rise in the rate of interest will not be sufficient to release sufficient cash from the speculative cash balances required to meet the excess transactions demand. As an indirect effect, the rise in the rate of interest will choke off the extra investment demand which will contain the inflationary pressure. However, since the rate of interest fails to rise high enough, only a part of the extra investment demand will be choked off which only moderates the inflationary pressure but does not eliminate it. Only if it is assumed that the speculative demand for cash balances is perfectly interest-inelastic will the rate of interest rise sufficiently to choke off the entire excess aggregate demand by reducing the investment and possibly to some extent also the consumption spending provided saving was interest elastic. In this case, Wicksell's analysis will hold.

Thus there is a significant difference between Wicksell's and Keynes' analysis of inflation. While according to Wicksell, any increase in the money supply (unless accompanied by a proportionate increase in the total productivity of money) is always inflationary, prices would rise even if the increase in M_s resulted from central bank's open market purchases of securities. Keynes, however, held that an increase in the M_s brought about through open market operations of the central bank may not raise prices if the economy started from the position of less than full employment.

Despite differences in their analytical reasoning, both Wicksell and Keynes explained inflation as an excess of the aggregate demand over the full employment output in the economy. Keynes merely removed the rigid relationship between the aggregate demand and the quantity of money and also considered less–than–full employment situation in which the increased aggregate demand was more likely to raise the level of economy's aggregate output than the prices.

Inflationary Gap

The excess demand inflationary analysis has been described in terms of the 'inflationary gap' measured as the excess of the aggregate demand $C + I + G$ over the full employment

aggregate supply Y_F shown in Figure 7.2. In other words, the situation of $C + I + G > Y_F$ shows the presence of inflationary gap in the economy. On the other hand, if $C + I + G < Y_F$ there is present a deflationary gap in the economy. Both the situations of inflationary gap and deflationary gap are situations of disequilibrium. The concept of the inflationary gap may be defined as the positive difference between the actual level of aggregate demand which exists in the economy at the full employment level of income and the amount of aggregate demand which is required to attain full employment. Since the amount of the aggregate demand required to attain full employment should be equal to the amount of the aggregate supply at full employment, we may also say that inflationary gap is the excess of the actual aggregate effective demand in the economy over the aggregate supply at the full employment level. If we erect a vertical line Y_FK at the full employment income $0Y_F$, we see that for full employment to exist in the economy the aggregate demand function $C + I + G$ in the diagram should cut the 45°-line at its intersection with the full employment line Y_FK. Since the aggregate demand schedule cuts the 45°-line $Y \equiv C + I + G$ at point E which is located to the right and above B which is the point of intersection of the 45°-line with the full employment line Y_FK, there is inflationary gap of the AB magnitude present in the economy which is measured as a distance between the aggregate demand schedule $C + I + G$ and the 45°-line $Y \equiv C + I + G$ at full employment level of income $0Y_F$.

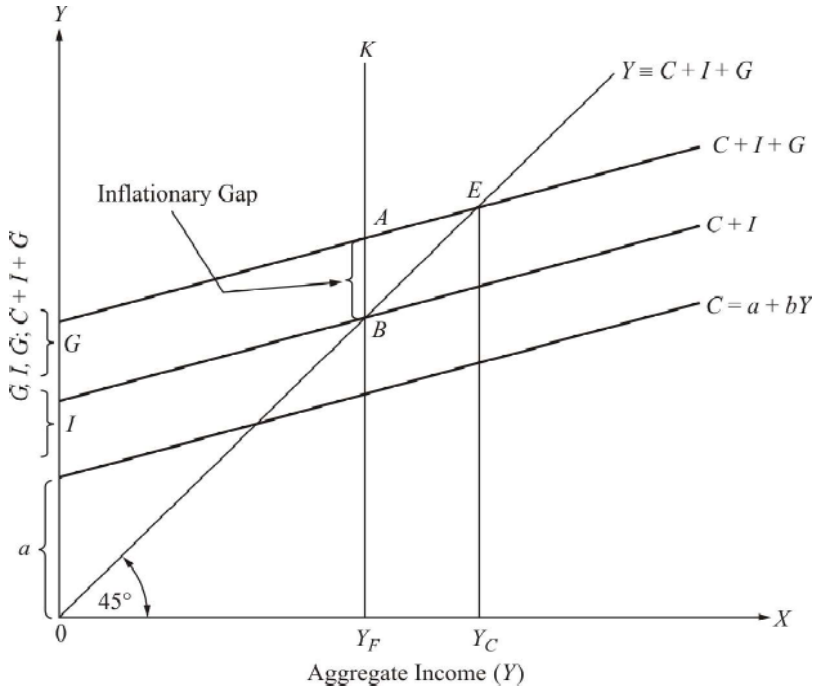


Fig. 7.2 *Aggregate Consumption as a Function of Aggregate Income*

In Figure 7.2, the aggregate consumption has been shown as a function of the aggregate income Y and the linear consumption function is of the form $C = a + bY$. The aggregate investment I and government spending G are both autonomously determined so that the aggregate demand function $C + I + G$ has been drawn parallel to the consumption function, the distance between the two being $I + G$. The line $C + I + G$ shows the aggregate amount of spending corresponding to each different level of the aggregate income. If the aggregate real output could increase without limit, the aggregate real income would increase to $0Y_C$ where the aggregate real output (supply) and the aggregate real expenditure (demand) are in equilibrium. If, however, the aggregate real

NOTES

output or supply cannot increase beyond $0Y_F$, the real income cannot also exceed the $0Y_F$. At $0Y_F$ the aggregate demand ($C + I + G$) exceeds the aggregate output by *AB*, i.e., $0Y_F + AB = C + I + G$. Thus, *AB* is the inflationary gap which raises the general price level. The inflationary gap disappears only when the aggregate money income increases from $0Y_F$ to $0Y_C$ raising the general price level. The increase in the money income is wholly due to the increase in prices since the aggregate real output is constant at full employment output $0Y_F$. The *AB* inflationary gap is eliminated and the inflationary process is halted when the new equilibrium aggregate money income is established at $0Y_C$. The increase in the aggregate money income from $0Y_F$ to $0Y_C$ is wholly accounted for by the increase in prices with no increase having taken place in the aggregate real output which remains fixed at the full employment output $0Y_F$. Knowledge regarding the magnitude of the inflationary gap and the value of the multiplier enables us to find out the equilibrium income. We can find out the new equilibrium aggregate income (which in the present case is $0Y_C$) in the following manner:

$$0Y_C = 0Y_F + \text{Inflationary Gap} \times \text{Multiplier}$$

Notwithstanding that demand-pull inflation may easily occur as a result of the strong investment boom resulting from massive innovations or the opening of new territory or any other such cause, the more frequently occurring demand-pull inflations have been due to the massive government outlays particularly those associated with the war or post-war reconstructions. Massive government outlays in the developing countries in building up 'social overhead' also generate strong inflationary pressures in the economy. While it is largely true that the demand-pull explanation of inflation shows that the rate of price rise accelerates in periods of low unemployment (the size of unemployment serving as a rough indicator of the strength of demand), i.e., when the demand forces are strong, it fails to explain the contemporary co-existence of high unemployment and high inflation situation of stagflation.

7.2.2 Cost-Push Inflation

Before the 1950s, the phenomenon of inflation was largely analysed in terms of the excess demand explained either in the classical quantity theory version or in terms of the Keynesian theory. The supply or cost analysis of inflation was revived in the 1950s. The cost-push inflation analysis, also known as the 'new inflation' theory, has been explained in its crude and sophisticated versions. In its crude version, the theory asserts that inflation occurs due to the increase in the costs or supply prices of goods caused by an increase in the cost of inputs. According to this explanation, rapidly rising money wages unaccompanied by corresponding increase in the labour productivity in certain key sectors of the economy become reflected in the higher prices in these same sectors, particularly as demand recovers. In this way, the purchasing power of wages becomes eroded causing the organized labour to seek redress in the form of further wage increase through their collective bargaining strength. The most common political expression of this view is based on the plea that the monopolistic trade unions cause inflation by pushing up wages through their excessive power in the field of collective bargaining. The sophisticated version of the cost-push theory differs from its crude counterpart in two main aspects. *First*, it deals with the question of demand which the cost-push inflationary pressures themselves generate through the needed increase in the money supply. *Second,* it strives to seriously analyse the motivations of the 'pushers' instead of resorting to name-calling as the crude version does.

There is nothing new in the cost-push inflation explanation. As Martin Bronfenbrenner and F D Holzman have stated, the 'cost inflation has been the layman's instinctive explanation of general price increase since the dawn of the monetary system. We know of no inflationary movement that has not been blamed by some people on 'profiteers', 'speculators', 'hoarders', or workers and peasants 'living beyond their station'.

Stated in terms of the aggregate demand and aggregate supply functions, cost-push inflation emerges in the economy, in the absence of excess demand, due to the pressure of various factors which shift the aggregate supply function which is subject to diminishing returns, upward. Figure 7.3 illustrates cost-push inflation. In the figure, $0Q_F$ is the full employment aggregate output which is determined at the point of intersection of the aggregate demand curve *DD* and the aggregate supply curve S_1S_1.

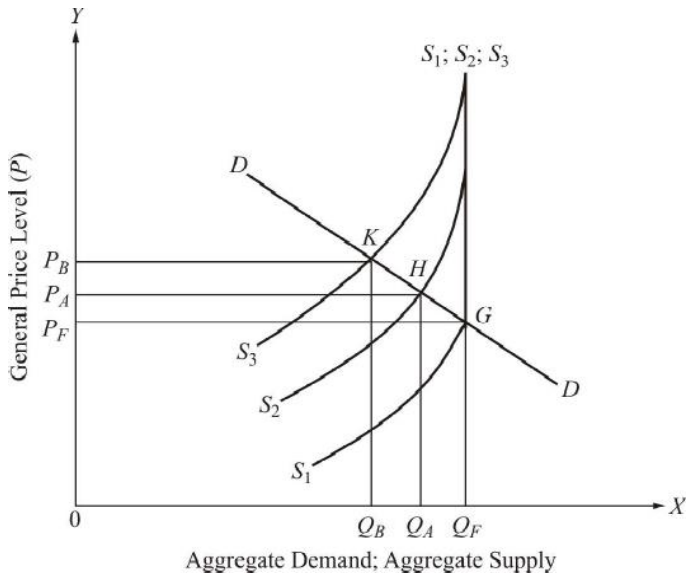


Fig. 7.3 *Cost-push Inflation*

Consequent upon the upward shift in the aggregate supply curve from the position of S_1S_1 to that of S_2S_2, the equilibrium aggregate output falls from $0Q_F$ to $0Q_A$ while the general price level rises from $0P_F$ to $0P_A$. A further upward shift in the aggregate supply function from S_2S_2 to S_3S_3 reduces the aggregate equilibrium output from $0Q_A$ to $0Q_B$ while the price level rises further from $0P_A$ $0P_B$. The increase in the general price level is the pure cost-push inflation. The two main factors responsible for the upward shift in the aggregate supply function are (i) the higher money wages secured for their members by the labour unions without any corresponding increase in their productivity, and (ii) the higher prices charged from consumers by the monopolistic and oligopolistic producers. The upward shift in the aggregate supply function and the resulting rise in the general price level due to the first factor is designated as 'wage-push' inflation to distinguish it from 'profit-push' inflation resulting from the operation of the second factor in the economy.

Generally speaking, cost-push inflation in the economy occurs as a result of the combination of both the wage-push and the profit-push factors. According to those who hold that prices are pushed up by rising costs rather than by the demand-pull forces, some control in the form of prices and incomes policy is necessary to bring the spiral of rising prices to a halt. Both the demand-pull and the cost-push explanations of inflation are closely linked with the now widely-held view that the problem of inflation is more sociological than economic in nature. According to this view, there appears an underlying

bias towards inflation in the economies of leading industrial nations of the world. In recent years, workers' and consumers' expectations have risen considerably far beyond the scope of existing productive capacity to meet their demand. A mounting stream of government transfer payments and increased government involvement have largely contributed to increased consumption and to sap both the individual initiative and the ability and incentive of the private enterprise sector to undertake the needed commitments to modernize and expand the economy's productive capacity.

7.3 PHILLIPS CURVE

The phenomenon of demand-pull inflation is frequently controlled by resorting to the instruments of monetary and fiscal policies. According to the classical quantity theory approach, the demand inflation can be controlled by resorting to an appropriate monetary policy so as to halt the expansion of the money supply. According to the monetarist explanation of inflation, money does matter significantly in the occurrence of inflation in the economy. According to the Keynesians, however, the monetary policy alone will not be able to check inflation. Consequently, it is suggested that we apply the restrictive fiscal policy instruments of curtailing the unproductive expenditure and widening and deepening the tax structure in the economy.

The anti-inflationary role of the monetary and fiscal policies is not, however, so clearly appropriate in the matter of controlling the supply or cost-push inflation. Restrictive monetary and fiscal measures have their immediate impact on the aggregate demand in the form of restricting it. But supply inflation is not the result of the aggregate demand rising in excess of the economy's full employment output.An important difference between the demand and supply inflations is that unlike the former, in the case of supply inflation the price rise takes place well before the full employment output has been hit.

Since cost-push inflation is largely caused by the rising cost, supply inflation can be controlled by maintaining wage-rate stability and by preventing those wage increases which are not related to the increase in labour productivity. A restrictive wage policy may check a wage-push inflation provided it reduces the aggregate demand and output sufficiently to create enough unemployment to prevent wage increases in excess of the increase in the labour productivity. It is, however, quite likely that the extent of unemployment which may be necessary to avoid wage-push inflation may be higher than what is considered socially and economically acceptable. In other words, the community may be forced to pay a certain price (and it might be fairly high) in the form of considerable social distress and slow rate of economic growth for purchasing price stability. If a sustained high percentage of 6 or 7 per cent of unemployment is necessary to achieve price stability, it might seem to the society the lesser of the two evils to accept moderate inflation with a low unemployment percentage of 2 or 3 per cent compared with a high unemployment percentage of 6 or 7 with zero rate of inflation in the economy.

Since wage costs constitute the backbone of the price structure, in recent years economists interested in the study of supply analysis have focussed their attention on the relationship between the rate of wage increase and the rate of labour unemployment in the economy. This analysis runs in terms of the 'Phillips curve', christened after A W Phillips whose pioneering work in this field was published in 1968. A Phillips curve showing the relationship between the percentage of wage increase and the percentage of unemployment of economy's labour force can be derived from an economy's data over a period of years by plotting on the graph for each year the percentage of money

Check Your Progress

1. State the basic characteristic of creeping inflation.
2. When is the climax of hyperinflation reached?
3. Define inflationary gap.

wage increase ($\Delta W/W$) against the percentage of unemployed labour force. Such a curve will be negatively sloping as shown in Figure 7.4 in which the *AA* curve slopes downwards to the right showing that the rate of increase in money wages is inversely related to the unemployment rate. The Phillips curve also suggests that a sufficiently high percentage of unemployment is needed for non-inflationary price stability. From this the conclusion emerges that the wage-push price inflation can be eliminated if the community is prepared to accept a high rate of unemployment.

It is not, however, true that every rate of increase in the money wages is inflationary; it is only such percentage increase in the money wages which does not represent a corresponding increase in the labour productivity to match the increase in the money wages which is inflationary. By drawing a horizontal line *WW* in Figure 7.4, we can represent the percentage increase in the labour productivity. Consequently, this much (0*W*) percentage increase in the money wages will be non-inflationary.

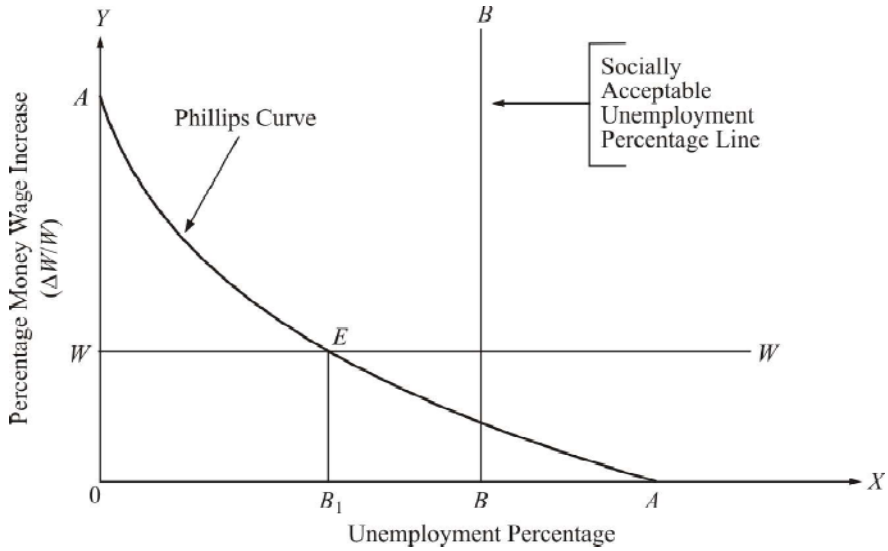


Fig. 7.4 *Percentage Increase in Labour Productivity*

According to Phillips, for the United Kingdom, a rate of 5.5 per cent unemployment was needed for wage stability and a rate 2.5 per cent unemployment was needed to hold prices stable. This would mean that wages would rise by the same percentage as the increase in labour productivity estimated by Phillips to be around 2 per cent for the United Kingdom. According to Paul ASamuelson and Robert M Solow, who have plotted a similar curve for the United States, unemployment rate has to be 5.5 per cent for maintaining the price stability, assuming a 5.5 per cent annual increase in the labour productivity. Obviously, Samuelson's and Solow's results are more distressing than those obtained by Phillips.

If *B* per cent of unemployment is taken as socially acceptable, and if *WW* line intersects the *AA* Phillips curve at point *E* which is located on the Phillips curve to the left of the socially acceptable unemployment percentage line *BB*, it indicates that the wage-push inflation can be eliminated through the monetary and fiscal policies without subjecting the economy to a socially unacceptable high rate of unemployment. Price stability can be maintained by inflicting only B_1 per cent unemployment on the economy. This unemployment percentage will be acceptable to the community because it is lower than the *B* per cent socially acceptable unemployment.

7.4 MONETARISM AND ITS BASIC ELEMENTS

The fiscal policy dominated approach which gathered strength under the protection of the Keynesian umbrella dominated the scene for about two decades. It was only in the mid-1950s that a reaction against the fiscalist approach set in, with the economists challenging the very existence of the Keynesian liquidity trap. Led by Professor Milton Friedman of the Chicago School of economists, a growingly large number of economists asserted that 'money does matter' in the economy. In fact, the 'hard core' within the group strongly asserted that 'money alone matters'. These economists who were labelled as the monetarists, pleaded the case for money. The monetarists have staged a grand comeback for the monetary policy and have advocated the view that the periods of hyperinflation and deep recession are exclusively the consequences of disturbances originating in the monetary sector of the economy resulting from expansion and contraction in the money supply.

The centerpiece of monetarism relates to the overwhelming influence of money on the long-run economic activity. According to the monetarists, proper growth rate of money supply is crucial for the stable growth of output and prices. The monetarists believe that money and prices are directly correlated. The central thesis of the monetarists is that money matters and consequently monetary policy has a crucial role to play in any programme of economic stabilization. The extreme monetarists have asserted that money alone matters. Their assertion is based on the joint validity of the particular assumptions regarding the demand for any supply of money.

According to these economists, the demand function for money and the supply function of money are perfectly interest-inelastic. This means that in terms of the *IS-LM* curves analysis, the *LM* curve is a vertical straight line. Consequently, any fiscal policy action denoted by the upward to the right or downward to the left shift in the *IS* curve will not at all affect the level of aggregate income and output in the economy. In short, fiscal policy will not matter at all and money alone matters because any rightward or leftward shift in the *LM* function caused by the expansionary or contractionary monetary policy action will change the level of aggregate income and output in the economy. Consequently, full employment in the economy can be achieved through the thrust of an expansionary monetary policy.

Origin and Contents of Monetarism

The origin of monetarism may be traced in the writings of a long line of the University of Chicago's distinguished students and teachers of money who are best represented by Professor Milton Friedman. The Chicago economists are the vehement supporters of the free market system. They are wedded to the belief that almost all economic problems have their solution in the free and unimpeded operation of the free market forces of supply and demand.

Milton Friedman's book *Capitalism and Freedom* published in 1962 admirably defends the free market enterprise system. It shows that the system is inevitable for the existence and enjoyment of political freedom. Friedman has cited many examples to show that government interference with the working of the market forces has mostly failed as evidence to assert his point that government should not intervene in the economy any more than is absolutely necessary.

However, even the most uncompromising supporter of the market feels that there is some justification for some governmental intervention. According to Milton Friedman, regulation of money supply is one such area where government control is called for. He is, however, of the firm belief that government intervention must be regulated by certain specified rules so that the scope for discretion is minimal. Rules, and not discretionary power, should govern the actions of the regulator.According to Henry Simons, the provision for specific rules to achieve specific targets ensures certainty implied by rules, avoids arbitrary actions by the regulator and removes speculation regarding the future actions of the regulator. Henry Simons was followed in this respect by Milton Friedman who observed that government must provide a monetary framework for a competitive system since the competitive system cannot provide one for itself. However, this monetary framework should operate under the 'rules of law' and not under the discretionary authority of the administration.

Apart from believing in the efficiency of the market system and the specific provision of the 'rules of law' for the government to operate the monetary framework, the Chicago economists adhere to the quantity theory of money as the basic framework for monetary analysis. Although the quantity theory of money suffered from disrepute and was generally ignored during the 30s, it continued to be used by the Chicago School of economists.

Furthermore, the supply of money is determined exogenously by the monetary authority quite independently of the demand for money. According to the monetarists, the central bank controls the money supply and the factors, at least the important ones, affecting the supply of money, do not affect the demand for money.

The monetarists also believe in the 'realness' of the rate of interest, being determined by the forces of thrift and productivity. They criticize the Keynesian approach which treats the rate of interest as a 'purely monetary phenomenon'. It is asserted that had the interest rate been a purely monetary phenomenon, the monetary authority could have pushed it to any chosen level. This might be regarded as enhancing the prestige of the monetary policy and consequently the appeal of the quantity theory of money. However, if it was to be believed that the interest rate was entirely determined in the money market, then manipulation of the interest rate would have no effect on the aggregate real economic activity.

Milton Friedman and Monetarism

The basic monetarist theory, as developed by Milton Friedman and his associates, asserts that money plays a crucial role in determining the level of aggregate economic activity. Monetarism is based on a revised version of the quantity theory of money and stresses the need for controlling the level or the rate of change of money supply for achieving economic stability. According to Friedman, the rate of growth of money supply should be kept stable and to achieve this, the monetary authority should use the open market operations. Speaking about the American economy, Milton Friedman asserts that the Federal Reserve System should ensure that the money supply grows at some predetermined constant annual rate which should be consistent with the long-run growth of the economy at stable prices. Only if this could be achieved, all the other monetary policy tools could be abandoned, asserts Milton Friedman. In support of his argument for a constant annual rate of growth of the money supply, Friedman has produced a good deal of historical and empirical evidence.

At the centre of monetarism is the key relationship between changes in the money supply and changes in the level of money income. The line of causation runs from changes in the money stock to changes in the money income so that money changes cause income changes. The mechanism starts with the demand for money stated in terms of the demand for real cash balances (*M/P*). The total real cash balances held by the public are measured as the total money supply *M* divided by general price level *P*. The total amount of real cash balances which the public desires to hold is a function of the real income and other variables. Starting from the equilibrium position in which the public's desired holdings of real cash balances are equal to its actual holdings of real cash balances, if the supply of money is increased by the monetary authority, the people will have a larger quantity of purchasing power with them than they actually want to hold. In other words, they will have surfeit of money supply with them of which they would like to be relieved. There is only one rational way of getting rid of the unwanted surplus cash balances and bringing their actual real cash balances equal to the desired level, i.e., by spending. However, as some people spend and reduce their excess cash holdings, the cash holdings of others in the economy increase. Consequently, they hold added cash balances. They then spend their excess cash balances and the process continues.

The excess cash balances are eliminated in one of the two ways. *First*, if as a result of increased spending, the real income and output increase, the demand for money increases and the excess cash balances are now absorbed by the higher demand for money at the higher real income level. *Second*, if only the prices increase due to additional spending, the real value of the total money supply falls. The increase in prices continues until the real value of the nominal money supply is restored to its original level. In the process of spending both *M* and *P* increase, leaving the aggregate real cash balances (*M*/*P*) unchanged.

A fall in the total money supply has the opposite effect. A decrease in the money supply reduces the real cash balances in the hands of the public below the desired level. To bring their actual real cash balance holdings equal to their desired cash holdings, people must either reduce their spending or sell their assets. The fall in spending causes either the prices or output or both to fall until the level of desired real cash balances is attained corresponding to lower prices or/and lower output.Afall in the prices accomplishes the objective by increasing the real money supply to the desired level while a decline in the output reduces the demand for the real cash balances until the reduced demand for the real cash balances is equal to the existing supply of the real cash balances.

While comparing his approach with that of Keynes, Friedman has argued that the Keynesian analysis first studies the effect of a given change in the supply of money on the rate of interest and then the effect of change in the interest rate on the investment spending and finally on the income. By channelling the effect of changes in the money supply via the interest rate mechanism, monetary policy affects only those forms of spending which are interest-elastic and not all forms of spending are interest-elastic. In contrast to this approach, Friedman has stated that his approach stresses a more direct relationship between changes in the money supply and income since the increase in the money supply causes the economic units to redistribute the added money throughout their balances sheets. The new money supply may be used to purchase investment goods, durable consumer goods, real estate or financial assets. Given the fact of flexible prices and interest rates, changes in these two follow from the distribution of the new purchasing power.

Friedman has overstated the difference between his analysis and that of the Keynesians. While the simple Keynesian analysis focusses attention on the speculative demand for money in terms of money versus bonds and relates the interest rate to business capital investment, the post-Keynesian analysis is broad enough, including as it does a relationship between money and all other financial assets and a relationship between interest rate and all forms of business and consumer spending. The real difference between the two approaches resides in the degree in which the change in the money supply affects the aggregate money spending directly as opposed to affecting the aggregate spending indirectly through the interest rate mechanism.

7.4.1 Demand Function for Money

In the Keynesian analysis, the demand for money is a function of the level of income and the rate of interest. According to Milton Friedman, the demand for money is a function of the following six factors.

1. The rate of return on the bonds–the higher the rate of return on bonds, the smaller is the demand for money.
2. The rate of return on equities (stock)–the higher the rate of return on stock, the lower is the demand for money.
3. The rate of change in prices–if prices are rising at a rapid rate, people will economize on their holdings of money in order to avoid a fall in the real purchasing power of their money holdings. Consequently, the demand for money holding is negatively related to the rate of change in prices.
4. The ratio of non-human wealth to human wealth–human capital is embodied in the individual in the form of investment made in education, skills, etc., which enables an individual to produce future returns. Non-human capital represents the ownership of income-yielding physical assets–ownership of land, house, machine, etc. A change in the proportions of the total wealth held in these two forms will change the demand for money.
5. Real income Y/P also affects the demand for the real cash balances M/P.
6. Tastes and preferences of the wealth-holders and economic and non-economic conditions also affect the demand for money by influencing the desire of people to hold money.

According to Milton Friedman, uncertainty and geographic mobility also act as possible causes for an increase in the public's liquidity preference or demand for money. In a subsequently developed simplified version of the demand for money, Milton Friedman makes the demand for money a function of the real per capita permanent income. In this study the demand function for money has been stated as:

$$M/NP_p = r\,(Y_p/NP_p)^{d}$$

where M = nominal stock of money N = population
Y_p = permanent income P_p = permanent prices

This latter version of the demand for money is not as different from the first version as it appears to be because the concept of permanent income is broad enough to include the several variables used in the earlier version of the demand for money.

Permanent income is affected by the yield on securities and human and non-human wealth holdings. According to Milton Friedman, on the basis of empirical evidence, the demand for nominal cash balances is represented by the following equation:

$$M = (0.00323)\,(y_p/N)_{1.81}\,(NP_p)$$

where all the variables have the same definition as stated above, except y_p which stands for the real permanent income. In the above form, the equation expresses that (assuming population and permanent prices to be constant) a 1 per cent increase in the real permanent income increases the demand for the nominal cash balances by 1.81 per cent. In other words, Milton Friedman's equation for the demand for money indicates that the real permanent income elasticity of demand for the nominal cash balances is 1.81.

Stability of the Velocity of Money

The monetarists also strongly hold that the velocity of money is highly stable. In fact, it is far more stable than the Keynesian expenditure multipliers. It is argued by the monetarists that the relationship between money and income is a more stable empirical relationship than the relationship between the autonomous expenditure investment and government spending and income. To prove his point, Friedman has offered a variety of empirical evidence. Of particular note is the study done by Friedman and David Meiselman for the Commission on Money and Credit and the subsequently published series of critical papers which this study evoked. In their study, the two authors have compared the simple monetary models and simple models of autonomous expenditures on the basis of each model's ability to predict consumption. The two basic models which have been employed in the study are (1) the Keynesian model–

$$C = a + KA$$

where C = consumption;
a = constant;
K = Keynesian multiplier; and
A = autonomous expenditure $[I + G + (X - M)]$

and (2) the quantity theory model

$$C = a + VM$$

where C = consumption;
a = constant;
V = money multiplier; and
M = supply of money

World War period, the empirical results indicate a better statistical relationship between the changes in the supply of money and changes in consumption than between changes in the autonomous expenditure and changes in consumption.

The Friedman-Meiselman study evoked great response from critics who in a series of ably written papers questioned the theory and the empirical evidence presented by Friedman and Meiselman. DePrano and Mayer in their study have concluded that better forecasts of consumption were obtained by using the alternative specifications of the autonomous expenditure than were obtained either by the Keynesian or by the

Barring the Second

monetarist models. Their finding was that both money and autonomous expenditure were important in the determination of income.

NOTES

Role of Monetary Policy and its Failure in the Past

Milton Friedman does not regard monetary policy as a mere tool of short-run economic stabilization. According to him, monetary policy 'can prevent money itself from being a major source of economic disturbance.' Furthermore, monetary policy would 'provide a stable background for the economy.' In Friedman's view, the stable rate of growth of the money supply would create stability of expectations with respect to wages and prices. Monetary policy could 'also contribute to offsetting major disturbances in the economic system arising from other sources.' In this objective, Milton Friedman refers to the long-run problems rather than to the short-run problems of the trade cycle.

Dwelling on the failure of monetary policy in the past, the monetarists have blamed the Federal Reserve System, which has been more concerned with interest rates and the conditions of the credit markets than with the control of money supply, for the dismal record of performance of monetary policy. The monetarists' major historical work is represented by the work titled *A Monetary History of the United States, 1867–1960* by Milton Friedman and Anna Jacobson Schwartz published in 1963. In this important work bearing on the monetary analysis, the authors have traced the development of money supply in the United States; have analysed changes in the determinants and components of the money supply; and have studied the monetary policy in terms of the effect on the monetary policy both prior to and after the establishment of the Federal Reserve System.

Milton Friedman and Anna Jacobson Schwartz have studied changes in the supply of money in terms of the following three 'proximate determinants of the money stock.'

1. The quantity of high-powered money (H) – This includes the supply of currency in the hands of the public plus the cash reserve deposits at the federal reserve banks. These items are called the *high-powered* because they serve as the basis for the expansion of money supply in the economy.
2. The ratio of bank deposits to bank reserves (D/R) –This ratio measures the extent to which the commercial banking system utilizes the available reserve assets.
3. The public's ratio of bank deposits to currency (D/C) – This ratio reveals that lower is the proportion of the money supply held as currency, higher the level of bank reserves and greater is the possible expansion of the money supply. Milton Friedman and Anna Jacobson Schwartz have used these ratios to determine the sources of changes in the money supply in the United States during the period 1867–1960.

The Friedman-Schwartz analysis of the great depression of the 1930s runs in the monetary terms and the Federal Reserve System has been mainly blamed for its failure to increase the money supply during the mid-and-late 1920s. Friedman and Schwartz have attributed the length and depth of the depression to the failure of the Federal Reserve System which instead of undertaking a massive expansion in the money supply contracted it by as much as 33 per cent during 1929-1933. The fall in the money supply allowed during 1929-1933 resulted from the sharp declines in the deposit-reserves ratio (D/R) and the deposit-currency ratio (D/C). Although the Federal Reserve System increased the money supply beginning in the late 1930s, the increase was too little and was made too late to be of any effective use.

In summary, the monetarists' interpretation of the economic history is based on the analysis of changes in the rate of growth of the money supply. The monetarists analyse the economic problems and offer their solutions in terms of the behaviour of the money supply.

Monetary Policy Lags

A major point in the monetarists' analysis concerns the lags in the effect of monetary policy. The problem of lags in the operation of monetary policy arises from the recognition of the fact that monetary policy actions of the central bank do not have a quick and immediate effect on the economic conditions. The effects of monetary policy do not work themselves out instantaneously. On the contrary, the monetary policy actions are subject to lags in time. The impact of what the central bank does today is not felt by the economy today itself but makes its impact on the economy with a time lag at some point in the future with the result that an appropriate monetary policy action may become inappropriate in the context of the time at which its impact is felt in the economy.

According to Milton Friedman, the monetary policy lag is both lengthy and of variable duration. In his view, today's policy measures meant to deal with today's problems will have their effect felt at some point of time in the future when the problem for which these were designed may not exist; instead, a new problem requiring the new policy measures may exist. For example, the anti-inflationary monetary policy measures taken today to curb the present-day inflationary pressures in the economy may have their contractionary effect felt a year or two later when inflation may no more be there and deflation mayinstead be the problem requiring a quick and effective solution. The monetary policy may consequently become a destabilizing policy instead of proving a stabilizing policy.

Regarding the duration of the monetary policy lag, Milton Friedman has stated: 'On the average of eighteen non-war cycles since 1870, peaks in the rate of change of the stock of money precede reference peaks by sixteen months and troughs in the rate of change of the stock of money precede reference troughs by twelve months.'According to Milton Friedman, the lag is long enough to justify his pleading for the use of rules regarding the growth rate of economy's money supply rather than the use of discretion of the monetary authority in the matter. The claimed superiority of the monetary rule depends, in part, on the length of lag in the effects of monetary policy. If the lag is as long as has been suggested by Friedman, a stable rate of growth of money supply policy might be an improvement. If monetary actions have their effects one to three years later in the future, the rule may be necessary. However, Friedman's lag is longer than the lags found by others and it seems that the use of the money supply growth rate rule would not represent a general improvement over the discretionary monetary policy.

Recent Developments

Monetarism has re-emerged as an alternative for the Keynesian prescriptions for economic growth in the early 1970s. This has thrown open the issue about what should be the objectives of a monetary policy–should monetary policy concentrate on the price stability or on ensuring economic growth? The issue has assumed added significance in the context of autonomy for the central banks.

It must be stressed that both price stability and growth have to be simultaneously pursued. Consequently, the central bank must maintain a balance between ensuring price stability and at the same time ensuring economic growth by making adequate credit available. While the case of price stability as an objective of monetary policy rests

on the fact that volatility in prices creates uncertainty in the decision making, without ensuring growth the benefits of price stability are elusive. In fact, a rigid price stability which ensures zero inflation may not be a good thing even in the long run. In short, a central bank should strive to achieve price stability while ensuring sustainable economic growth through the judicious use of its monetary policy.

Among the world's central banks, the greatest devotee to monetarism for long has been the German Bundesbank. The Bundesbank had fixed a target for the M_3 growth of between 4 to 7 per cent for 1998 allowing a gap of 75 per cent between the minimum and maximum growth in the money supply.

In India, the Reserve Bank of India's monetary policy stance has been monetarist in as much as it continues to place its faith in the efficacy of monetary aggregates, the favoured number being M_3 – the broad money supply for which a target of 15 to 15.5 per cent growth had been fixed for 1998-99. During 1999–2000, the broad money supply (M_3) increased by 17.1 per cent against 19.7 per cent in 1998–99. The reserve money increased by 11.9 per cent in 1999–2000 against 12.2 per cent in 1998–99 while M_1 increased by 14.7 per cent against 12 per cent in 1998–99. The targets for the annual growth of M_3 during the earlier years in the 1990s have hovered around the 14–16 per cent band. Notwithstanding the fact that in its recent statements the RBI has tried to clarify that it is not totally monetarist although it believes in the primacy of the monetarist principles; it, however, continues to remain more a monetarist than a Keynesian. The economic orthodoxy in India is still Friedmanite believing in the stability of the monetary demand function although the velocity has declined from nearly 4.2 to 1.9 during the past two and half decades.

Evaluation

The ascendancy of monetarism during the past three and a half decades has condemned much of the developed world and some of the developing countries to low economic growth and high unemployment, albeit with low inflation. It is a high social price which these countries have paid in terms of the denial of growth and perpetration of unemployment in order to achieve some degree of price stability. Eminent Keynesians like Nicholas Kaldor and James Tobin have found pure monetarist theory a flawed piece of reasoning. Citing Milton Friedman's public admission of the failure of the United Kingdom's experiment with monetarism and attributing it to the gross incompetence of the Bank of England, Nicholas Kaldor has asked: 'does the Quantity Theory hold good only in countries with competent central banks?'

Milton Friedman's rediscovery of monetarism and the Keynesian perspective present the two differing views of economic reality. Both these approaches were formulated and developed in the context of world's developed affluent economies. What is imperative is the need to see how and to what extent, monetary policy in the context of the developing countries needs to be recalibrated, given these countries' economic compulsions of promoting the economic growth and reducing unemployment.

Check Your Progress

4. State the central thesis of monetarists.
5. What is the basic theme of monetarism?
6. How are the excess cash balances eliminated?

7.5 MONETARIST THEORY OF INCOME DETERMINATION

According to Keynes, rate of interest is determined by equilibrium between demand for money and supply of money (i.e., through money market equilibrium).The effect of money supply on rate of interest and the effect of rate of interest on aggregate demand provides a mechanism through which changes in money supply affect the goods market.

This determines level of economic activity in the economy – i.e., level of output and employment.

Monetary policy has a profound effect on the rate of interest. If the rate of interest is reduced as a result of an increase in money supply, the rate of investment will rise and the increase in investment will lead to increase in income and employment via the multiplier.

Hence, when in times of recession, money supply in the economy is increased it will cause investment to increase. As a result, there will be an increase in aggregate expenditure (i.e., aggregate demand) which will lead to the increase in real national income (aggregate output) and employment will increase.

The change in money supply leads to the increase real income output and employment is shown in the following figure.

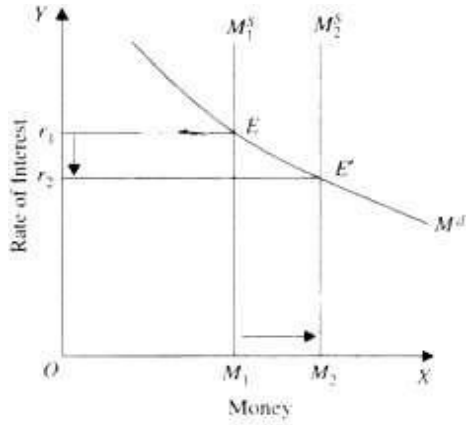


Fig. 7.5 *Money Supply and the Rate of Interest*

The first link in the transmission mechanism is the effect of expansion in money supply on the rate of interest. This depends on how far demand for money holdings is sensitive or elastic to the changes in rate of interest. The expansion in money supply (M_S) causes the rate of interest to fall.

The influence of change in rate of interest on the rate of investment, which is determined by the elasticity of investment with respect to rate of interest is the second step in the transmission mechanism. The fall in rate of interest leads to the increase in investment in the economy.

The next step in the process is the effect of increase in investment on aggregate demand and therefore on national income (aggregate output) and employment in the economy. The effect of investment on income, output and employment is determined by the size of multiplier.

The expansion in money supply which leads to the increase in aggregate demand will affect both the real national income (i.e., GNP) and the price level jointly.

7.5.1 Monetarist Theory of Inflation

Modern monetarists, led by Milton Friedman, revived and modified the classical monetary theory of inflation. The modern monetarists hold that the general level of price rises only due to increase in money supply but not proportionately. According to Milton Friedman, 'Inflation is always and everywhere a monetary phenomenon and can be produced only by a more rapid increase in the quantity of money than in output.'

More importantly, while classical economists, especially Irving Fisher, considered increase in the stock of money as the sole cause of inflation and price rise being proportional to increase in money supply, modern monetarists do not agree with proportionality of increase in price level. Recall here Fisher's quantity theory equation,

$$MV = PT$$

and $$P = MV/T$$

where MV = money supply = currency × velocity of money; P = general price level; and T = total number of transactions (sale and purchase).

Clearly, the volume of transactions (T) remaining constant, P increases in proportion to increase in MV, the total supply of money. This proposition is not acceptable to modern monetarists. In Friedman's own words, 'In its most rigid and unqualified form the quantity theory asserts strict proportionality between the quantity of what is regarded as money and the level of prices. Hardly any one has held the theory in that form.' That is, modern monetarists do not agree that there is proportional relationship between the supply of money and the price level.

7.6 FRIEDMAN-PHELP CRITIQUE OF PHILLIPS CURVE AND MONEY SUPPLY RULE

Generally, in the case of economies that are industrialized, inflation is seen to be pro-cyclical. In other words, when there is high economic activity, inflation is high. When we measure economic activity in terms of unemployment rate, the statistical relationship that is formed is referred to as Phillips curve. At times, the Phillips curve is looked upon as being a basis for monetary policymakers. This means that choice can be made between high employment and low inflation or low employment and high inflation. In this interpretation of the Phillips curve the assumption is made that there is a structural relationship between inflation and unemployment and this relationship will not crumble when policymakers try exploiting the perceived tradeoff.

During the 1970s, several economies were faced with high inflation and this caused discredit to the structural interpretation of the Phillips curve. Nevertheless, the 1980s and initial years of 1990s saw a period of low inflation and this lead the economists to restart work on a structural interpretation of the Phillips curve.

Phillips curve is the representation of the relationship that exists between unemployment rate and rate of inflation. The study done by A. W. H. Phillips for United Kingdom of the inflation of wage and of unemployment during 1861 and 1957 is looked upon as being a landmark for the development of macroeconomics.

A consistent inverse relationship was found by Phillips between the two. This showed that in times of high unemployment there was a slow increase in wages and in times of low unemployment there was a rapid rise in wages.

It was estimated by Phillips that the unemployment rate being low lead to tight labour market and as unemployment rate became lower the labour market became tighter. So, this lead to the need for firms to be faster at raising wages to attract and retain labour which was not scarce. When the rates of employment were high, pressure decreased. In the Phillips's "curve" there is the representation of the average relationship between wage behaviour and unemployment over the business cycle. It depicts at what rate of wage inflation will a particular level of unemployment persisted for a period of time.

Check Your Progress

7. How is the rate of interest determined according to Keynes?
8. How is the rate of interest affected by the monetary policy?

NOTES

In case of developed economies, economists quickly formulated Phillips curves. Majority of them formed a relationship between general price inflation, and not wage inflation, with unemployment. It is true that the prices charged by a company have a close link with the wages paid by the company.

In the figure given below (Fig. 7.6), we find a Phillips curve that has been fitted to data of United States between 1961 and 1969. Such a close fit as seen in the figure between the estimated curve and the data was an encouragement to several economists to look upon the Phillips curve as being some kind of 'menu' of policy options in the process of policy making.

To take an example, for 6 per cent unemployment rate, the economy might be stimulated by the government to bring down unemployment to 5 percent. It can be seen in Figure 7.6 that the cost, as higher inflation, would be a little more than ½ percentage point. In case initially the government saw lower unemployment rates, the costs would be much higher. That is, if unemployment was brought down from 5% to 4%, there would be greater than twice as big a rise in inflation rate (approximately 1.25 percentage points).

When the popularity of the Phillips curve as being an apt guide to policy was at its peak, Edmond Phelps and Milton Friedman challenged its theoretical foundation, independently. The argument that they put forth was that workers and employers who were rational and well-informed would only care for real wages (inflation-adjusted purchasing power of money wages). They opined that there would be an adjustment in real wages such that supply of labour became equal to the demand for labour, and this would move the rate of unemployment to a level uniquely associated with that real wage ("natural rate" of unemployment).

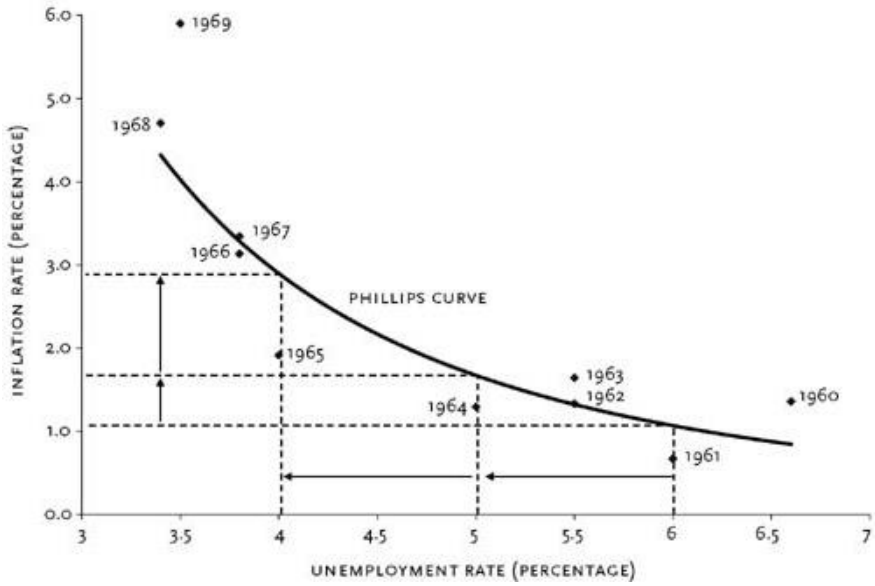


Fig 7.6 *The Phillips Curve, 1961–1969*

Source: *Bureau of Labor Statistics*

It was argued by Phelp as well as Friedman that it was not possible for a government to go trading higher inflation for lower unemployment. If it is considered that unemployment is at the natural rate. The real wage is constant: workers expect that

there will be a given rate of price inflation, push for an increase of their wages at same rate as inflation so that their purchasing power faces no erosion. Let us consider that an expansionary fiscal or monetary policy is used by the government to try to bring unemployment to a level lower than its natural rate. The rise in demand that results will encourage firms to raise their prices faster than anticipated by the workers. As revenues are higher, there is willingness on the parts of firms for employing more workers at the old wage rates and they are also ready to increase the rate a little. For a little while, workers will face money illusion, meaning that they find their money wages to have gone up and they readily provide more labour. So, there is a fall in the rate of unemployment. The workers do not immediately realize that there has been a decline in their purchasing power since there has been a more rapid increase in prices than was expected by the workers. Nevertheless, with time as workers begin anticipating higher rates of price inflation, they will supply less labour and insist on wage increases that will match up with inflation. There is a restoration of the real wage to its old level, and the unemployment rate returns to the natural rate. The wage and price inflation caused due to expansionary policies will keep going on at the new, higher rates.

The analysis by both Friedman and Phelps shows the difference in the 'long-run' and 'short-run' Phillips curves. Till the average rate of inflation stays reasonably constant (like it was during the 1960s), unemployment and inflation will remain related inversely. Yet, if there is change in the average rate of inflation, like in the case when policymakers continue to attempt to push unemployment below the natural rate, post a period of adjustment, there will be a return of unemployment back to its natural rate. That is, when there has been time for the adjustment of workers' expectations of price inflation, unemployment's natural rate becomes compatible with any rate of inflation. In figure 7.6, it is possible to show the long-run Phillips curve in the form of a vertical line above the natural rate.

Then, the original curve will be applicable just to periods that were brief and transitional and it would shift in the face of any persistent change in the average rate of inflation. It is possible to combine the short-run and long-run relations in one 'expectations-augmented' Phillips curve. The faster do workers' expectations of price inflation adapt to changes in the actual rate of inflation, the faster will unemployment get back to its natural rate, and the lower the success that the government shall attain in lowering unemployment with fiscal and monetary policies.

The basic point that was made by Friedman and Phelps was confirmed by the 1970s in contrast with the initial Phillips curve, when there was an increase in the average rate of inflation from approximately 2.5% in the 1960s to approximately 7% in the 1970s, there was no decline in unemployment rate and it went up to 6% from 4%.

Economists in general, do not support the central tenet of the analyses of Friedman and Phelps, which is that there exists a rate of unemployment which, when maintained, will be compatible with inflation that is at a stable rate. Several economists also look at this as the "nonaccelerating inflation rate of unemployment" (NAIRU) due to the fact that, unlike the term "natural rate," NAIRU is not suggestive of an unemployment rate being 'socially optimal, unchanging, or impervious to policy'.

It is possible for a policymaker to want to place a value on NAIRU. For a simple estimate we can look at Figure 7.7 which covers 1976 to 2002 and plots the changes in the rate of inflation (acceleration of prices) against the unemployment rate for this period.

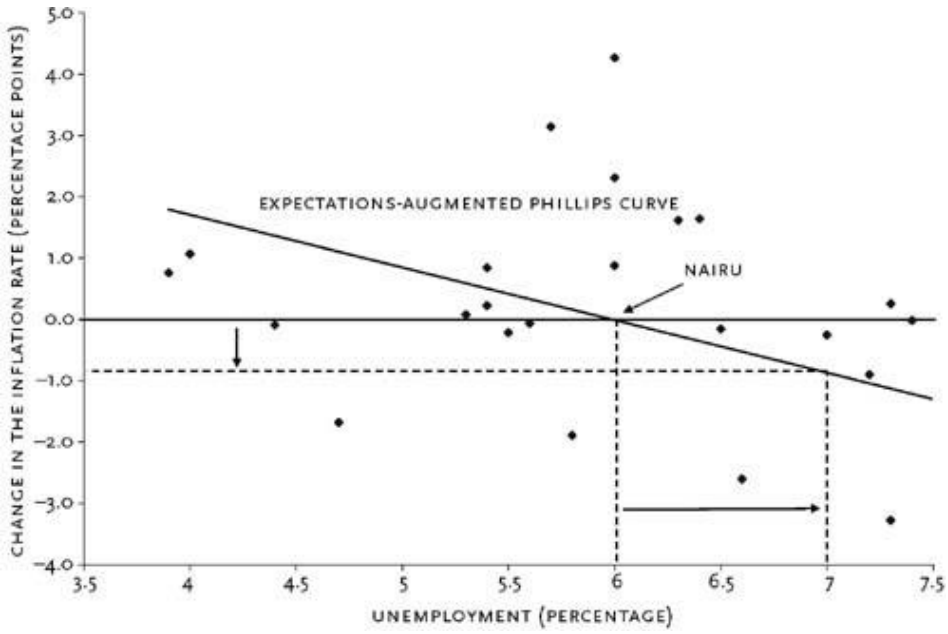


Fig 7.7 *The Expectations-Augmented Phillips Curve, 1976–2002*

Source: *Bureau of Labor Statistics*

The expectations-augmented Phillips curve will be that straight line which is the best fit for the points on the graph (regression line). This provides a summarization of a rough inverse relationship. Based on the regression line, NAIRU (rate of unemployment for which the change in the rate of inflation is zero) is approximately 6%. A Phillips curve's slope is indicative of the speed of adjustment of price. If we consider the economy to be at NAIRU and with an inflation rate of 3% and the government wants to bring down the inflation rate to 0 (zero). It is suggested by Figure 7.7 that contractionary fiscal and monetary policies that pushed average rate of unemployment as high as 7% (a point more than NAIRU) would bring reduction in inflation of approximately a percentage point per year. So, if due to the policies of the government the unemployment rate hovered around 7 per cent, a 3 per cent inflation rate would, on average, get decreased by one point every year and reach 0 (zero) in approximately 3 years.

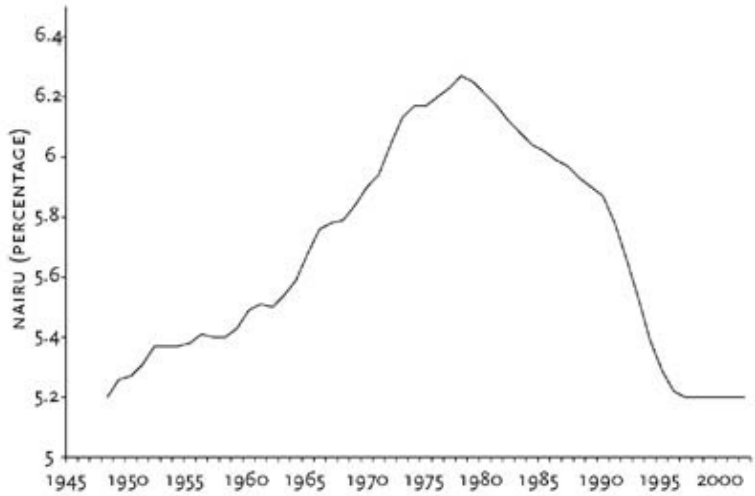


Fig. 7.8 *Nonaccelerating Inflation Rate of Unemployment*

Source: *Congressional Budget Office*

With the use of similar, though more refined, methods, it was estimated by the Congressional Budget Office estimated (as shown in Figure 7.8) that in 1950 NAIRU stood at approximately 5.3 percent, and rising steadily, it peaked at approximately 6.5% in 1978 before steadily falling all the way down to approximately 5.2% till century's close. It became obvious that NAIRU is not constant. There is a change in NAIRU when there is change in the 'real factors' affecting the demand and supply of labour, like relative prices, taxation structure, power of labour unions, technology and demographics. Fiscal and monetary policy should not affect a change in NAIRU, though they affect aggregate demand and do not change these real factors.

7.7 PRICE BEHAVIOUR AND EXPECTATION

Ever since the beginning of economics as a science, economists have been concerned with the theory of value explaining that value is determined by the intersection of the supply and demand curves and that the elasticities of these curves influence prices significantly. However, we are also told that the value of money depends on the quantity of money. The quantity theory of money, which constitutes an important pillar of the classical macroeconomic theory, states that any change in the quantity of money produces a proportionate change in the same direction in the general level of prices. In other words, the quantity theory of money states that the value of money is a function of the supply of money such that when the supply of money is doubled, its value is halved and *vice versa.* This conclusion is based on the assumption of full employment, i.e., given aggregate real output. The causal relationship between changes in the supply of money and the general price level constitutes the core of the quantity theory of money. In the classical economic theory, money has no inherent utility and it is exclusively demanded for the transactions purposes.

Criticizing the classical quantity theory of money, Keynes has observed thus: 'If we reflect on what we are being taught and try to rationalize it, in the simpler discussions it seems that the elasticity of supply must have become zero and demand proportional to the quantity of money; whilst in the more sophisticated we are lost in a haze where nothing is clear and everything is possible. We have all of us become used to finding ourselves sometimes on the one side of the moon and sometimes on the other, without knowing what route of journey connects them, related, apparently, after the fashion of our waking and our dreaming lives.'

Classical economics is faulty because there is a dichotomy between the theory of the relative prices and the theory of the absolute price level. According to the classical economic analysis, the real and the monetary sectors of the economy were completely self-contained and independent of one another. This being so, it was argued by the classicists that the relative prices which were determined by the intersection of the real supply and demand curves in the real or expenditure sector were not affected by changes in the quantity of money. The classical economists were guilty of failing to recognize that changes in the aggregate money supply could never influence the general price level in the economy without first affecting the relative prices because the former was merely an aggregation of the latter. Any change in the general price level must necessarily be traceable to prior changes in the relative prices of goods and services in the economy. Consequently, it is necessary to relate the quantity theory of money with the theory of relative prices in order to find out the real chain of causation between changes in the quantity of money and changes in the general price level.

Check Your Progress

9. What is a Phillips curve?
10. According to Phillips, what happens when the unemployment rate is low?

Keynes criticized the classical approach which regarded money as neutral having no influence on the economy's real equilibrium—on the equilibrium relative prices, interest rate and output. The classical theory was a theory of the stationary equilibrium while Keynes was largely concerned with developing a theory of the shifting equilibrium—the theory of an economy in which changing views about the future influence the present situation. In such a world, money is important because it serves as a vital link between the present and future.

In Chapter 21 of *The General Theory*, Keynes has developed a reformulated quantity theory of money integrating the theory of prices with the general theory of value and output. He has denied any direct causal relationship between increases in the quantity of money and the rise in the general price level so long as unemployed resources are present in the economy. According to Keynes, the effect of changes in the quantity of money on the general price level is composed of the effect on the wage-unit and the effect on employment. Assuming that all the unemployed resources are homogeneous and interchangeable in their efficiency to produce what is wanted and that the factors of production entering into the marginal cost are content with the same money wage so long as there is a surplus of them unemployed, there will be constant returns and a fixed wage-unit as long as unemployment exists.

In this situation, an increase in the quantity of money will not raise prices so long as there is unemployment and employment will increase in the exact proportion to increase in the aggregate effective demand brought about by an increase in the quantity of money. But after full employment has been reached, wage-unit and prices will rise in the same proportion in which the aggregate effective demand increases as a consequence of the increase in the quantity of money. It, therefore, follows that if the aggregate supply function is perfectly elastic, so long as there is unemployment and perfectly inelastic after full employment has been reached and if the aggregate effective demand changes in the same proportion as the quantity of money, the quantity theory of money according to John Maynard Keynes, can be enunciated thus: 'So long as there is unemployment, employment will change in the same proportion as the quantity of money; and when there is full employment, prices will change in the same proportion as the quantity of money.'

This reformulated quantity theory of money is based upon a sufficient number of simplifying assumptions which will not always hold. The following possible complications will, in fact, influence the events and upset the operational importance of the theory:

1. The aggregate effective demand will not change in the same proportion in which the quantity of money will change.
2. Since resources are not homogeneous, diminishing returns will result as employment increases.
3. Since resources lack perfect substitutability, the supply of some commodities will become inelastic while there are still unemployed resources available for the production of other goods.
4. The wage-unit will rise before full employment has been reached.
5. The factor rewards entering into the marginal cost will not change in the same proportion.

Due to the above complicating factors, the increase in the aggregate effective demand occasioned by the increase in the total money supply will, generally speaking, spend itself partly in increasing the volume of employment and partly in raising the level of prices in the economy. Thus instead of assuming constant prices before full employment

and prices rising after full employment, we have a situation of prices rising gradually as employment increases. 'The Theory of Prices, that is to say, the analysis of the relation between changes in the quantity of money and changes in the price level with a view to determining the elasticity of prices in response to changes in the quantity of money, must, therefore, direct itself to the five complicating factors set forth above.'

A change in the quantity of money affects the aggregate effective demand by influencing the rate of interest. By how much will the rate of interest change as a consequence of a given change in the quantity of money will depend upon the nature of the liquidity preference schedule which shows by how much the rate of interest must fall so that the new money supply may be demanded by the willing hoarders. Given the fall in the rate of interest, the total increase in aggregate investment which this fall in the rate of interest will produce will be determined by the schedule of the marginal efficiency of investment. Finally, by how much will a given increase in the aggregate investment increase the aggregate effective demand will be determined by the investment multiplier.

However, this analysis presents an over-simple view because we should not forget that the liquidity preference schedule, the marginal efficiency of investment, and the investment multiplier are partly themselves dependent on the complicating factors 2, 3, 4 and 5 set forth above. 'For the schedule of liquidity preference itself depends on how much of the new money is absorbed into the income and industrial circulation, which depends in turn on how much effective demand increases and how the increase is divided between the rise of prices, the rise of wages, and the volume of output and employment. Furthermore, the schedule of marginal efficiency will partly depend on the effect which the circumstances attendant on the increase in the quantity of money have on expectations of the future monetary prospects. And finally the multiplier will be influenced by the way in which the new income resulting from the increased effective demand is distributed between different classes of consumers. Nor, of course, is this list of possible interactions complete. Nevertheless, if we have all the facts before us, we shall have enough simultaneous equations to give us a determinate result. There will be a determinate amount of increase in the quantity of effective demand which, after taking everything into account, will correspond to, and be in equilibrium with, the increase in the quantity of money. Moreover, it is only in highly exceptional circumstances that an increase in the quantity of money will be associated with a decrease in the quantity of effective demand.'

According to Keynes, the ratio between the quantity of the aggregate effective demand and the quantity of money closely corresponds to the 'Income velocity of money' except that the effective demand corresponds to the income the expectation of which has set the production moving, not to the actually realized income, and to the gross, not to the net income. The income velocity of money will not be constant because it depends upon many complex and variable factors.

The quantity theory of money asserts a direct and proportionality relationship between the quantity of money and the general price level. In other words, the money elasticity of the general price level is unity, i.e. $e = \frac{dP}{dM} \cdot \frac{M}{P} = 1$. There is, however, no direct relationship between changes in the quantity of money and changes in the general price level. The general price level will not rise in the same proportion in which the quantity of money increases unless the aggregate effective demand increases in the

NOTES

same proportion in which the money supply increases, i.e., unless the money elasticity of the aggregate effective demand is unity, i.e.,

$$e_d = \frac{dD}{dM} \cdot \frac{M}{D} = 1$$

Furthermore, the general price level has to rise in the same proportion as the aggregate effective demand if the prices have to rise in the same proportion as the increase in the quantity of money. In other words, the elasticity of money prices with respect to the aggregate effective demand as measured in terms of money should be unity or one, i.e., $e_p = \frac{dP}{dD} \cdot \frac{D}{P} = 1$. This condition will be satisfied if the elasticity of aggregate output or supply with respect to the aggregate effective demand is zero, i.e. if $e_p = \frac{dO}{dD} \cdot \frac{D}{O} = 1$ or if the elasticity of money wages with respect to the aggregate effective demand in terms of money is unity, i.e. if $e_w = \frac{dW}{dD} \cdot \frac{D}{W} = 1$. The condition $e_0 = 0$ means that the aggregate output does not respond to an increase in the aggregate effective demand while the condition $e_w = 1$ means that the wage-unit in terms of money, increases in the same proportion in which the aggregate effective demand increases. The aggregate output in either case will be unaltered. According to Keynes, the traditional quantity theory of money will be valid if $e = e_p \cdot e_0 = 1$, i.e., if,

$$\frac{dP}{dM} \cdot \frac{M}{P} = \left(\frac{d_P D}{DP}\right) \cdot \left(\frac{dD}{dM} \cdot \frac{M}{D}\right) = 1$$

This means that both the e_d and e_P must separately be unity or if one is less than unity, the other must be greater than unity such that their product is unity to give the quantity theory of money result, namely that $e = 1$. In general, e is not unity and, as Keynes has stated, 'It is, perhaps, safe to make the generalization that on plausible assumptions relating to the real world, and excluding the case of a flight from currency in which e_d and e_w become large, e is, as a rule, less than unity.'

Keynes' theory of money and prices is superior to the old quantity theory of money. While according to the old quantity theory of money, every increase in the money supply is necessarily the cause of a price rise, Keynes' theory of prices exposes the quantity theory of money's fallacy by stressing the fact that money inflation will result in price inflation only after full employment has been reached. So long as there exist unemployed resources in the economy, increases in the quantity of money will increase employment and not prices. As a guide to practical policy, Keynes' theory of money and prices stresses the desirability of deficit financing by creating and releasing more money in circulation to remove unemployment of resources from the economy. The theory relieves the policy-makers of the unduly false fear of inflation when the economy is caught in the whirlpool of depression. But it warns us to guard against inflation as soon as full employment is reached. Keynes' theory of money and prices is superior to the old

NOTES

7.8 ADAPTIVE AND RATIONAL EXPECTATIONS

Adaptive Expectations

Adaptive expectations, in economics is a hypothesized process by which people form their expectations about what will happen in the future based on what has happened in the past. Let's take an example: if inflation has been higher than expected in the past, people would revise expectations for the future.

One simple version of adaptive expectations is stated in the following equation,

where p^e is the next year's rate of inflation that is currently expected; p^e_{-1} is this year's rate of inflation that was expected last year; and p is this year's actual rate of inflation:

$$p^e = p^e_{-1} + \lambda(p - p^e_{-1})$$

where λ is between 0 and 1. According to the equation, current expectations of future inflation reflect past expectations and an 'error-adjustment' term, in which current expectations are raised (or lowered) according to the gap between actual inflation and previous expectations. This error-adjustment is also called "partial adjustment."

The theory of adaptive expectations can be applied to all previous periods so that current inflationary expectations equal:

$$p^e = \lambda \sum_{j=0}^{\infty} \left((1-\lambda)^j \, p_j \right)$$

where p_j equals actual inflation j years in the past. Thus, current expected inflation reflects a weighted average of all past inflation, where the weights get smaller and smaller as we move further in the past.

In this model, once a forecasting error is made by agents, due to a stochastic shock, they will be unable to correctly forecast the price level again even if the price level experiences no further shocks since they only ever incorporate part of their errors. Due to this problem, the economist John Muth developed a different way of how expectations were formed, that is, the rational expectations theory. This theory will be discussed subsequently.

Adaptive expectations were instrumental in the Phillips curve outlined by Milton Friedman. For Friedman, workers form adaptive expectations, so the government can easily surprise them through unexpected monetary policy changes. As agents are trapped by the money illusion, they are unable to correctly perceive price and wage dynamics, so, for Friedman, unemployment can always be reduced through monetary expansions. The result is an increasing level of inflation if the government chooses to fix unemployment at a low rate for an extended period of time.

Rational Expectations

Rational expectations, in economics are model-consistent expectations. Agents inside the model on average assume that the model's predictions are valid. To obtain consistency within a model, the predictions of the future value of economically relevant variables are optimal given the decision-maker's information set and model structure. The rational expectations assumption is used especially in many contemporary macroeconomic models. Rational expectations does not imply individual rationality. It should not be confused with

rational choice theory, which is used extensively in game theory. This type of modelling expectation was first proposed by the American economist John Muth in 1961.

Most macroeconomic models today study decisions over many periods. Due to this the expectations of workers, consumers and firms about future economic conditions are an essential part of the model. There has been a controversy since long about how to model these expectations. It is well-known that the macroeconomic predictions of the model may differ depending on the assumptions made about expectations. To assume rational expectations is to assume that agents' expectations may be wrong, but are correct on average over time. In other words, although the future is not fully predictable, agents' expectations are assumed not to be systematically biased and use all relevant information in forming expectations of economic variables.

In rational expectations theory it is assumed that outcomes that are being forecast do not differ systematically from the market equilibrium results. As a consequence, rational expectations do not differ systematically or predictably from equilibrium results. That is, it assumes that people do not make systematic errors when predicting the future, and deviations from *perfect foresight* are only random. In an economic model, this is typically modelled by assuming that the expected value of a variable is equal to the expected value predicted by the model.

For example, suppose that P is the equilibrium price in a simple market, determined by supply and demand. The theory of rational expectations says that the actual price will only deviate from the expectation if there is an 'information shock' caused by information unforeseeable at the time expectations were formed. In other words, *ex ante* the price is anticipated to equal its rational expectation:

$$P = P^* + \varepsilon$$

$$E[P] = P^*$$

where P^* is the rational expectation and ε is the random error term, which has an expected value of zero, and is independent of P^*.

These theories were formulated in response to the problems in adaptive expectations theory. In adaptive expectations, the expectations of the future value of an economic variable are based on past values. For example, people would be assumed to predict inflation by looking at inflation last year and in previous years. Under adaptive expectations, if the economy suffers from constantly rising inflation rates (perhaps due to government policies), people would be assumed to always underestimate inflation. Many consider this an unrealistic assumption believing that rational individuals would sooner or later realize the trend and take it into account in forming their expectations.

The hypothesis of rational expectations addresses this criticism of adaptive expectation theory by assuming that individuals take all available information into account in forming expectations. Though expectations may turn out incorrect, they will not deviate systematically from the realized values.

Criticisms of rational expectations theory

Some of the criticisms of rational expectation theory are as follows:

- Rational expectations are expected values in the mathematical sense. In order to be able to compute expected values, individuals must know the true economic model, its parameters, and the nature of the stochastic processes that govern its evolution. If these extreme assumptions are violated, individuals simply cannot form rational expectations.

- Muth's model assumes that at any specific time, a market or the economy has only one equilibrium (which was determined ahead of time), so that people form their expectations around this unique equilibrium. Muth's hypothesis assumed that P^* was unique. If there is more than one possible equilibrium at any time then the more interesting implications of the theory of rational expectations do not apply. In fact, expectations would determine the nature of the equilibrium attained, reversing the line of causation posited by rational expectations economists.

7.9 NEW-CLASSICAL ECONOMICS

The new-classical theory grew largely as a result of the criticism of the Harrod-Domar growth theory which assumed a single production process carried on by employing capital and labour in a rigidly fixed ratio. The neo-classical growth theorists attacked this assumption as being unrealistic. In its place they employed a production function that allowed for changes in the proportion in which both labour and capital may be used in the production of goods and services in the economy. In other words, while the Harrod-Domar theory assumed non-substitutability between labour and capital, the neo-classical theory allowed for finite substitutability between the factors of production used in the production process. The new-classical theory also assumes that the rate of capital accumulation depends on the thriftiness of the economy at full employment.

According to the new-classical growth theory, capital is a unique abstract factor of production which can be adjusted at any time to absorb any size of labour force into employment. Since labour and capital can be combined in the varying proportion, the theory assumes an indefinitely large number of production processes, each process being characterized by a different labour-capital ratio. It, therefore, follows that in place of the fixed output-capital ratio assumed by Harrod and Domar, the neo-classical growth theorists assume that the output-capital ratio can be continuously varied. With a given capital stock, the employment of more labour into production would entail a diminishing output-labour ratio and a higher output-capital ratio. Conversely, the smaller amount of labour employed with the given capital stock would raise the productivity of labour and lower the productivity of capital. All this follows from the law of diminishing returns.

The theory also adopts the classical approach to saving and investment equilibrium which is required for the continued full utilization of factors of production in the economy. The theory assumes perfect competition in the product and factor markets. With the flexible prices of output and inputs, the aggregate output depends on the supply of available inputs which under the assumption of flexible prices find full employment. Thus, having assumed full employment of resources through the classical approach, the theory analyses the growth path which will be followed by a fully resource-employed economy as the endowment of the economy's resources grows over time. Unlike the Harrod-Domar theory, there is no need in the neo-classical growth theory to distinguish between the growth rate of the potential or capacity output and the growth rate of the actual or realized output of the economy because the latter becomes identical with the former.

Output Growth without Technological Progress

The rate of growth of the economy's output basically depends on the rate of growth of the economy's capital stock, labour force and technology. This relationship for any particular time period may be expressed in the following aggregate production function.

$$Y = f(K, L, T)$$

Check Your Progress

11. How has Keynes criticised the classical theory of money?
12. What are adaptive expectations? State an example.
13. Why is Keynes' theory of money and prices seen superior to the old quantity theory of money?

where Y is the total output, K is total capital stock, L is total labour force in the economy and T is the index of technical know-how which grows at some given rate over time. Assuming no or zero growth of technology in the economy, the above aggregate production function can be written as–

$$Y = f(K, L)$$

How would the economy's aggregate output Y change when, *ceteris paribus*, K and L change. According to the Harrod-Domar growth model, Y would increase in the same proportion in which both K and L change. For example, if both K and L increase by 2 per cent, Y would also increase by 2 per cent. The neo-classical theory also gives the same result because the assumption in both the theories is that of the constant returns to scale. However, there is one difference between the two theories. In the Harrod-Domar theory, a 3 per cent increase in the economy's labour force L, with no increase in the total capital stock K, will cause no increase in the aggregate output Y due to the requirement of fixed proportion between L and K. Unless K also increases by 3 per cent, the increase in L would be useless. However, this is not quite so in the neo-classical growth theory where it is possible to vary the labour-capital ratio. Consequently, the increased labour force could be absorbed into employment with the same fixed capital stock and some increase in Y could be obtained. If the increase in the economy's aggregate labour force in any given time period was not very large, the increase in output (ΔY) would be equal to the increase in the economy's total labour force (ΔL) times the marginal physical product of labour MPP_L, i.e.,

$$\Delta Y = MPP_L \cdot \Delta L$$

Similarly, with constant labour force L, an increase of ΔK amount in the economy's total capital stock would cause the economy's aggregate output Y to increase to the extent of increase in the capital stock times the marginal physical product of capital, i.e.,

$$\Delta Y = MPP_K \cdot \Delta K$$

when both L and K change in any given time period, the change in the output may be written as:

$$\Delta Y = MPP_K . \Delta K + MPP_L . \Delta L$$

Dividing both sides by Y gives–

$$\Delta Y/Y = (MPP_K/Y)\, \Delta K + (MPPL/Y)\, \Delta L$$

We may also write–

$$\Delta Y/Y = (MPP_K . K/Y) . \Delta K/K + (MPP_L . L/Y) . \Delta L/L \quad ...(7.1)$$

Under the assumption of perfect competition in all markets, together with the marginal productivity theory of factor pricing, the total output will be exactly absorbed by the total factor payments when the production function shows the constant returns to scale, i.e.,

$$MPP_K \cdot K + MPP_L \cdot L = Y$$

Dividing through by Y we obtain–

$$MPP_K \cdot K/Y + MPP_L \cdot L/Y = Y/Y = 1$$

Since $MPP_K \cdot K/Y + MPP_L \cdot L/Y = 1$, we may write a and $1 - a$ for $MPP_K \cdot K/Y$ and $MPP_L \cdot L/Y$ respectively in equation (7.1) and may rewrite the equation in the following form:

$$\Delta Y/Y = a\, (\Delta K/K) + (1 - a)\, \Delta L/L \quad ...(7.2)$$

The size of a shows the proportion or percentage of the economy's total output or income which would accrue as returns on the capital if capital was paid according to its

marginal physical product. In other words, a measures the capital elasticity of output. If $a = 0.5$ and if capital is paid its marginal product, capital would share 50 per cent of the economy's total product or income. Expressed differently, a 2 per cent increase in the economy's total capital stock would increase the economy's total output by 1 per cent. The same holds good with respect to labour whose share in the total product is represented by $(1 - a)$.

Assuming $a = 0.5$, the percentage change in total output resulting from a given percentage change in both capital and labour can be found out from equation (7.2). If capital and labour both increase by 10 per cent, the resulting increase in the total output will also be 10 per cent because the underlying production function shows constant returns to scale. The fact that the total output will increase by 10 per cent has been shown below:

$$\Delta Y/Y = 0.05 \times 0.1 + 0.05 \times 0.1 = 0.1 \text{ or } 10 \text{ per cent}$$

In the neo-classical theory, with its assumption of factor substitutability, the rate of growth of labour has its own influence on the rate of growth of the economy's output. Unlike the Harrod-Domar growth theory where the equilibrium growth rate of the economy's output was defined by the growth equation $\Delta Y/Y = s$, in the neo-classical theory the term s (propensity to save times the productivity of capital) does not explain the equilibrium rate of growth of economy's output. In fact, it may well be that the rate of growth of capital, apart from being insufficient in itself to explain the rate of growth of economy's output, may not be as important as is the rate of growth of labour force for the growth of economy's output. This is so because the influence of capital accumulation depends heavily on the value of a and it any be relatively small.

Output Growth and Technological Progress

It is an established fact that technological progress contributes significantly to the growth rate of economy's total output. In this respect it is a sort of input or factor independent of capital or labour with which it collaborates in production. If we designate the rate of technological progress by $\Delta T/T$ and assume that a 3 per cent rate of technological progress contributes to a 3 per cent increase in the output which would not have been possible in the absence of technological progress, the growth rate of the economy's output will be given by the following equation.

$$\frac{\Delta Y}{Y} = \frac{\Delta T}{T} + a\left(\frac{\Delta K}{K}\right) + (1-a)\frac{\Delta L}{L} \qquad ...(7.3)$$

The growth of output per worker is denoted by the following equation.

$$\frac{\Delta Y}{Y} - \frac{\Delta L}{L} = \frac{\Delta T}{T} + a\left(\frac{\Delta K}{K} - \frac{\Delta L}{L}\right) \qquad ...(7.4)$$

The growth of output per worker no longer depends only on the growth of capital per worker. Even if $\Delta K/K = \Delta L/L$, the growth of output per worker will not be zero, it will be positive and equal to $\Delta T/T$, i.e., $\Delta Y/Y - \Delta L/L = \Delta T/T$. The per worker output growth rate will, of course, be higher if $\Delta K/K > \Delta L/L$. However, the condition that $\Delta K/K > \Delta L/L$, i.e. the growth rate of capital must be higher than labour's growth rate if the output has to grow is no longer necessary for a positive growth rate of output per worker which was necessary in the absence of technological progress. Consequently, in the new situation,

of technological progress. Treating technological progress as a residual or

catchall for all the other factors of growth, the rate of growth of output per worker will depend on the growth of capital and technology. Since the growth rate of output per worker is $\Delta K/K - \Delta L/L$, the ratio of a $(\Delta K/K - \Delta L/L)$ to $\Delta Y/Y - \Delta L/L$ gives the proportion of total growth rate of output per worker which is due to the growth of capital per worker. The proportion of the total growth rate of output per worker due to the technological progress can be found out by deducting the either proportion from 1, i.e., the proportion of economy's output growth rate which is due to the technological progress will be equal to–

$$1 - [a(\Delta K/K - \Delta L/L) \div (\Delta Y/Y - \Delta L/L)]$$

In contrast to the Harrod-Domar growth theory which assumes rigidly fixed labour-capital ratio, the neoclassical growth theory regards this ratio as variable due to the substitutability between labour and capital. The theory considers capital stock as a uniquely homogeneous resource which is capable of being reshaped and adjusted to be employed with any amount of labour. Capital is a jelly which can take any form to suit the changing amount of labour in employment. But once we consider the embodied technological progress, capital ceases to be homogeneous and one capital product will differ from the other in terms of the different technological progress that it embodies. Consequently, the amount of total labour required per unit of capital will differ in the case of each capital good of different vintage. Moreover, in practice the labour required per existing capital unit of each vintage is fairly fixed over a wide range.

7.10 REAL BUSINESS CYCLE AND ECONOMIC POLICY

The real business-cycle theory also referred to as RBC theory, is a part of the New classical macroeconomics models according to which it is possible to mostly account for business-cycle fluctuations with using real (in contrast to nominal) shocks.

In contrast with some other prominent theories pertaining to business cycle, the RBC theory looks upon fluctuations in business cycle to be 'efficient response to exogenous changes in the real economic environment'. It means that the level of national output necessarily maximizes expected utility, and due to this governments must have greater focus on long-run structural policy changes instead of intervening with discretionary monetary or fiscal policy which is formulated for the active smoothening of short-term economic fluctuations.

RBC theory puts forth that business cycles are 'real' as they are not representative of a failure of markets to clear and instead represent the economy's most efficient possible operation, given the economy's structure.

In the real business cycle theory, there is a categorical rejection of the Keynesian economics as well as the real effectiveness of monetary policy as promoted by monetarism and by New Keynesian.

Business Cycles

If snapshots of different points in time were taken of an economy, each snapshot would be different from the other. There are two reasons why this would be so:

- Over time, a number of advanced economies display sustained growth. Due to this if snapshots are taken several years apart, in the later periods there will most probably be seen higher economic activity levels.

- There exist seemingly random fluctuations around this growth trend. Thus given two snapshots in time, predicting the latter with the earlier is nearly impossible.

Fig. 7.9 *Real GNP (billions)*

One method that is commonly used for observing behaviour of this nature is studying a time series of an economy's output (more specifically of gross national product (GNP)). It will be the value of the services and goods that are created by the workers and businesses of the country.

In the above figure, there is the depiction of the time series of real GNP for the United States of America between 1954 and 2005. It depicts a continuous growth in the output, but the increase is not steady with fast and slow growth periods. The figure given below depicts these levels as growth rates of real GNP and brings out a growth trend that is much smoother. The method commonly used to arrive at this trend is the Hodrick–Prescott filter. What is needed is finding the balance between the extent to which general growth trend follows the cyclical movement. With the HPfilter is possible to identify the longer term fluctuations as a part of the growth trend and looking at the more jumpy fluctuations as belonging with the cyclical component.

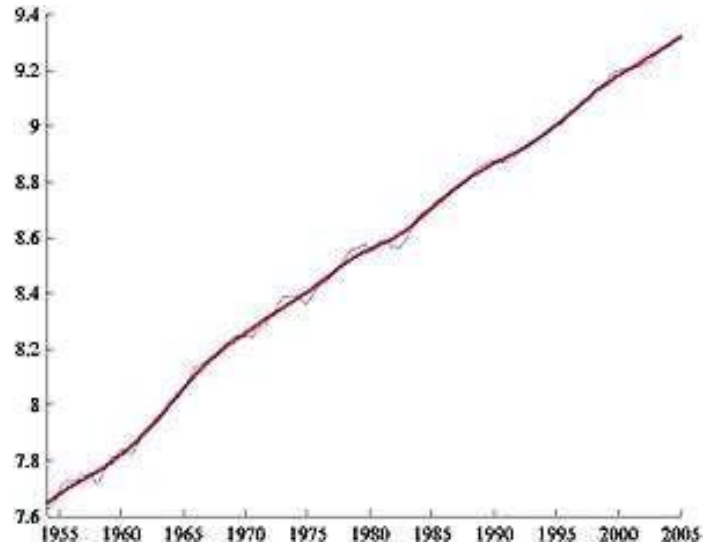


Fig. 7.10 *Log Real GNP and Trend*

The cyclic movements that are based on the trends are referred to by economists as business cycles. The figure given below depicts such deviations. If focus is kept on the horizontal axis at 0, each point on that line indicates that in that year, no deviation from the trend occurred. All other pints whether they lie above or below that horizontal line represent deviations. Through employing log real GNP the distance between the 0 line and any other point is roughly equal to percentage deviation from the growth trend of the long run.

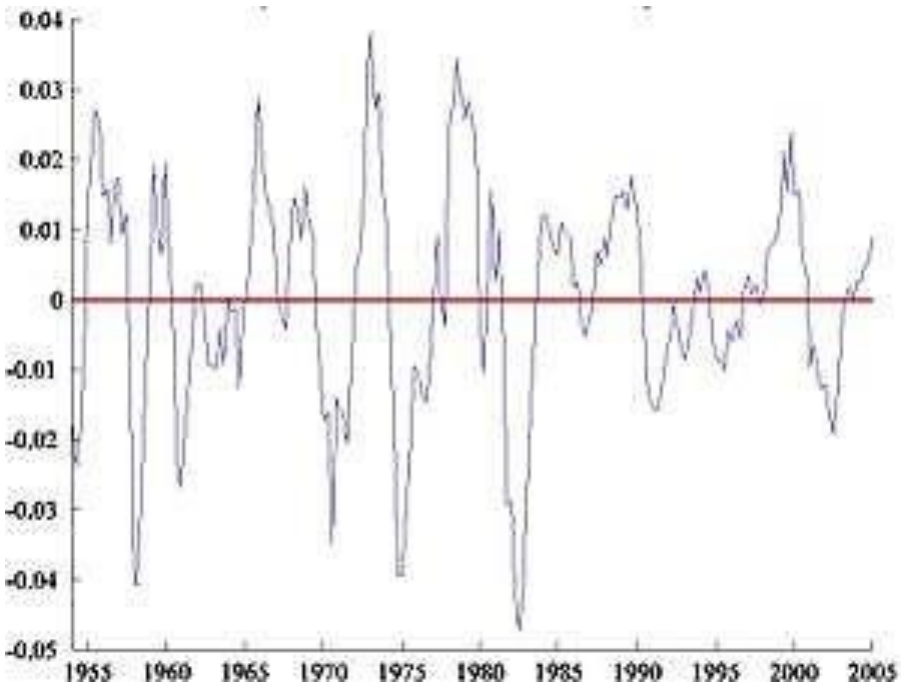


Fig. 7.11 *Deviations from Trend in Log GNP*

Large positive deviations that are above the 0 axis are referred to as peaks. The negative deviations are the ones lying below the 0 axis and the relatively larger amongst them are referred to as troughs. Booms are a series of positive deviations that cause peaks and recessions are a series of negative deviations that create troughs.

When macroeconomic variables are considered, patterns are seen in the deviations and the irregularities. To take an example, let us look at the figure given below. This figure depicts fluctuations in consumption spending and in output. It can be seen that troughs and peaks align at almost the same places and there is a coinciding of downturns and upturns.

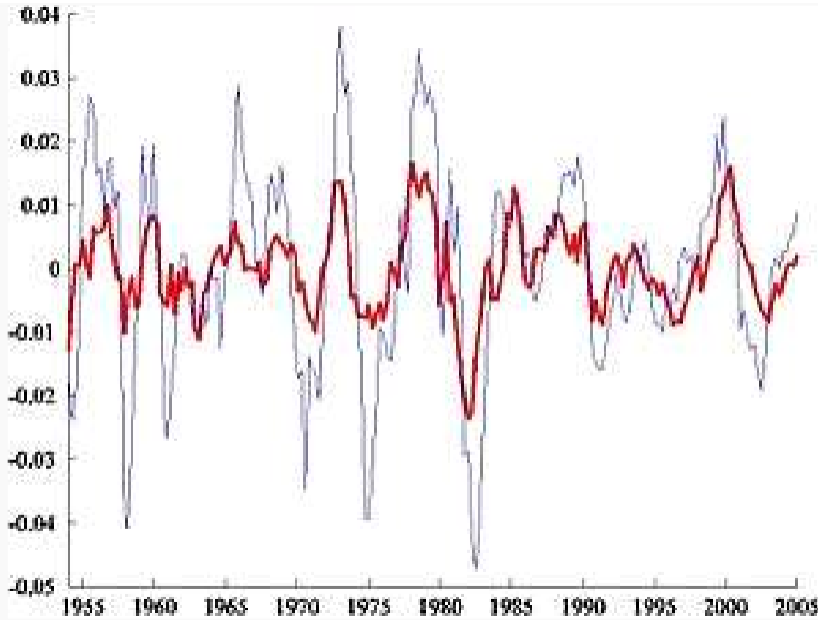


Fig. 7.12 *Deviations from Trend in Log GNP and Consumption*

So, it would not be wrong to say that if some other similar data is taken, similar qualities may be displayed by it. Some examples of similar data would be: (a) labour, hours worked (b) productivity, how effective firms use such capital or labour, (c) investment, amount of capital saved to help future endeavours, and (d) capital stock, value of machines, buildings and other equipment that help firms produce their goods.

Though in the graph given below, the peaks and trough are coinciding, that is not the case in the next graph following it. To be able to find more similarity, there is need to look closely at statistics.

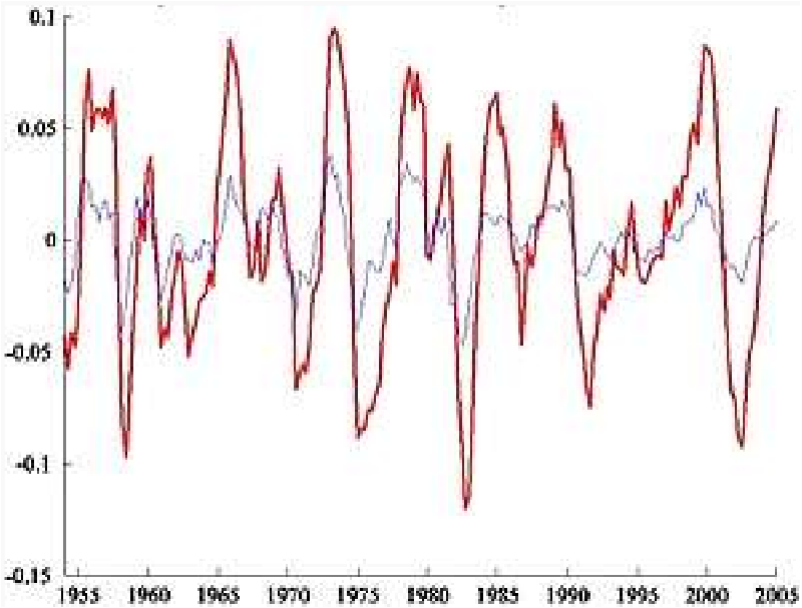


Fig. 7.13 *Deviations from Trend in Log GNP and Investment*

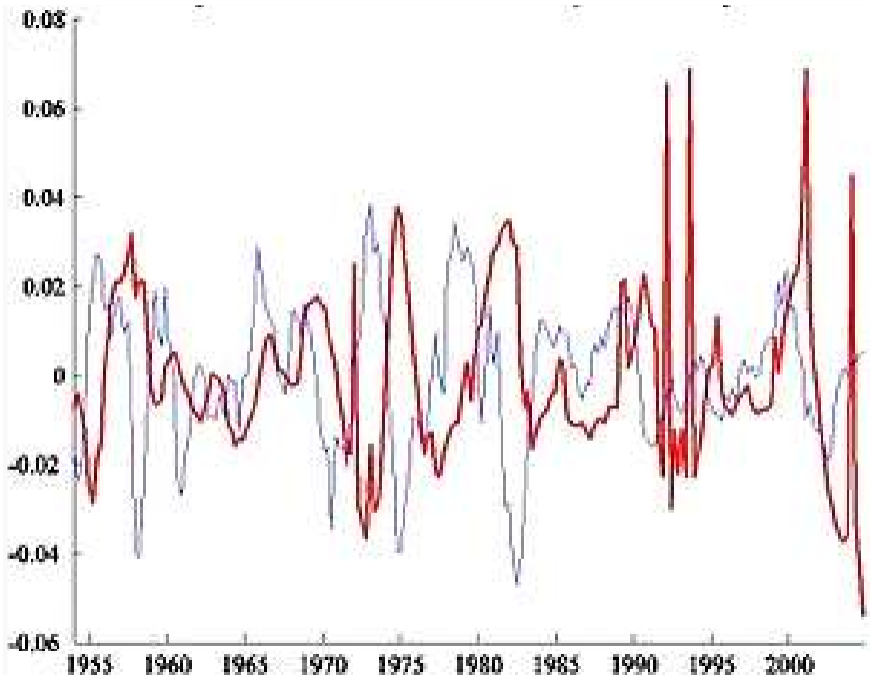


Fig. 7.14 *Deviations from Trend in Log GNP and Capital*

Stylized Facts

Even with a quick look at data it is possible to locate various regularities, and these are at times referred to as stylized facts. One such irregularity is persistence.

7.11 SUMMARY

In this unit, you have learnt that,

- Inflation, which today confronts the economic policy-makers throughout the world in the form of a most dominant economic problem, is not a new phenomenon because from the earliest days of recorded history, mankind has been puzzled and discomfited by rising prices.
- According to the public understanding, inflation means a condition which produces a rising trend in the general price level in the economy.
- Inflation may, however, be present in the economy if the sustained price rise, which would have otherwise occurred, is prevented from occurring by imposing the price and physical controls in the economy.
- For a milder sustained price rise, economists have used the terms walking inflation and running inflation.
- In walking inflation, a sustained price rise may be of about 8-10 per cent yearly. For the higher two-digit sustained annual price rise the term 'running inflation' is used.
- According to the classical economic analysis, the general price level depends directly and proportionately on the supply of money.
- According to the classicists, inflation occurs when the quantity of money increases and comes to a halt when the quantity of money becomes stable.

Check Your Progress

14. On what factors does the rate of growth of the economy's output depend upon?
15. Define real business cycle theory.
16. If snapshots of different points in time were taken of an economy, why would each snapshot be different from the other?

- Keynes severed the close relationship between the quantity of money and the level of aggregate demand by showing that even with constant money supply, some inflation may be experienced.
- Despite differences in their analytical reasoning, both Wicksell and Keynes explained inflation as an excess of the aggregate demand over the full employment output in the economy.
- Keynes merely removed the rigid relationship between the aggregate demand and the quantity of money and also considered less–than–full employment situation in which the increased aggregate demand was more likely to raise the level of economy's aggregate output than the prices.
- Cost-push inflation in the economy occurs as a result of the combination of both the wage-push and the profit-push factors.
- The phenomenon of demand-pull inflation is frequently controlled by resorting to the instruments of monetary and fiscal policies.
- According to the classical quantity theory approach, the demand inflation can be controlled by resorting to an appropriate monetary policy so as to halt the expansion of the money supply.
- Since cost-push inflation is largely caused by the rising cost, supply inflation can be controlled by maintaining wage-rate stability and by preventing those wage increases which are not related to the increase in labour productivity.
- The centerpiece of monetarism relates to the overwhelming influence of money on the long-run economic activity. According to the monetarists, proper growth rate of money supply is crucial for the stable growth of output and prices.
- Milton Friedman does not regard monetary policy as a mere tool of short-run economic stabilization. According to him, monetary policy 'can prevent money itself from being a major source of economic disturbance.'
- A major point in the monetarists' analysis concerns the lags in the effect of monetary policy. The problem of lags in the operation of monetary policy arises from the recognition of the fact that monetary policy actions of the central bank do not have a quick and immediate effect on the economic conditions.
- Monetary policy has a profound effect on the rate of interest. If the rate of interest is reduced as a result of an increase in money supply, the rate of investment will rise and the increase in investment will lead to increase in income and employment via the multiplier.
- Modern monetarists, led by Milton Friedman, revived and modified the classical monetary theory of inflation. The modern monetarists hold that the general level of price rises only due to increase in money supply but not proportionately.
- Phillips curve is the representation of the relationship that exists between unemployment rate and rate of inflation. The study done by A. W. H. Phillips for United Kingdom of the inflation of wage and of unemployment during 1861 and 1957 is looked upon as being a landmark for the development of macroeconomics.
- In case of developed economies, economists quickly formulated Phillips curves. Majority of them formed a relationship between general price inflation, and not wage inflation, with unemployment.

- The quantity theory of money states that the value of money is a function of the supply of money such that when the supply of money is doubled, its value is halved and vice versa.
- A change in the quantity of money affects the aggregate effective demand by influencing the rate of interest.
- Keynes' theory of money and prices is superior to the old quantity theory of money.
- While according to the old quantity theory of money, every increase in the money supply is necessarily the cause of a price rise, Keynes' theory of prices exposes the quantity theory of money's fallacy by stressing the fact that money inflation will result in price inflation only after full employment has been reached.
- Adaptive expectations, in economics is a hypothesized process by which people form their expectations about what will happen in the future based on what has happened in the past.
- Rational expectations, in economics are model-consistent expectations. Agents inside the model on average assume that the model's predictions are valid.
- The real business-cycle theory also referred to as RBC theory, is a part of the New classical macroeconomics models according to which it is possible to mostly account for business-cycle fluctuations with using real (in contrast to nominal) shocks.
- If snapshots of different points in time were taken of an economy, each snapshot would be different from the other.

7.12 KEY TERMS

- **Demand-pull inflation**: Demand-pull inflation is asserted to arise when aggregate demand in an economy outpaces aggregate supply.
- **Inflationary gap**: An inflationary gap, in economics, is the amount by which the actual gross domestic product exceeds potential full-employment GDP.
- **Cost-push inflation:** Cost-push inflation is caused by an increase in prices of inputs like labour, raw material, etc.
- **Phillips curve**: In economics, the Phillips curve is a historical inverse relationship between rates of unemployment and corresponding rates of inflation that result in an economy.

7.13 ANSWERS TO 'CHECK YOUR PROGRESS'

1. The basic characteristic of creeping inflation is that the annual price rise is almost imperceptible so as to be lost sight of by the casual observers.
2. The climax of hyperinflation is reached when the flight from currency becomes so fantastically high that the velocity of money in circulation approaches infinity.
3. The concept of the inflationary gap may be defined as the positive difference between the actual level of aggregate demand which exists in the economy at the full employment level of income and the amount of aggregate demand which is required to attain full employment.

4. The central thesis of the monetarists is that money matters and consequently monetary policy has a crucial role to play in any programme of economic stabilization. The extreme monetarists have asserted that money alone matters. Their assertion is based on the joint validity of the particular assumptions regarding the demand for any supply of money.
5. At the centre of monetarism is the key relationship between changes in the money supply and changes in the level of money income. The line of causation runs from changes in the money stock to changes in the money income so that money changes cause income changes.
6. The excess cash balances are eliminated in one of the two ways. First, if as a result of increased spending, the real income and output increase, the demand for money increases and the excess cash balances are now absorbed by the higher demand for money at the higher real income level. Second, if only the prices increase due to additional spending, the real value of the total money supply falls. The increase in prices continues until the real value of the nominal money supply is restored to its original level.
7. According to Keynes, rate of interest is determined by equilibrium between demand for money and supply of money (i.e., through money market equilibrium).
8. Monetary policy has a profound effect on the rate of interest. If the rate of interest is reduced as a result of an increase in money supply, the rate of investment will rise and the increase in investment will lead to increase in income and employment via the multiplier.
9. When we measure economic activity in terms of unemployment rate, the statistical relationship that is formed is referred to as Phillips curve.
10. It was estimated by Phillips that the unemployment rate being low leads to tight labour market and as unemployment rate becomes lower the labour market becomes tighter. So, this leads to the need for firms to be faster at raising wages to attract and retain labour which was not scarce.
11. Keynes criticized the classical approach which regarded money as neutral having no influence on the economy's real equilibrium—on the equilibrium relative prices, interest rate and output. The classical theory was a theory of the stationary equilibrium while Keynes was largely concerned with developing a theory of the shifting equilibrium—the theory of an economy in which changing views about the future influence the present situation.
12. Adaptive expectations, in economics is a hypothesized process by which people form their expectations about what will happen in the future based on what has happened in the past. Let's take an example: if inflation has been higher than expected in the past, people would revise expectations for the future.
13. Keynes' theory of money and prices is superior to the old quantity theory of money. While according to the old quantity theory of money, every increase in the money supply is necessarily the cause of a price rise, Keynes' theory of prices exposes the quantity theory of money's fallacy by stressing the fact that money inflation will result in price inflation only after full employment has been reached.
14. The rate of growth of the economy's output basically depends on the rate of growth of the economy's capital stock, labour force and technology.

15. The real business cycle theory also referred to as RBC theory, is a part of the New classical macroeconomics models according to which it is possible to mostly account for business-cycle fluctuations with using real (in contrast to nominal) shocks.
16. If snapshots of different points in time were taken of an economy, each snapshot would be different from the other. There are two reasons why this would be so:
    - Over time, a number of advanced economies display sustained growth. Due to this if snapshots are taken several years apart, in the later periods there will most probably be seen higher economic activity levels.
    - There exist seemingly random fluctuations around this growth trend. Thus given two snapshots in time, predicting the latter with the earlier is nearly impossible.

7.14 QUESTIONS AND EXERCISES

Short-Answer Questions

1. Differentiate between adaptive and rational expectations.
2. State the six factors that affect the demand function of money.
3. State in brief the concept of business cycle and the fluctuations.
4. How is the sophisticated version of the cost-push theory different from its crude counterpart?

Long-Answer Questions

1. Inflation is an excess of the aggregate demand over the aggregate supply. Discuss the statement and analyse the factors which cause inflation.
2. Explain demand-pull and cost-push inflation.
3. What is an inflationary gap? How does it arise and how can it be removed in the economy?
4. Analyse the role of monetary policy and its failure in the past.
5. Discuss the concept of output growth with and without technological progress.

7.15 FURTHER READING

Mankiw, N Gregory. 2010. *Macroeconomics.* Worth Publishers, New York.

Shapiro, Edward. 1996. *Macroeconomic Analysis.* Galgotia Publication, New Delhi.

Branson, WH. 1979. *Macroeconomic Theory and Policy*. Universal Book Stall, New Delhi.

Jha, R. 1999. *Contemporary Macroeconomic Theory and Policy*. New Age International, New Delhi.

Gupta, SB. 1995. *Monetary Economics: Instruments and Policy*. S Chand & Co. New Delhi.

UNIT 8 PROBLEMS IN ASSET PRICING

Structure

8.0 INTRODUCTION

A portfolio is a bundle or a combination of individual assets or securities. Portfolio theory provides a normative approach to investors to make decisions to invest their wealth in assets or securities under risk. It is based on the assumption that investors are *risk-averse*. This implies that investors hold well-diversified portfolios instead of investing their entire wealth in a single or a few assets. One important conclusion of the portfolio theory is that if the investors hold a well-diversified portfolio of assets, then their concern should be the expected rate of return and risk of the portfolio rather than individual assets and the contribution of individual asset to the portfolio risk. The second assumption of the portfolio theory is that the returns of assets are normally distributed. This means that the mean (the expected value) and variance (or standard deviation) analysis is the foundation of the portfolio decisions. Further, we can extend the portfolio theory to derive a framework for

valuing risky assets. This framework is referred to as the capital asset pricing model (CAPM). An alternative model for the valuation of risky assets is the arbitrage pricing theory (APT). In this unit, we discuss the problems in asset pricing and risks involved in a portfolio and show how CAPM and APT work in valuing assets.

NOTES

8.1 UNIT OBJECTIVES

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

- Assess the returns from assets and risk
- Discuss the types of risk measurement and the statistical techniques used in measuring risk
- Evaluate the various types of risk in a portfolio
- Assess the theories of cost of capital
- Analyse the capital structure theories
- Evaluate the traditional theory of cost and Modigliani-Miller approach to cost structure
- Explain the Capital Asset Pricing Model (CAPM) and its value in the stock market
- Distinguish between CAPM and Arbitrage Pricing Theory (APT)

8.2 RETURNS FROM ASSETS AND RISK

These two terms — risk and uncertainty — are interchangeably used in practice. '*Risk* is that uncertainty which is predictable, and to which probability can be assigned.' This definition has two conditions in it, (a) the decision maker is aware of all possible outcomes, which may happen and affect the decision, and (b) the decision-maker is in a position to assign probability to each possible outcome envisaged by him. *Uncertainty* is that event which cannot be predicted, and therefore, no probability can be assigned to it.

Consider two projects, A and B. Project A is likely to offer the rate of return in the range of 25 to 45 per cent, whereas Project B may offer 15 to 35 per cent. If the cut-off rate of the firm is 20 per cent one would say that Project B is riskier than Project A, though in both cases the uncertainty (difference of 20 per cent between highest and lowest rate of return) level is equal.

A layman's understanding of the two terms is quite different from statisticians' definitions. A layman looks at the down side risk (a possibility of earning rate of return less than the minimum acceptable rate). We shall also use the term 'uncertainty' to denote 'risk'.

Probability

Probability is the likelihood of happening of some event. Merriam Webster online dictionary defines probability as 'the chance that a given event will occur'. Probability is expressed as a proportion of one or percentage. The sum of probabilities of all possible occurrence is always one or one hundred per cent.

For example, an investor, who buys shares of Company A and Company B, estimates the occurrence of outcome (rate of return) from the investment as given in Table 8.1:

Table 8.1 *Occurrence of Rate of Returns and Probabilities of Two Shares*

State of Economy	Probability	Rate of Returns from Shares of...	
		Company A	Company B
Strong	0.20	50%	30%
Good	0.60	10%	8%
Weak	0.20	–20%	–10%
Total	1.00		

The above occurrences can be plotted on the probability graph given in Figure 8.1A and Figure 8.1B.

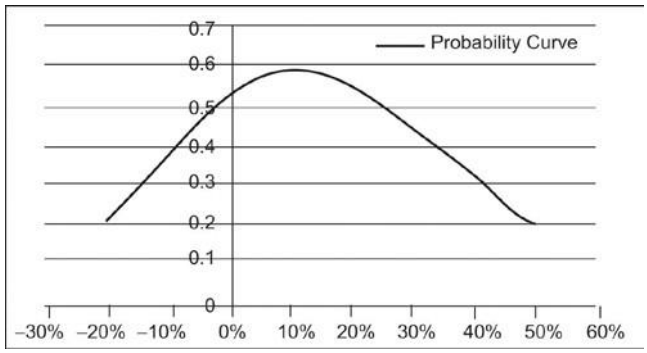


Fig. 8.1A *Probability Distribution of Outcomes for Shares of Company A*

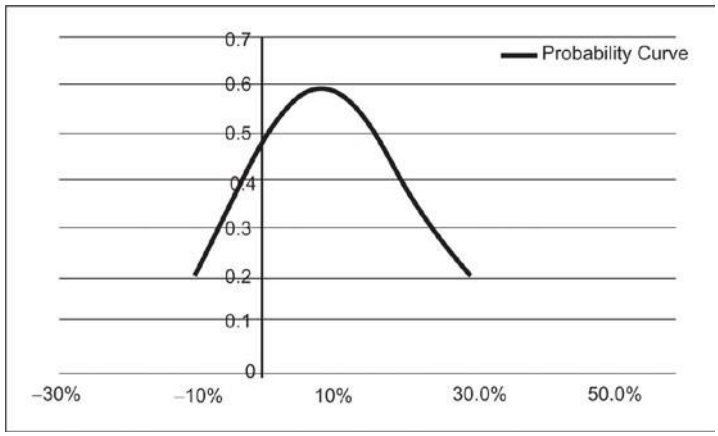


Fig. 8.1B *Probability Distribution of Outcomes for Shares of Company B*

If any occurrence is assigned Probability 1, then there is only one occurrence and it is a certain outcome and not risky.

Volatility

Risk causes volatility. However, not all volatilities are indicators of risk. For example, seasonality causes ups and downs in sales revenue and profits from one quarter to another.

Economic cycle causes cyclical movement in sales revenue and profits from year to year. If volatility is predictable (because seasonality and cyclicality are expected) then it does not indicate riskiness in sales revenue and profits of the firm. The readers must have noticed that share prices and market index do not show seasonality and cyclicality, because the changes in sales revenue and profit are expected outcomes and they are well factored as normal. However, volatility in share prices is an indication of the risk.

Estimating Returns and Probability

The real test of a decision maker is in estimation of returns (occurrences) and probabilities for each occurrence. Returns on financial securities are derived from their underlying assets. Returns from equity shares would depend on earnings of a firm from its business. One needs to access relevant information available from various sources and demonstrate high degree skills to interpret them before returns from equity shares are anticipated. Returns on bonds are easier to forecast as dividend flow is contractual. However, if one wants to sell the binds in open market, the market price can be an estimate based on anticipated changes in the market interest rate.

Assigning probabilities to the outcome is more difficult. Past experience about a similar situation is useful in estimating probabilities. Opinion of expert economists and various agencies like government, rating agencies and others can be taken into account in estimating the probabilities of each outcome.

Expected Returns

Expected return from an investment with more than one expected outcomes is the weighted average returns of all outcomes. Weights of probabilities are considered as given in the example in Table 8.2.

Table 8.2 *Calculation of Expected Returns*

State of Economy	Probability	Company A		Company B	
		Rate of Return	Rate of Return × Probability	Rate of Return	Rate of Return × Probability
Strong	0.20	50%	10.00%	30%	06.00%
Good	0.60	10%	06.00%	8%	04.80%
Weak	0.20	–20%	–04.00%	–10%	–02.00%
Total	1.00		12.00%		08.80%

The equation will be,

$$\bar{R} = R_1 \times P_1 + R_2 \times P_2 + R_3 \times P_3 + ... + R_n \times P_n \quad (8.1)$$

This can be rewritten as,

$$\bar{R} = \sum_{i=1}^{n} R_1 \times P_1 \quad (8.2)$$

Where, $\bar{R}$ = Expected return

R = Returns

R_i = Returns from the occurrence 1 to n

P = Probability of returns

Expected Return (A) = 50% × 0.2 + 10% × 0.6 + (–20%) × 0.2

= 10% + 6% - 4% = **12%**

Expected Return (B) = 30% × 0.2 + 8% × 0.6 + (–10%) × 0.2

= 6% + 4.8% – 2% = **8.8%**

8.2.1 Risk Free Assets

The features of risk free asset are:

(a) Absence of default risk and interest risk and

(b) Full payment of principal and interest amount.

The return from the risk free asset is certain and the standard deviation of the return is nil. The relationship between the rate of return of the risk free asset and risky asset is zero. These types of assets are usually fixed income securities. But fixed income securities issued by private institutions have the chance of default. If the fixed income securities are from the government, they do not possess the default risk and the return from them are guaranteed. Further, the government issues securities of different maturity period to match the length of investors holding period. The risk free assets may be government securities, treasury bills and time deposits in banks.

Inclusion of risk free asset: If a risk free asset is included in a portfolio, the investor can invest part of his money on risk free asset and the remaining amount on the risky asset. It is also assumed that the investor would be able to borrow money at risk free rate of interest. When risk free asset is included in the portfolio, the feasible efficient set of the portfolios is altered. This can be explained in the Figure. 8.2.

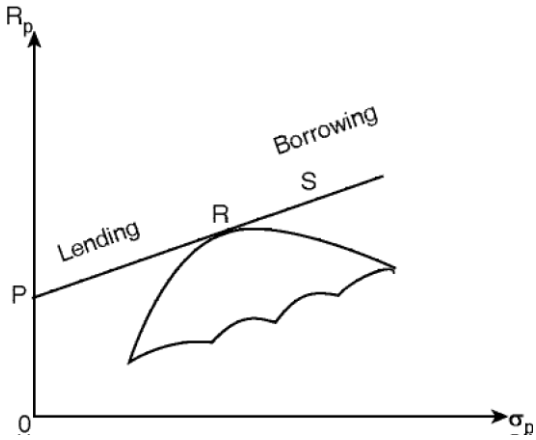


Fig. 8.2 *Efficient Frontier with Borrowing and Lending*

In the Figure 8.2, OP is gained with zero risk and the return is earned through holding risk free asset. Now, the investor would attempt to maximize his expected return and risk relationship by purchasing various combinations of riskless asset and risky assets. He would be moving on the line connecting attainable portfolio R and risk free portfolio P i.e. the line PR. When he is on the PR, part of his money is invested in fixed income securities i.e. he has lent some amount of money and invested the rest in the risky asset within the point PR. He is depending upon his own funds. But, if he moves beyond the point R to S, he would be borrowing money. Hence, the portfolios located between the points RP are lending portfolios and beyond the point R consists of borrowing portfolios. Holding portfolio in PR segment with risk free securities would actually reduce risk more than the reduction in return.

Check Your Progress

1. What are the two conditions inherent in the definition of risk?
2. Define probability.

8.3 MEASURE OF RISK AND DIVERSIFICATION

- Mean absolute deviation

Types of risk measurement are discussed in three parts: techniques of measuring risk, decision rules, and finally, some more tools of measure cum decisions.

Measures of dispersion are used in measuring project risk. Some of them use probabilities, others do not. Some consider default probability equal, if probabilities are not assigned. The following statistical techniques are useful in measuring risk.

- Range

- Variance
- Semi-variance
- Standard deviation
- Co-efficient of variation

NOTES

The following section introduces all these techniques of measuring risk. While interpreting the techniques some decision rules will also be discussed for better understanding of techniques. However, those decision rules ignore the important step of incorporating risk in the project data analysis. Therefore, those decision rules only deal with mutually exclusive projects. The application of these decision rules directly to the measures of risk, therefore, is not the right way of deciding about the risky project. It is only meant for explanation.

(i) Range

Range is the difference between the highest value of outcome and the lowest value of outcome as given in Equation 8.3.

$$R_g = R_h - R_1 \tag{8.3}$$

Where, R_g = Range of a distribution

R_h = Highest value in a distribution

R_i = Lowest value in a distribution

The range does not consider probabilities, nor does it consider other possible outcomes between the highest and lowest values. Sensitivity analysis uses range values in its application.

(ii) Mean Absolute Deviation (MAD)

The term 'mean absolute deviation' is quite self-explanatory. The differences between the mean and observations is taken at absolute value (sign is ignored), they are added up, and then the sum is divided by the number of observations to get mean absolute deviation of a distribution. Equation 8.4 gives the MAD formula.

$$MAD = \sum_{i=1}^{n} | R_1 - ER | \tag{8.4}$$

Where, n = Number of observations

P_i = Probability of ith possible value

R_i = ith possible value of variable

ER = Mean (expected value) of the distribution

$R_i - ER$ = Only the absolute value is considered, negative is ignored

If probabilities are not given then Equation 8.5 should be used (it assumes in a way that each outcome has equal probability):

$$MAD = \frac{1}{n} \times \sum_{i=1}^{n} | R_1 - ER | \tag{8.5}$$

Expected return (ER) is calculated using the following Equation 8.2,

$$MAD = \sum_{i=1}^{n} R_1 \times P_1$$

NOTES

(iii) Variance

Variance measures the dispersion of data using Equation 8.6:

$$\text{Variance} = \sum_{i=1}^{n} P_i(R_i - ER)^2 \quad (8.6)$$

The difference of the observed value from the mean of distribution is squared to avoid negative sign. By squaring the difference, the values which are far away from the mean are attached more weight. Due to squaring, the variance cannot be compared with expected (mean) return.

(iv) Semi-Variance

It is same as variance but it considers R values only if $R_i < ER$ (i.e. only if there is a chance of getting less than expected result). If $R_i > ER$ then zero value is taken for R_i. Where, SV = Semi-variance

$$SV = \sum_{i=1}^{n} P_1(R_i - ER)^2 \quad (8.7)$$

$(R_i - ER)$ is considered only if it is positive and negative $(R_i - ER)$ is taken as zero.

(v) Standard Deviation (SD or σ)

Variance cannot be compared with the expected (mean) return. Therefore, its root is taken and standard deviation (SD) is calculated so that it can be compared with the return.

$$SD = \sigma = \sqrt{\text{Variance}} = \sqrt{\sum_{i=1}^{n} P_1(R_i - ER)^2} \quad (8.8)$$

Standard deviation ignores the size of the project, therefore, comparison of two projects becomes difficult.

(vi) Coefficient of Variation (CV)

The size of the project is considered in co-efficient of variation (CV).

$$CV = \frac{\sigma}{ER}$$

When standard deviation (s) is divided by expected returns (i.e., mean or *ER*), we get the coefficient of variation. It is, therefore, useful in comparing risks of two projects with different sizes. Lower coefficient of variation is considered good.[1]

Example 8.1: Risk measurement

Compute the six measures of risk for the outcome of an investment given below:

Outcome (₹)	Probability
1,100	0.2
700	0.5
600	0.3

Solution:

Range = 1,100 – 600 = ` 500

Since the probabilities are given we should use Equation

MAD = 0.2 × |1,100 – 750| + 0.5 × |1,100 – 750| + 0.3 × |1,100 – 600|

= 0.2 × 350 + 0.5 × 50 + 0.3 × 150 = 70 + 25 + 45

= 140

For the rest of the calculations the following work will be useful:

P_i	R_i	$P_i \times R_i$	$(R_i - ER)$	$(R_i - ER)^2$	$P_i\ (R_i - ER)^2$
0.2	1,100	220	350	122,500	24,500
0.5	700	350	–50	2,500	1,250
0.3	600	180	–150	22,500	6,750
ER		750		Variance	σ^2 32,500
				Standard deviation	σ 180

$$\text{Variance} = \sum_{i=1}^{n} P_i\left(R_i - ER\right)^2$$

= 0.2 × (1,100 – 750)2 + 0.5 × (700 – 750)2 + 0.3 × (600 – 750)2

= 32,500

Semi-variance = 0.2 × (1,100 0 – 750)2 = 24,500 (Using Equation 8.7)

This is upside risk because we took observation higher than *ER*. We can calculate semi- variance and determine the downside risk by taking values less than *ER*. Note that the sum of both semi-variances is equal to the variance.

Semi-variance = 0.5 × (700 - 750)^2 + 0.3 × (600 – 750)2 = 1,250 + 6,750 = 8,000

$$SD = \sigma = \sqrt{\text{Variance}} = \sqrt{\sum_{i=1}^{n} P_1\left(R_i - ER\right)^2}$$

$$= \sqrt{32,500} = 180.28 \text{ (Using variance) or}$$

$$= \sqrt{24,500} = 156.52 \text{ (Using first semi-variance)}$$

$$CV = \frac{\sigma}{ER} = \frac{180.28}{750} = 0.24 \text{ (Using variance) or}$$

$$CV = \frac{\sigma}{ER} = \frac{156.52}{750} = 0.21 \text{ (Using first semi-variance)}$$

8.3.1 Simple Diversification

Portfolio risk can be reduced by the simplest kind of diversification. Portfolio means the group of assets an investor owns. These assets may vary from stocks to different types of bonds. Sometimes, the portfolio may consist of securities from different industries. When different assets are added to the portfolio, the total risk tends to decrease. In the case of common stocks, diversification reduces the unsystematic risk or unique risk. Analysts opine that if 15 stocks are added to the portfolio of an investor, the unsystematic risk can be reduced to zero. But at the same time, if the number exceeds 15, additional risk reduction cannot be ensured. However, diversification cannot reduce systematic risk.

This naive kind of diversification is known as simple diversification. Here, securities are selected at random and no analytical procedure is used. This sort of diversification reduces risk only to a certain extent.

The total risk of the portfolio consists of systematic and unsystematic risks and is measured by the variance of the rates of return over time. Many studies have shown that systematic risk accounts for one quarter of the total risk.

Asimple random diversification reduces the total risk. Aportfolio with randomly selected 10 securities has less risk than a portfolio with three securities. The reason is that unsystematic price fluctuations are not correlated with the market's systematic fluctuations. Figure 8.3 shows how simple diversification reduces the risk. The standard deviations of the portfolios are plotted on the Y-axis and the number of randomly selected portfolio securities, on the X-axis.

The standard deviation for each portfolio is calculated and plotted. As the portfolio size increases, the total risk line starts declining. It flattens out after a certain point. Beyond that risk cannot be reduced. This implies that spreading out assets beyond a certain level cannot be expected to reduce the portfolio's total risk below the level of the undiversifiable risk.

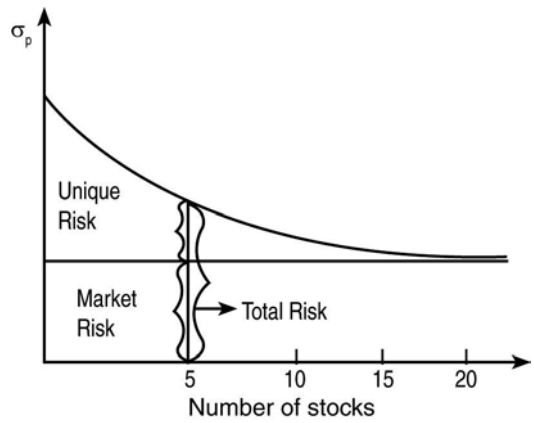


Fig. 8.3 *Diversification and Portfolio Risk*

But spreading the investment across too many assets can give rise to such problems as the inclusion of poor performers, information inadequacy, high research, and high transaction costs. These are discussed below:

Poor performers: When numerous stocks are involved, the investor may sometimes also buy stocks that will not yield adequate return.

Information inadequacy: If there are too many securities in a portfolio, it is difficult for the portfolio manager to have all information about their individual performance. He has to be well versed with the details of each company's performance. But access to such information, all at the same time, can be quite difficult.

High research costs: When a large number of stocks are included in a portfolio, the returns and risks associated with individual stocks should be analysed before their inclusion. For this, a lot of information has to be gathered and kept and this involves high costs.

High transaction costs: When small quantities of stocks are purchased frequently, the investor has to incur higher transaction costs than for the purchase of large blocks at less frequent intervals. In spite of all these difficulties, big financial institutions purchase hundreds of different stocks. Likewise, mutual funds also invest in many different stocks.

Check Your Progress

3. 'Types of risk measurement are discussed in three parts'. What are the three parts?
4. Why is standard deviation useful in comparing risks of two projects with different sizes?
5. What are the problems that arise on spreading the investment across too many assets?

NOTES

8.4 RISK IN A PORTFOLIO

If you invest all your money in shares of one company, and if that company falters, you lose your money. But if you have split your investment in shares of two different companies, the chances of losing money is reduced. You have built a portfolio of shares of two companies and diversified your investment. Likewise, companies also diversify their investment in projects. This principle is easy to understand. The diversification made logically can be more effective in improving risk-return trade-off. In this section, we will discuss the portfolio effect, types of risk and calculation of portfolio risk and portfolio return.

8.4.1 Portfolio Effect

Risk was defined as variability in the returns. Portfolio effect occurs when the variability of returns from the existing investment and that from the new investment are different. That means if the new investment is in a different risk class from existing investment, the portfolio effect will occur. Moreover, the risk-return trade-off for the portfolio may improve if the new investment has the same degree of risk, but its cash flows are negatively correlated with those of existing investments. Figures 8.4A, 8.4B and 8.4C give an idea of the portfolio effects when two investments with a particular cash flow streams combine in equal proportion.

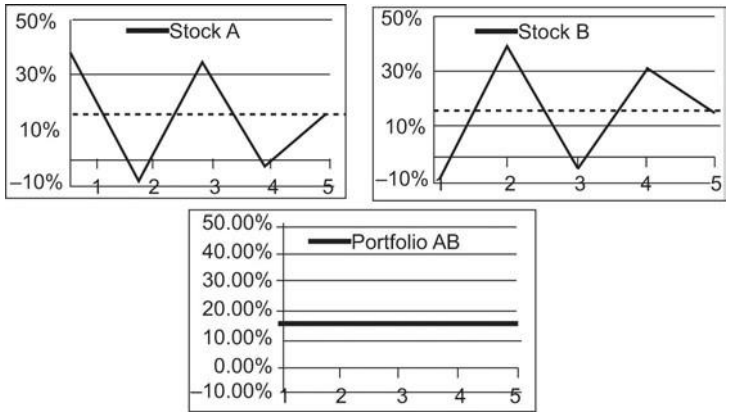


Fig. 8.4A *Portfolio Effect of Two Perfectly and Negatively Correlated Projects*

Figure 8.4A is based on shares of Company A and Company B in equal amount, which are expected to give returns over the next five years as given in the Table 8.3.

Table 8.3 *Risk and Return Profile of Stock A and Stock B and Portfolio Effect*

Year	Returns		
	Stock A	Stock B	Portfolio AB
1	38%	–8%	15.00%
2	–10%	40%	15.00%
3	35%	–5%	15.00%
4	–2%	32%	15.00%
5	15%	15%	15.00%
Expected Returns	15%	15%	15%
Standard Deviation	21.5%	21.5%	0.0%

Please note that in Table 8.3, both the shares offer equal returns (15 per cent) and carry equal risk (standard deviation of 14.9 per cent). When equal amount is invested in both, the shares returns have not changed but the risk is reduced to zero, as can be read in the Figure 8.4A.

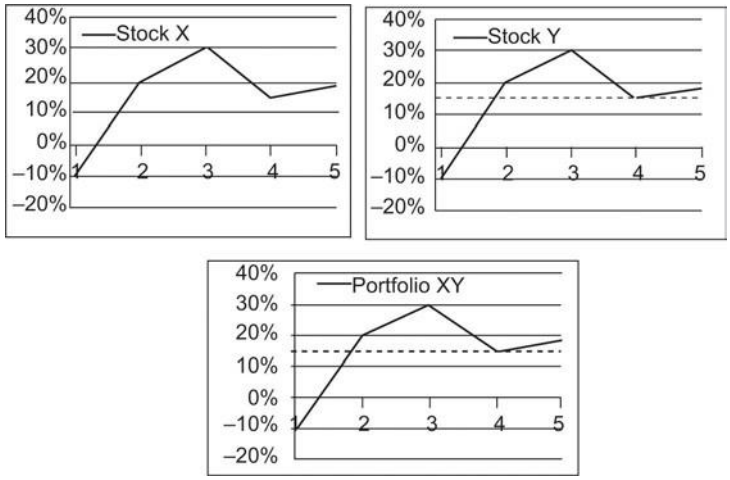


Fig. 8.4B *Portfolio Effect of Two Perfectly and Positively Correlated Projects*

Table 8.4 *Risk and Return Profile of Stock X and Stock Y and Portfolio Effect*

Year	Returns		
	Stock X	Stock Y	Portfolio XY
1	–10%	–10%	–10%
2	20%	20%	20%
3	30%	30%	30%
4	15%	15%	15%
5	18%	18%	18%
Expected Returns	15%	15%	15%
Standard Deviation	14.9%	14.9%	14.9%

Table 8.4 gives the data of Stock X and Stock Y in terms of expected returns, with equal probability over the years. A portfolio of both the shares in equal proportion has neither changed returns nor the risk. That is because the returns profile of both the shares is equal. This is observed in Figure 8.4B.

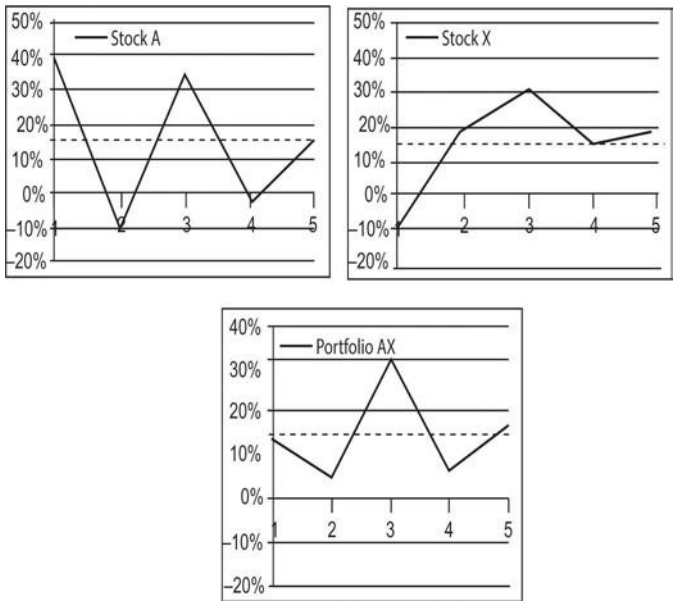


Fig. 8.4C *Portfolio Effect of Two Partially and Positively Correlated Projects*

Table 8.5 *Risk and Return Profile of Stock A and Stock X and Portfolio Effect*

Year	Returns		
	Stock A	Stock X	Portfolio AX
1	38%	−10%	14%
2	−10%	20%	5%
3	35%	30%	33%
4	-2%	15%	7%
5	15%	18%	17%
Expected Returns	15%	15%	15%
Standard Deviation	21.5%	14.9%	11.0%

Table 8.5 shows that when a portfolio of equal amount of Stock A and Stock X is prepared, the overall returns remain unchanged at 15 per cent. However, the portfolio risk is lower than even Stock X, which is lower than that of Stock A. You can see this in Figure 8.4C.

Thus, when the cash flows of two investments (Stock Aand Stock B) are perfectly and negatively correlated, the perfect portfolio effect is achieved and the risk is reduced to zero. However, when cash flows of two investments (Stock X and Stock Y) are perfectly and positively correlated there is no advantage in building a portfolio, as portfolio risk does not reduce. Stock Aand Stock X have partially correlated cash flows; therefore, when a portfolio is made of them overall risk is reduced. The visual effect of this can be observed in Figures 8.4A, 8.4B and 8.4C.

This implies that the firm can improve its risk-return profile by introducing (diversifying into) new projects in the portfolio, which have less or negative correlation with the existing projects. Unsystematic risk is, therefore, reduced.

8.4.2 Types of Risk

This section is discussed from the angle of business projects. However, one can replace the term project with securities investment and apply the same concepts and tools to investment risk. Risk can be classified into: (a) systematic risk, and (b) unsystematic risk. Figure 8.5 gives an idea of these two types of the risk.

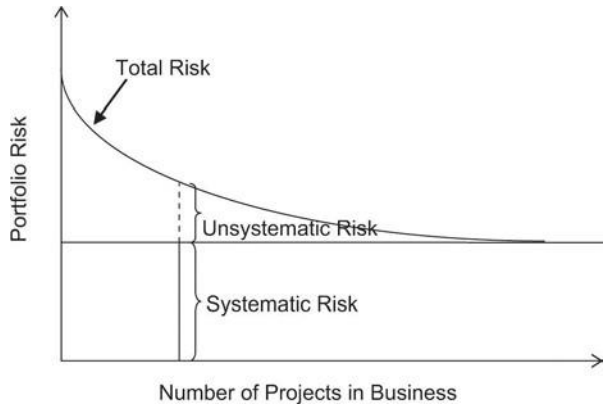


Fig. 8.5 *Effect of Diversification of Total Risk and its Components*

Acompany with a single project is exposed to external vagaries. With the addition of a new project in the business (project portfolio), the total risk may reduce. The process of risk reduction continues for some time with the acceptance of more and more projects in the portfolio. Beyond a point, the overall risk will not be reduced further. This implies that the project portfolio has become balanced. Till the portfolio of the project becomes

balanced, the business takes the total risk higher than what is minimum risk under a balanced portfolio. This higher risk is *unsystematic risk*, which the firm can diversify away. However, *systematic risk* cannot be diversified away. Product diversification and market diversification both give similar results and reduce the unsystematic risk. A company prefers to launch a product in more than one market in the market diversification programme, whereas in the product diversification strategy the same market is offered more than one products.

When it comes to investment in corporate projects, systematic risk is only a concept. Total risk can be reduced through diversification, but that can rarely touch the level of systematic risk, unlike that in security investment. Security markets are well-developed ones, where with very little cost investors can sell their investments and buy other investments, which would fit best in their portfolios. Also, in security market since unit value of investment is small one can practically invest in small fractions. The process of buying, holding and selling securities at a minimum cost and minimum time and that too in any amount enables the investors to attain a balanced portfolio.

In the project investment area, such ease, flexibility and fractional investment are not imaginable. Despite that, the concept that the risk can be reduced through well-thought diversification is useful in the corporate investment field.

Risk is also classified in different ways, business risk, financial risk and market risk. Figure 8.6 gives the pictorial interfaces among these three types of risks.

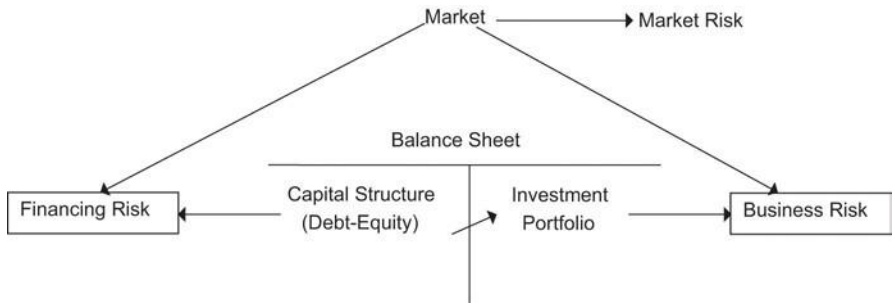


Fig. 8.6 *Interfaces among Business Risk, Financing Risk and Market Risk*

Financing risk arises out of the debt-equity structure chosen by the company. Knowledge of capital structure theory for setting an appropriate debt-equity target, and designing of securities can jointly influence the total financing risk. The variability in shareholders' net income due to the change in the net operating income (business risk) is called the financial risk.

Market risk gets reflected on the cost of individual funds for the reasons related to market factors. A perfect capital market would offer zero market risk. Imperfection in capital market observes the change in investor psychology and re-allocation of market funds over competing securities. The market process of re-allocation of funds changes the cost of capital of a company, even though the business risk and financial risk of the company may not have changed.

Business risk is that part of the total risk, which occurs due to a particular project portfolio. A diversified portfolio may have a low business risk. Business risk is a function of the product market conditions faced by a firm for all the businesses collectively, and variability of operating income due to those factors. This is also called the project-portfolio risk. The risk arising out of product market conditions and their impact on a single project is called project risk.

Business risk can be divided into two categories, *external* and *internal*. External business risk arises out of the external conditions faced by the firm due to circumstances beyond its direct control. Variation in sales quantity, variation in sales price, uncertainty of input price and their productivity, government policy and taxes, and competition are external factors causing business risk. Internal factors emanate from the efficiency of the business. Operating leverage, project design, process control, culture of organizations and such other internal factors determine the impact of variation in external factors on the operating profits of the company. Financial risk can also be divided into two categories, *external* and *internal*. Interest rate is largely caused by external factors, whereas degree of financial leverage is caused by financing decision and, therefore, it is internal risk.

The following points have emerged from the above discussion on the types of risk:

- Companies take business risk and financial risk. (Remember, all other risks, like risk due to fire and natural calamity, are insured by the companies.)
- Market risk affects all the businesses in general, and that gets reflected in financial risk through adjustment in the cost of capital. (Remember, that the debt-equity structure is a dynamic decision; the structure has to be reset for attaining the minimum overall cost of funds, whenever market risk changes.)
- Business risk contains two parts in it, systematic risk and unsystematic risk.
- Unsystematic risk represents the imbalance in the portfolio of projects.
- Unsystematic risk cannot be totally removed through diversification especially in case of corporate projects, because projects are not in perfect market conditions. But, it can surely be reduced through thoughtful diversification, without reducing average return. This is called the portfolio effect.

8.4.3 Mathematical Calculations

Portfolio effect is measured through the calculation of:

- **Calculating the Covariance**

The covariance between two projects is the product of three terms: the coefficient of correlation (r_{ij}), and the standard deviation of two projects (σ_i and σ_j), as expressed in the following equation:

$$COV_{i,j} = \sigma_{ij}\,\sigma_i\,\sigma_j \qquad (8.9)$$

The correlation coefficient measures the nature and strength of the relationship between the two projects, and it takes value in the following range,

$$-1.0 \le \sigma_{ij} \le 1.0 \qquad (8.10)$$

- **Calculating the Portfolio Return**

The expected return of the portfolio is the weighted average returns of all projects included in the portfolio. The following formula is used for the calculation of portfolio returns:

$$ER_p = \sum_{j=1}^{n} X_j ER_j$$

Where, X_j = Proportion of project j in the portfolio
ER_j = Expected rate of return on project *j*.

- **Calculating the Portfolio Risk**

The portfolio risk, denoted by standard deviation, is not the weighted average of the standard deviations of all individual projects included in the portfolio.

Portfolio risk = Risk of projects in isolation + Manner in which projects interrelate.

$$\sigma_p - \sqrt{\sum_{j=1}^{n} X_j^2 \sigma_j^2 + 2\sum_{j=1}^{n-1} \sum_{i=j+1}^{n} X_i X_j Cov_{ij}} \qquad (8.11)$$

Where, *i* and *j* represent all projects in the portfolio, paired off for the purpose of computing covariance.

If only two projects are considered in the portfolio then Equation 8.9 can be used for calculating covariance; and it can be incorporated in Equation 8.11 to get Equation 8.12 given below for the calculation of portfolio risk:

$$\sigma_p - \sqrt{\sum_{j=1}^{n} X_j^2 \sigma_j^2 + 2\sum_{j=1}^{n-1} \sum_{i=j+1}^{n} X_i X_j \sigma_{ij} \sigma_i \sigma_j} \qquad (8.12)$$

Example 8.2: A firm's existing project's data and proposed project's data are given below:

	Existing Projects	Proposed Project
Expected return (ER_j)	30%	26%
Risk (σ_j)	7%	4%
Proportion (X_j)	0.5	05
Correlation Coefficient (ρ_{ij})	–0.27	

Compute the portfolio return and the portfolio risk.

Solution:

$$ER_p = \sum_{j=1}^{n} X_j ER_j = (0.5 \times 0.3) + (0.5 \times 0.26) = 28\%$$

$$\sigma_p = \sqrt{\sum_{j=1}^{n} X_j^2 \sigma_j^2 + 2\sum_{j=1}^{n-1} \sum_{i=j+1}^{n} X_i X_j \sigma_{ij} \sigma_i \sigma_j}$$

$$\sigma_p = \sqrt{0.5^2 \times 0.07^2 + 0.5^2 \times 0.04^2 + 2 \times 0.5 \times 0.5 \times (-0.27) \times 0.07 \times 0.04}$$

$$\therefore \quad \sigma_p = \sqrt{0.001247} = 3.53\%$$

Notice that the portfolio risk is lower than the risk of any individual project, whereas portfolio return is the weighted average return. The small negative correlation between the two projects has changed the risk-return trade-off of the portfolio.

8.5 COST OF CAPITAL

Firms borrow money and also take owners' money. The lenders and owners both have to be rewarded for sacrificing the purchasing power of money and must be reimbursed for the inflation, tax and risk they take. The lenders of money will charge interest, which is the cost of borrowing for a firm. The owners expect returns on their investment,

Check Your Progress

6. When does portfolio effect occur?
7. How can risk be classified?
8. What is business risk?

which is the cost of equity for a firm. The risk of a firm emanates from the volatility of its profit, which depends on the variability of sales revenue and firm's investment and financing decisions. In this section, we will discuss theories of cost of capital and then the triangular relations among the investment decisions, financing decisions and intervention of market factors.

8.5.1 Capital Structure Decision

How to finance the total requirement of funds? This is a question of capital structure. The capital structure decision is the decision regarding the proportion of ownership funds and borrowed funds in total funds needed. What should be the size of debt vis-à-vis equity? Does it matter in managing the cost of capital? Do they affect each other? Before we address these issues we need to visit a small but important point that emerges from the different terms 'capital structure' and 'financial structure'.

Capital Structure vs Financial Structure Decision

Should we include the current liabilities and short-term debt when we consider debt and equity for calculating the cost of capital? If only the long term debt is included and compared with the ownership funds, then it is called a capital structure. When total borrowed funds (including short-term debt) are compared with the ownership funds, it is known as financial structure. Current liabilities are excluded in capital structure as well as in financial structure. Should we take a capital structure or a financial structure for the calculation of cost of funds? A pragmatic answer is needed for this question.

The types of borrowed money in the question here include: (a) current liabilities, (b) short-term loans and (c) long-term loans. The current liabilities are generated in a normal course of business operations. Accounts payables happen because we buy resources and enjoy business-customary credit. Unpaid wages are normal because of wage-payment cycle. Unpaid bills are also the results of business practices, like electricity bill is paid every month. These current liabilities are effortless and cost-free financing of the current assets. Since it is a result of routine operations, it should not be counted as borrowing when we calculate cost of debt.

The firms borrow short-term funds from various sources like overdraft, cash credit, commercial papers and others. The short-term funds are needed for funding uneven operations due to seasonality and other factors during the year. Thus, the cost of short-term borrowing is an operating cost. Interest expense on short-term funds is estimated and included in the operating cash flow of a project. Business practices are also on the same line. The short-term borrowing should not be included in the calculation of cost of capital. Only the long-term borrowings and equity are included in the calculation of cost of capital. Thus, the capital structure, and not the financial structure, is an appropriate consideration in the cost of capital theories.

8.5.2 Capital Structure Theories

Does it really matter how the capital is raised, either through debt or equity? Divergent views are expressed. They can be summarized in three approaches of capitalization; namely, net operating income (NOI) approach, net income (NI) approach, and traditional approach. The traditional approach will be dealt with in the next section. These three theories of capital structure are briefly explained because it is necessary for building blocks required for developing an understanding of the cut-off decision for corporate projects.

Net Operating Income (NOI) Approach

This approach is based on the hypothesis that if two firms have earned the same amount of operating income (EBIT) and if they are in the same risk class, then the value of both the firms will be equal irrespective of the debt-equity ratio of each firm. This implies that the overall cost of funds (K_o) remains constant at all debt-equity proportions, as depicted in Figure 8.7.

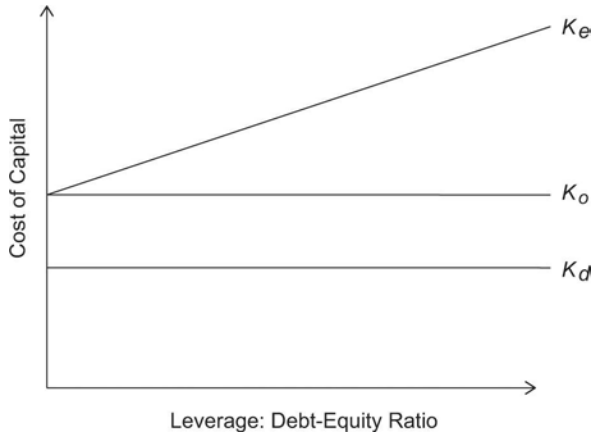


Fig. 8.7 *Net Operating Income (NOI) Approach*

Initial propositions of Modigliani and Miller (M&M) supported this view with the help of series of assumptions. No tax, no transaction cost and personal leverage as perfect substitutes of corporate leverage are important assumptions in M&M's proposition. According to this approach, the cost of equity will increase just enough to consume the savings due to the higher amount of low cost debt and overall cost of funds will remain constant. The adjustment in the cost of equity would take place through a process of arbitraging by the investors.

Net Income (NI) Approach

According to the NI approach, the cost of debt and the cost of equity, both remain unchanged for any level of debt-equity ratio. This leads to a corner solution, because with the increase in debt K_O falls, and that process continues till K_O equals K_d. Almost all debt in capital structure proves to be ideal as per the net income approach as can be seen in the Figure 8.8.

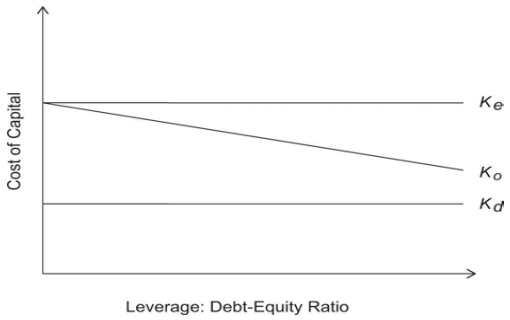


Fig. 8.8 *Net Income (NI) Approach*

NI approach introduces tax in the theory. Interest paid on debt is tax exempt. Therefore, post-tax Kd is low, giving the true advantage of leverage to the equity

leverage advantage due to tax ensures that the equity cost remains constant and K_o declines. Shareholder wealth, therefore, increases with increase in the debt-equity ratio.

Traditional Approach

Many realistic factors have culminated in the pragmatic theory of capital structure called the traditional approach of capital structure. The net operating income approach and the net income approach both prove that debt-equity decision is a redundant decision. The NOI approach clearly proves the irrelevance of capital structure decision, and the NI approach favours almost infinite debt-equity structure. The traditional approach, however, establishes that the capital structure is an active decision variable, because there exists an optimum debt-equity ratio, very unique to a company, at which the overall cost of funds is minimum. Figure 8.9 graphically depicts this approach.

When the debt-equity ratio is low, a small increase in debt will not increase the cost of debt; and cost of equity may either remain the same or decline slightly, because the debt gives leverage advantage together with the tax shelter. Beyond a particular point of debt-equity structure the cost of debt starts rising, and so does the cost of equity. Chances of not getting timely return (default risk) increases beyond a point of debt-equity ratio, and hence investors expect a higher return. The rate of increase in cost of funds suddenly shows a steep rising tendency at a very high debt-equity ratio. Chances of default and chances of bankruptcy are reflected in the steep rise in the cost of funds. Consequently, K_o which is equal to K_e at zero debt structure, initially declines with the increase in debt-equity ratio, and after some debt-equity ratio starts rising. Tax advantage, initial positive effect of leverage and subsequent chance of bankruptcy bring a situation that at some point an optimum capital structure is experienced, and Ko is lowest. Debt-equity ratio is ideal between points D/E1 and D/E2 in Figure 8.9, as that will cost the least. Each firm must identify its unique optimum capital structure and maintain it. Debt-equity structure thus, is an active decision variable.

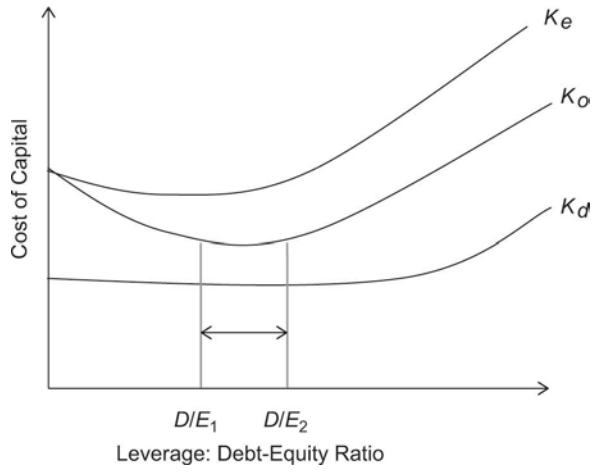


Fig. 8.9 *Traditional Approach*

Signalling Theory

The traditional approach of capital structure does not precisely determine the optimum capital structure. One would have at the most a judgment about the optimal level within

shareholders. Net

a range that can be understood from the traditional theory. Because the optimum structure cannot be determined precisely, market looks for signals from the company, which leads to the development of signalling theory of capital structure.

Managerial actions send signals to the market and investors interpret the managerial actions to judge about the valuation of firm and quality of capital structure. Firms like to keep some borrowing capacity in reserve by keeping a lower debt ratio and retaining profits, so that it can access cheaper funds when growth opportunities come up. Thus, retained profit sends a signal to the market about optimum capital structure. Of course, this will hold true in case of symmetric information. In asymmetric information scenario wherein managers have more and better information, the signalling theory may not be very effective.

8.5.3 Views about Capital Structure

How frequently must capital structure decision be reviewed? Three views are expressed regarding the capital structure as a decision variable.

- **Static view:** The capital structure, once decided for a minimum cost of capital, remains static forever, according to this view. This view assumes that the quality of earnings and capital market conditions are not variable affecting the cost of capital. This may be true in a period of no competition and primitive financial market.
- **Comparative static view:** This is based on the assumptions that some of the underlying factors might change rendering the earlier choice of debt-equity sub-optimal. Debt-equity needs to be adjusted every time a change in underlying factor (say, tax, inflation, market allocation etc.) is experienced. But such changes are only occasional.
- **Dynamic view:** The quality of earnings, capital market conditions, incidence of tax, and inflationary conditions are constantly changing variables. The capital structure should be viewed as the by-product of these changes. The firm must constantly adjust its debt-equity proportion, in the light of changes in these variables, and aim at constantly keeping the low cost of capital. Thus, debt-equity is a very dynamic decision variable.

Very competitive product market and efficient capital market prove the appropriateness of the dynamic view of capital structure. Finance managers are expected to constantly review the capital structure of a firm in the light of dynamically changing external economic factors like money supply, tax and inflation; and changing business risk profile with the constant pursuit of growth.

8.5.4 Triangular Interface

The discussion on the pragmatic capital structure theory aptly proved that the cost of capital is a function of debt-equity ratio. The quality of earnings and market factors affect it too. The firms raise funds for investment in projects, and projects as well as debt-equity ratio together with the variability of sales revenue, create risk-profile of business, which in turn has the effect on the cost of capital and thereby on project profitability. All these factors are inter-dependent, and they create a triangular relationship. Figure 8.10 depicts these interfaces among these factors, which collectively affect the overall cost of funds and value creating potential for a business.

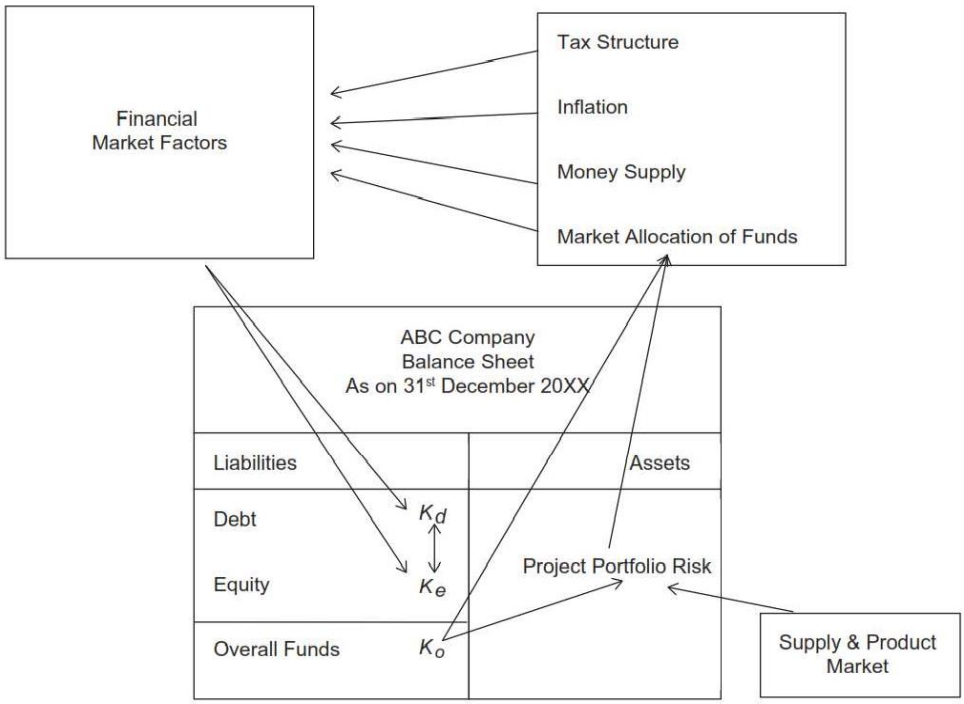


Fig. 8.10 *Triangular Interface and Cost of Capital*

These sources of risk can be divided into internal company risk and market risk. Company risk is further divided into: (a) business risk and (b) financial risk. Business risk emanates from uncertainty of sales revenue and variability of expenses. They together create variability in profits which is also caused by the degree of operating leverage or investment decisions of the firm.

Market factors comprise the tax structure in the economy, inflation, money supply, and allocation of money on the securities of various companies. Decrease (increase) in tax rate increases (decreases) the cost of funds. Increase (decrease) in the inflation increases (decreases) the cost of capital. The market has limited amount of funds, for allocation on various market securities. Tax is a fiscal policy of government. The central bank (Reserve Bank of India) attempts to control inflation and money supply in the economy. If money supply increases the cost of capital reduces. Financial market rationally and competitively allocates available money supply over the market securities and other demands for money.

The firms take calculated risk and make investments in business projects. When a firm adds new projects risk profile (variability of income) of business change, forcing market to reallocate funds over the market securities and other demands for money. The risk profile of a business depends on supply markets and product markets. If the markets for resources bought by a firm are volatile, and if markets for products produced by the firm are volatile, the firm will be riskier. The financial market will allocate fewer funds for the securities of riskier firms forcing the cost of capital go up. Investors would park their money elsewhere. Investors' appetite for risk change with changing market conditions, and they have several choices available to them. This makes imperative on the part of a firm to constantly adjust its variables to take risk as much market wants. The market factors are very dynamic in a competitive economy, making the interfaces very dynamic, and as a result making financing decision a very dynamic variable.

This proves that not only the traditional theory of capital structure is evident but it is also very dynamic. The ideal debt-equity ratio constantly changes and firms need to chase it constantly, so that the cost of capital can remain low, make projects viable and create value for shareholders. The traditional approach gives a conceptual understanding, but is not helpful in determining the ideal debt-equity ratio.

8.6 TRADITIONAL THEORY OF COST AND MODIGLIANI-MILLER THEOREM

According to this approach a firm's cost of capital depends upon the method and level of financing or its capital structure. A firm can, therefore, change its overall cost of capital by increasing or decreasing the debt-equity mix. For example, if a company has 9 per cent debentures (issued and payable at par) the cost of funds raised from this source comes to only 4.5 per cent (assuming a 50 per cent tax rate). Funds from other sources, such as equity shares and preference shares, also involve cost. But the raising of funds through debentures is cheaper because of the following reasons:

(i) Interest rates are usually lower than dividend rates.

(ii) Interest is allowed as an expense resulting in a tax benefit while dividend is not allowed as an expense while computing taxable profits of the company.

The traditionalist theorists, therefore, argue that the weighted average cost of capital will decrease with every increase in the debt content in the total capital employed. However, the debt content in the total capital employed should be maintained at a proper level because cost of debt is a fixed burden on the profits of the company. It may have adverse consequences in periods when a company has low profitability. Moreover, if the debt content is raised beyond a particular point, the investors will start considering the company too risky and their expectations from equity shares will go up.

8.6.1 Modigliani-Miller Approach

According to this approach, the corporation's total cost of capital is constant and it is independent of the method and level of financing. In other words, according to this approach a change in the debt-equity ratio does not affect the total cost capital. According to traditional approach, as explained above, the cost of capital is the weighted average cost of the debt and the cost of equity. Each change in the debt-equity ratio automatically offsets change in one with the change in the other on account of change in the expectation of equity shareholders. For example, the capital structure of a company is as follows:

9% Debentures	` 1,00.000
Equity Share Capital	1,00,000

The company has at present even debt-equity ratio. It has been paying dividend at the rate of 12 per cent on equity shares. In case, the debt-equity ratio changes to say 60 per cent debt and 40 per cent equity, the following consequences will follow:

- The debt being cheaper, the overall cost of capital will come down.
- The expectation of the equity shareholders from present dividend of 12 per cent, will go up because they will find the company now more risky.

Thus, the overall cost of capital of the company will not be affected by change in the debt-equity ratio. Modigliani and Miller, therefore, argue that within the same risk class, mere change of debt-equity ratio does not affect cost of capital. Their following

Check Your Progress

9. What is capital structure decision?
10. What are the sources for the firms' borrowing of short-term funds?
11. What does the static view express regarding the capital structure as a decision variable?

observations in their article, 'Cost of Capital, Corporation Finance and Theory of Investment', need careful consideration:

(i) The total market value of the firm and its cost of capital are independent of its capital structure. The total market value of the firm can be computed by capitalizing the expected stream of operating earnings at discount rate considered appropriate for its risk class.

(ii) The cut-off rate for investment purposes is completely independent of the way in which investment is financed.

8.6.2 Assumptions under Modigliani-Miller Approach

The Modigliani-Miller Approach is subject to the following assumptions:

(i) **Perfect capital market:** The securities are traded in perfect capital markets. This implies that:

(a) The investors are free to buy or sell securities.

(b) The investors are completely knowledgeable and rational persons. All information and changes in conditions are known to them immediately.

(c) The purchase and sale of securities involve no costs such as broker's commission, transfer, fees, etc.

(d) The investors can borrow against securities without restrictions on the same terms and conditions as the firms can.

(ii) **Firms can be grouped in homogeneous risk classes:** Firms should be considered to belong to a homogeneous class if their expected earnings have identical risk characteristics. In other words, all firms can be categorized according to the return that they give and a firm in each class is having the same degree of business and financial risk.

(iii) **Same expectation:** All investors have the same expectation of firm's net operating income (EBFI) which is used for evaluation of a firm. There is 100 per cent dividend pay out, i.e., to firms distribute all of their net earnings to the shareholders.

(iv) **No corporate taxes:** In the original formulation Modigliani and Miller hypothesis assumes that there are no corporate taxes. This assumption has been removed later.

MM's Proposition I: Based on the above assumptions, the MM's first proposition is as under:

'The value of a firm is equal to its expected operating income divided by the discount rate appropriate to its risk class. It is independent of its capital structure.' Symbolically, it may be put as under:

$$V = D + E$$

Or, $V = O/r$

Where, V = market value of the firm

D = market value of debt

E = market value of equity

O = expected operating income

r = discount rate applicable to the risk class to which the firm belongs.

MM's Proposition II: According to the MM's first proposition, an increase in the financial leverage increases the expected earnings per share but not the price of the

share. This is because, as discussed above the expected earnings is offset by a corresponding change in the return expected by the shareholders. The expected return can be computed as under:

$$\text{Expected return on assets} = r_a = \frac{\text{Expected Operating Income}}{\text{Market Value of all Securities}}$$

NOTES

In other words, the expected return by an investor on his portfolio is equal to the weighted average of the expected return on his individual securities. Symbolically this can be put as under:

$$r_a = \left[\frac{D}{D+E}\right] r_d + \left[\frac{D}{D+E}\right] r_e$$

Expected return on assets	=	Proportion of debt	×	Expected return on debt	+	Proportion of equity	×	Expected return on equity

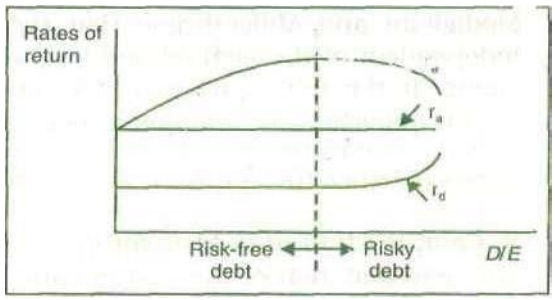
Rearranging this equation, we get:

r_e	=	r_a	+	$(r_a - r_d)$	(D/E)

Expected return on equity	=	Expected return on assets	×	Expected return on assets	–	Expected return on debt	×	Debt-equity return on

According to the second proposition, the expected return on equity is equal to, the expected rate of return on assets, plus a premium. The premium is equal to the debt-equity ratio times the difference between the expected return on assets and the expected return on debt, in case the firm has low levels of debt, it is considered risk free. As the debt of the firm crosses the threshold limit, the risk of default increases and the expected return on debt also rises. To compensate this, the expected return on equity decreases. In other words, at higher level of debts, return on equity become less sensitive to further borrowings. This is because the debt-equity ratio increases beyond the threshold level, a portion of the firm's business risk is borne by the suppliers of debt capital. The more the firm borrows, the more of its business risk is shifted from the shareholders to the creditors.

The relationship between the cost of equity and financial leverage in accordance with the MM Proposition II can be shown by the following diagram:



The above diagram shows that the return on equity goes on increasing up to the threshold limit with more debt content. But after the threshold limit, the expected return *on* equity declines.

Similarly, up to the threshold limit the debt content in the capital structure does not affect the expected return on debt but after the threshold limit, the expected return on debt increases. In other words, lower expected return on equity is compensated by a higher expectation on debt.

NOTES

In conclusion, it may be said that in spite of the correctness of the basic reasoning of Modigliani and Miller, the traditional approach is more realistic on account of the following reasons:

(i) The corporations are subject to income-tax and, therefore, due to tax effect, the cost of debt is lower than cost of equity capital.

(ii) The basic assumption of Modigliani and Miller hypothesis that capital markets are perfect, is seldom true.

On account of the above reasons the Modigliani and Miller approach has come under severe criticism. Mr, Ezra Soloman has observed, 'The thesis that the company's cost of capital is independent of its financial structure is not valid. As far as the leverage effect alone is concerned (and ignoring all other considerations that might affect the choice between debt and equity) there does exist a clearly definable optimum position-namely, the point at which the marginal cost of more debt is equal to or greater than a company's average cost of capital.' However, Mr. E. W. Walker's remarks very aptly explains the utility of Modigliani and Miller's approach. According to him, 'the criticisms lodged against Modigliani and Miller's thesis are valid thus limiting its use in actual situations. Nevertheless, the propositions as well as their criticisms should be carefully studied, since they will serve as an aid to understanding capital structure theory.'

8.7 CAPITAL ASSET PRICING MODEL

Jack Treynor (1961), William Sharpe (1964), John Lintner (1965) and Jan Mossin (1965) developed this model independently. The theory was based on the work of Harry Markowitz. It is a model of linear general equilibrium return. In CAPM theory, the required rate return of an asset has a linear relationship with an asset's beta value i.e., its undiversifiable or systematic risk.

Assumptions of CAPM

CAPM is based on the following assumptions:

- An individual seller or buyer cannot affect the price of a stock. This is the basic assumption of a perfectly competitive market.
- Investors make their decisions only on the basis of the expected returns, standard deviations and covariance of all pairs of securities.
- Investors have homogenous expectations during the decision-making period.
- An investor can lend or borrow any amount of funds at the risk free rate of interest. This rate of interest is the one offered for treasury bills or government securities.
- Assets are infinitely divisible. Thus, an investor can buy shares in any amount i.e., an investor can even buy ` 10 worth of Reliance Industry shares.
- There is no transaction cost i.e., no cost is involved in the buying and selling of stocks.
- There is no personal income tax. Hence, the investor is indifferent to the form of return, whether it accrues as a capital gain or dividend.
- Unlimited quantum of short sales is allowed. An individual can sell any amount of shares short.

Check Your Progress

12. Why are raising of funds through debentures cheaper?
13. What do the traditionalist theorists argue with regard to the cost of capital?
14. Why is the traditional approach more realistic in spite of the correctness of the basic reasoning of Modigliani and Miller?

NOTES

Lending and Borrowing

CAPM assumes that the investor can borrow or lend any amount of money at a risk-free rate of interest. When this opportunity is given to investors, they can mix risk-free assets with risky ones in a portfolio for the desired rate of risk-return combination.

R_p = portfolio return

X_f = proportion of funds invested in risk-free assets

$1\text{-}X_f$ = proportion of funds invested in risky assets

R_f = risk-free rate of return

R_m = return on risky assets

The expected return on the combination of risky and risk-free assets is:

$$R_p = R_f X_f + R_m(1 - X_f)$$

This formula can be used to calculate the expected returns in different situations, like mixing risk-free assets with risky ones, investing only in the risky asset, or using borrowed money for the risky assets.

Now let us assume that borrowing and lending rates are 12.5 per cent and the return from the risky asset is 20 per cent. There is a trade-off between the expected return and risk. If an investor invests in risk free assets and risky assets, his risk may be less than what he invests in the risky asset alone. If he borrows to invest in risky assets, his risk will increase more than he invests his own money in the risky assets. When he borrows to invest, we call it financial leverage. If he invests 50 per cent in risk-free assets and 50 per cent in risky assets, his expected return of the portfolio would be:

$$R_p = R_f X_f + R_m (1 - X_f)$$
$$= 12.5 \times 0.5 + 20 (1 - 0.5)$$
$$= 6.25 + 10$$
$$= 16.25\%$$

If there is zero investment in risk-free assets and 100 per cent in risky ones, the return is:

$$R_p = R_f X_f + R_m (1 - X_f)$$
$$= 0 + 20\%$$
$$= 20\%$$

If –0.5 is in the risk-free asset, and 1.5 in the risky asset, the return is:

$$R_p = R_f X_f + R_m (1 - X_f)$$
$$= (12.5 \times -0.5) + 20 \times 1.5$$
$$= -6.25 + 30$$
$$= 23.75$$

The variance of the above portfolio can be calculated by using the equation:

$$\sigma^2_p = \sigma^2_f X_f^2 + \sigma^2_m (1 - X_f)^2 + 2\, Cov_{fm} X_f (1 - x_f)$$

The previous example can be taken for the calculation of the variance. The variance of the risk-free asset is zero. The variance of the risky asset is assumed to be 15. Since the variance of the risk-free asset is zero, the portfolio risk depends solely on the portion of investment in the risky asset.

NOTES

Table 8.6 *A Sample Portfolio*

Proportion in the risky asset ($1 - X_f$)	Portfolio risk (in %)
0.5	7.5
1.0	15.0
1.5	22.5

The risk, at 22.5 per cent, is more in the borrowing portfolio, but the return is also high. In the portfolio with lending, the risk is 7.5 per cent, but the return is also the lowest. The risk premium is in proportion to the risk. The risk premium of a portfolio is the difference between R_p and R_f i.e., the amount by which a risky rate of return exceeds the risk-free rate of return.

Table 8.7 *The Risk-Return Tradeoff*

Portfolio return	Risk-free return	Risk premium	Portfolio risk	Factor of proportionality
R_p	R_f	$R_p - R_f$	σ_p	$(R_p - R_f)/\sigma_p$
16.25	12.5	3.75	7.5	0.5
20.0	12.5	7.5	15.0	0.5
23.75	12.5	11.25	22.5	0.5

The risk-return proportionality ratio remains constant at 0.5. It shows that one unit of risk premium is accompanied by 0.5 units of risk.

8.7.1 Basic Concept of CAPM

According to CAPM, all investors hold only a market portfolio and risk-free securities. The market portfolio comprises all stocks in the market. Each asset is held in proportion to its market value as compared to the total value of all risky assets. For example, if Reliance shares represent 20 per cent of all risky assets, then the market portfolio of the individual investor will have 20 per cent of Reliance shares. At this stage, the investor has the ability to borrow or lend any amount of money at the risk-free rate of interest. The efficient frontier of the investor is given in Figure 8.11. The investor prefers any point between B and C because with the same level of risk that he faces on line BA, he can earn higher profits. The ABC line shows the investor's portfolio of risky assets. He can include risk-free assets either by lending or borrowing. This is shown in Figure 8.12.

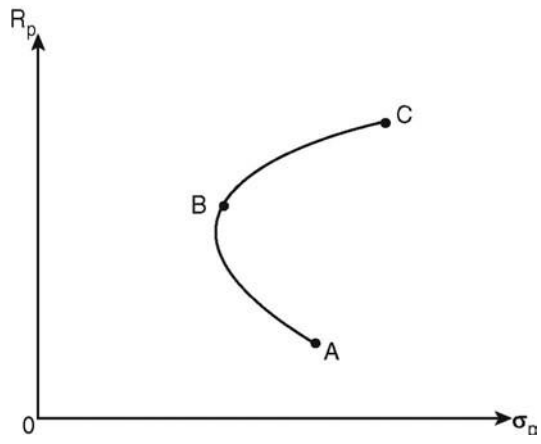


Fig. 8.11 *Efficient Frontier*

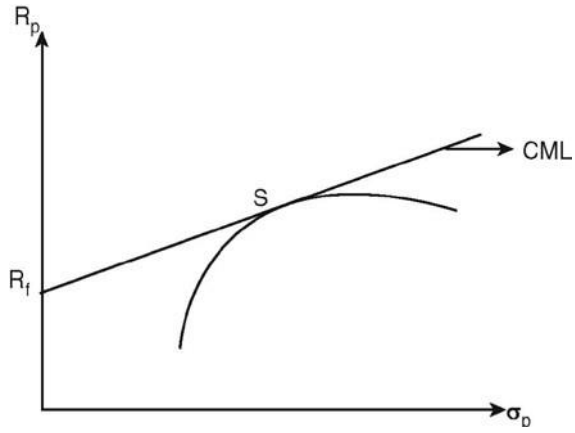


Fig. 8.12 *Capital Market Line*

The line R_fS represents all possible combinations of risk-free and risky assets. The S portfolio does not represent any risk-free asset, but the line R_fS gives the combination of both. The portfolio on the R_fS line in Figure 8.12 is the lending portfolio i.e., some money may be invested in the risk-free asset or may be deposited in the bank for a fixed rate of interest. If it crosses the point S, it becomes a borrowing portfolio. Money is borrowed and invested in the risky asset. The straight line is called capital market line (CML). It represents the desirable set of investment opportunities between risk-free and risky investments. The CML represents a linear relationship between the required rates of return for efficient portfolios and their standard deviations.

$$E(R_p) = R_f + \left(\frac{R_m - R_f}{\sigma_m}\right)\sigma_p$$

$E(R_p)$ = portfolio's expected rate of return

R_m = expected return on market portfolio

σ_m = standard deviation of a market portfolio

σ_p = standard deviation of the portfolio

For a portfolio on the capital market line, the expected rate of return in excess of the risk-free rate is in proportion to the standard deviation of the market portfolio. The price of the risk is given by the slope of the line. The slope equals the risk premium for the market portfolio $R_m - R_f$ divided by the risk or standard deviation of the market portfolio. Thus, the expected return of an efficient portfolio is,

Expected return = Price of time + (Price of risk × Amount of risk)

Price of time is the risk-free rate of return. Price of risk is the premium amount over and above the risk-free return.

8.7.2 Security Market Line (SML)

The risk-return relationship of an efficient portfolio is measured by the capital market line. It does not show the risk-return tradeoff for other portfolios and individual securities. Inefficient portfolios lie below the capital market line and the risk-return relationship cannot be established with the help of the capital market line. Standard deviation includes systematic and unsystematic risk. Unsystematic risk that can be diversified is not related to the market. If the unsystematic risk is eliminated, then systematic risk is alone of concern. This systematic risk can be measured by beta. Beta analysis is useful for

individual securities and portfolios whether efficient or inefficient.

When a security is added to the market portfolio, risk is also added to it. The variance of a portfolio is equal to the weighted sum of the covariances of the individual securities in the portfolio. If we add a security to the market portfolio, its marginal contribution to the variance of the market is the covariance between the security's return and a market portfolio's return. If the security is included, the covariance between the security and the market measures the risk. Covariance can be standardized by dividing it by the standard deviation of a market portfolio (Cov_{im}/σ_m). This shows the systematic risk of the security. Then, the expected return of the security is given by the equation:

$$R_i - R_f = \frac{R_m - R_f}{\sigma_m} Cov_{im} / \sigma_m$$

This equation can be rewritten as follows.

$$R_i - R_f = \frac{Cov_{im}}{\sigma_m^2}[R_m - R_f]$$

The first term of the equation is nothing but the beta coefficient of the stock. The beta coefficient of the equation of security market line (SML) is the same as the beta of the market (single index) model. In equilibrium, all efficient and inefficient portfolios lie along the security market line. The SML line helps to determine the expected return for a given security beta. In other words, when betas are given, we can generate expected returns for the given securities. This is explained in Figure 8.13.

If we assume the expected market risk premium to be 8 per cent and the risk-free rate of return to be 7 per cent, we can calculate expected return for A, B, C and D securities using the formula:

$E(R_i) = R_f + ß_i [E(R_m) - R_f]$

If ß = 1

= 7 + 1 (8)

= 15%

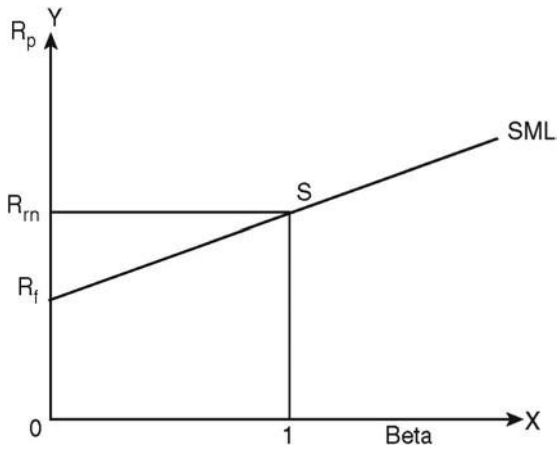


Fig. 8.13 *Security Market Line*

Security A

Beta = 1.10

E (R) = 7 + 1.10(8)

= 15.8

Security B

Beta = 1.20

E (R) = 7 + 1.20(8)

= 16.6

Security C

Beta = 0.7

E (R) = 7 + 0.7(8)

= 12.6

The same can be found out easily from the Figure 8.13 too. All we have to do is mark the beta on the horizontal axis and draw a vertical line from the relevant point to touch the SML line. Then, from the intersection, draw another horizontal line to touch the Y-axis. The expected return could be very easily read from the Y-axis. The securities A and B are aggressive securities, because their beta values are greater than one. When beta values are less than one, they are known as defensive securities. In our example, security C has a beta value of less than one.

Evaluation of Securities

The relative attractiveness of a security can be found with the help of the security market line. Stocks with a high risk factor are expected to yield more return and vice-versa. The investor is interested in knowing whether the security offers a return that is greater or less than proportional to its risk.

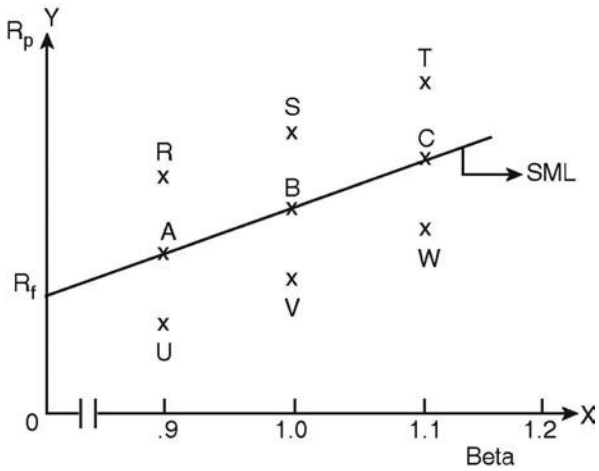


Fig. 8.14 *Evaluation of Securities with SML*

Figure 8.14 provides an explanation for the evaluation. There are nine points in the diagram. A, B and C lie on the SML, R, S and T above it and U, V and W below the SML. A, R and U have the same beta of 0.9. Likewise, beta values of S, B and V are 1 and that of T, C and W is 1.1. The stocks lying above the SML yield higher returns for

NOTES

the same level of risk. They are underpriced compared to their beta values. We can prove that they are undervalued by using the simple rate of return formula:

$$R_i - \frac{P_i - P_0 + Div}{P_0}$$

P_i is the current price, P_o, the purchase price and the dividend. When the purchase price is low, i.e., when the denominator value is low, the expected return could be high. Applying the same principle, the stocks U, V and W can be classified as overvalued securities and are expected to yield lower returns than stocks of comparable risk. When the denominator value is high i.e., the purchase price is high, the prices of these stocks may fall and lower the denominator. This will then increase the securities' returns.

The securities A, B and C are on the line. This means they are appropriately valued. They offer returns in proportion to their risk. They have average stock performance as they are neither undervalued nor overvalued.

Market Imperfections and SML

Information regarding the share price and market conditions may not be immediately available to all investors. Imperfect information may affect the valuation of securities. In a market with perfect information, all securities should lie on the SML. Market imperfections would lead to a band of SML rather than a single line. They affect the width of the SML band. If there are more imperfections, the width will also be more. Figure 8.15 shows an SML in an imperfect market.

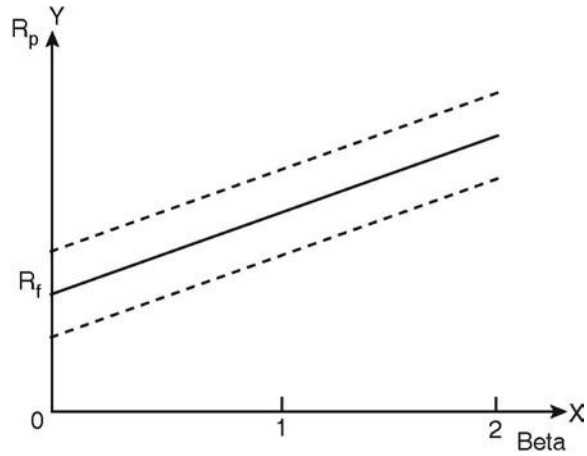


Fig. 8.15 *SML in an Imperfect Market*

8.7.3 Empirical Tests of the CAPM

In the CAPM, beta is used to estimate the systematic risk of a security and reflects the future volatility of the stock in relation to the market. Future volatility of the stock is estimated only through historical data. Historical data are used to plot the regression line or the characteristic line and calculate the beta value. If historical betas are stable over a period, they are a good proxy for their ex-ante or expected risk.

Robert A. Levy (1971), Marshall E. Blume (1975) and others have studied the question of beta stability intensively. Levy calculated betas for both individual securities and portfolios. The main conclusions of his study were as follows:

- The betas of individual stocks are unstable; hence past betas are not good estimators of future risk.
- The betas of portfolios of ten or more randomly selected stocks are reasonably stable; hence, the betas of past portfolios are good estimators of future portfolio volatility. This is because errors in the estimates of betas of individual securities tend to offset one another in a portfolio.

Various researchers have attempted to find out the validity of the model by calculating the beta and realized rate of return. They have attempted to test: (1) whether the intercept is equal to R_f, i.e., the risk-free rate of interest or the interest rate offered for treasury bills and (2) whether the line is linear and passes through beta = 1, which is the required rate of return of the market. In general, the studies have produced the following results.

- A significant positive relationship between the expected return and systematic risk. The slope of the relationship is usually less than that predicted by the CAPM.
- The risk and return relationship appears to be linear. Empirical studies give no evidence of significant curvature in the risk/return relationship.
- The CAPM theory implies that unsystematic risk is not relevant, but unsystematic and systematic risks are positively related to security returns. Higher returns are needed to compensate both the risks. Most of the observed relationship reflects statistical problems rather than the true nature of the capital market.
- According to Richard Roll (1977), the ambiguity of the market portfolio leaves the CAPM untestable. The practice of using indices as proxies is loaded with problems. Different indices yield different betas for the same security.
- If the CAPM is completely valid, it should apply to all financial assets including bonds. When bonds are introduced into the analysis, they do not fall on the security market line.

Current Validity of CAPM

The CAPM is greatly appealing at an intellectual level, as it is logical and rational. Although its basic assumptions raise some doubts in the minds of the investors, investment analysts have been creative in adapting CAPM for their use. The following points about CAPM should be noted:

- By focusing on the market risk, it makes the investors think about the riskiness of assets in general. It provides the basic concepts and these are of fundamental value.
- It is useful for the selection of securities and portfolios. Securities with higher returns are considered to be undervalued and attractive buys. Overvalued securities, whose returns are lower than the normal return, are suitable for sale.
- In the CAPM, it is assumed that investors consider only the market risk. Given the estimate of the risk-free rate, the beta of the firm and the required market rate of return, one can find out the expected return for a firm's security. This expected return can be used as an estimate of the cost of retained earnings.
- Even though CAPM is regarded as a useful tool for financial analysts, it has its critics. They point out that when the model is ex-ante, the inputs also should be ex-ante, i.e., based on the expectations of the future. Empirical tests and analyses are ex-post, i.e., use only past data.

NOTES

- Historical data regarding the market return, risk-free rate of return and betas vary differently for different periods. The various methods used to estimate these inputs also affect the beta value. Since the inputs cannot be estimated precisely, the expected return determined through the CAPM model is also subject to criticisms.

8.8 ARBITRAGE PRICING MODEL

Arbitrage pricing theory (APT) is one of the tools used by investors and portfolio managers. Capital asset pricing theory explains the returns of the securities on the basis of their respective betas. According to the models discussed earlier, an investor chooses investments on the basis of the expected return and variance. The alternative model in asset pricing developed by Stephen Ross is known as arbitrage pricing theory. APT explains the nature of equilibrium in asset pricing in a less complicated manner with fewer assumptions than CAPM.

8.8.1 Arbitrage

Arbitrage refers to the process of earning a profit by taking advantage of differential pricing for the same asset. The process generates risk-free profit. In the security market, it involves selling a security at a high price and the simultaneous purchase of the same security at a relatively lower price. Since the profit earned through arbitrage is risk-free, investors have the incentive to undertake this whenever an opportunity arises. In general, some investors indulge more in this type of activity than others. However, the buying and selling activities of the arbitrageur reduce and eliminate the profit margin, bringing the market price to the equilibrium level.

APT is based on the following assumptions:

- Investors have homogenous expectations
- They are risk averse and utility maximizers
- Perfect competition prevails in the market, and there is no transaction cost

APT does not assume: (1) a single period investment horizon, (2) no taxes, (3) investors can borrow and lend at the risk-free rate of interest and (4) the selection of the portfolio is based on the mean and variance analysis. These are the assumptions of the CAPM theory.

Check Your Progress

15. Name the scholars who developed the capital asset pricing independently.
16. Fill in the blanks with appropriate terms.
    (i) In CAPM theory, the required rate return of an asset has a _____ with an asset's beta value.
    (ii) According to CAPM, all investors hold only a market portfolio and _____ securities.

Arbitrage Portfolio

According to APT, an investor tries to determine the possibility of an increase in the returns from his portfolio without increasing the funds in it. He also likes to keep the risk at the same level. For example, say the investor holds A, B and C securities and wants to change the proportion of his investment in the securities without any additional financial commitment. Let the change in the proportions be denoted by X_A, X_B and X_C. The investor can increase his investment in security A if he reduces the proportions of his investments either in B or C, given that he tries to earn a higher income without increasing his financial commitments. Thus, the changes in different securities will add up to zero. This is the basic requirement of an arbitrage portfolio. If X indicates the change in proportion,

$$\Delta X_A + \Delta X_B + \Delta X_C = 0$$

The factor sensitivity indicates the responsiveness of a security's return to a particular factor. The sensitiveness of the securities to any factor is the weighted average of the sensitivities of the securities, with the weights being the changes made in the proportion. For example, b_A, b_B and b_C are the sensitivities; in an arbitrage portfolio the sensitivities become zero.

$$b_A \Delta X_A + b_B \Delta X_B + b_C \Delta X_C = 0$$

An investor holds A, B and C stocks with the returns and sensitivity to changes in industrial production is as follows. The total amount invested is `1,50,000.

1	R (%)	b	Original weights
Stock A	20	0.45	0.33
Stock B	15	1.35	0.33
Stock C	12	0.55	0.34

Say the proportions change to:

$$\Delta X_A = 0.2$$

$$\Delta X_B = 0.025$$

$$\Delta X_C = -0.225$$

For an arbitrage portfolio:

$$\Delta X_A + \Delta X_B + \Delta X_C = 0$$

$$0.2 + 0.025 - 0.225 = 0$$

The sensitivities also become zero:

$$\Delta X_A b_A + \Delta X_B b_B + \Delta X_C b_C = 0$$

$$0.2 \times 0.45 + 0.025 \times 1.35 - 0.225 \times 0.55 = 0$$

In an arbitrage portfolio, the expected return should be greater than zero.

$$\Delta X_A R_A + \Delta X_B R_B + \Delta X_C R_C > 0$$

$$= 0.2 \times 20 + 0.025 \times 15 - 0.225 \times 12$$

$$4.375 - 2.7 > 0$$

i.e., 1.675%

The investor would increase his investment in stock A and B by selling C. The new compositions of weights would be:

$$X_A = 0.53$$

$$X_B = 0.355$$

$$X_C = 0.115$$

The portfolio allocation to stocks A, B and C is as follows:

$$= 1,50,000 \times 0.53 + 1,50,000 \times 0.355 + 1,50,000 \times 0.115$$

$$= `(79,500 + 53,250 + 17,250)$$

$$= `1,50,000$$

The sensitivity of the new portfolio will be:

$$= 0.45 \times 0.53 + 1.35 \times 0.355 + 0.55 \times 0.115$$

$$= 0.239 + 0.479 + 0.063$$

$$= 0.781$$

This is same as the old portfolio sensitivity:

i.e., $0.45 \times 0.33 + 1.35 \times 0.33 + 0.55 \times 0.34 = 0.781$

The return of the new portfolio is higher than that of the old portfolio.

Old portfolio return:

$= 20 \times 0.33 + 15 \times 0.33 + 12 \times 0.34$

$= 6.6 + 4.95 + 4.08$

$= 15.63\%$

The new portfolio return:

$= 20 \times 0.53 + 15 \times 0.355 + 12 \times 0.115$

$= 10.6 + 5.325 + 1.38$

$= 17.305\%$

This is equivalent to the old portfolio return plus the return that occurred due to the change in portfolio:

$= 15.63 + 1.675 = 17.305\%$

The variance of the new portfolio's change is only due to changes in its non-factor risk. Hence, the change in the risk factor is negligible. From the analysis, it can be concluded that:

- The return in the arbitrage portfolio is higher than that of the old portfolio.
- The arbitrage and old portfolio sensitivities remain the same.
- The non-factor risk is small enough to be ignored in an arbitrage portfolio.

Effect on the Price

To buy stock A and B the investor has to sell stock C. The buying pressure on stock A and B would lead to an increase in their prices. Conversely, selling of stock C will result in a fall in the price of stock C. At the lower price, there would be a rise in the expected return of stock C. For example, if stock C at the price of `100 per share earns 12 per cent return, at `80 per share the return would be $12/80 \times 100 = 15$ per cent. At the same time, return rates would decline in stock A and B with the rise in price. This buying and selling activity will continue until all arbitrage possibilities are eliminated. At this juncture, an approximate linear relationship arises between expected returns and sensitivities.

8.8.2 The APT Model

According to Stephen Ross, the returns of the securities are influenced by a number of macroeconomic factors. These are growth rates of industrial production, rate of inflation, spread between long-term and short-term interest rates and spread between low-grade and high-grade bonds. The arbitrage theory is represented by the equation:

$R_i = \lambda_0 + \lambda_1 bi_1 + \lambda_2 bi_2 + \ldots + \lambda_j bi_j$

R_i = average expected return

λ_1 = sensitivity of return to bi1

b_{i1} = beta co-efficient relevant to the particular factor

The equation is derived from the model

$R_i = \alpha_1 + b_{i1} I_1 + b_{i2} I_2 + \ldots + b_{ij} I_j + e_i$

Let us take the two-factor model,

$R_i = \lambda_0 + \lambda_1 b_{i2} + b_{i2}$

If the portfolio is a well diversified one, unsystematic risk tends to be zero and systematic risk is represented by b_{i1} and b_{i2} in the equation.

NOTES

Let us assume the existence of three, well-diversified portfolios as shown here:

Portfolio	Expected return	b_{i2}	b_{i2}
A	12.0	1	0.5
B	13.4	3	0.2
C	12	3	–0.5

The equation $Ri = \lambda_0 + \lambda_1 b_{i1} + \lambda_2 b_{i2}$ can be determined with the help of the above-mentioned details. By solving the following equations,

$12 = \lambda_0 + 1\lambda_1 + 0.5\lambda_2$

$13.4 = \lambda_0 + 3\lambda_1 + 0.2\lambda_2$

$12 = \lambda_0 + 3\lambda_1 - 0.5\lambda_2$

we can get:

$R_i = 10 + 1b_{i1} + 2 b_{i2}$

The expected return is:

$$R_p \sum_{i=1}^{N} X_i R_i$$

The risk is indicated by the sensitivities of the factors:

$$b_{p1} = \sum_{i=1}^{N} X_i b_{i1}; b_{p^2} = \sum_{i=1}^{N} X_i b_{i\,2}$$

All the portfolios constructed from portfolios A, B and C lie on the plane described as A, B and C. Assume there exists a portfolio D with an expected return of 14 per cent, $b_{i1} = 2.3$ and $b_{i2} = 0.066$. This portfolio can be compared with portfolio E having equal portion of A, B and C portfolios. Every portfolio would have a share of 33 per cent. The portfolio b_{pj} are:

$b_{p1} = 1/3 \times 1 + 1/3 \times 3 + 1/3 \times 3 = 2.33$

$b_{p2} = 0.5 \times 1/3 + 0.2 \times 1/3 + (-0.5 \times 1/3) = 0.066$

$1/3\,(12) + 1/3\,(13.4) + 1/3\,(12)$

$= 12.46$

$R = 10 + 1\,(2.33) + 2\,(0.066)$

$= 12.46\%$

Portfolios D and E have the same risk but different returns. At this juncture, the arbitrageur enters and buys portfolio D by selling portfolio E short. Thus, buying portfolio

D with the funds generated by selling E would provide risk-free profit with no investment and no risk. Let us assume that the investor sells `1,000 worth of portfolio E and buys `1,000 worth of portfolio D. The cash flow is as follows:

NOTES

	Initial cash fl w	End of period	b_{i1}	b_{i2}
Portfolio D	– `1000	+1140.0	+2.33	+0.06
Portfolio E	+ `1000	–1124.6	–2.33	–0.06
Arbitrage portfolio	0	15.4	0	0

The arbitrage portfolio involves zero investment, has no systematic risk (b_{i1} and b_{i2}) and earns `15.4. Arbitrage would continue until portfolio D lies on the same plane.

8.8.3 Arbitrage Pricing Equation

In a single factor model, the linear relationship between the return R_i and sensitivity b_i can be given in the following form.

$R_i = \lambda_0 + \lambda_i b_i$

R_i = return from stock A

λ_0 = risk-free rate of return

b_i = sensitivity related to the factor

λ_i = slope of the arbitrage pricing line

The above model is known as a single factor model since only one factor is considered. Here, only industrial production is considered. The APT one-factor model is shown in Figure 8.16.

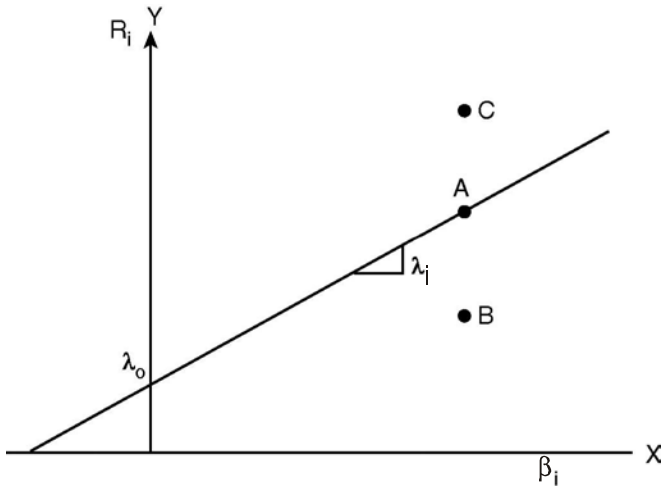


Fig. 8.16 *APT One-factor Model*

In the graph (Fig. 8.16), risk is measured along the horizontal axis (X) and return on the vertical axis (Y). Stocks A, B and C stocks are considered to be in the same risk class. The arbitrage pricing line intersects the Y axis on λ_O, which represents the risk-free rate of interest i.e., the interest offered for the treasury bills. Here, the investments involve zero risk and it is appealing to investors who are highly risk averse. λ_i stands for the slope of the arbitrage pricing line. It indicates the market price of risk and measures the risk-return trade off in the securities market. β_i is the sensitivity coefficient or factor beta that shows the sensitivity of the asset or stock A to the corresponding risk factor.

NOTES

The Constants of the APT Equation

The existence of a risk-free asset yields a risk-free rate of return that is a constant. The asset does not have sensitivity to the factor, for example, industrial production.

If $b_i = 0$,

$R_i = \lambda_O + \lambda_i 0$

$R_i = \lambda_O$

$R_i = \lambda_O$

In other words, λ_O is equal to the risk-free rate of return. If the single factor portfolio's sensitivity is equal to one i.e., $b_i = 1$, then:

$$R_i = \lambda_O + \lambda_i 1$$

This can be rewritten as

$$R_i = \lambda_O + \lambda_i$$

$$R_i - \lambda_O = \lambda_i$$

Thus, λ_i is the expected return in excess of the risk-free rate of return for a portfolio with a unit sensitivity to the factor. The excess return is known as the risk premium.

8.8.4 Factors Affecting the Return

Many financial analysts have spelt out the factors affecting return. Chen, Roll and Ross (1986) took four macroeconomic variables and tested them. According to them, the factors are inflation, the term structure of interest rates, risk premia and industrial production. Inflation affects the discount rate or the required rate of return and the size of future cash flows. Short-term inflation is measured by monthly percentage changes in the consumer price index. The interest rates on long-term bonds and short-term bonds differ. This difference affects the value of payments in the future relative to short-term payments. The difference between the return on high-grade bonds and low-grade (more risky) ones indicates the market reaction to risk. Industrial production represents the business cycle. Changes in industrial production have an impact on the expectations and opportunities of the investor. The real value of the cash flow is also affected by it.

Burmeister and McElroy (1988) have estimated the sensitivities with some other factors. These are given below:

- Default risk
- Time premium
- Deflation
- Change in expected sales
- Market return not due to the first four variables

The default risk is measured by the difference between the return on long-term government bonds and the return on long-term bonds issued by corporations plus one-half of one per cent. Time premium is measured by the return on long-term government bonds minus the one-month treasury bill rate of one month ahead. Deflation is measured by the expected inflation at the beginning of the month minus the actual inflation rate during the month. According to them, the first four factors accounted for 25 per cent of the variation in the Standard & Poor composite index and all the four coefficients were significant.

NOTES

Salomon brothers (1985) identified five factors in their fundamental factor model. Inflation was the only factor they had in common with the others. Their factors are given below:

- Growth rate in gross national product
- Rate of interest
- Rate of change in oil prices
- Rate of change in defence spending

All the three sets of factors have some common characteristics. They all affect macroeconomic activities. Inflation and interest rate are identified as common factors. Thus, the stock price is related to aggregate economic activity and the discount rate of future cash flows.

8.8.5 APT and CAPM

The simplest form of the APT model is consistent with the simple form of the CAPM model. When only one factor is taken into consideration, the APT can be stated as:

$R_i = \lambda_O + b_i\lambda_i$

It is similar to the capital market line equation

$R_i = R_f + ß_i (R_m - R_f)$, which is similar to the CAPM model.

APT is more general and less restrictive than CAPM. In APT, the investor has no need to hold the market portfolio because it does not make use of the market portfolio concept. The portfolios are constructed on the basis of the selected factors to eliminate profits from arbitrage. APT is based on one price holding for all possible portfolio combinations.

The APT model takes into account the impact of numerous factors on the security. The macroeconomic factors are taken into consideration, and it is closer to reality than CAPM.

The market portfolio is well defined conceptually. In the APT model, factors are not well specified. Hence, the investor finds difficult to establish an equilibrium relationship. The well-defined market portfolio is a significant advantage of the CAPM and this has led to the wide use of this model in the stock market.

The factors that have an impact on one group of securities may not affect another group of securities. There is a lack of consistency in the measurements of the APT model.

Further, the influences of the factors are not independent of each other. It may be difficult to identify the influence that corresponds exactly to each factor. Apart from this, not all variables that exert an influence on a factor are measurable.

Check Your Progress

17. What is arbitrage?
18. List the macroeconomic factors that influence the returns of the securities.

8.9 SUMMARY

In this unit, you have learnt that,

- The two terms—risk and uncertainty—are interchangeably used in practice. 'Risk is that uncertainty which is predictable, and to which probability can be assigned.'
- Uncertainty is that event which cannot be predicted, and therefore, no probability can be assigned to it.

- Probability is the likelihood of happening of some event. Merriam Webster online dictionary defines probability as 'the chance that a given event will occur'.
- Risk causes volatility. However, not all volatilities are indicators of risk.
- The real test of a decision maker is in estimation of returns (occurrences) and probabilities for each occurrence. Returns on financial securities are derived from their underlying assets. Returns from equity shares would depend on earnings of a firm from its business.
- Expected return from an investment with more than one expected outcomes is the weighted average returns of all outcomes.
- Types of risk measurement are discussed in three parts: techniques of measuring risk, decision rules, and finally, some more tools of measure cum decisions.
- Range is the difference between the highest value of outcome and the lowest value of outcome.
- The range does not consider probabilities, nor does it consider other possible outcomes between the highest and lowest values. Sensitivity analysis uses range values in its application.
- Standard deviation ignores the size of the project, therefore, comparison of two projects becomes difficult.
- When standard deviation (s) is divided by expected returns (i.e., mean or *ER*) we get the coefficient of variation. It is, therefore, useful in comparing risks of two projects with different sizes. Lower coefficient of variation is considered good.
- Portfolio risk can be reduced by the simplest kind of diversification. Portfolio means the group of assets an investor owns. These assets may vary from stocks to different types of bonds.
- In simple diversification securities are selected at random and no analytical procedure is used. This sort of diversification reduces risk only to a certain extent.
- Spreading the investment across too many assets can give rise to such problems as the inclusion of poor performers, information inadequacy, high research, and high transaction costs.
- Risk was defined as variability in the returns. Portfolio effect occurs when the variability of returns from the existing investment and that from the new investment are different.
- Risk can be classified into (a) systematic risk, and (b) unsystematic risk.
- Acompany with a single project is exposed to external vagaries. With the addition of a new project in the business (project portfolio), the total risk may reduce. The process of risk reduction continues for some time with the acceptance of more and more projects in the portfolio.
- Financing risk arises out of the debt-equity structure chosen by the company. Knowledge of capital structure theory for setting an appropriate debt-equity target, and designing of securities can jointly influence the total financing risk.
- Market risk gets reflected on the cost of individual funds for the reasons related to market factors. A perfect capital market would offer zero market risk.
- Business risk is that part of the total risk, which occurs due to a particular project portfolio.

NOTES

- Portfolio effect is measured through the calculation of:
  - Covariance
  - Portfolio return
  - Portfolio risk
- The capital structure decision is the decision regarding the proportion of ownership funds and borrowed funds in total funds needed.
- The firms borrow short-term funds from various sources like overdraft, cash credit, commercial papers and others.
- According to the Net Income approach, the cost of debt and the cost of equity, both remain unchanged for any level of debt-equity ratio.
- Many realistic factors have culminated in the pragmatic theory of capital structure called the traditional approach of capital structure.
- The traditionalist theorists argue that the weighted average cost of capital will decrease with every increase in the debt content in the total capital employed.
- According to Modigliani-Miller (MM) approach the corporation's total cost of capital is constant and it is independent of the method and level of financing.
- According to the MM's first proposition an increase in the financial leverage increases the expected earnings per share but not the price of the share.
- Jack Treynor (1961), William Sharpe (1964), John Lintner (1965) and Jan Mossin (1965) developed Capital Asset Pricing Model (CAPM) independently.
- CAPM assumes that the investor can borrow or lend any amount of money at a risk-free rate of interest. When this opportunity is given to investors, they can mix risk-free assets with risky ones in a portfolio for the desired rate of risk-return combination.
- The risk-return relationship of an efficient portfolio is measured by the capital market line. It does not show the risk-return trade-off for other portfolios and individual securities.
- Historical data regarding the market return, risk-free rate of return and betas vary differently for different periods. The various methods used to estimate these inputs also affect the beta value. Since the inputs cannot be estimated precisely, the expected return determined through the CAPM model is also subject to criticisms.
- Arbitrage pricing theory (APT) is one of the tools used by investors and portfolio managers. Capital asset pricing theory explains the returns of the securities on the basis of their respective betas.
- The alternative model in asset pricing developed by Stephen Ross is known as arbitrage pricing theory. APT explains the nature of equilibrium in asset pricing in a less complicated manner with fewer assumptions than CAPM.
- According to Stephen Ross, the returns of the securities are influenced by a number of macroeconomic factors. These are growth rates of industrial production, rate of inflation, spread between long-term and short-term interest rates and spread between low-grade and high-grade bonds.
- APT is more general and less restrictive than CAPM. In APT, the investor has no need to hold the market portfolio because it does not make use of the market portfolio concept. The portfolios are constructed on the basis of the selected factors to eliminate profits from arbitrage. APT is based on one price holding for all possible portfolio combinations.

NOTES

8.10 KEY TERMS

- **Risk:** It is that uncertainty which is predictable, and to which probability can be assigned.
- **Uncertainty:** It is that event which cannot be predicted, and therefore, no probability can be assigned to it.
- **Probability:** It is the likelihood of happening of some event.
- **Range:** It is the difference between the highest value of outcome and the lowest value of outcome.
- **Portfolio:** It means the group of assets an investor owns.
- **Financial risk:** The variability in shareholders' net income due to the change in the net operating income (business risk) is called the financial risk.
- **Business risk:** It is a function of the product market conditions faced by a firm for all the businesses collectively, and variability of operating income due to those factors.
- **Project risk:** The risk arising out of product market conditions and their impact on a single project is called project risk.
- **Capital structure decision:** It is the decision regarding the proportion of ownership funds and borrowed funds in total funds needed.
- **Arbitrage:** It refers to the process of earning a profit by taking advantage of differential pricing for the same asset.

8.11 ANSWERS TO 'CHECK YOUR PROGRESS'

1. The two conditions inherent in the definition of risk are: (a) the decision maker is aware of all possible outcomes, which may happen and affect the decision, and (b) the decision-maker is in a position to assign probability to each possible outcome envisaged by him.
2. Probability is the likelihood of happening of some event. Merriam Webster online dictionary defines probability as 'the chance that a given event will occur'.
3. Types of risk measurement are discussed in three parts: techniques of measuring risk, decision rules, and finally, some more tools of measure cum decisions.
4. When standard deviation (s) is divided by expected returns (i.e., mean or *ER*) we get the coefficient of variation. It is, therefore, useful in comparing risks of two projects with different sizes.
5. Spreading the investment across too many assets can give rise to such problems as the inclusion of poor performers, information inadequacy, high research, and high transaction costs.
6. Portfolio effect occurs when the variability of returns from the existing investment and that from the new investment are different.
7. Risk can be classified into: (a) systematic risk, and (b) unsystematic risk. Risk is also classified in different ways, business risk, financial risk and market risk.
8. Business risk is a function of the product market conditions faced by a firm for all

the businesses collectively, and variability of operating income due to those factors.

9. The capital structure decision is the decision regarding the proportion of ownership funds and borrowed funds in total funds needed.
10. The firms borrow short-term funds from various sources like overdraft, cash credit, commercial papers and others.
11. The capital structure, once decided for a minimum cost of capital, remains static forever, according to the static view. This view assumes that the quality of earnings and capital market conditions are not variable affecting the cost of capital. This may be true in a period of no competition and primitive financial market.
12. The raising of funds through debentures is cheaper because of the following reasons:
    - Interest rates are usually lower than dividend rates.
    - Interest is allowed as an expense resulting in a tax benefit while dividend is not allowed as an expense while computing taxable profits of the company.
13. The traditionalist theorists argue that the weighted average cost of capital will decrease with every increase in the debt content in the total capital employed.
14. In spite of the correctness of the basic reasoning of Modigliani and Miller, the traditional approach is more realistic on account of the following reasons:
    - The corporations are subject to income-tax and, therefore, due to tax effect, the cost of debt is lower than cost of equity capital.
    - The basic assumption of Modigliani and Miller hypothesis that capital markets are perfect, is seldom true.
15. Jack Treynor (1961), William Sharpe (1964), John Lintner (1965) and Jan Mossin (1965) developed capital asset pricing model independently.
16. (i) linear relationship

    (ii) risk-free
17. Arbitrage refers to the process of earning a profit by taking advantage of differential pricing for the same asset.
18. According to Stephen Ross, the returns of the securities are influenced by a number of macroeconomic factors. These are growth rates of industrial production, rate of inflation, spread between long-term and short-term interest rates and spread between low-grade and high-grade bonds.

8.12 QUESTIONS AND EXERCISES

Short-Answer Questions

1. How are risks and volatility related? Give examples.
2. What are expected returns from an investment?
3. List the statistical techniques used in measuring risk.
4. How can portfolio risk be reduced?
5. Write a short note on simple diversification.
6. 'Risk is also classified in different ways, business risk, financial risk and market risk.' Describe them in brief.

7. What is the net operation income approach to capital structure?
8. What is the traditional approach of capital structure?
9. List the assumptions under Modigliani-Miller approach to cost structure.
10. Distinguish between the security market line and capital market line.
11. What are the advantages of APT over CAPM?
12. What is meant by factor sensitivity? How is it estimated?

Long-Answer Questions

1. Assess the returns from assets and risk.
2. What are risk-free assets? What are its features?
3. Discuss the types of risk measurement and the statistical techniques used in measuring risk.
4. What is portfolio effect? When does it occur?
5. Evaluate the various types of risk in portfolio.
6. 'The capital structure, and not the financial structure, is an appropriate consideration in the cost of capital theories.' Discuss.
7. Critically analyse the capital structure theories.
8. Evaluate the traditional theory of cost and Modigliani-Miller approach to cost structure.
9. Explain the CAPM theory and its value in the stock market.
10. Distinguish between CAPM and Arbitrage Pricing Theory (APT).

8.13 FURTHER READING

Mankiw, N Gregory. 2010. *Macroeconomics.* Worth Publishers, New York.

Shapiro, Edward. 1996. *Macroeconomic Analysis.* Galgotia Publication, New Delhi.

Branson, WH. 1979. *Macroeconomic Theory and Policy*. Universal Book Stall, New Delhi.

Jha, R. 1999. *Contemporary Macroeconomic Theory and Policy*. New Age International, New Delhi.

Gupta, SB. 1995. *Monetary Economics: Instruments and Policy*. S Chand & Co. New Delhi.

Endnote

1. The rule 'lower coefficient of variation is better' assumes that the decision maker is risk neutral. That means the utility curve of the decision maker is neither convex nor concave, but it is a linear sloping line. Risk neutral characteristic implies that the rate of risk-premium is constant for every unit of incremental risk.

UNIT 9 FINANCIAL MARKET

Structure

NOTES

9.0 INTRODUCTION

A financial market is a broad term describing any marketplace where buyers and sellers participate in the trade of assets such as equities, bonds, currencies and derivatives. Financial markets are typically defined by having transparent pricing, basic regulations on trading, costs and fees, and market forces determining the prices of securities that trade.

In economics, typically, the term market means the aggregate of possible buyers and sellers of a certain good or service and the transactions between them. The term 'market' is sometimes used for what are more strictly exchanges, organizations that facilitate the trade in financial securities, e.g., a stock exchange or commodity exchange.

9.1 UNIT OBJECTIVES

market

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

- Discuss the role, structure and characteristics of money and capital market
- Analyse the features of the primary and secondary market for securities
- Describe the different kinds of treasury bills and the advantages of treasury bill

NOTES

- Discuss the functioning of commercial bill market and discount market
- Analyse the role of government securities market
- Discuss the market of derivatives, their different types, uses and pricing techniques
- Describe the construction of sensex and nifty

9.2 MONEY AND CAPITAL MARKET: ROLE, STRUCTURE AND CHARACTERISTICS

Let us discuss the role, structure and characteristics of money and capital market.

9.2.1 Money Market

Money market includes the trade in short-term loans between banks and other financial institutions.

Role of Money Market

Monetary assets, which are short term in nature and less than one year are traded in the money market. Money market facilitates RBI's conduct of monetary policy. There was a paucity of instruments in the money market for a long time. After the mid 1990s the money market in India experienced significant development in terms of refinements of existing money market instruments and introduction of new instruments. The money market instruments in general are close substitutes to money. The various money market instruments are given below.

- Call/ Notice/ Term Money
- Repos
- Treasury Bills
- Certificate of Deposits (CD)
- Commercial Papers (CP)
- Inter Bank Participation Certificates
- Inter Bank Term Money
- Interest Rate Swaps/Forward Rate Agreements and
- Bills Rediscounting

Structure of Money Market

The money market has two components namely the organized and the unorganized. The participants of the organized money market are the Reserve Bank of India, Life Insurance Corporation, General Insurance Corporation, Unit Trust of India, Securities Trading Corporation of India Ltd., Discount and Finance House of India, other primary dealers, commercial banks and mutual funds. The mainstay of the money market is the inter-bank call money market where short-term money borrowing/lending is carried out to manage temporary liquidity mismatches.

Money market is a very important segment of the Indian financial system. It is the market for dealing in monetary assets of short-term nature. Short-term funds up to one year and for financial assets that are close substitutes for money are dealt with in the money market. Money market instruments have the characteristics of liquidity (quick

conversion into money), minimum transaction cost and no loss in value. Excess funds are deployed in the money market which in turn are availed of to meet temporary shortages of cash and other obligations. Money market provides access to providers and users of short-term funds to fulfil their borrowings and investment requirements at an efficient market clearing price. It performs the crucial role of providing an equilibrating mechanism to even out short-term liquidity and in the process, facilitating the conduct of monetary policy. Short-term surpluses and deficits are evened out. The money market is the major mechanism through which the Reserve Bank of India (RBI) influences liquidity and the general level of interest rates. The RBI's interventions to influence liquidity serve as a signalling device for other segments of the financial system.

The Indian money market was segmented and highly regulated and lacked depth till the late 1980s. It was characterized by a limited number of participants, regulation of entry and limited availability of instruments. The instruments were limited to call (overnight) and short notice (up to 14 days) money, interbank deposits and loans and commercial bills. Interest rates on market instruments were regulated. Sustained efforts for developing and deepening the money market were made only after the initiation of financial sector reforms in early nineties.

Characteristics of Money Market

The money market is a wholesale market. The volumes are very large and generally transactions are settled on a daily basis. Trading in the money market is conducted over the telephone, followed by written confirmation from both the borrowers and lenders.

There are a large number of participants in the money market: commercial banks, mutual funds, investment institutions, financial institutions and finally the RBI. The RBI's operations ensure that the liquidity and short-term interest rates are maintained at levels consistent with the objective of maintaining price and exchange rate stability. The Central bank occupies a strategic position in the money market. The money market can obtain funds from the Central bank either by borrowing or through sale of securities. The bank influences liquidity and interest rates by open market operations, repo transactions, changes in bank rate, cash reserve requirements and by regulating access to its accommodation. A well-developed money market contributes to an effective implementation of the monetary policy.

9.2.2 Capital Market

The capital market provides long-term funds for corporates, central and state governments.

Role and Structure of Capital Market

The capital market has sub-markets like debt market, equity market and derivative market.

- Fixed income securities of various types and features are issued and traded in the debt market. Debt markets are therefore markets for fixed income securities issued by Central and state governments, municipal corporations, government bodies and commercial entities like financial institutions, banks, public sector undertakings, public limited companies and structured finance instruments.
- In the equity market, variable income securities like equity shares are issued and traded. Equity shares are issued by the companies and financial institutions.
- In the derivative market, financial products and commodity based derivatives like options and futures are traded. Commodity options and futures on precious metals,

NOTES

oil and agricultural products are traded on the commodity exchanges. Financial product derivatives like stock options and futures, stock index futures and options, and interest rate futures are traded on the stock exchanges.

Equity and debt instruments are issued in the primary market. The primary market is also known as new issue market. An already listed company or unlisted company can make a fresh issue of securities to the public. The primary market has no physical location. A host of intermediaries like lead manager, registrar, underwriter, custodian and depository are involved in the primary market. SEBI regulates the new issue market. Under SEBI guidelines, no company shall make a public issue or right issue of debt instruments (whether convertible or not) unless credit rating is obtained from at least one credit rating agency registered with SEBI. The issuer also has to prominently display the ratings in all the marketing literature and advertisements related to the particular debt instrument. The issued securities are listed in the stock exchanges.

Stock exchanges are the backbone of both the equity and debt markets. The issued securities are listed in the stock exchanges. The secondary market provides a trading place to buy and sell the securities already issued. It also provides liquidity to the initial buyers in the primary market to re-offer the securities to any interested buyer at any price, if mutually accepted. An active secondary market promotes the growth of primary market and capital formation because investors in the primary market are assured of a continuous market and can liquidate their investments. The secondary market consists of 23 stock exchanges including the National Stock Exchange (NSE), Over-the-Counter Exchange of India (OTCEI) and Inter Connected Stock Exchange of India Ltd. (ISE).

Foreign Exchange Market

The foreign exchange market is the market for foreign currencies. Foreign currency transactions take place in the case of exports, imports, capital movement, interest and repayment of loans. In the foreign exchange market, central-banking institutions, commercial banks and merchant bankers act as dealers under the control of RBI and Foreign Exchange Dealers Association of India (FEDAI). Star hotels, travelling agencies, individuals and institutions are permitted to act as money changers under the provisions of RBI and FEDAI.

Characteristics of an Efficient Financial Market

An efficient financial market leads the economy towards efficient use of factors and allocation of factors for socially productive purposes. An efficient financial market has the following features.

- **Operational efficiency:** Efficient market structure and their linkages minimize the administrative and transaction costs of transfer of funds.
- **Information efficiency:** All the available information is absorbed by the prices of securities. The quickness with which the prices reflect all the relevant information indicates the information efficiency. The probability of getting an abnormal profit is much limited in a perfect market.
- **Valuation efficiency:** In an efficient market, the market price of the security should be equal to its intrinsic value. The true economic worth of the asset is its intrinsic value. In a perfectly competitive market, the price of an asset reflects its intrinsic value.
- **Hedging efficiency:** Efficient market provides hedging facilities to reduce risk. Hedging lowers the risk of incurring loss. Options and futures markets provide hedging facilities.

Check Your Progress

1. State the various money market derivatives.
2. What are the sub-markets of capital market?
3. What is valuation efficiency?

9.3 PRIMARY AND SECONDARY MARKET FOR SECURITIES

NOTES

Stock market or securities market is a market where securities issued by companies in the form of shares, bonds and debentures can be bought and sold freely. The components of stock market are primary market and secondary market. Primary market or new issues market is concerned with the issue of new securities.

Features of Primary Market

The features are as follows:

- It is a market for long-term capital where the securities are sold for the first time. Hence, it is also called New Issue Market (NIM).
- Funds are collected and securities are issued directly by the company to the investors.
- Primary issues are carried out by the companies for the purpose of inception and functioning of business.

Benefits of Primary Market

The benefits are as follows:

- Company need not repay the money raised from the market.
- Money has to be repaid only in the case of winding up or buyback of shares.
- There is no financial burden, because it does not involve interest payment. If the company earns profit, dividend may be paid.
- Better performance of the company enhances the value for the shareholders.
- It enables trading and listing of securities at stock exchanges.
- There is greater transparency in the corporate governance.
- If the company performs well; the image of the company brightens.

Disadvantages of the Public Issues

The disadvantages are as follows:

- It is a time consuming process involving the fulfillment of legal formalities.
- It is expensive and many intermediaries are involved in it. It necessitates constant adherence to listing of agreements and legal requirements.
- Cornering of shares and hostile takeover may take place.
- Speculative trading of the company's equity affects the reputation of the company.

Types of Issues

Issues are classified into three types:

- **Public Issue**

When a company wants to raise capital by issuing shares to general investors, the process is generally called public issue of shares or a primary market issue. Public issues can be further classified as:

- Initial public offering
- Follow-on public offering
- Fast-track issue

NOTES

Initial Public Offering

If the company is a new entrant to the capital markets, the issue made by such a company is called Initial Public Offering (IPO). In other words, if the company makes the public issue for the first time, it is an IPO. These issues are listed and traded on stock exchanges as specified in the offer documents.

Follow on Public Offering (FPO)

If a company whose shares are already listed on a stock exchange, issues shares, such an offering is called a further public offering. These are also called follow-on public issues. Sometimes, the term IPO is interchangeably used with FPO. Listing or continuous listing has to be satisfied by the FPO company. In 2006-2007, out of the 85 public issues, 77 were IPOs and the remaining were FPOs.

Fast-track Issue

The fast-track system introduced by SEBI in November 2007, was suggested by the SEBI's Primary Market Advisory Committee (PMAC). In the Fast-Track Issue (FTI), well established and compliant listed companies need to make only rationalized disclosures, rather than comprehensive ones, for follow-on public offers and rights issues. The compliant company here means the company which complies with the required rules and regulations. Provisions of the fast-track issue are on the lines of the well-known seasoned issuers model of the US. This facility is available to the companies that are listed on the Bombay Stock Exchange (BSE) or the National Stock Exchange (NSE) for at least three years.

Companies eligible for a fast-track issue of shares must have an average market capitalization of ` 10,000 crore of public shareholding for a period of one year. The trading on the securities market should constitute at least 2 per cent of the total listed shares of the last one year. No prosecution proceedings or show-cause notice issued by SEBI should be pending against the company, its promoters or whole-time directors.

- **Preferential Issue**

Listed companies issue securities to a select group of persons under section 81 of the Companies Act, 1956. The select group may be financial institutions, mutual funds, or high net worth individuals. It is not a rights issue or a public issue. The issuer company has to comply with the Companies Act and SEBI (Disclosures and Investor Protection) guidelines regarding the preferential issue. This helps the companies to raise capital quickly as compared to the public issue. The use of PAN has been made mandatory for preferential allotment of shares, as in the case of most of the transactions in the capital markets.

- **Rights Issue (RI)**

It means an issue of capital to be offered by the company to its existing shareholders through a Letter of Offer, under section 81(1) of the Companies Act, 1956. In simple terms, a listed company issues fresh securities to its existing shareholders only. The rights are offered in a specified ratio to the number of securities held prior to the issue by

the shareholder. The ratio is fixed on the basis of the capital requirement of the company. The stake of the existing shareholders is not diluted in the right issue. When the right issue size is more than ` 50 lakh, the company has to file a draft offer document with SEBI for observations. SEBI's observation letter is valid for three months.

Right issues are seen as a way to reward the company's shareholders as the issue of new shares is normally made at a price which is lower than the current market price. Indian companies raised over ` 14085 crore through various right issues in 2007. The Tata Group had announced a mega right issue of about ` 9,134 crore in November 2007.

Stock market has two components: the primary market and the secondary market. Primary market is a channel for the sale of new securities whereas secondary market provides a platform for sale of the already issued and listed securities. It has been defined as, 'a body of individuals, whether incorporated or not, constituted for the purpose of assisting, regulating and controlling the business of buying, selling and dealing in securities'. It helps the market participants to adjust their holdings in response to their risk and return perspective. However, both the markets are interdependent and inseparable.

Features of a secondary market

The features are as follows:

- Trading of securities in the secondary market does not provide any funds to the company.
- The investors as well as the speculators trade in securities.
- Securities of listed public limited companies are traded on a recognized stock-exchange.
- Secondary market provides liquidity to the investors.
- The market prices in the secondary market reflect the investors' perceptions of a company's performance.

Market Segments

The secondary market has the following three segments.

- Capital Market (CM) segment where equity, preference shares and warrants are traded.
- Wholesale Debt Market (WDM) segment where state and Central Government securities, T-bills, PSU bonds, corporate debentures, commercial papers, certificate of deposits, mutual funds etc., are traded.
- Futures and Options (F and O) segment where derivatives based on equity and indices are traded. Index option, index futures stock options and stock futures are bought and sold in this segment.

Participants

Participants of the secondary market mainly consist of:

- Investors
- Market intermediaries
- Regulatory bodies

NOTES

Investors

The investors can be broadly classified into:

- **Retail Investors**

They are individual investors with a limited access to funds. They park their surplus funds in securities to earn returns. Equity investment is considered to be high risk high return proposal as compared to other investment instruments like fixed deposits and post-office savings schemes. The term 'High Net worth Individual' (HNI) is used to refer to individuals and families who are affluent in their wealth holdings and consequently have a higher risk profile. It is a relative term and its meaning is different in different financial markets and regions.

- **Foreign Institutional Investors (FIIs)**

They are venture capital funds, pension funds, hedge funds, mutual funds and other institutions registered outside the country of the financial market in which they take an investment exposure. They are allowed to invest in the primary and secondary capital markets in India through the Portfolio Investment Scheme (PIS). Under this scheme, FIIs can acquire shares/debentures of Indian companies through the stock exchanges in India. They have to be registered with SEBI.

The ceiling for the overall investment for FIIs is 24 per cent of the paid up capital of the Indian company. In the case of public sector banks, including the State Bank of India, the limit is 20 per cent of the paid up capital. The ceiling of 24 per cent for FIIs' investment can be raised up to sectoral cap/statutory ceiling, subject to the approval of the board and the general body of the company passing a special resolution to that effect.

The number of FIIs registered with the Securities and Exchange Board of India (SEBI) has doubled to 1,050 between March 2001 and June 2007. The Indian capital market has attracted many global majors like HSBC, Citigroup, Merrill Lynch, Crown Capital, Fidelity, Goldman Sachs, Morgan Stanley, UBS, Capital International and ABN Amro. FIIs have gained a lot of importance as market participants as they invest huge sums of money. They are often instrumental in giving direction to the stock market movements in the short term.

Stock Market Review

Robust economic growth and strong fundamentals of the corporates made the FIIs to show interest in the stock market in 2007. Foreign investors choose India as one of their top investment destinations. India posted a relativity higher GDP growth (9.6 per cent) among the emerging economies. The market experienced large inflows from foreign institutional investors and wider participation of domestic investors' particularly institutional investors. FII's have invested a large amount of ` 7606510 crore in 2007.

The appreciation of value of rupee by more than 12 per cent in 2007 created a natural hedge for FII's investments. This was one of the reasons for the market rally in that year taking the Sensex and Nifty to a record index levels of over 20,000 and 6,000, respectively. However, this rupee appreciation adversely affected exporters and IT companies. The effect was seen in the falling values of SCRIPS of these companies. Yet the year 2007 belonged to the bulls.

NOTES

FIIs were net buyers in ten out of twelve months in 2007. They turned net sellers in the rest, mainly to make up the losses on account of the sub-prime crisis in the US economy. FIIs had pulled out an estimated $13 billion from the stock market in 2008. It means that the FIIs had withdrawn nearly three-fourth of over $17 billion invested in this year.

The sub-prime crisis in the US resulted in a decline of more than 2000 points from the Sensex on two days on 21 and 22 January, leading to a massive erosion of over ` 6 trillion in investor's wealth. The housing boom in the US market led many of the mortgage banks to disburse loans to many sub-prime (less credit worthy) borrowers. When recession set in the economy, many borrowers defaulted and mortgage banks made huge losses. The Citigroup wrote down billions of dollar, which made the global stock market nervous. Bear Stearns was the first mortgage bank to declare losses on mortgage lending in March 2008. The institution was sold to JP Morgan Chase. After that Fannie Mae and Freddie Mac were other two mortgage banks which reported loss. Followed by these institutions, collapse of Lehman Brothers and Merrill Lynch brought uncertainty to the Indian stock market. Lehman Brother holdings amounted to about ` 1000 crore as on 30 June 2008. Lehman's holding were concentrated mostly in the mid and small-cap stocks. Merrill Lynch and AIG had their presence felt in the Indian mutual fund industry. Thus the global crisis affected the market pushing the Sensex level below 10,000. The movement of Sensex is given in the next figure:

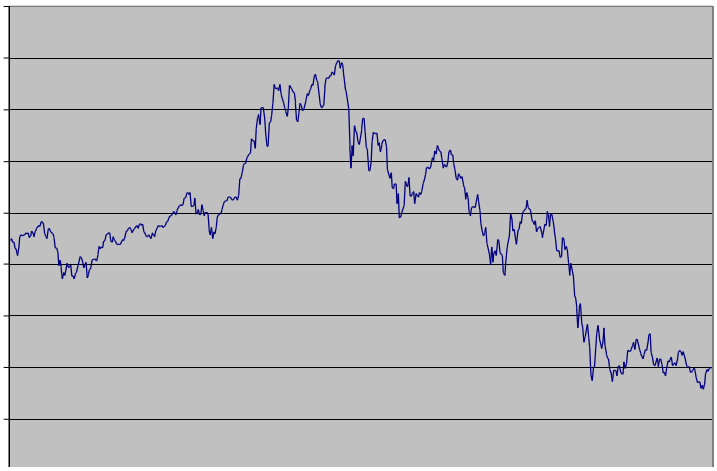


Fig. 9.1 *Movement of Sensex*

The Sensex reached all time low of 8198 on 5 March 2009 lowest since 2 November 2005. Even the US benchmark Dow Jones finished below 7000 on 2 March 2009 for the first time since 7 May 1997. The market breadth was weak with declines in 1,686 stocks and advances in only 699 stocks. Bankex, the banking index fell by 4.15 per cent at 3728.22 and in the NSE Bank Nifty lost 3.92 per cent at 3446.35. The continuous cut in interest rates, which pressures margins, asset quality worries and slowdown in credit offtake have been major concerns of the bank stocks.

The major reasons cited for the fall in Sensex are:

- FIIs were net sellers. They sold equities for ` 591 crore on 5 March 2009.
- Weak European markets
- Fall in the value of rupee (the rupee lost 0.25 paise against the dollar on 5 March 2009. The rupee touched a low of 52.18 in day trade)

NOTES

- Widening fiscal deficit at 6 per cent of GDP
- Announcement of $62 billion loss for the fourth quarter of 2008-09 by the insurance major AIG
- Reported 70 per cent drop in the earnings for the year 2008 by the banking major HSBC
- IMF's cut in the prediction of global growth to 0.5 percent in 2009 from an earlier prediction of 2.2 per cent.

9.4 TREASURY BILL MARKET

Treasury bills are short-term (up to one year) borrowing instruments of the government. In other words, these are short-term rupee-denominated government securities, i.e., promissory notes issued by the government. Since the government guarantees the payment of the treasury bills, they are more liquid than the trade bills. They are like zero-coupon bonds. They do not pay interest prior to maturity; instead they are sold at a discount on the par value to create a positive yield to maturity. Treasury bills are considered by many to be the most risk-free investments. The Bank of England first issued treasury bills in 1877. In India, these were first issued with an aim at mobilizing resources for the first world war and mop excess liquidity in the economy due to the heavy war expenditure.

Issuing Authority

Originally, the government used to issue the treasury bills directly. They were sold through the tender method and the periods of maturity were three months, six months, nine months and twelve months. In 1935, the Reserve Bank of India took over their issue. Till 1950, state governments occasionally issued treasury bills.

Characteristics of Treasury bills

The characteristics of treasury bills are as follows:

- Treasury bills have a short-term maturity period
- They are highly liquid in nature
- Their default risk is absent
- Their transaction cost is very low
- They are regularly available (The RBI auctions them on weekly and fortnightly basis)
- Their yield is assured
- They are eligible for inclusion in securities for Statutory Liquidity Ratio (SLR) purposes
- Their capital depreciation is negligible

Check Your Progress

4. State the different features of primary market.
5. State the disadvantages of public issues.
6. What are the three types of investors?

9.4.1 Types of Treasury Bills

After the establishment of RBI, the periodicity of the treasury bills was changed according to the financial needs and maturity policy of the RBI. Table 9.1 gives a chronology of treasury bills of various maturities introduced and discontinued subsequently.

Table 9.1 *Treasury Bills–Chronology of Development*

Type of T-bill	*Introduced*	*Discontinued*
91-day T-bill on weekly auction	Before 1950s	Mid-1950s
91-day ad-hoc T-bill	Mid-1950s	April 1997
91-day T-bill on tap	Mid-1950s	March 1997
182-day T-bill on weekly auction	November 1986	April 1992
14-day T-bill on weekly auction	April 1997	May 2001
365-day T-bill on fortnightly auction	April 1992	
91-day T-bill on weekly auction	January 1993	
182-day T-bill on weekly auction	Re-introduced in June 1999	May 2001
182-day T-bill on weekly auction	Re-introduced in April 2005	

Source: RBI statistics on government securities market 1 August 2007

1. 14-day treasury bill

This treasury bill was introduced in May 1997, on an auction basis. The objectives of introduction of 14-day treasury bills are to meet the cash management needs of various segments of the economy and facilitate the emergence of a comprehensive yield curve. These 14-day T-bill are different from the 14-day intermediate treasury bills. The former were issued through auctions on weekly basis. During 2001–02 the total issues aggregated to ` 1100 crore with 64 per cent being allowed to competitive bids. These bills were discontinued after 14 May 2001.

However, the latter was mainly introduced for investing the surplus funds of state governments, foreign central banks and other specified bodies. They were issued at a discount rate, equivalent to the ways and means advances to the Central Government. They were transferable and not issued through auctions.

2. 91-day treasury bill

Since 1932, treasury bills with a maturity of 91-day have been introduced. There are two types of 91-day T-bills, i.e., ordinary and ad hoc. The ordinary T- bills were issued on tap basis at the discount rates varying between 2 to 5 per cent, and 4 to 6 per cent. Under the tap sale, bills can be purchased on any day of the week. The actual rate of discount is not subject to fluctuations like the weekly auctions. The main purpose of this bill is essentially to park its surplus funds. However, the Tap T-bills were discontinued with effect from 1 April 1997 and uniform price auction method was introduced. In the tap issue, after the initial primary auction of a security, the issue remains open to further subscriptions. The period for which the issue remains open may be time specific or volume specific. They are now auctioned on weekly basis.

3. 182-day treasury bill

The 91-day T-bills were not able to smoothen the short-term liquidity requirement. Further, the money market has to be widened to provide outlets for the surplus funds. For this purpose, 182-day T-bill was introduced in November 1986 and was auctioned on a

NOTES

fortnightly basis from 1988. However, the issue of 182-day T-bill was discontinued in April 1992. The 182-day T-bills were reintroduced on 26 May 1999, and notified amount remained constant at ` 100 crore at the fortnightly auctions. The yield at cut-off price has remained stable at about 10 per cent. During 2001–2002, a notified amount of ` 300 crore was fully absorbed by competitors' bids with no devolvement on the RBI. However, the issue of these bills has been discontinued from 14 May 2001.

4. 364-day treasury bill

In April 1992, the 182-day treasury bill was discontinued and 364-day T-bill was introduced. The auctions are held at fortnightly basis. The 364-day T-bills are a flexible yield or variable yield instrument. The discount rates quoted by the market participants determine the rate of return. The 364-day bill became a safe avenue for investors and the auction of bill received a good response from the market participants. The bill offers short-term investment opportunity for the bank and other market participants. This bill could not be rediscounted with the RBI. The non-competitive bidders were allowed to participate from 1999 in the auction of 364-day T-bill.

5. Ad hoc Treasury bills

These were created in favour of RBI. It was a standing arrangement to provide the Central Government with enough funds to restore its working balance. These were phased out over a period of three years, i.e., from 1994–95, 1995–96, and 1996–97. At any time of the financial year, the net issue of the ad hoc treasury bills should not exceed ` 9000 crore for more than three continuous working days. The system of ad hoc T-bills was totally discontinued in 1977, and the existing bills were converted into special securities with out any specific maturity period.

Size of the Treasury Bills

The minimum denomination of 91-day treasury bills is ` 25000, i.e., the minimum amount of bids for the 91-day treasury bills are to be made for a minimum amount of ` 25000 and in multiples thereafter. The notified amount for the 91-day treasury bills was ` 250 crore but the amount has been increased to ` 1000 crore from December 2002.

The minimum amount of the 365-day treasury bills is ` 10,00,00. The bidding of treasury bills normally takes place in multiples of ` 1 crore.

Auctions

While 91-day T-bills are auctioned every week on Wednesdays, 182-day and 364-day T-bills are auctioned every alternate week on Wednesdays. The Reserve Bank of India issues a quarterly calendar of T-bill auctions which is available at its website (www.rbi.org.in). It also announces the exact dates of auction, the amount to be auctioned and payment dates by issuing press release prior to every auction. The following table shows this.

Table 9.2 *Auction Dates and Payment Dates*

Type of T-bill	*Day of auction*	*Day of payment**
91-day	Wednesday	following Friday
182-day	Wednesday of non-reporting week	following Friday
364-day	Wednesday of reporting week	following Friday

* If the day of payment falls on a holiday, the payment is made on the day after the holiday.

Price based or French auction

Under this method, all bids equal to or above the cut-off price are accepted. However, the bidder has to obtain the treasury bills at the price quoted by him. This results in differential price for different bids. Those who have offered higher prices in the bids have to pay more. This is known as 'winners curse'.

For example, the RBI conducted an auction to sell 91-day T-bills on 23 May 2007. The notified amount for the auction was ` 20 billion out of which ` 15 billion was an amount of MSS (Market Stabilization Scheme). At a multiple price-based auction, RBI received 66 bids worth ` 53.24 billion for the bills. However, it accepted only forty-five bids for the notified amount. The cut-off price determined at ` 98.13, offered a ytm (yield to maturity) of 7.6435 per cent. The weighted average price worked out to ` 98.14.

Uniform price based or Dutch auction

Under this system, all the bids equal to or above the cut-off price are accepted at the cut-off level. However, unlike the multiple price based method, the bidder obtains the treasury bills at the cut-off price and not the price quoted by him. This method is applicable in the case of 91-day treasury bills only. The RBI has done away with the system of Dutch auction, w.e.f., 8 December 2002 for the 91-day treasury bill.

The auction of treasury bills is done only at Reserve Bank of India, Mumbai. Bids are to be submitted on Negotiated Dealing System (NDS) by 2:30 pm on Wednesday. If Wednesday happens to be a holiday, then the bids are submitted on Tuesday. Bids are submitted in terms of price per ` 100. For example, a bid for 91-day treasury bill auction could be for ` 97.50. Auction committee of Reserve Bank of India decides the cut-off price and the results are announced on the same day. Bids above the cut-off price receive full allotment; bids at cut-off price may receive full or partial allotment, and bids below the cut-off price are rejected.

Partial pro-rata allotments are made to bids submitted at cut-off prices. The weighted average of the successful bids is calculated and allotments are made to non-competitive bidders like state governments and pension funds.

The notified amount is not fixed in the case of 365-day treasury bills. Bids are accepted on the basis of the yields that are bid for. There is no limit to the bid size. The bids must be quoted on a price basis with a precision of two decimal points. In 364-treasury bills, the bid amount has to be deposited within forty-eight hours. Purchase and sale of treasury bills takes place through subsidiary general ledger (SGL) account. The RBI maintains the SGL for investors like commercial banks, Discount and Finance House of India (DFHI), Securities Trading Corporation of India (STCI) and other financial institutions. If an investor does not have the facility of a SGL account, he may operate through the DFHI.

Payment

Payment by the allottees at the auction is required to be made by debit to their custodian's current account.

Purchase method

T-bills auctions are held on the NDS and the members electronically submit their bids on the system. Non-competitive bids are routed through the respective custodians or any bank or Primary Dealer (PD), which is an NDS member.

NOTES

Participants

All entities registered in India like banks, financial institutions, primary dealers, firms, companies, corporate bodies, partnership firms, institutions, mutual funds, foreign institutional investors, state governments, provident funds, trusts, research organizations, Nepal Rashtra Bank and even individual treasury bills are the participants.

Provident funds can participate in all the T-bill auctions either as competitive bidders or as non-competitive bidders. Participation as non-competitive bidders means that provident funds need not quote the price at which they desire to buy these bills. The Reserve Bank allots bids to the non-competitive bidders at the weighted average price of the competitive bids accepted in the auction. Allocations to non-competitive bidders are in addition to the amount notified for sale. In other words, provident funds do not face any uncertainty in purchasing the desired amount of T-bills from the auctions.

In the purchase of treasury bills, investors who do not have SGL account facility can buy it through the DFHI. DFHI keeps the treasury bills on behalf of their clients, pays the investor and collects the proceeds on the maturity day. Treasury bills can be purchased from DFHI at the DFHI counters by providing a cheque drawn on the RBI for the purchase price. DFHI provides the duly completed and signed SGL transfer form to the investor. If the investor wishes to sell, he has to tender a duly completed SGL transfer form at the DFHI counter. In return, a cheque drawn on the RBI is given by the DFHI. The treasury bills are not popular with corporate entities and the public. The commercial banks and RBI account for about 90 per cent of the holdings of the treasury bills.

Yield on Treasury Bills

The yield of a T-bill is determined by the rate of discount at which a particular T-bill is sold. The yield depends on the rate of discount, period of maturity and the issue price. The yield on a T-bill can be calculated with the help of the following formula:

$$D = \frac{FV - P}{V} \times \frac{360}{N} \times 100$$

D = discount

Fv = face value or nominal value or redemption value of T-bill.

P = purchase price or issue price of T-bill

N = maturity in days of T-bill.

Example: If the 91-day T-bills were issued at a fixed price of ` 97.50 for face value of ` 100, the discount rate will be:

$$\text{Discount} = \frac{100 - 97.5}{100} \times \frac{360}{91} \times 100 = 9.89\%.$$

Coupon equivalent can be found as follows:

$$\text{Coupon equivalent} = \frac{FV - P}{P} \times \frac{365}{N} = \frac{100 - 97.5}{97.5} \times \frac{365}{91} \times 100$$

$$= 10.28\%.$$

Effective yield also can be calculated as follows:

$$= \frac{(1 + FV - P)^{365/N} - 1}{P} \times 100$$

$$= \frac{(1 + 100 - 97.5)^{365/91} - 1}{97.5} \times 100$$

$$= (1.02564)^{4.011} - 1 \times 100 = (1.1068 - 1) \times 100$$
$$= 10.68\%$$

NOTES

T-bill Index

The need for T-bill index was felt because of the growth of secondary market operations and the increasing interest of the market participants to park their surplus funds in short-term treasury bills. T-bill index helps to measure the returns to the investors. The available T- bill indices are:

(a) JP Morgan Indian T- bill index

(b) NSET-bill index

(c) CCILT-bill index

JPMorgan India Treasury bill index is the first treasury bill index introduced by JP Morgan India in April 1996. It uses all the trades in T-bills. NSET-bill index was introduced in 1997. The construction of this index is based on zero-coupon yield curve methodology.

The CCIL T-bill index is built by Clearing Corporation of India Ltd. The index is based on individual T-bill trades of face value of ` 50 million or more. The CCIL has constructed two indices. CCILEqual Weight T-bill index and CCI Liquidity Weight T-bill index. The latter is constructed because the maturity period affects the liquidity of the bills. Some maturity buckets are more liquid than the others. Maturity buckets are different time intervals (for example, 1-14 days, 14-90 days, 91-180 days and 181-365 days), in which the value of a particular T-bill is placed depending upon its residual maturity. Hence, the trading volumes of T-bills with different maturity buckets are taken as weights for calculating the index. The weights are reviewed every month according to the trading volumes. The index time series data are available from 1 January 2004.

9.4.2 Advantages of Treasury Bill Market

A treasury bill is a major money market instrument. The issuer (the Central Government) and the market participants like banks and financial institutions are benefited by the treasury bills.

Advantages to Government

The RBI issues treasury bills on behalf of the government. These help the government to raise short-term funds to meet the temporary budget deficit, mop the excess liquidity in the economy and ease the pressure of inflation. Investors other than the RBI purchase the treasury bills but they cannot be monetized.

Advantages to Market Participants

1. Treasury bills are liquid in nature. DFHI offers two quotes every day, which makes the investors sell and purchase the treasury bills with ease as there is certainty in terms of availability, entry and exit.
2. They can be discounted with the RBI. RBI also provides refinance facility against them. Thus, market for the treasury bills is readily available.
3. They provide safety regarding the payment of interest and the repayment principal. These being sovereign papers, there is zero default risk. Treasury bills are one of

NOTES

the eligible securities to be held by commercial banks for statutory liquidity requirement (SLR). While maintaining the SLR requirement, the banks can earn the interest income from the treasury bills. This is one of the reasons for the commercial banks to invest in the treasury bills.

4. They are available on weekly and fortnightly auctions. The time mix of the investment can be planned properly by buying treasury bills of different maturity dates. It is ideal for fund management particularly for short-term tenors of less than fifteen days.
5. They can be parked in the constituent SGL account with the Discount and Finance House of India itself till maturity or resale. These transactions are hassle free. There is no need to handle securities in the physical form. There is transparency in operations as the transactions are put through the Reserve Bank of India's SGL or the client's gilt account only.
6. Investment in treasury bills provides a hedge against the interest rate fluctuations and volatility of call loan market.
7. They ensure better returns especially in the short-term and market related yields. The yield on 364-day treasury bills is issued as a benchmark for several floating rate bonds issued by the public sector undertakings and financial institutions and the RBI issued Floating Rate Bonds (FRBs) of four-year maturity with half-yearly coupon rates. The coupon rate was linked with the yield on 364-day treasury bills plus a margin of 1.25 per cent. However, the floor coupon rate was 13 per cent.
8. High degree of tradability and active secondary market facilities are available. Two way quotes are offered by primary dealers for purchase and sale of treasury bills.
9. Buy-back transactions for surplus treasury bills are available whenever there is a temporary difficulty for funds. However, the transactions can be reversed once the financial need is met. These transactions are available only for treasury bills. Treasury bills are also eligible for repo facilities.

Reforms in the T-Bill Market

- The practice of automatic monetization of the Central Government budget deficit through ad hoc treasury bills was eliminated with effect from 1 April 1997 and Ways and Means Advances (WMA) scheme was introduced.
- The rate of interest on WMA is linked to the bank rate. Consequent upon this, the outstanding ad hoc and tap treasury bills were converted and added to the stock of special securities with the RBI, without any specified maturity, at an interest rate of 4.6 per cent.
- Investors have the option to invest in 91-day, 182-day and 364-day treasury bills.
- To bring about more transparency in auctions, the RBI is now pre-announcing notified amounts in auctions of treasury bills (91-day, 182-day, and 364-day) and dated securities.
- Treasury bills are traded in the NSE and also through the brokers in the government securities.

Check Your Progress

7. State any four characteristics of treasury bills.
8. State the objectives of the fourteen day treasury bill.

9.5 COMMERCIAL BILL MARKET AND DISCOUNT MARKET

NOTES

Websites of many banks and financial institutions provide such information, 'Through our overseas branches and subsidiaries, we are very actively participating in financing national and international trade bills by way of discounting the same. The bills under the Letter of Credits (LCs) of most of the Indian as well as international banks can be discounted at competitive rates with us.'

Factoring and Bill Discounting

Like factoring, bill discounting is also a short-term finance. Bills of account receivables are discounted in bills discounting. However, the two financing arrangements of account receivables are different from each other. The dissimilarities are given below.

Table 9.3 *Differences between Bill Discounting and Factoring*

Bill Discounting	*Factoring*
1. Finance alone is provided	Both finance and integrated receivables management are provided
2. Advance is made against bills	Advance is made against trade debt
3. It is always with recourse	It is with or without recourse
4. Charge is registered with Registrar of Companies (ROC)	Owner of trade debt
5. Individual transaction oriented	Whole turnover-bulk finance is provided
6. Drawee has to accept the bill	Notification is sufficient for acceptance and all future transaction
7. It is not an off-balance-sheet mode of financing	The non-recourse factoring provides an off-balance-sheet mode of financing

Bill financing is an important mode of financing working capital requirements of trade and industry. Discounting of bills by the banks ensures adequate flow of working capital to the industry from the banks. While discounting a bill, the bank buys the bill, i.e., bill of exchange or promissory note before it is due and credits the value of the bill after a discount charge to the customer's account. The transaction is practically an advance against the security of the bill and the discount represents the interest on the advance from the date of purchase of the bill, until it is due for payment. Commercial bills are short term in tenor, self-liquidating in nature and can be rediscounted at RBI.

According to the Indian Negotiable Instrument Act, 1881, it is a written instrument containing an unconditional order signed by the maker, directing a certain person to pay a certain sum of money only to or to the order of a certain person or to the bearer of the instrument. In simple terms, a bill of exchange is a freely exchangeable instrument evidencing commercial transactions. This instrument evidences the liability to make payment on a fixed date when goods are bought on credit. The term 'good commercial bill of exchange' refers to a bill of exchange which exhibits a degree of creditworthiness, since it bears the signature of two or three addresses who are regarded as being solvent.

NOTES

Bill of exchange

The seller sells goods to the buyer. The buyer is not in a position to pay immediately and the seller wants the payment to be made. In this case, the seller writes the bill of exchange for a specified amount and period of maturity on the buyer. Thus, he draws a bill on the buyer and is called drawer. The drawer takes the role of the creditor. The buyer is the drawee. The seller sends the bill to the buyer for his acknowledgment in writing. The buyer acknowledges his responsibility for the payment of amount specified in the bill. The third party who is willing to assume credit risk can also sign the bill. Generally, acceptance houses give acceptance for a fee after due diligence. This makes the bill more secure. Now, the seller or the drawer can take the bill to any bank or financial institution and discount it. Thereby he gets the payment. Difference between the bill value and the actual amount is the discount charge.

Historical perceptive of bill market

Greeks were known to have used the bills for settling commercial transaction in the 4^{th} century BC. According to historical evidence, written document akin to the promissory note was prevalent in 12th century AD in Italy. Bonds came in to existence before the appearance of bills. Bonds had to be executed by a debtor before a court or public notary to make it legal. It was time consuming and expensive. This led to the invention of bills by the merchants. A bill was an informal document acknowledging the debt. Then the document came to be transferred from one person to another. In the course of time, it developed into a full fledged commercial and financial instrument.

The bills became negotiable during the sixteen century. The endorser of bills became liable on debt represented by the instrument in order of its endorsement. The endorsement gave a guarantee to the bill. However, it was in eighteen century in England that full legal sanctity was accorded to the bill exchange. A bill of exchange as *hundi* was used in India to finance the movement of agricultural produce in India. *Hundies* were drawn and payable at sight (*Darshan hundi*) or payable after a stipulated period mentioned in *hundi* (*Muddati hundi*). They were similar to the bill of exchange. Indigenous bankers used these *hundies* to raise money, remit funds and finance inland trade. Predominantly, *Multani Shraff* and *Chattays* dealt in *hundies*. Based upon the creditworthiness of the indigenous bankers, the bills were discounted by the commercial banks. With the structural changes that took place in the financial markets, *hundies* lasted their predominance as a tool of financing trade. The growth of commercial banking and financial institutions further paved the way for a more formal bill of exchange.

Features

- Three parties are involved in the bill exchange namely the drawer, the drawee and the payee. The writer of the bill of exchange who directs to pay is the drawer. The person, to whom the direction is given, is the drawee. The person to whom the payment is made is the payee. In some cases the drawer and the payee may be the same person.

  The drawer or the payee may keep the bill of exchange and he is the holder of the bill. If he endorses the bill, he is called the endorser. The person to whom the bill is endorsed is called the endorsee. If the endorsement gives the name of any person other than drawee who can be resorted in times of need, he is called the drawer in case of need. If the bill is dishonored by non-acceptance or non-payment, the drawee in need has to pay the committed amount.

- The sum payable and the person to whom it has to be paid is given in writing.
- Conditional payment is not possible in a bill. It is an order to pay and not a request.
- The order to pay should be in legal tender.
- The bill of exchange should be signed by the drawer and presented to the drawee for acceptance.
- Even though it is not legally specified, the date, number, place and consideration are written on the bill of exchange.

Direct discounting of bill

Letter of credit backed bill discounting and clean bill discounting are the convenient modes of financing for domestic trade transactions. Banks and financial institutions provide bill finance facility to its corporate/non-corporate clients. The facility of purchasing/discounting of bill is made available to those parties who have sanctioned limits for that purpose. Limits considered for purchase/discount of bills commensurate with the business requirements of the borrowers based on the normal credit assessment/ appraisal.

In the fund based bill finance facilities, post-sale credit against bills is offered as follows:

- Purchase of bills drawn under LC or confirmed order.
- Discounting of usance bills drawn under LC or confirmed order.
- Negotiation of documents under LC.

Assistance would be 100 per cent of the total value including insurance, taxes and freight. Interest rate/discount rate would be as prevalent at the time of discounting of bills, depending on monthly/quarterly/half-yearly/yearly payments and according to temporal profile of bills.

In the case of usance bills, normally the usance period should not exceed ninety days, unless otherwise specifically permitted, as in the case of foreign documentary usance bills, in which case usance period is extended up to a maximum of 180 days. In the case of demand documentary bills, documents of title to goods should be of a recent date, covering consignments of approved goods in which the customer deals in and drawn on places to which the customer usually consigns goods. The borrower has to deal with the same commodity/goods represented by bills and not others.

Discounting of export bills

After submission of the export documents, the exporter requests the bankers to purchase/ discount the export bill. The bankers will convert the entire bill amount taking the spot buying rate prevalent on the date of purchase and the premium for the tenor. In case there is any outstanding amount in the packing credit account, the transit period of the bill and the total bill amount will be adjusted to the packing credit account, else it will be credited to the exporter's current account. All the exporters are eligible to cover their bills drawn under letters of credit and non-credit bills under sanctioned limits in the bill rediscounting scheme.

Interest rate

The rate of interest is determined by the respective bankers, as per RBI directives. The interest rates charged by the banks depend on the tenure, amount, the letter of credit and the ratings of the borrower, e.g., Bank of Baroda charges for bills outside letter of credit

NOTES

having tenor of more than one hundred and eighty days but up to one year 1.50 per cent above benchmark prime lending rate, i.e., 14.00 per cent p.a. whereas it charges 0.50 per cent above BPLR, i.e., 13.00 per cent for bills drawn under letter of credit of other prime banks for the same tenure (Rate per annum w.e.f. 1 Jan 2009). The rates charged by Vijaya Bank are also given below:

(***Source:*** www. vijayabank.com)

(a) Rates for discount on bills covered by LC

(with effect from 12 Jan 2009)

DA up to 90 days	BPLR – 1 = 11.75% p.a.
Above 90 days to 180 days	BPLR – 0.50 = 12.25% p.a.
Above 180 days up to 270 days	BPLR = 12.75% p.a.

(exclusively for borrowers enjoying sanctioned bill discount limit)

(b) Discounting of bills other than mentioned under (a) above on standalone basis

DA up to 90 days	BPLR = 12.75% p.a.
Above 90 days to 180 days	BPLR +1 = 13.75% p.a.

For the borrowers whose rating is done by Accredited Domestic Rating Agencies whose Risk weight is rated at 20% (lowest Credit Risk], concession of 1% p.a. on the above applicable rate may be granted by the Competent Authority

(c) Discounting of external bills

Post shipment credit–(Rupee)

Against DP Bills – Rupee Credit

(a) For Transit Period	BPLR–2.50 = 10.25%
(b) Beyond transit period	BPLR + 2 = 14.75%

DA Bills - Documents are delivered on acceptance.

DP Bills - Documents are delivered on payment.

Importance of Bill Market

- Discounting of bills helps to meet the credit needs of trade and industry.
- Bills have the advantage of being self-liquidating in character because they mature after a fixed period.
- Next to cash, call loans and treasury bills, discounted bills are highly liquid in nature and backed by trade document.
- It creates a culture of discipline in the mindset of borrowers, as they have to honour the bills on the due date.
- Rediscounting facility is available for discounted bills. This helps the banks to meet their short-term financial needs.
- Surplus funds of the banks and financial institutions can be profitably invested in bills. The banks have to select bills with appropriate maturities to suit their financial plans.
- Efficient functioning of the bill market provides flexibility in the money market and banking system.

NOTES

Features of developed bill market

A well-developed bill market has the following features:

- Widespread culture of borrowing against commercial bills
- Presence of a large number of genuine trade bills
- Continuous and adequate supply of commercial bills in the money market and high velocity of circulation.
- Commercial banks and financial institutions have the facility for acceptance of bills at low cost.
- Commercial banks are ready to discount the bills and legal procedures are simple.
- Central bank is willing to rediscount the bills with minimum restrictions.
- Secondary market for commercial bills is developed to facilitate the rediscounting of the bills by the banks.
- There is a high degree of financial discipline in the case of drawers and drawees of the bills.

9.6 GOVERNMENT SECURITIES

The government's requirement of funds for developmental and non-developmental expenditure led to the origin of government securities market. Like any other body corporate, the government also issues securities of different maturity dates. These securities are traded in the debt market and serve as a benchmark for pricing of other debt instruments. The funds mobilized through government and semi-government securities are larger as compared to the funds mobilized by the industrial securities. The market for government securities is larger than the industrial securities in the UK and India, but not in the US.

Government Security (G-Sec)

A government security is a claim on the government and is commonly referred to as gilt. British government entered the details of government securities in a book with golden edges. This made the G-Secs to be called gilt edged securities. Public Debt Act defined gilts as securities issued by the Central or state government. Abroader definition includes the securities that are unconditionally guaranteed by the Central or state government.

All G-Secs in India currently have a face value of ` 100 and are issued by the RBI on behalf of the Government of India. G-Secs are normally coupon (interest rate) bearing and have semi-annual coupon or interest payments with a tenure ranging from five to thirty years. This may change according to the structure of the instrument. These securities are generally issued in the form of entries in RBI's Subsidiary General Ledger (SGL). However, it can be issued in physical form for retail lots. Government securities are highly liquid instruments available both in primary and secondary market.

Check Your Progress

9. How is the post-sale credit against bills offered?
10. State any five features of a well-developed bill market.

Instruments

Most of the government securities were in the form of plain vanilla, i.e., fixed coupon securities prior to 1990s. However, presently the coupon on such securities is announced before the date of flotation. Specified coupon rate remains fixed during the term of the security. The securities may be issued at par or at a discount and redeemed at par. Later

on, the government issues a wide variety of bonds to meet the diverse funding and hedging needs of its participants. Some of these are given below.

NOTES

Dated Securities

These are issued at the face value. Generally, they have a fixed maturity period and carry semi-annual coupons. Coupon or interest rate is fixed at the time of issuance and remains constant till the redemption of security. The security is redeemed at par (face value) on its maturity date. These securities are called dated securities because they are identified by their date of maturity and the coupon, e.g., 11.03 per cent GOI 2012, is a Central Government security which matures in 2012 and carries a coupon of 11.03 per cent which is payable half yearly.

Partly Paid Stock

It is floated when the government does not need funds immediately. The first issue of partly paid stock was made on 15 November 1994, for ` 2000 crore with eight years of maturity. Thereafter, it has been issued a few more times.

These are issued at the face value, but this amount is paid in instalments over a specified period. Coupon or interest rate is fixed at the time of issuance and remains constant till redemption of the security. The tenor of the security is also fixed. Interest/ coupon payment is made on a half-yearly basis on its face value. The security is redeemed at par (face value) on its maturity date.

Floating Rate Bonds (FRBs)

Debt instruments that have variable coupon rates rather than fixed coupon rates but are based on a predefined benchmark are known as floating rate debt instruments. For example, the +1.50 per cent Floating Rate Bond (FRB) 2020 (the maturity year) means that +1.5 per cent coupon will be over and above a benchmark rate. The benchmark rate may be a six-month average of the implicit cut-off yields of 364-day treasury bill auctions. If this average works out to be 6.50 per cent per annum, then the coupon will be established at 6.50 per cent + 1.5 0 per cent, i.e., 8.0 per cent per annum. Normally FRBs (floaters) also bear a floor and cap on interest rates. Interest so determined is intimated in advance before the semi-annual coupon payment.

Zero Coupon Bonds

These bonds bear no coupon payment. They are issued at a discount to the face value. The discount provides the implicit interest payment. In effect, zero coupon bonds are like long duration T-bills. RBI determines the cut-off price on the basis of the bids received through tenders of the investors.

Embedded Bonds

Here, the bonds have call and put option. If the call option is specified in the notification, the securities are repaid to the holder of the security, at the option of the government, before the specified redemption date. However, if neither a call option nor a put option is specified, the prefixed amount is repaid on the date as notified in the issue.

RBI issued a bond with call and put option in 2002, which is due for redemption in 2012, carries a coupon of 6.72 per cent and has call and put option after five years, i.e., in the year 2007. In other words, the holder of the bond can sell it back (put option) to the government in 2007 or the government can buy back (call option) the bond from the holder in 2007. This bond has been priced in line with five year bonds.

Capital Indexed Bonds

In these, G-Sec interest rate is a fixed percentage over the wholesale price index. This provides investors with an effective hedge against inflation. These bonds were first floated on 29 December 1997, on tap basis. The principal redemption is linked to the Wholesale Price Index (WPI). They are issued at face value. Coupon or interest rate is fixed as a percentage over the wholesale price index at the time of issuance. Therefore, the actual amount of interest paid varies according to the change in the wholesale price index. The tenor of the security is fixed. Interest/coupon payment is made on a half-yearly basis on its face value.

Strips

Separate Trading of Registered Interest and Principal of Securities is known as STRIPS. A conventional security is stripped into a number of zero coupon securities and is traded separately. Such newly created securities are called STRIPS. For example, a five year government dated security can be stripped into eleven zero coupon securities; ten carrying half yearly coupons with maturities six months, twelve months, eighteen months and so on and one carrying final redemption amount with a maturity of five years. A ` 100 crore gilt carrying a coupon of 10 per cent with 5 years maturity has a cash flow of 10 semi-annual payments of ` 5 crore each and repayment of principal of ` 100 crore after five years. Each of this eleven cash flows can be treated as a zero coupon instrument which can be traded at varying yields. These eleven instruments are the STRIPS of the underlying government dated security.

STRIPS help to expand the breadth of debt market, as one underlying security is converted into eleven zero coupon securities. Increased supply of securities of various maturities provides a continuous market and consequently improves liquidity. Since STRIPS are issued at a discount, it is easy to arrive at the yield of these instruments. When STRIPS are derived from gilts, they are free from default risk.

State Government Securities

State government securities are securities/loans issued by the Reserve Bank of India on behalf of various state governments for financing their developmental needs. The Reserve Bank of India auctions these securities from time to time. These auctions are of fixed coupon, with pre-announced notified amounts for different states.

The Issuer

These securities are mainly issued by the Central Government, state governments and semi government authorities like city corporations and municipalities. The Reserve Bank of India mainly issues G-Secs on behalf of the Government of India. This forms a part of the borrowing programs approved by the Parliament in the union budget.

Mode of Issue

The government securities are issued in two forms. These are:

Physical Form

Securities are issued in the form of a certificate. An investor can dematerialize the physical certificate into an electronic mode.

NOTES

SGL Form

It is issued in the account form. Account is maintained by the client with the RBI or any other approved entity. Account is maintained in the SGL. SGLmeans Subsidiary General Ledger. Investment made by the SGL entity on its own account is held in SGLI account. Investments of all the constituents are held in the SGL-II or CSGL account.

The Public Debt office undertakes the responsibility of issuing the gilt edged securities. Issuing of government securities is notified through government notification and press communique. Prospectus is not issued. The SEBI gives a no objection certificate for the issue of government securities as a routine procedure. There are no separate listing requirements for them, since they are deemed to be listed in the stock market.

The issue is kept open for two to three days. The applications for securities are received at the RBI office and at the branches of State Bank of India. To enable the subscription of new loans, RBI generally suspends the sale of existing loans till the closure of the fresh loans. The government has the right to retain a certain percentage of over subscription up to 10 per cent excess of the notified amount. In case the issue is approximately subscribed to the issue amount before the date of closure, it will be closed.

In the case of state government securities, over subscription of one state government can be transferred to another state government loan, which is open for subscription. This can be carried out only with the option of the subscriber. The RBI follows either one of the following trading practices.

Sale of G-Secs

The various modes of sale of G-Secs have been listed.

Auctioning

The Government of India notifies the auction of government securities. It also notifies the amount, whether it will be a new loan or reissue of an existing loan. It also announces whether the bidders have to bid for the price or the coupon (interest rate). Auctions for government securities are either yield-based or price-based. In a yield-based auction, the bidders submit bids in terms of the yield at which they are ready to buy the security. In a price-based auction, the Reserve Bank of India announces the issue size (or notified amount), the tenor of the paper to be auctioned, as well as the coupon rate. The bidders submit bids in terms of the price. This method of auction is normally used in the case of reissue of existing government securities.

The auction can be either a multiple price or a uniform price. In the multiple price or French auction, participants get allotments at their quoted prices/yields. However, in the uniform price or Dutch auction, all the participants get allotments at the same price. For example, Government of India announced to sell (reissue) 7.50 per cent Government Stock 2034 for a notified amount of ` 2,000 crore through a price based auction using multiple price method dated 31 December 2004.

The competitive bidders put in bids for the price or the coupon. Then the RBI announces the cut-off price or yield on the basis of the bids received. Cut-off yield is the rate at which the bids are accepted. Bids which are at higher yields than the cut-off yield are rejected and the ones which are lower than the cut-off are accepted. Cut-off price is the minimum price accepted for the security. Bids at prices lower than the cut-off are rejected and higher than the cut-off are accepted. Coupon rate for the security remains unchanged. Bidders who bid at higher than the cut-off price pay a premium on the

security, thereby getting a lower yield. Price-based auctions lead to finer price discovery than yield-based auctions. All successful bidders are allotted the securities auctioned either in full or in part.

NOTES

Non-competitive Bidding

The Reserve Bank of India announced a facility of non-competitive bidding in dated government securities on 7 December 2001 for small investors. In non-competitive bidding, the bidder is able to participate in the auctions of dated government securities without having to quote the yield or price in the bid. Non-competitive bidding is open to individuals, HUFs, firms, companies, corporate bodies, institutions, provident funds, trusts and any other entity prescribed by RBI. As the focus is on the small investors lacking market expertize, the scheme is open to those who:

- Do not have a Current Account (CA) or Subsidiary General Ledger (SGL) account with the Reserve Bank of India
- Do not require more than ` 1 crore (face value) of securities per auction

As an exception, Regional Rural Banks (RRBs), Urban Cooperative Banks (UCBs) and Non-banking Financial Companies (NBFCs) can also apply under this scheme in view of their statutory obligations. However, the restriction regarding the maximum amount of one crore per auction per investor will remain applicable.

Non-competitive bids are allowed up to 5 per cent of the notified amount in the specified auctions of dated securities. The reserved amount is within the notified amount. Eligible investors cannot participate directly. They have to necessarily come through a bank or Primary Dealer (PD) for auction. The minimum amount for bidding is ` 10,000 (face value) and in multiples of ` 10, 000. An investor can make a single bid through any bank or PD under this scheme, in each specified auction. In case, the aggregate amount bid is more than the reserved amount through non-competitive bidding, allotment would be made on a pro rata basis. The non-competitive bidders pay the weighted average price, which emerges in the auction.

Tap Sales

When the issue of the bonds is announced, a part of the issue is taken over by the RBI. Gradually, the RBI sells the bond in the ensuing period. As the bonds are continuously available, they are called ‘taps’. The closure of tap issue is also notified, e.g., 10.79 percent government stock with maturity in 2015, was issued on 19 May 2000 and was closed on 23 May 2000. Sometimes, the tap issues are closed on the same day, e.g., on 2 July 1999, 11.90 per cent government stock was closed on the same day.

Grooming

The securities are issued in blocks with a single maturity date. If bonds are issued, the Reserve bank of India buys it and gradually sells it in the stock market. Gradually, the RBI repurchases the bonds through the stock market leaving a small portion to be redeemed at the maturity date. This gradual purchase of bonds is called ‘grooming’. This leads to a continuous process of issue and redemption.

Switching

In its open market operations, RBI purchases one security against the sale of another security. This is called switching. This method helps the banks and institutions to improve their yield on their investments on government dated securities. RBI fixes the annual quota for each bank.

NOTES

Private Placement

The Central Government may also privately place government securities with Reserve Bank of India. This is usually done when the Ways and Means Advance (WMA) is near the sanctioned limit and the market conditions are not conducive to an issue. The issue is priced at market-related yields. RBI may later offload these securities to the market through Open Market Operations (OMO).

Open Market Operations (OMO)

Government securities that are privately placed with the RBI are sold in the market through its open market operations. The yield at which these securities are sold may differ from the yield at which they were privately placed with RBI. Open market operations are used by the Reserve Bank of India to infuse or reduce liquidity from the system. Whenever the RBI wishes to infuse liquidity in the system, it purchases government securities from the market and whenever it wishes to suck out the liquidity from the system, it sells government securities in the market.

Repayment

Government securities are repaid at par on the expiry of their tenor. Different repayment methods are followed. For SGL account holders, the maturity proceeds are credited to their current accounts with the Reserve Bank of India. For gilt account holders, the bank/primary dealers receive the maturity proceeds and pay the gilt account holders. For entities having a demat account with NSDL, the maturity proceeds would be collected by their DP's and they in turn would pay the demat account holders.

Market Players

All entities registered in India like banks, financial institutions, primary dealers, firms, companies, corporate bodies, partnership firms, institutions, mutual funds, foreign institutional investors, state governments, provident funds, trusts, research organizations, Nepal Rashtra bank and even individuals are eligible to purchase the government securities.

Banks, provident funds, insurance companies and financial institutions are the main buyers and traders of gilts. These securities are bought by the banks to manage their cash flows and meet their statutory obligations as per the RBI guidelines. Earlier RBI used to hold a large quantum of central government securities. However, at present it has declined it to minimum.

9.6.1 Role of Government Securities Market

The role of government securities market can be explained as follows:

1. To finance developmental expenditure

It is considered appropriate to finance a part of the developmental expenditure through issue of government securities. The increasing outlay in the successive plans cannot be met only through taxation, surpluses of public undertaking and deficit financing, since these sources of finances are also subject to certain limitations. Therefore, these sources of finance have to be supplemented by borrowing a large scale from the public, banks and other financial institutions.

2. To refund the existing loans

The issue of securities may be undertaken for refunding of the existing securities. This may be for conversion or refunding of maturing securities and advance refunding of securities that have not yet matured. The issue undertaken for repayment of the existing loan is also known as reissue of loans. Refunding can be carried out by two ways, either by selling new securities for cash settlement or by offering the right to exchange old securities for new issues to the holders. The conversion and reissue of loans lengthens the maturity structure of the government debt and reduces the cash repayment of the loans.

NOTES

3. To promote savings and investment

Investment in government securities has been recognized to be a very safe investment. If an investor puts a part of his funds in government securities, which are safe yielding assets, he can invest the rest of his funds in risky ventures, because he has already invested in government securities. Government securities thus, provide a safe investment outlet to the investors and enable them to contribute to the development of the economy by investing these funds in the equity capital of the various industries.

4. To strengthen the debt market

The growth of government securities helps the development of the debt market. The introduction of different types of securities like capital index bonds and floating rate bonds has helped to establish an expanding debt market. The growth of the G-Secs has also given birth to the gilt edged mutual funds. Issuing of securities is not only for securing loans with a minimum interest rate but also for fostering the growth of the money and capital markets.

5. To implement monetary policy

The issue of government securities provides an opportunity to make the monetary policy an effective instrument of economic policy. The structure of interest rates of the G-Secs influences the general interest rate structure of the capital market. Further, to contain inflation, the bank liquidity has to be controlled. Through the change in the Cash Reserve Ratio (CRR) and Statutory Liquidity Ratio (SLR), the banks are made to buy the G-Secs. This affects the advances and liquidity of the commercial banks.

It helps the government to meet short-term and long-term fund requirement. It assists the government to make payments to the international organizations like IMF, World Bank, and Asian Development Bank. It also facilitates the conduct of the open market operations effectively.

9.7 MARKETS FOR DERIVATIVES: TYPES, USES AND PRICING

Check Your Progress

11. What are floating rate debt instruments?
12. What are STRIPS?
13. State the concept of grooming securities.

Options, like futures, are also derivatives. An option is a legal contract, which gives the holder the right to buy or sell the underlying asset, at a specified price, on a specified date. Although it gives the holder the right to buy or sell the underlying asset, he is not obligated to do so. This is the basic difference between option and futures. In futures, the holder is obligated to sell or buy the underlying asset according to the nature of contract entered into by both the parties.

NOTES

Two parties are involved in the options contract, a buyer and a seller. The buyer who buys an option takes a long position. The seller who sells an option takes a short position, i.e., he writes the option and is the writer of the option. The buyer is the holder of the option. Physical certificates are not created when options are written. Book keeping entry is the form of transaction in option trading. The underlying assets of the options are; selected agricultural products, foreign currencies and stocks.

Call option

It gives the buyer the right to buy (call away) a specified underlying asset from the option writer at a specific price at any time up to or on a specific date. If it is in the case of stocks, the contract consists of the:

- Name of the company whose shares can be bought
- Number of shares to be bought (lot size)
- Exercise price or strike price of the shares at which the shares can be bought
- Expiration date, i.e., the date at which the right to buy expires
- Premium or purchase price

It can be explained clearly with an example. Consider 'B' and 'C' who are willing to sign a call option contract. 'B' is the buyer of the call option and 'C' is the writer of the call option. If we take the underlying asset as a share, then this contract allows B to buy 100 ICICI Bank shares for ` 400 at any time during the next three months. Assuming that ICICI Bank share is selling at ` 350 per share on NSE, 'B' thinks that the share price will rise in future. However, 'C' has a different opinion and believes that its stock price will not rise above ` 400 during the next three months. If the price increases above ` 400, 'B' can call the shares, which means that he has the right to buy 100 ICICI Bank shares. At the same time, if the price falls below ` 400, he has the option of not exercising the option, i.e., he may not buy the stock. What he loses is only the premium, which he paid to the writer of the call option.

Put option

A put option gives the right to sell the underlying asset at the specified price on or before the specified date. For example, X is a farmer who is not certain about the prices of apples which are now ` 400-500 per box. He fears that the price of apples would fall below ` 300 per box. The dealer too is not sure of the price, but is willing to have the deal because in his opinion, the price would remain between ` 320-350. The dealer writes the option and the farmer buys it. The contract states that the dealer is willing to buy apples per box at ` 300 after three months. At the same time, if the farmer gets a better price, he may not sell the apples. To compensate, the dealer asks for a token money and the farmer pays it. If the farmer does not sell the apples, the token money is a loss to him but he can sell the apples at a better price in the market. The farmer has no obligation to sell. However, the dealer has the obligation to buy, if the apples are sold to him.

In the above example, the farmer would sell the apples only if the price falls below ` 300. He can walk out of the deal if the prevailing price is above ` 300 and he can sell it in the open market. Thus, puts have the following salient features.

- A put owner has the right to sell an asset at a certain price within the specified period of time.
- Put owner who is the buyer of the put, is not obligated to sell. He has the choice to exercise his right to sell.

- The seller of the put, i.e., the writer of the put has an obligation to buy.
- Puts are very similar to having a sell position on an asset.

This can be further explained with the help of a stock market example. Assume that ICICI Bank stock is selling at ` 390 and X purchases August put option with the strike price of ` 400; thus paying a premium of ` 10. The option will be exercised on the expiry date if the spot price of ICICI Bank stocks on that date is less than ` 400. Since the options are cash-settled, the put holder receives the difference on exercising the option. In case the price of ICICI stock is more than ` 400, the put holder may not exercise the option and would rather sell it at a higher price in the open market.

Assets of the Option

The asset that can be bought and sold with an option is called an underlying asset or underlying. Options are written on the following assets:

- Stocks
- Stock indices
- Foreign currencies
- Commodities
- Options on the futures, etc.

Option Terminology

The option terminology is as follows:

Position of the Option

The option may be at-the-money, in-the-money or out-of-the-money. When the strike price is equal to the spot price of the option, it is known to be at-the-money in European style option, which can be exercised only at the maturity date.

The call option is in-the-money if the strike price is lower than the spot price. The put option is in- the-money if the strike price is higher than the spot price.

Call option is out-of-the-money when the strike price is higher than the spot price. Put option is out-of-the-money when the strike price is lower than the spot price.

This can be indicated in a tabular form in Table 9.4 where St indicates the current stock price and X is the strike price.

Table 9.4

Situation	*Call option*	*Put option*
St > X	In-the-money	Out-of-the-money
St < X	Out-of-the-money	In-the-money
St = X	At-the-money	At-the-money

Premium

It is the value or price of an option that option buyer pays to the option seller at the time of signing the contract. It is not paid back even when the option is not exercised. It is also known as option value or option price. Premium consists of intrinsic value and time value.

NOTES

Intrinsic Value

It denotes the amount by which the call is in-the-money. It represents the gain accruing to the holder on exercise of the option. An out-of-the-money option has no intrinsic value. The excess of current price over strike price is the intrinsic value of the call option.

Intrinsic value of call option) = Max (Stock Price-Strike Price,0)

The intrinsic value of the put option is the excess of strike price over the current price.

Intrinsic value of put) = Max (Strike Price-Stock Price,0)

When the options are in-the-money, they have an intrinsic value. An option which is out-of-the-money has no intrinsic value. Intrinsic value is always positive and cannot be negative because the out-of-the- money option is not exercised.

Time Value

It is the difference between the premium and the intrinsic value of an option. In other words, time value of an option represents the amount of money that the buyer is willing to pay over and above the intrinsic value. For a call option in-the-money, time value exists if the call price is greater than the intrinsic value St-X. Longer the time of a call to mature, the greater is the time value.

For a put option which is in-the-money, time value exists if the premium exceeds the intrinsic value X- St. The entire premium is the time value for put and call options when they are at-the-money or out-of-the-money.

Time Value ——→ Maximum at the start of the contract

——→ Zero at expiry

9.7.1 Factors Affecting Option Premium

It is affected by the volatility in the stock prices and some other factors as indicated below:

1. Stock Price

A call option is exercised when the stock price is higher than the exercise price. It is the pay-off for the holder. The call option is more valuable when the stock price increases and less valuable when it decreases. However, it is reverse in the case of put option.

2. Exercise Price

In case of two calls on a stock with same exercise date with varying exercise price, the call option with a higher price cannot be expected to be valued higher than the call option with lower exercise price. Suppose the two calls on a stock with same exercise dates have the exercise price of ` 200 and ` 210 respectively. The holder of the call option of ` 200 is in a better position to buy the stock at ` 200 than the other person who holds the option at ` 210. It is thus clear that the calls with a higher exercise price are not valued higher than the calls with the lower exercise price. The reverse is true with the put option.

3. Maturity Period

Here, it means the length of time to expiration. The effect of time depends on whether the option is American or European. The call and put options of American style are more

valuable if the time to expiration increases. The call option with a longer life offers the investor more opportunities to exercise than the one with the shorter life. If there are two calls on same stock with same exercise price, the call with a longer life is preferred to the call with a shorter life.

European options can be exercised only at maturity. If the dividend is declared within the maturity period, the price of the underlying stock would decline after the declaration of the dividend. This has a negative influence on the pay-off of the holder. In this situation, the option with a limited time to expiration is of more value than the call option with a longer time to expiration.

4. Dividend

It affects the value of stock and options. The prices of the stock are low in the ex-dividend period than in the pre-dividend period. In the ex-dividend period, the call option price is adversely affected and put option price is positively affected because the stock prices fall during that period. The quantum of dividend expected also affects the option price. Higher the dividend expected, higher is the price of the call option.

5. Interest Rate

Risk free rate of interest affects the option prices indirectly. Increase in the risk free rate of interest leads to an increase in the price of call options. The higher the risk free rate, the higher is the value of call option. The reverse is true with the put option.

6. Volatility

It shows the range of fluctuation in the underlying asset price within a certain period. There are two kinds of volatility, viz., historical volatility and implied volatility.

- **Historical Volatility**

It measures the price changes in the underlying asset in a particular period. The historical volatility affects the pricing of the option. Investor analyses the movement of the stock prices and the direction of the movement before buying the options.

- **Implied Volatility**

The implied volatility (IV) is the market's perception of the volatility of the underlying security. The pricing models help to find out the implied volatility. The Black-Scholes Model uses standard deviation(s), which indicate(s) the historical volatility to calculate the price of the option. For a given market price, one can calculate the standard deviation, by using the formula by trial and error process. The derived standard deviation () is the IV. This IV helps to forecast how volatility is likely to move in the future.

The calculated IVs are compared with the past values of IVs. If the present implied volatility is higher than its past values, the prices may return to the normal level unless there is a fundamental change in the economy, market or company.

High-implied volatility shows that the option prices are higher than the theoretical price and provide positive premium to the call and put holders. At the same time, call and put writers receive a negative premium. Low implied volatility indicates a negative premium to the call and put holder and a positive premium to the call and put writer.

9.7.2 Valuation Models

Several models are available for valuing the options. However, the most important models are binomial option pricing model and Black-Scholes option pricing model. Both the

models rely upon changes in the prices of the underlying asset or stock and are suitable for European style.

1. Binomial Model

NOTES

The binomial model assumes that the percentage change in share price follows a binomial distribution. Here, the option type is assumed to be of European type and the underlying stock does not pay any dividend during the life of the option.

Consider a stock whose present price is ` 100. Suppose that over the next year, the stock price can go either up by 10 per cent or down by 5 per cent, so that the stock price at the end of the year is either ` 110 or ` 95. There also exists a call on the stock with exercise price ( E) ` 100 and the time to expiration is of one year. If the price increases to ` 110, the buyer gains ` 10. If the stock price declines to ` 95, the gain is ` 0. If the risk free rate of interest on the one year bond is 8 per cent, then these three assets will have the following payoff patterns:

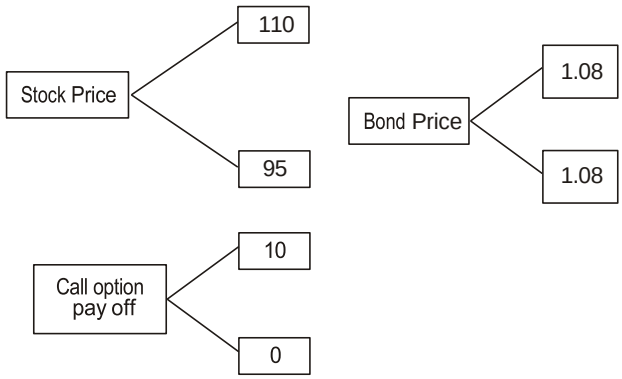


Fig. 9.2 *Bionomial Model*

The option payoffs can be replicated by a linear combination of the stock and the bond. This combination helps to define the price of the call option. The number of shares may be denoted by 'S' and the number of bonds by 'B', which exactly replicate the option's payoffs. This gives the following system of linear equations to solve:

$$110\ S + 1.08\ B = 10$$

$$95\ S + 1.08\ B = 0$$

Solving these equations gives S= 0.67, B = – 58.94, thus, purchasing 0.67 of a share of the stock and borrowing ` 58.94 at 8 per cent for one period will give payoffs of ` 10 if the stock price goes up and ` 0 if the stock price goes down. It follows that the price of the option must be equal to the cost of replicating its payoffs, i.e., call option price:

$$= 0.67 \times 100 - 58.94$$

$$= ` 8.06$$

Thus, $C = N_s P_s + N_b P_b$

C denotes call value of the option, P_s denotes price of the underlying stock Pb denotes price of the risk free bond, N_s and N_b are the number of shares and risk free bonds required to replicate the options pay-off. The cost of replicating portfolio gives the fair value of the option.

This logic is called 'pricing by arbitrage'. If two assets or sets of assets (the call option and the portfolio of 0.67 of the stock and – 58.64 of the bonds) have the same payoffs, they must have the same market price as well.

2. Black-Scholes Model

Fisher Black and Myron Scholes gave a model for valuation of options. According to Black and Scholes, the formula is,

$$C = St\ N(d_1) - X\ e^{-rt}\ N(d_2)$$

$$d_1 = \frac{\ln(St / X) + (r + s^2 / 2)t}{s\sqrt{t}}$$

$$d_2 = \frac{\ln(St / X) + (r - s^2 / 2)t}{s\sqrt{t}}$$

or $d_2 = d_1 - s\sqrt{t}$

C = value of the call option

r = continuously compounded risk free rate of return

St = current price of the stock

X = exercise price of the option

t = time remaining before the expiration date (expressed as a fraction of a year)

σ = Standard deviation of continuously compounded annual rate of return

ln = natural logarithm

N(d) = value of the cumulative normal distribution evaluated at d

Put option value can also be found out using the following formula

$$P = X\ e^{-rt}\ N(-d_2) - St\ N(-d_1)$$

Assumptions of the Black-Scholes model

- Options do not have the possibility of early exercise. The option is written on European style.
- There are no transaction costs and taxes.
- The risk free rate of interest is known and is constant over the life of the option.
- The market is assumed to be efficient.
- The underlying security pays no dividend during the life of the option.
- The volatility of the underlying stock is known and is constant over the life of the option.
- The continuously compounded rate of return of the shares follows a normal distribution pattern.

Example: The following information is given about Apex company stock.

Current price of the share St = ` 470

Exercise price of the option X = ` 460

Time period to expiration = 2 months t = 2/12 years

s of the rates of return = 0.4

continuously compounded risk free interest rate r = 0.06

Calculate the value of the call and put options. The call option is currently traded at ` 33.

NOTES

NOTES

Solution:

$$C = St\ N(d_1) - X\ e^{-rt}\ N(d_2)$$

$$d_1 = \frac{\ln(St/X) + (r + s^2/2)t}{s\sqrt{t}}$$

$$d_2 = \frac{\ln(St/X) + (r - s^2/2)t}{s\sqrt{t}}$$

$$d_1 = \frac{\ln(470/460) + (0.06 + (0.5 \times 0.4^2))(2/12)}{0.4\sqrt{2/12}} = 0.2746$$

$$d_2 = \frac{\ln(470/460) + (0.06 - (0.5 \times 0.4^2))(1/12)}{0.4\sqrt{2/12}} = 0.1113$$

$N(d_1) = Z = 0.2746$ the area $= 0.6064$

$N(d_2) = Z = 0.1113$ the area $= 0.5438$

$C = St\ N(d_1) + Xe^{-rt}\ N(d_2) = 470 \times (0.6064) - 460 \times e - 0.06 \times 2/12\ (0.5438)$

$= ` 37.36$

If the call is selling at ` 33, the call is undervalued. In such a case, the investor can purchase a call.

Likewise the value of put option value can also be determined.

$$P = Xe^{-rt}\ N(-d_2) - St\ N(-d_1)$$

$$= 460\ e^{-(0.06)(2/12)}\ 0.4562 - (470 \times 0.3936) = ` 22.76$$

9.7.3 Types of Options

The various types of options are as follows:

1. Options on Futures

In this option, the underlying asset is a future contract. Many exchanges provide for trading of options on futures. The future contract matures after the expiration of the option. In the options on futures, if the call option holder decides to exercise it, the investor acquires a long (buy) position in the underlying future contract along with the cash amount equal to the excess of the futures price over the exercise price. If a put option on futures contract is exercised, then the writer must accept the delivery of the appropriate futures contract from the buyer. The holder of the put option assumes the short position and earns a cash amount equal to the excess of exercise price over the future price.

There is a distinction between the future and options on futures. If the investor buys a future contract, he gains if the price increases and loses if there is a substantial drop in price. In case he has bought a futures call option on the asset, his gains are similar if the price increases. However, if the price of the underlying asset falls, he may not exercise his option and thus only the premium paid is lost.

Likewise, if an investor is selling a future contract, he can make money if the price of the asset declines. However, if the price of the asset increases, the investor loses money. In an option contract, if an investor buys a futures put option on an asset and if the price of the asset declines, he makes profit. Unlike futures, if the price of the asset increases, the investor may not exercise his option and he loses only his premium.

2. Options on Indices

The options contracts, which are based on some index, are known as index options contract. In the beginning, futures and options were permitted only on S&P Nifty and BSE Sensex. Then, sectoral indices were also permitted for derivatives trading subject to fulfilling the eligibility criteria. If 80 per cent of the index constituents are individually eligible for derivatives trading, then derivative contract is permitted on an index. However, no single ineligible stock in the index shall have a weightage of more than 5 per cent in the index. The index is required to fulfill the eligibility criteria even after a derivative trading on the index has begun. If the index does not fulfill the criteria for three consecutive months, then derivative contracts on such index would be discontinued.

NOTES

Mini options contracts on the leading Indian equity index like Sensex and Nifty are also offered to make the retail investors participate in the ever growing derivatives market.

- It requires lower capital outlay and lower trading costs.
- Smaller size allows for more precise hedging and flexible trading.

The market lot for the mini Sensex option is five. Market lots for NSE indices are given below:

Table 9.5 *Lot Size of NSE Indices*

Underlying	*Symbol*	*Market Lot*
BANK Nifty	BANKNIFTY	25
CNX 100	CNX100	50
CNX IT	CNXIT	50
CNX Nifty Junior	JUNIOR	25
Nifty Midcap 50	NFTYMCAP50	75
S&P CNX Nifty	NIFTY	50
S&P CNX Nifty	MINIFTY	20

3. Options on Individual Securities

NSE was the first exchange to launch trading in options on individual securities in India. Trading commenced from 2 July 2001. Option contracts are of American style and cash-settled. At present, options on 338 securities are permitted by the SEBI. Option contracts have a maximum of three month trading cycle near month (one), the next month (two) and the far month (three) contracts. New contract prices are introduced at new strikes for both call and put options on the trading day following the expiry of the near month contract. These contracts are introduced for a duration of three months.

Strike Price

Generally, stock exchange provides a minimum of five strike prices for every option type (call or put) during the trading month. At any time, there are two contracts in-the-money (ITM), two contracts out-of-the-money (OTM) and one contract at-the-money (ATM). The NSE provides seven contracts:

- three in-the-money
- one at-the-money
- three out-of-the money

NOTES

New contracts with new strike prices for existing expiration date are introduced for trading on the next working day, based on the previous day's underlying close values, as and when required. The price steps in respect of individual securities option is ` 5. Base price of the option contract is the theoretical value of the options contract, based on the Black-Scholes model calculation of option premium.

Price Bands

There is no day-minimum and maximum price range applicable for an option contract. However, in order to prevent erroneous order entry, operating ranges for option contracts are kept 99 per cent of the base price. In view of this, members are not allowed to place orders at prices beyond 99 per cent of the base price. Members desiring to place orders in option contract beyond the day minimum range have to send a request to the exchange. The exchange may change the base price based on the request of the trading members.

Contract Size

Lot size refers to number of underlying securities in one contract. Based on the recommendation of the Standing Committee on Finance, SEBI has specified that the value of a derivative contract should not be less than ` 2 lakh at the time of introducing the contract in the market. The lot size is determined keeping in mind the minimum contract size requirement at the time of introduction of derivative contracts on a particular stock.

For example, if shares of ABC Ltd. are quoted at ` 1600 each and the minimum contract size is ` 2 lakh, then the lot size for that particular scrip stands to be 200000/1600 = 125 shares, i.e., one contract in ABC Ltd. covers 125 shares.

However, over a period of time, there has been variation in the prices of the underlying stocks; the minimum contract size/value of most derivative contracts has far exceeded the prescribed value of ` 2 lakh. In case of some derivative contracts, due to the fall in the price of the underlying stock, the minimum contract size/value has fallen below ` 2 lakh. The minimum value of the contract as on a particular day is determined by multiplying the market lot by the closing price of the underlying security on that day. In February 2004, the exchanges were advised to realign the contract sizes of the existing derivative contracts to ` 2 lakh. Subsequently, the exchanges were authorized to align the contract sizes, as and when required, in line with the methodology prescribed by SEBI.

The lot size of the option contracts on individual securities varies from security to security. For example, the lot size is 75 for Bharat Heavy Electricals Ltd., 50 for Larsen & Toubro Ltd., 200 for Infosys Technologies Ltd., 1000 for Sun Tv Network Ltd. and 2100 for Union Bank Of India, etc.

Members

The various types of membership in the derivatives market are as follows:

- **Trading member:** He can trade on his own behalf and on behalf of his clients.
- **Clearing member:** He is permitted to settle his own trades as well as the trades of the other non-clearing members known as trading members who have agreed to settle the trades through them.
- **Self-clearing member:** He can clear and settle his own trades only.

The financial transactions and asset liability positions of the traders are exposed to price risk. The price may fluctuate due to change in exchange rate, interest and market conditions. This has led to emergence of currency futures, interest rate futures, equity futures and commodity futures.

Futures

A futures contract calls for the delivery of either a physical asset or a financial instrument at a specified date or during a specific period of time at an agreed-upon price. Futures contracts obligate a person to buy or sell a particular commodity at a particular price on a particular day. If 'X' buys a futures contract, he is obligated to buy the commodity at the stated price. On the other hand, if 'X' sells such futures, he is obligated to sell the commodity at the stated price. The terms of the futures contracts are always specified in detail. The contract includes:

- The amount and type of the asset to be delivered: the type and quantity of asset or set of assets must be indicated precisely.
- The delivery date or maturity date: the date or period of time at which the exchange is to be consummated
- The exact place and the process of delivery. Apart from this, regulatory authorities specify the margin amount to be deposited to ensure that each party follows through with his or her side of transaction. Restrictions are laid on the extent of price changes that can occur within a trading day and the size of possession that can be taken.

9.7.4 Forwards and Futures

A forward contract is a contract between two people who agree to buy/sell a specified quantity of a financial instrument/commodity at a certain price at a certain date in future. For example, Anil is a wholesale sugar dealer and Sunil is the prospective buyer. Sunil agrees to buy 25 kg of sugar at ` 20 per kg after three months. The price is fixed on the basis of the supply and demand in the market and future perceptions about the price of sugar. If after three months, the market price of sugar is ` 25 per kg, then Sunil gains and if the price of sugar is ` 15 per kg, then Anil earns profit. Forwards are similar contracts to futures but customizable in terms of contract size, expiry date and price, as per the needs of the user. Though futures markets are regulated by regulatory agencies, forward markets are self-regulated. Some key features of the futures market are:

- Futures are traded in organized exchanges.
- The futures contracts are of a standardized size.
- Clearing houses guarantee that all the traders in futures market will honour their obligations.
- Margin payment and daily settlement is required.
- Future positions are closed easily.

Advantages of futures

They are useful to the hedgers and speculators, to discover the price of the underlying asset.

Check Your Progress

14. State the various kinds of membership in the derivatives market.
15. State the key features of the futures market.

NOTES

- **Hedgers**

They buy and sell futures to offset the anticipated risk position in the spot market. They may be either producers of the asset or users of the asset. For example, a treasury manager who probably has a dollar exposure after six months from now with regard to a settlement on account of exports, would seek to limit the risk of the depreciation in the US dollar by selling futures. It would help him to lock in the entire quantum of inflow on that future date.

Likewise, a sugarcane growing farmer, who is expecting a fall in the price of sugarcane, can sell sugarcane futures today and lock in a particular price as selling price. Hedging thus helps him to reduce the risk. The sugar mill owner is also assured of sugarcane supply at a particular price.

- **Speculators**

They may not be producing and using the underlying assets. They can buy and sell futures based on the forecast price of the underlying assets. They are after short-term profits. If they anticipate a substantial rise in the price of sugarcane futures, they buy it. Later, they would reverse the trade by selling the futures. If the forecast is accurate, they would make profit. They can carry out their business with margin money and are not required to have the entire sum.

- **Price Discovery**

The futures market reveals anticipations regarding the prices of the underlying assets. A comparison of the forecast of prices in the futures market with the spot prices helps the investors and the traders to take appropriate decisions.

- **Types of Futures Contracts**

The futures contracts fall into two general categories, namely, commodity futures and financial futures. Even though commodity futures were the first to be traded in the futures market, in this chapter financial futures are discussed at first. The important types of financial futures are currency futures, interest rate futures, stocks and stock index futures.

1. Currency Futures

The market for currency futures is of recent origin. The Chicago Mercantile Exchange set up its international monetary market division for trading of currency futures in 1972. Some of the notable currency futures exchanges are Philadelphia Board of Trade, London International Financial Futures Exchange, Singapore International Monetary Exchange and Sydney Futures Exchange.

The currency futures contract involves trading of an underlying currency at a specified exchange rate for a fixed maturity. The currency futures are traded only in a limited number of currencies like Australian Dollar, Japanese Yen, Swiss Franc, British Pound, US Dollar, South African Rand and Euro.

The size of contract is standardized, involving a fixed amount of different currencies. The size differs for different currencies and exchanges in which the contracts are traded. In Chicago Mercantile Exchange, the size is 62,500 British Pounds per contract. The size of CME Canadian Dollar futures contract trading unit is 100,000

Canadian Dollars. The size of CME Euro FX contract is 125,000 Euro and the size of Japanese Yen trading contract is 12,500,000 Japanese Yen.

NOTES

Clearing House

The deals are settled through the clearing house. If 'X' is a buyer of currency futures, he acquires a long position with the clearing house. At the same time, if 'Y' is a seller, he acquires a short position with the clearing house, not with the actual buyer. The clearing house becomes the buyer to every seller and seller to every buyer. It guarantees every deal that takes place in the exchange. The buyer of a futures contract can close his position before the settlement date by selling an identical futures contract. Similarly, a seller also can close his position before the settlement date by purchasing a currency futures contract. The difference between the buying and selling price of the trader determines his gain or loss which is settled through the clearing house.

Margin

Like any other futures, the currency futures also require margin amount to be remitted to the clearing house. Margin is required to cover the risk of non-payment of money at the time of settlement. Generally, the margin is in the form of cash deposits and liquid securities like treasury bills. Interests earned by the liquid securities are also paid to the traders. The margin amount is decided by the concerned futures exchange.

Marking to Market

The currency futures deal is settled on a day-to-day basis. The rates of futures are matched everyday with spot rates. The gain or losses are accordingly settled everyday. This process is known as marking to market. The gain is paid and the loss is deducted from the margin money. If the loss is heavy and the margin falls below a certain level, additional amount known as maintenance margin is required to be paid by the trader within a specified period. On the maturity day, the trader receives the amount of the contract after the adjustment of profit or loss.

Costs Involved

Dealing in currency futures involves certain costs. These are brokerage commission, clearing fee and delivery cost. Brokerage commission is paid for both opening and reversing the trade to the commission brokers. The clearing fee is charged by the stock exchange and clearing house. If the trade takes place through brokers, the commission includes the trading and clearing fee. When the traders carry out trade for themselves, they have to pay this fee. The delivery cost is charged if the currencies are actually delivered. However, mostly the delivery of currency does not happen and the gain/loss is settled through the clearing house.

Hedging

Foreign currency futures help to reduce foreign exchange risk. For example, if an American trader imports goods from Japan for one million Yen, he needs this amount for making payments to the exporter in Japan. Therefore, he buys Yen currency futures worth one million Yen for this payment. The currency futures free him from the loss that he may incur if the value of Yen appreciates.

NOTES

In the same manner, the Japanese exporter sells the Yen futures and locks the price of export to be received in terms of Yen. It protects him from the loss that may occur due to the depreciation of the value of Yen in terms of dollars at a future date.

Spread

It is of two types, i.e., intra-currency spread and inter-currency spread. Intra currency spread occurs when the buy and sell rates of the same currency differ for two delivery dates. Suppose the spot rate for dollar is £ 0.59, the June delivery rate is £ 0.58 and the September delivery date is £ 0.55. If there is anticipation about fall in the value of dollar than shown by the futures, the trader will buy the two futures for the above dates. Prior to maturity, he can reverse the two contracts at £ 0.59 and £ 0.54. In the original contract, the price difference is £ 0.58 - £ 0.55 = £ 0.03, while in the reverse contracts, the difference is £ 0.57 – £ 0.53 = £ 0.04. The net difference thus amounts to 0.01 and the profit amounts to 0.01 × 1,25,000 = £1250.

Inter Currency spread

This spread occurs when the deal involves purchase and sale of futures of two different currencies with the same delivery date. If the trader expects an appreciation of British Pound in relation to US Dollar, he will buy British Pound futures and sell dollar futures. Before maturity, he will reverse the two contracts. When the price difference of the two reverse contracts is less than the price difference of the original contract, the trader is able to gain money.

2. Interest Rate Futures

The underlying assets for the interest rate futures are fixed income securities. The prices of the fixed income earning securities are influenced by the current and forecast interest rates. Their pricing is related to the term structure of the interest rates. The assets are treasury bills bonds, euro dollar deposits and municipal bonds in the developed market. Contracts on foreign debt instruments are traded on foreign futures exchanges. London International Financial Futures Exchange trades contracts on British Government bonds. Chicago Board of Trade had launched a futures contract on Japanese government bonds in 1990.

In India, the interest rate futures were introduced in the year 2003. SEBI, in consultation with Reserve Bank of India, introduced the Exchange Traded Interest Rate Derivative Contracts in the futures and option segment.

Product Characteristics

Contract underlying	Notional 10 year bond (6% coupon)	Notional 10 year zero coupon bond	Notional 91 day T-bill
Contract descriptor	NFUTINT NSE10Y06 26 JUNE 2003	NFUTINT NSE10YZC 26 JUNE 2003	FUTINT NSET891D 26 JUNE 2003
Contract Value	` 2,00,000		
Lot size	2000		
Tick size	Re 0.01		
Expiry date	Last Thursday of the month		
Contract months	The contracts shall be for a period of a maturity of one year with three continuous contracts for the first three months and fixed quarterly contracts for the entire year		
Pri			
Settlement	As may be stipulated by NSCCL in this regard from time to time		

Trading Cycle

The interest rate futures contracts are for a period of one year with three months continuous contracts for the first three months. Fixed quarterly contracts are available for the entire year. New contracts are introduced on the trading day following the expiry of the near month contract. Interest rate futures contracts expire on the last Thursday of the expiry month. If the last Thursday is a trading holiday, the contracts shall expire on the previous day.

Contract Size

The permitted lot size for the interest rate futures contracts is 2000. At present, the minimum value of interest rate futures contract is ` 2 lakh. The price steps in respect of all interest rate future contracts admitted to dealings on the exchange is Re 0.01. The different types of orders are:

(a) Regular lot order
(b) Stop loss order
(c) Immediate or cancel
(d) Good till day
(e) Good till cancelled
(f) Spread order

Good till cancelled (GIC) orders are those which can be cancelled at the end of the period of seven calendar days from the date of entering an order.

Settlement Procedure

Daily mark to market settlement and final mark to market settlement are carried out in admitted deals in interest rate futures contracts. They are cash settled by debiting or crediting of the clearing accounts of clearing members with respective clearing bank.

Settlement Price

The daily settlement price for an interest rate futures contract is the closing price of such interest rate futures contract on the trading day. The closing price is calculated on the basis of the last half-an hour weighted average price of such interest rate futures contract. In the absence of trading in the last half-an-hour, the theoretical price is taken or such other price as may be decided by the relevant authority from time to time. Theoretical price for unexpired futures contract is computed with the spot price arrived at from the applicable zero-coupon yield curve. In the case of notional T-bill, the settlement price is 100 minus the annual yield for the specific period, using the zero-coupon bond.

The scheduled commercial banks and the financial institutions have to settle their trades directlywith the clearing corporation/clearing house. Regulated entities participating through approved futures and option (F and O) members have to settle trades as a participant clearing member or through approved professional custodial clearing member.

Risk Containment Measures

The portfolio-based margining approach applicable to equity derivative contracts is extended to interest rate derivatives. The margins would be computed taking an integrated view on the risk on a portfolio of an individual, taking into consideration all the derivative contracts.

NOTES

Initial Margin

The initial margin covers 99 per cent of value at risk due to the price changes over one day. It is based on the volatility estimate and volatility changes. Prof. J.R. Varma Committee had specified exponentially moving average methodology for calculation of the volatility or standard deviation. The standard deviation at the end of day t (s_t) is estimated using the previous volatility estimate. The formula is given below.

$$= -(\sigma)^2_t - 1(\sigma)^2_{t-1} + (I - 1)(r)^2_r.$$

σ_t = Standard deviation of the present day returns

σ_{t-1} = Standard deviation of the previous day returns

(return $r_t = \ln [I_t/I_{t-1}]$ where I_t is the interest rate futures price at time t. The return r_t used in the formula is computed from the prices of the near month interest rate futures contract)

r_t = the return is defined as the logarithmic return

I is the parameter which determines how rapidly volatility changes. The value of I is fixed at 0.94.

In case of long bond futures, the price scan range shall be 3.5 standard deviation and in no case the initial margin shall be less than 2 per cent of the notional value of futures contract. For notional T-bill futures, the price scan range shall be 3.5 standard deviation and in no case the initial margin shall be less than 0.2 per cent of the notional value of the futures contract.

Exposure Limits

The notional value of gross open positions at any point of time in futures contracts on the notional bonds shall not exceed 100 times the available net worth of a member. Therefore, the exchange would be required to ensure that 1 per cent of the notional value of the gross open positions in futures contracts on the notional ten year bond is collected or adjusted from the liquid net worth of a member on a real time basis.

In the case of T-bills, the notional value of the gross open positions at any point of time in the contract shall not exceed 1000 times the available liquid net worth of a member. Therefore, the exchange would be required to ensure that 0.1 per cent of the notional value of gross open positions in futures contracts on the notional T-bill is collected/ adjusted from the liquid net worth of a member on a real time basis. Exposure limits are in addition to the initial margin requirements.

Zero-coupon yield curve is computed at the end of the day. Zero-coupon yields on a real time basis are also calculated several times during the course of the day. Margins computed on the basis of the latest available yield curve are applied to the clients' portfolios on a real time basis. Exchange also computes the end-of-day margins on the basis of the provisional yield curve, for example, based on (T + 0) trades, because the final day yield curve becomes available only later in the evening. In such a case, the exchanges specify and disclose the basis on which the margins are calculated.

Capital Adequacy

The net notional principal amount in respect of futures positions with the same underlying and settlement dates is multiplied by the conversion factor given below to arrive at the credit equivalent.

Original maturity		*Conversion factor*
Less than one year	-	0.5%
One year and less than two years	-	1.0%
For each additional year	-	1.0%

NOTES

The credit equivalent thus obtained shall be multiplied by the applicable risk weight of 100 per cent.

The existing norm of 5 per cent of the total transaction during a year shall be observed by scheduled commercial banks and financial institutions who participate through approved futures and option members of the exchanges.

Disclosures and Reporting

The regulated entities undertaking interest rate derivatives on exchanges have to disclose, details in specified format as a part of the notes on accounts to balance sheets. Banks and specified financial institutions have to submit a monthly statement to DBS or DBS (FID) respectively.

3. Stock and Stock Index Futures

Stock futures contracts are written on specified stocks. Stock index futures contracts are written on major stock indexes like New York Stock Exchange Index, the Standard Pool's 500 Index Value Line Index. In India, the stock index futures are available for Sensex of the Bombay Stock Exchange and Nifty of the National Stock exchange.

Some of the other important stock index futures in other countries are: ASX Securities Index in Australia, Osaka Stock Futures 50 in Japan, Nikkei Stock Average in Japan and Singapore, Hang Seng stock Index in Hong Kong, Barclays share price index in New Zealand, OMX index in Sweden and FTSE 100 index in UK. Stock futures are traded until the expiry of the contract. It allows the trader to recognize profit or loss from price changes without receiving delivery of the actual underlying stock. Of course, a transaction can be settled by delivery of the underlying stock. However, generally, the futures trades are cash-settled.

A trader buys futures if he expects a price rise, but he sells if he anticipates the price decline. For example 'XYZ's, share futures operate for a period of three months. January series of a year ends in March of that particular year. The market price of the 'XYZ' share is ` 310 per share. 'A' anticipates the 'XYZ' share price to increase to ` 360 by March 31. At the same time 'B' expects a fall in the price of 'XYZ' stock. 'A' will buy futures, while 'B' will sell futures at a negotiated price say ` 320. Both 'A' and 'B' have to enter into this transaction through futures and option trade brokers who in turn will route the transaction through the clearing house of futures exchange. 'A' can sell the futures at any time during the contract period or can hold them till expiry. For example, if the 'XYZ' futures in Feb are at ` 345, 'A' can sell the futures and earn profit. The following terms are used to indicate the position of the trader in the market.

- Long position-Outstanding/unsettled purchase position at any point of time
- Short position-Outstanding/ unsettled sales position at any point of time
- Open position-Outstanding/unsettled long or short position at any point of time

NOTES

Lot Size

It refers to the number of underlying securities in one contract. The lot size is determined keeping in mind the minimum contract size requirement at the time of introduction of derivative contracts on a particular underlying. SEBI has specified that the value of a derivative contract should not be less than ` 2 lakh at the time of introducing the contract in the market. In February 2004, the exchanges were advised to re-align the contract sizes of the existing derivative contracts to ` 2 lakh. Subsequently, the exchanges were authorized to align the contract sizes, as and when required, in line with the methodology prescribed by SEBI.

The lot size differs from one stock to another. For example, the lot size for Canara Bank is 800, Central Bank of India is 2000, HDFC Bank Ltd. is 200 and ICICI Bank Ltd. is 175.

Contract Size of Stock Under Futures

In case of Nifty contracts, the value of the contract is equal to the index value multiplied by fifty. If Nifty is 4556, then the contract size is 4156 × 50 = ` 207800. On the Sensex futures, the value of the contract is equal to the index value multiplied by twenty-five. If the Sensex is 15041 then the contract value is 15041 × 25=375525.

Fair Value

The fair value is calculated on the basis of the spot index value which represents the closing value of the index on which the futures are based. The closing value of the spot index is multiplied by a compounding factor. The 91-day treasury bill rate is used as a compounding factor.

Contract Period

The life time of each contract is generally three months. At any point of time, there are three series open for trading. For example, on 29 August 2008, there were three contracts to be traded.

- One month September contract that matures on 25 September
- Two months October contract that matures on 30 October
- Three months November contract that matures on 27 November

On the expiry of the one month September contract, a new series of three month December contract will be traded. Then, the two month October contract will automatically become one month October contract and three months November futures will become August two month futures.

Margin Requirement

Like any other futures contract, margin is required for index futures and stock futures. The margin is not a down payment as neither loan nor margin interest is involved.

Initial Margin

It is the margin amount initially required to open a margin account for trading. For most futures contracts, the initial margin may be 5 per cent to 10 per cent of the contract value. It may seem to be small relative to the value of the contract. However, the amount is reasonable because of the prevalence of the daily settlement of mark to market system. The initial margin needs to cover only one day's price fluctuations because

any loss will be covered by posting an additional margin. It must be noted that any excess amount in the margin account, i.e., in excess of the initial margin may be withdrawn by the trader.

NOTES

Maintenance Margin

It is the minimum amount of margin money that must be maintained in a margin account. If the balance in margin account falls below this level, a margin call is made and the trader is required to deposit an additional amount. It helps to restore the balance in the margin account back to the level of initial margin.

Variation Margin

It is the amount margin call required to be deposited by the trader if the balance in margin account falls below the maintenance margin level.

Client Margin

Clearing members and trading members are required to collect initial margins from all their clients. The collection of margins at client level in the derivative markets is essential as derivatives are leveraged products and non-collection of margins at the client level would provide zero cost leverage. In the derivative markets, all the money paid by the client towards margins is kept in trust with the clearing house. In the event of default of the trading or clearing member, the amounts paid by the client towards margins are segregated and not utilized towards the default of the member.

Therefore, clearing members are required to report on a daily basis, the details in respect of such margin amounts due and collected from their trading members' clearing and settling through them. Trading members are also required to report on daily basis the details of the amount due and collected from their clients. Reporting the collection of margins by the clients is done electronically through the system at the end of each trading day. This plays a crucial role not only in ensuring that the members collect margins from the clearing corporation, but also in having a record of the quantum of funds that it has to keep in trust for the clients.

Settlement of Futures Contracts

An open index futures position held by an investor is cash-settled on the settlement day. Cash settlement means that there is no physical delivery of securities. Only the difference between the contracted values and the closing value of the index/security futures is collected. All derivative contracts are currently cash-settled. For example, consider that 'A' has bought 100 units of 26 December 2008 expiry contract at ` 1400 on 12 December 2008 and has not closed out his position till 26 December 2008. The closing value of the December expiry contract on a day before the last trading day is 1425. The final settlement price is ` 1425. 'A's' position is thus closed out at ` 1425. Difference between the closing value of the previous day and the final settlement is paid T + 1, i.e., ` 500 (100 × 1420 – 100 × 1425).

Settlement Price

Daily settlement price on a trading day is the closing price of the respective futures contracts on such day. At present, the closing price for a futures contract is the weighted average of the last half-an-hour price of the contract in the futures and option segment of NSE or BSE. The final settlement price is the closing price of the relevant underlying index/security in the capital market segment on the last trading day of the contract. The

closing price is the weighted average of the last half-an-hour price of the underlying asset in the capital segment.

Trading in futures

NOTES

For an investor to trade futures on NSE, he has to pay an initial upfront margin to his brokers as a security deposit. For example, if 'A'has bought 50 units of NIFTYDecember expiry contract at ` 4523 and if the daily margin is 5 per cent, he has to pay 4523 × 50 × 5% = ` 11307.5 as upfront margin. The exchange assesses the value of the investor's position on a daily basis. The position of the investor is marked to market every day. If 'A' has bought 100 units of December expiry contract at 4550 on 2 Januray 2008, the settlement price of December expiry contract on 2 January 2008 is 4570. 'A' makes a mark to market profit of (` 4570 – 4550) × 100 = ` 2000. However, at the same time, if the index falls to ` 4540, he incurs a loss of ` 1000. If the market value of the investor's position shows a loss, then the investor has to pay the difference. If his position has gained in value, he receives the profit amount. The loss is debited to the margin account. This brings down the balance in the margin amount and so variation call of the margin amount is made to maintain the required margin.

Eligibility criteria for the stocks

The stocks to be included in the derivative trade segment have to fulfil certain eligibility criteria.

The security should be amongst the top 500 securities in terms of the average daily market capitalization and average traded value during the previous six months, which is calculated on 15th of every month, on a selling basis.

The median quarter-sigma order size of the securities over the last six months should be at least ` 5 lakh. The order size is calculated by taking four snapshots in a day from the order book of the underlying stock in the past six months. Exponentially weighted moving average method, which gives more weight to the recent price movements and less weight to the historical price movement, is used to calculate the standard deviation. The volatility estimate is available for each stock on a daily basis. This Value at Risk (VaR) is multiplied by 0.25 to get quarter sigma value. Then the quarter sigma is multiplied by best buy and best sell value. The quarter sigma price is calculated for all the four snapshot values.

The median for buy-side and sell-side is calculated for all the four times a day value for the last six months. The average of the median order sizes for buy and sell-side is taken as quarter-sigma order size of the security. The so completed median quarter size sigma order size should be equal or greater than ` 5 lakh. It ensures that the stocks on the derivative segment are liquid.

If an existing security fails to meet the eligibility criteria for continuously three months, fresh contracts are not permitted to be issued. However, the existing contracts on the stocks are permitted to be traded.

Eligibility for Trading Members

Members or brokers are eligible to trade the futures contract in exchange subject to certain conditions. These are:

(a) Members have to pay an approval fee as prescribed by the exchange from time to time.

NOTES

(b) They have to be registered with SEBI in addition to their registration with any of the exchanges.

(c) Members have to pass a certification programme which has been approved by the SEBI.

(d) The trading member is required to deposit security either in the form of cash deposit receipts, bank guaranty of an approved banker or others, subject to the conditions as the exchange may specify from time to time.

(e) The member should have a minimum security net worth as may be prescribed by the exchange from time to time.

Fair price of the stock index futures

Theoretical or fair price of the stock is calculated from the cost of carry model.

$$F = S(1 + C) - \sum_{i=1}^{N} D_i \ (1 + r_i)$$

F = price of the stock index futures
S = the value of stock underlying stock index
C = the percentage of cost of carrying the stocks from a particular date to the expiration date
D_i = the dividend to be received from i^{th} stock
r_i = the interest earned on carrying the i^{th} dividend from its time of receipt until the expiration date of the futures

The stock index futures do not receive dividends; it is simply a price index. Holding a stock gives dividends to the stock holder. The value of the index depends on the prices of stocks, not on the dividends of the underlying stocks. In turn, the stock index futures depend on the index value, which does not include the dividends. However, the futures price must be adjusted to include the dividends that would be received between the present and the expiration date of the futures.

Dividends lower the cost of carrying the stocks. Dividends received from the stocks reduce the value of the stocks (purchase price – dividend). This can be explained with an example. For example a trader borrows ` 100 and buys 'ABC' stock which is currently trading at ` 100. Let us assume that the 'ABC' stock pays ` 15 as dividend in six months. The trader will receive ` 15 as dividend and he will invest the proceedings for the remaining six months at rate of 8 per cent. The future value of the stock P_1 is unknown to the trader. Table 9.6 gives the cash flow of the trader.

Table 9.6 *Cash Flows from Carrying Stock*

Situation 1	
Borrow ` 100 for 1 year at 10%	+ 100
Buys ABC stock	100
Situation 2	
Received dividend	+ 15
Invested ` 15 for 6 months at 8%	15
Situation 3	
Collect proceeds from dividend invested (after six months)	15.60
Sell ABC stock	P_t
Repay the debt	– 110.00

Total Profit = P_t + 15.60 – 110

NOTES

This cash and carry trading opportunity requires that the price of the futures must be less than or equal to the cash inflows at the futures expiration. Thus, the stock index futures price must be equal to the cost of underlying stock index, plus the cost of carrying those stocks to expiration S (1 + C) minus the future value of all the dividends to be received D $(1 + r)$. The stock futures price has its fair value when the futures price fits the cost of carry model.

9.7.5 Depth, Breadth and Resilience of Markets

There are three liquidity-related characteristics of a secondary market that investors find desirable: depth, breadth, and resiliency. First, a secondary market is said to have depth if there are orders both above and below the price at which it is currently trading. When a security trades in a deep market, temporary imbalances of purchase or sale orders that would otherwise create substantial price changes are offset with corresponding orders. Second, a secondary market is said to have breadth if its orders give its market depth in a significant volume. The broader the market for a stock, the greater the potential for stabilization of temporary price changes that may arise from order imbalances. Third, a market is resilient if orders promptly respond to price changes.

9.8 CONSTRUCTION OF SENSEX AND NIFTY

Let us discuss the construction of Sensex and Nifty

9.8.1 Sensex

The BSE Sensitive index has long been known as the barometer of the daily temperature of Indian bourses. In 1978-79 stock market contained only private sector companies and they were mostly geared to commodity production. Hence, a sample 30 was drawn from them. With the passage of time more and more companies private as well as public came into the market. Even though the number of scrips in the Sensex basket remained the same, representations were given to new industrial sectors such as services, telecom, consumer goods and 2 and 3 wheeler auto sector. The continuity and integrity of the index are kept intact, so that a comparison of the current market condition with those of a decade ago is made easy and any distortion in the market analysis is avoided. The criteria adopted in the selection of 30 scrips are listed below.

1. **Industry representation:** The index should be able to capture the macro-industrial situation through price movements of individual scrips. The company's scrip should reflect the present state of the industry and its future prospects. Companies chosen should be representative of the industry. For example, company like ACC in the Sensex is a representative of the cement industry. The logic here is that ACC reflects the fortunes of the cement industry that inturn is discounted by the market in the scrip's pricing. Care is taken in selecting scrips across all the major industries to make the index act as a real barometer to the economy.
2. **Market capitalisation:** The market capitalisation of the stock indicates the true value of the stock, as the outstanding number of shares is multiplied by the price. Price indicates the demand and growth potential for the stock. The outstanding shares depend on the equity base. The selected scrip should have a wide equity base too.
3. **Liquidity:** The liquidity factor is based on the average number of deals .

Check Your Progress

16. State the important types of financial futures.
17. What is daily settlement price?

The market fancy for the share can be found out by the trading volumes. The Financial Express Equity Index is weighted by trading volume and not by market capitalisation.

4. **The market depth:** The market depth factor is the average deal as a percentage of company's shares outstanding. The market depth depends upon the wide equity base. If the equity base is broad based then number of deals in the market would increase. For Example Reliance Industries has a wide equity base and larger number of outstanding shares.
5. **Floating stock depth:** The floating stock depth factor is the average number of deals as a percentage of floating stock. Low floating stock may get overpriced because the simple law of demand and supply apply here. For example MRF with its low floating stock is able to command high price. Its sound finance and internal generation of funds led growth may be the reason for the low flotation. Though the public holding is fairly high at around 40 per cent due to small equity of `4.24 Cr, the free float of the company stock is low.

Trading volumes are directly linked to the public holding in the equity of the company. Wide public holding is a pre-requisite for high trading volume. Reliance Industries is a good example. The free float of company is 45 per cent and it has its positive effect on the trading volume.

Revision of Sensex Scrips

In 1998, the index committee of the BSE decided to give a wide representation to the four market favourites at its meeting. The need to have a broad-based and liquid index was felt with index futures on its way. The revised index has a representation of 15 industries. It has assigned the food and beverages sector a weight of 12.75 per cent with the market cap at `21,113 crore. Health care industry has been given a weight of 4.70 per cent, Consumer non-durables — 21.74 per cent and auto industry 7.92 per cent. IT accounts for a weight of 4.33 per cent, oil and gas –6.22 per cent, petrochemicals –6.70 per cent, telecom –7.71 per cent, power –1.98 per cent and hotel –1.10 per cent.

The Recast—BSE Sensitive Index

Company	*Market Cap (₹Cr.)*	*Weightage (%)*
ACC	1358.6	0.89
Andhra Valley	464.4	0.30
Castrol India	3603.2	2.35
BHEL	5640.5	3.68
BSES	2142.5	1.40
Bajaj Auto	6507.9	4.25
Colgate	2349.2	1.53
Infosys	3696.8	2.41
Glaxo India	2766.6	1.81
Grasim Industries	1110.7	0.72
Gujarat Ambuja	1381.6	0.90
HPCL	5294.7	3.46
Hindaclo	4226.7	2.76
Hindustan Lever	33577.1	21.91
ICICI	2030.7	1.33
IDBI	2668.8	1.74

NOTES

Novartis	2158.8	1.41
ITC	17155.4	11.20
Larsen & Toubro	3984.7	2.60
MTNL	11645.6	7.60
Mahindra & Mahindra	1628.1	1.06
Nestle	4220.3	2.75
Ranbaxy	2942.9	1.92
Reliance Industries	10009.8	6.53
State Bank of India	8478.7	5.53
TELCO	3144.0	2.05
Tata Chemicals	1574.1	1.03
Tata Power	1022.0	0.67
TISCO	3085.0	2.01
NIIT	3349.9	2.19
Total	153219.8	100.0
The Outgoing Stocks		
SAIL	2292.2	1.58
IPCL	1322.4	0.91
Arvind Mills	330.8	0.23
GE Shipping	549.2	0.38
Total	4494.7	3.10
The Incoming Stocks		
NIIT	3349.9	2.19
Novartis	2158.8	1.41
Infosys Tech	3696.8	2.41
Castrol India	3603.2	2.35
Total	12808.7	8.36

9.8.2 Nifty

Nifty index is built by India Index Services Product Ltd (IISL) and Credit Rating Information Services of India Ltd. (CRISIL). The CRISIL has a strategic alliance with Standard and Poor Rating Services. Hence, the index is named as S & P CNX Nifty. NSE - 50 index was introduced on 22 April 1996 with the objectives given below:

- Reflecting market movement more accurately
- Providing fund managers a tool for measuring portfolio returns vis-market return.
- Serving as a basis for introducing index based derivatives.

Nifty replaced the earlier NSE - 100 index, which was established as an interim measure till the time the automated trading system stabilised. To make the process of building an index as interactive and user driven as possible an index committee is appointed. The composition of the committee is structured to represent stock exchanges, mutual fund managers and academicians. To reflect the dynamic changes in the capital market, the index set is reduced and modified by the index committee based on certain predetermined entry and exit criteria.

There has been a recast of basket of Nifty stocks and the new basket came into effect on October 9, 1998. The accompanying Table 9.7 shows the earlier and present composition of the Nifty index. IT stocks are included. The Nifty composition in April 2000 is given below.

NOTES

Table 9.7 *The Nifty 2000*
Composition of S & P CNX Nifty

ACC	I P C L
Asea Brown Boveri	ITC
Asian Paints	Infosys Technologies
BHEL	Indian Hotels
Britannia	Larsen & Toubro
BSES	MTNL
Bank of India	Mahindra & Mahindra
Bajaj Auto	Nestle India
Castrol	NIIT
Cipla	Novarits
Cochin Refineries	Oriental Bank
Colgate	Procter & Gamble
Dabur	Ranbaxy
Dr. Reddys	Reckitt & Colman
Glaxo India	Reliance Industries
Grasim	Reliance Petroleum
Gujarat Ambuja	Satyam Computers
HDFC	Smithkline Beechem
HDFC Bank	State Bank of India
HCL Infosystems	TELCO
Hero Honda	Tata Chemicals
Hindustan Petroleum	Tata Power
Hindalco	TISCO
Hindustan Lever	Tata Tea
ICICI	Zee Telefilms

Selection criteria

The selection criteria are the market capitalisation and liquidity. The selection criterion for the index was applied to the entire universe of securities admitted on NSE. Thus, the sample set covers a large number of industry groups and includes equities of more than 1200 companies.

The market capitalisation of the companies should be ` 5 billion (US $ 118 Million) or more. The selected securities are given weights in proportion to their market capitalisation.

Liquidity (Impact Cost): Here the liquidity is defined as the cost of executing a transaction in security in proportion to the weightage of its market capitalisation as against the index market capitalisation at any point of time. This is calculated by finding out the percentage mark up suffered while buying/selling the desired quantity of security compared to its ideal price (best buy + best sell)/2.

Order book

	Buy	***Price***	***Sell***	***Price***
	2000	90	2000	91
	3000	91	2500	94
	2000	92	1000	96
To buy	2500			

NOTES

$$\text{Ideal Price} = \frac{90+91}{2} = 90.5$$

$$\text{Actual buy Price} = \frac{(2000\times 91)+(500\times 94)}{2500}$$

$$= \frac{182000+47000}{2500} = 91.6$$

$$\text{Impact cost for 2500 Shares} = \frac{91.6-90.5}{90.5}\times 100 = 1.21$$

Impact cost for selling price also can be calculated. The company scrip should be traded for 85% of the trading days at an impact cost less than 1.5%.

Base period: The base period for the S & P. CNX Nifty index is the closing prices on 3 November 1995. The base period is selected to commensurate the completion of one year operation of NSE in the stock market. The base value of index is fixed at 1000 with the base capital of `2.06 of trillion. Its unique features are:

- S & P. CNX Nifty provides an effective hedge against risk. The effectiveness of hedging was compared with several portfolios that consist of small cap, midcap and large cap companies and found to be higher.
- The index represents 45 per cent of the total market capitalisation.
- The impact cost of S & P. CNX Nifty portfolio is less compared with other portfolios.
- Nifty index is chosen for derivative trading.

Cnx nifty junior

The Nifty Junior also consists of fifty stocks, but these stocks belong to the midcap companies. Stocks that are having market capitalisation greater than `2 billion are included with the objective of measuring the performance of stock in the midcap range. The liquidity criterion is same as that of S & P. CNX Nifty. The impact cost should not be greater than 2.5% for 85% of the traded days. The base date is the same for Nifty and Nifty Junior but the base capital is `0.42 trillion. Nifty Junior represents about 7 per cent of the total market capitalisation and it is an ideal index to be used in derivative trading.

There is a recast in the Nifty Junior in 1998 with the number of stocks going up to the Nifty. The composition of the Nifty Junior has also been overhauled. Apart from the six that moved to the Nifty, Ispat Industries, Hindustan Powerplus, Alstom India, Kotak Mahindra and Lakme have been excluded. The eleven stocks replacing these in the Nifty Junior are: Bank of Baroda, Tata Infotech, Dr. Reddy's Labs, Satyam Computers, Zee Telefilms, Pentafour Software, Nirma, Nicholas Piramal, ICI India, ICICI Bank and GSFC. Nifty Junior turns out to be as nimble as its predecessor. The odds are high because of the sluggish nature of five of the excluded stocks as well as the quality of the new entrants.

S & P CNX 500

It is a broad based index consisting of 500 scrips. The companies are selected on the basis of their market capitalisation, industry representation, trading interest and financial performance. The market capitalization is used as weights. The companies influence on the index depends upon their market capitalisation. The companies selected are either leaders or representative of their industries. They should reflect the movement of their

industry. The industry groups included in the S & P.CNX 500 are 79. The number of representation from each industry group is changed to reflect the market.

The selected companies should have minimum record of three years of operation with positive net worth. The base year is 1994 because it is considered to be closer the post liberalisation era.

Since the index is a broad based one, it represents 72 per cent of the total market capitalisation and 98 per cent of the total traded value. As it is weighted with market capitalisation, it mirrors the market movement more effectively. The broad base of the index provides a bench mark for comparing portfolio return with market return.

NOTES

9.9 SUMMARY

In this unit, you have learnt that,

- Monetary assets, which are short term in nature and less than one year are traded in the money market.
- Money market facilitates RBI's conduct of monetary policy. There was a paucity of instruments in the money market for a long time.
- The money market has two components namely the organized and the unorganized.
- The participants of the organized money market are the Reserve Bank of India, Life Insurance Corporation, General Insurance Corporation, Unit Trust of India, Securities Trading Corporation of India Ltd., Discount and Finance House of India, other primary dealers, commercial banks and mutual funds.
- Money market is a very important segment of the Indian financial system. It is the market for dealing in monetary assets of short-term nature.
- The Indian money market was segmented and highly regulated and lacked depth till the late 1980s. It was characterized bya limited number of participants, regulation of entry and limited availability of instruments.
- The money market is a wholesale market. The volumes are very large and generally transactions are settled on a daily basis.
- Trading in the money market is conducted over the telephone, followed by written confirmation from both the borrowers and lenders.
- The capital market provides long-term funds for corporates, central and state governments. The capital market has sub-markets like debt market, equity market and derivative market.
- The foreign exchange market is the market for foreign currencies. Foreign currency transactions take place in the case of exports, imports, capital movement, interest and repayment of loans.
- Stock market or securities market is a market where securities issued by companies in the form of shares, bonds and debentures can be bought and sold freely. The components of stock market are primary market and secondary market.
- Treasury bills are short-term (up to one year) borrowing instruments of the government.
- Originally, the government used to issue the treasury bills directly. They were sold through the tender method and the periods of maturity were three months six months nine months and twelve months.

Check Your Progress

18. State the objectives of NSE-50 index.
19. What is the base period of Nifty index?

NOTES

- In the purchase of treasury bills, investors who do not have SGL account facility can buy it through the DFHI. DFHI keeps the treasury bills on behalf of their clients, pays the investor and collects the proceeds on the maturity day.
- Bill financing is an important mode of financing working capital requirements of trade and industry. Discounting of bills by the banks ensures adequate flow of working capital to the industry from the banks.
- Letter of credit backed bill discounting and clean bill discounting are the convenient modes of financing for domestic trade transactions.
- A government security is a claim on the government and is commonly referred to as gilt. British government entered the details of government securities in a book with golden edges.
- State government securities are securities/loans issued by the Reserve Bank of India on behalf of various state governments for financing their developmental needs.
- The issue of government securities provides an opportunity to make the monetary policy an effective instrument of economic policy. The structure of interest rates of the G-Secs influences the general interest rate structure of the capital market.
- The secondary market trades are either carried out through Negotiated Trading System or on the National Stock Exchange (NSE) and Bombay Stock Exchange (BSE).
- The price of security in the stock market has three components. These are: principal, coupon and notional tax components.
- Two parties are involved in the options contract, a buyer and a seller. The buyer who buys an option takes a long position. The seller who sells an option takes a short position, i.e., he writes the option and is the writer of the option.
- A futures contract calls for the delivery of either a physical asset or a financial instrument at a specified date or during a specific period of time at an agreed-upon price.
- There are three liquidity-related characteristics of a secondary market that investors find desirable: depth, breadth, and resiliency.
- Nifty index is built by India Index Services Product Ltd (IISL) and Credit Rating Information Services of India Ltd. (CRISIL). The CRISIL has a strategic alliance with Standard and Poor Rating Services.

9.10 KEY TERMS

- **Capital market:** The part of a financial system concerned with raising capital by dealing in shares, bonds, and other long-term investments.
- **Treasury bill**: Ashort-dated UK or US government security, yielding no interest but issued at a discount on its redemption price.
- **Commercial bill**: A bill of exchange issued by a commercial organization to raise money for short-term needs.
- **Sensex**: An abbreviation of the Bombay Exchange Sensitive Index (Sensex) - the benchmark index of the Bombay Stock Exchange (BSE).
- **NIFTY**: The Nifty, is an indicator of the 50 top major companies on the National Stock Exchange (NSE).

9.11 ANSWERS TO 'CHECK YOUR PROGRESS'

NOTES

1. The various money market derivatives are:
   - Call/ Notice/ Term Money
   - Repos
   - Treasury Bills
   - Certificate of Deposits (CD)
   - Commercial Papers (CP)
   - Inter Bank Participation Certificates
   - Inter Bank Term Money
   - Interest Rate Swaps/Forward Rate Agreements and
   - Bills Rediscounting
2. The sub-markets of capital markets include debt market, equity market and derivative market.
3. In an efficient market, the market price of the security should be equal to its intrinsic value. The true economic worth of the asset is its intrinsic value. In a perfectly competitive market, the price of an asset reflects its intrinsic value. This is known as valuation efficiency.
4. The different features of primary market are:
   - It is a market for long-term capital where the securities are sold for the first time. Hence, it is also called New Issue Market (NIM).
   - Funds are collected and securities are issued directly by the company to the investors.
   - Primary issues are carried out by the companies for the purpose of inception and functioning of business.
5. The disadvantages of public issues are as follows:
   - It is a time consuming process involving the fulfillment of legal formalities.
   - It is expensive and many intermediaries are involved in it. It necessitates constant adherence to listing of agreements and legal requirements.
   - Cornering of shares and hostile takeover may take place.
   - Speculative trading of the company's equity affects the reputation of the company.
6. The investors can be broadly classified into:
   - Retail investors
   - Institutional investors
   - Foreign institutional investors
7. Four characteristics of treasury bills are:
   - They have a short-term maturity period
   - They are highly liquid in nature
   - Their default risk is absent
   - Their transaction cost is very low
8. The objectives of the 14-day treasury bill are to meet the cash management needs of various segments of the economy and facilitate the emergence of a comprehensive yield curve.
9. Post-sale credit against bills is offered as follows:
   - Purchase of bills drawn under LC or confirmed order.

NOTES

- Discounting of usance bills drawn under LC or confirmed order.
- Negotiation of documents under LC.

10. Five features of a well-developed bill market are:
    - Widespread culture of borrowing against commercial bills
    - Presence of a large number of genuine trade bills
    - Continuous and adequate supply of commercial bills in the money market and high velocity of circulation.
    - Commercial banks and financial institutions have the facility for acceptance of bills at low cost.
    - Commercial banks are ready to discount the bills and legal procedures are simple.
11. Debt instruments that have variable coupon rates rather than fixed coupon rates but are based on a predefined benchmark are known as floating rate debt instruments.
12. Separate Trading of Registered Interest and Principal of Securities is known as STRIPS. Aconventional security is stripped into a number of zero coupon securities and is traded separately. Such newly created securities are called STRIPS.
13. The securities are issued in blocks with a single maturity date. If bonds are issued, the Reserve bank of India buys it and gradually sells it in the stock market. Gradually, the RBI repurchases the bonds through the stock market leaving a small portion to be redeemed at the maturity date. This gradual purchase of bonds is called 'grooming'.
14. The various kinds of membership in the derivatives market are:
    - Trading member
    - Clearing member
    - Self-clearing member
15. Some key features of the futures market are:
    - Futures are traded in organized exchanges.
    - The futures contracts are of a standardized size.
    - Clearing houses guarantee that all the traders in futures market will honour their obligations.
    - Margin payment and daily settlement is required.
    - Future positions are closed easily.
16. The important types of financial futures are currency futures, interest rate futures, stocks and stock index futures.
17. The daily settlement price for an interest rate futures contract is the closing price of such interest rate futures contract on the trading day.
18. The objectives of NSE-50 index are as follows:
    - Reflecting market movement more accurately
    - Providing fund managers a tool for measuring portfolio returns vis-market return.
    - Serving as a basis for introducing index based derivatives.
19. The base period for the S & P. CNX Nifty index is the closing prices on 3 November 1995. The base period is selected to commensurate the completion of one year operation of NSE in the stock market.

9.12 QUESTIONS AND EXERCISES

NOTES

Short-Answer Questions

1. State the characteristics of the money market.
2. What are the characteristics of an efficient financial market?
3. Write a short note on the auctions of treasury bills.
4. Briefly state the concept of sensex and nifty.
5. What are the advantages of treasury bills to market participants?

Long-Answer Questions

1. Discuss the features of the primary and the secondary market of securities.
2. Analyse the various kinds of treasury bills.
3. Discuss the features of commercial bill market and discount market.
4. Analyse the role of government securities market.
5. Describe the factors affecting the option premium and the various valuation models of options.

9.13 FURTHER READING

Mankiw, N Gregory. 2010. *Macroeconomics.* Worth Publishers, New York.

Shapiro, Edward. 1996. *Macroeconomic Analysis.* Galgotia Publication, New Delhi.

Branson, WH. 1979. *Macroeconomic Theory and Policy*. Universal Book Stall, New Delhi.

Jha, R. 1999. *Contemporary Macroeconomic Theory and Policy*. New Age International, New Delhi.

Gupta, SB. 1995. *Monetary Economics: Instruments and Policy*. S Chand & Co. New Delhi.

UNIT 10 INDIAN FINANCIAL INSTITUTIONS

NOTES

Structure

10.0 INTRODUCTION

The financial sector plays an indispensable role in the overall development of a country. The important constituents of this sector are the financial institutions, act as that conduit for the transfer of resources from net savers to net borrowers, that is, from those who spend less than their earnings to those who spend more than their earnings. The financial institutions have traditionally been the major source of long-term funds for the economy. These institutions provide a variety of financial products and services to fulfil the varied needs of the commercial sector. Besides, they provide assistance to new enterprises, small and medium firms as well as to the industries established in backward areas. Thus, they have helped in reducing regional disparities by inducing widespread industrial development.

The Government of India, in order to provide adequate supply of credit to various sectors of the economy, has evolved a well-developed structure of financial institutions in the country. These financial institutions can be broadly categorized into All India institutions and State level institutions, depending upon the geographical coverage of their operations. At the national level, they provide long and medium term loans at

reasonable rates of interest. They subscribe to the debenture issues of companies, underwrite public issue of shares, guarantee loans and deferred payments, etc. Though, the State level institutions are mainly concerned with the development of medium and small scale enterprises, but they provide the same type of financial assistance as the national level institutions.

This unit will deal with the financial institutions of India like Securities and Exchange Board of India (SEBI), commercial, cooperative and regional rural banks, self-help groups, non-bank financial intermediaries (NBFIs), insurance institutions and mutual funds.

10.1 UNIT OBJECTIVES

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

- Evaluate the working of capital market and Securities and Exchange Board of India (SEBI)
- Assess the functions and services provided by commercial, cooperative and regional rural banks
- Discuss the problems of microfinance and the emergence of self-help groups
- Explain the nature of non-bank financial intermediaries (NBFIs)
- Describe the concept of insurance and insurance institutions
- Analyse the policies included in life and general insurance
- Assess the concept, types and operation of mutual funds

10.2 WORKING OF CAPITAL MARKET AND SEBI

The capital market is susceptible to fraudulent and unfair practices. It is important to protect the trader/investor from such practices. In India too, there have been several incidents like the Ketan Parekh and the Harshad Mehta scams that have intensified the need for a regulatory mechanism.

A comprehensive legal framework was provided by the Securities Contract Regulation Act, 1956 and the Securities and Exchange Board of India Act, 1992. A three-tier regulatory structure comprising Ministry of Finance, Securities and Exchange Board of India and Governing Boards of the stock exchanges regulate the functioning of stock exchanges.

1. **Ministry of Finance:** The Stock Exchange division of the Ministry of Finance has powers related to the application of the provision of the SCRAct and licensing of dealers in the other area. According to the SEBIAct, the Ministry of Finance has appellate and supervisory powers over SEBI. It has the power to grant recognition to the stock exchanges and regulate their operations. The Ministry of Finance has the power to approve the appointments of executive chiefs and the nominations of public representatives on the Governing Boards of stock exchanges. It has the responsibility of preventing undesirable speculation.
2. **Securities and Exchange Board of India:** Even though established in 1988, the Securities and Exchange Board of India received statutory powers only on 30 January 1992. Under the SEBI Act, wide powers are vested in SEBI. It has the powers to regulate the business of stock exchanges, other security markets and mutual funds. Registration and regulation of market intermediaries are also carried out by SEBI. It has the responsibility to prohibit fraudulent,

unfair trade practices and insider dealings. Takeovers are also monitored by SEBI. Stock exchanges have to submit periodic and annual returns to SEBI. SEBI has the duty of promoting the healthy growth of the capital market and protecting the interests of investors.

3. **Governing Board:** The Governing Board of a stock exchange consists of elected member directors, government nominees and public representatives. Rules, bylaws and regulations of the stock exchange provide substantial powers to the Executive Director for maintaining efficient and smooth day-to-day functioning of the stock exchange. The Governing Board also has the responsibility of maintaining an orderly and well-regulated market.

The Governing Body of a stock exchange consists of 13 members of which: (a) six members are elected by the members of the stock exchange, (b) the Central Government nominates not more than three members, (c) the Board nominates three public representatives, (d) SEBI nominates persons not exceeding three, and (e) the stock exchange appoints one Executive Director.

One-third of the elected members retire during the annual general meeting. The retired member can offer himself for re-election if he is not elected for two consecutive years. If a member serves on the governing body for two years consecutively, he should refrain from offering himself for another re-election.

Members of the Governing Body elect the president and vice-president. It needs no approval from the Central Government or the Board. The tenure of office for the president and vice-president is one year. They can offer themselves for re-election, if they have not held office for two consecutive years. In that case they can offer themselves for re-election after a gap of one year.

The Securities and Exchange Board of India (SEBI) is an autonomous body created by the Government of India in 1988 to act as a regulator of the capital market in India. It was given statutory form in 1992 with the SEBI Act of 1992. Its head office is in Mumbai, while its other offices are in Chennai, Kolkata and New Delhi.

The organization of SEBI is structured in a manner so as to fulfil its objectives. To suit its scope of activities, it is divided into several departments (see Table 10.1).

Table 10.1 *Departments of SEBI*

Department	*Responsibility*
1. Primary Department	Policy matters related to primary market, intermediaries and self regulatory organizations, redressal of investors' grievances and guidance.
2. Issue Management & Intermediaries Department	Registration, regulation and monitoring of the intermediaries and scrutiny of offer document.
3. Secondary Market Department	Policy matters related to major stock exchanges, price monitoring market surveillance, prevention of insider trading and brokers' registration.
4. Institutional Investment	Mutual funds, FIIs, mergers, acquisitions.

Apart from these departments, SEBI has legal and investigation departments. It has separate advisory committees for primary and secondary markets to assist policy formulation. SEBI has regulated:

- Primary market
- Secondary market
- Mutual funds
- Foreign institutional investments

NOTES

10.2.1 Critical Review of SEBI

SEBI has made progress in many areas, but there are other areas within SEBI's domain which need attention. For example, the vetting of issues by SEBI leaves much to be desired and allows large number of issues of dubious credibility to mobilize public money between 1994 and 1996, leading to a dull primary market. Out of the 6,200 odd companies listed on the BSE, less than one-third are actively traded.

1. **Disclosures:** Even though SEBI has made enormous efforts to make disclosures more transparent, the quality of information flow leaves much to be desired. Timeliness of disclosures is another area of concern. Despite the time limit of 48 hours to inform SEBI about any deal that would affect the shareholders' interest, it is not adhered to. For example, IFCI has reported a deal concluded in late August 1998 to sell the shares of Sri Vishnu Cement at ` 100 each, and which was informed in October. There is a need for SEBI/NSE/BSE/to ensure the timely dissemination of information.
2. **Dissemination process:** The list of the information specified for disclosure is exhaustive. But the mode is through SEBI. The companies have to provide the information to the SEBI and SEBI to the investing public. Considerable delay is involved in this process. To avoid delay, websites can be created to download the information. The information can be classified and downloaded. In this process, information would reach the participants quickly.
3. **Settlement:** The National Stock Exchange of India (NSE) has a Wednesday–Tuesday settlement cycle while BSE has a Monday–Friday cycle. This creates more arbitrage opportunities. The prices of securities experience fluctuations on the opening and closing days. The implementation of a uniform settlement cycle would improve the quality of price-formation and reduce the costs imposed on investors due to arbitrage-related trade. The rolling settlement with T + 5 would reduce arbitrage.
4. **Badla trade:** The carry forward system is mainly used by speculative brokers and large traders. Speculative deals may lead to abnormal price rise and payment crises and affect the investors' confidence. Badla should be considered as a lending and borrowing system with suitable checks and balances. Index futures offers an opportunity for hedgers and speculators.
5. **Special watch:** The sophistication of the Indian stock market demands better surveillance and regulatory capabilities. Stock exchanges across the country are shifting towards online screen-based trading. These stock exchanges account for almost 99.80 per cent of the aggregate business transacted. A stock watch system parallel to the system at the New York Stock Exchange should be developed. The stock watch system will alert the exchange administration about abnormal price and volume fluctuations. This will provide transparent and fair dealings in the market.
6. **Capital adequacy:** Capital adequacy of registered market participants is low. SEBI requires stock brokers of BSE/NSE to maintain an absolute minimum of ` 500,000 which is the largest for all stock exchanges. The additional or optional capital, including the basic minimum capital, has to be maintained at 8 per cent or more of the gross outstanding trade. This adds to the default risk of the brokers. Further, SEBI is forced to register many small participants and regulate them.

7. **Single authority:** The regulation of the capital market is carried out by multiple regulators such as RBI and the Ministry of Finance. This multiple-regulator system has undermined the role of SEBI. A case in point is the order of the Ministry of the Finance negating SEBI's order in the Hindustan Lever Ltd. insider trading case while agreeing that information on the merger was price sensitive. The regulatory aspect should be properly coordinated and super regulator can be created. The UK has created a super regulator which regulates the capital and forex markets.
8. **Stricter registration of brokers:** SEBI's registration system is far from adequate. The registration of sub-broker is nominally in place but lack of enforcement power permits the functioning of many hundreds of unregistered sub-brokers. The registered sub-brokers are also not effectively controlled. This would add to the risk involved in the trading of securities.

Objectives of SEBI

The promulgation of the SEBI ordinance in the parliament gave statutory status to SEBI in 1992. According to the preamble of the SEBI, the three main objectives are:

- To protect the interests of the investors in securities
- To promote the development of securities market
- To regulate the securities market

10.2.2 Powers and Functions of SEBI

The main functions entrusted with SEBI are as follows:

- Regulating the business in stock exchanges and any other securities market
- Registering and regulating the working of stock brokers, sub-brokers, share transfer agents, bankers to the issue, trustees of trust deeds, registrars to an issue, merchant bankers, underwriters, portfolio managers, investment advisers and such other intermediaries who may be associated with securities market in any manner
- Registering and regulating the working of collective investment schemes including mutual funds
- Promoting and regulating self-regulatory organizations
- Prohibiting fraudulent and unfair trade practices in the securities market
- Promoting investors education and training of intermediaries in securities market
- Prohibiting insiders trading in securities
- Regulating substantial acquisition of shares and take-over of companies
- Calling for information, undertaking inspection, conducting enquiries and audits of the stock exchanges, intermediaries and self-regulatory organisations in the securities market

Role of SEBI in the Primary Market

To protect the interest of the investors and to bring back the small investors to the market several measures have been undertaken by SEBI. Entry and disclosure norms are tightened to prevent the exploitation of investors by the unscrupulous promoters. Allocation of shares and promoters' contribution are regulated.

NOTES

1. **Entry norms:** SEBI has issued guidelines to tighten the entry norms for companies accessing the capital market.
   (a) A company should have a track record of dividend payments for a minimum period of 3 years preceding the issue.
   (b) A company whose shares are already listed would fulfill the entry level requirement only if the post issue net worth becomes more than five times the pre issue net worth.
   (c) If a manufacturing company does not have such a track record, it could access the public issue market, provided its project is appraised by a public financial institution or a scheduled commercial bank. The appraising entity should also participate in the project fund.
   (d) It would be necessary for a corporate body making a public issue to have at least five public shareholders for every `1 lakh of the net capital offer made to the public.
   (e) SEBI would not vet offer documents of companies having a track record of 3 years consistent dividend payment. Likewise, offer documents of companies seeking listing on OTCEI would not be vetted by SEBI. Further banks are required to satisfy the criteria of two years of profitability for issues above par.
2. **Promoters' contribution:** Promoters' contribution means contribution by those described in the prospectus as promoters, directors, friends, relatives and associates.
   (a) Promoters' contribution should not be less than 20 per cent of the issued capital irrespective of the issue size.
   (b) The entire promoters' contribution should be received before the public issue. If the issue size exceeds `100 crores, the promoters can bring in not less than 50 per cent of their contribution before opening of the issue and bring in the balance before the calls are made on the shareholders.
   (c) SEBI announced that not more than 20 per cent of the entire contribution brought in by promoters cumulatively in public or preferential issue would be locked in for 5 years. SEBI lifted the provision of lock in period for promoters, contribution in case of listed companies with 3 years track record of dividend payment.
   (d) According to the decision taken by the SEBI Board, in case of non-under written public issue promoters could bring their own money or procure subscription from elsewhere within 60 days of the closure of the issue subject to such disclosures in the offer document.
3. **Disclosure:** The draft prospectus filed with SEBI is made as a public document to enhance transparency. The draft prospects should provide all the needed information to the investor regarding: the present position of the company, the future prospect and the risk factors associated with the investment of the company. In August 1997, SEBI introduced compendium of disclosure and investor protection guidelines, deleting the provisions that have lost relevance and rationalizing the over lapping provision. In accordance with the recommendation of C. B. Bhave committee, SEBI has advised all the listed companies to publish unaudited financial results on a quarterly basis.
4. **Book building:** Book building has been accepted as one of the modes of public issue. SEBI issued guidelines relating to 100 per cent book-building in an issue of

security to the public through prospectus. It recommended a two-tier underwriting system for book built issues. The syndicate members would be responsible for the primary underwriting and the book runners would take on their liability in case of default. SEBI also stipulated that there should be at least 30 book building centres with the syndicate members being present at each centre.

5. **Allocation of shares:** To bring back the small investors to the primary market, the minimum application of share has been reduced from 500 to 200. Proportionate allotment of shares is made. A reservation of minimum 50 per cent of net offers to the small investors is being made. Small investors mean those who have applied for 1000 or fewer shares or securities. The companies were required to complete the allotment of securities within 30 days of the closure of the public issue. There after, they would be required to pay an interest at the rate of 15 per cent per annum, if refund of application money is not made within the specified period.
6. **Market intermediaries:** Licensing of merchant bankers or authorization by SEBI was the first step undertaken to regulate the intermediaries. This licensing of merchant bankers is based on the capital adequacy as well as the track record of the capital market related activities. In course of time, other financial intermediaries such as underwriters, registrars and transfer agents came to be licensed. SEBI has the right to inspect the records of the intermediaries.

The merchant bank categories of II, III and IV are abolished. According to SEBI guidelines, a merchant banker is one who handles public issue and manages them. He has to show a network of `5 crores by the time of his renewal. From 7 December 1997 SEBI advised the merchant bankers to segregate the fund based activities from the fee based activities. This would prevent the mismanagement of investors funds by the merchant bankers.

Role of SEBI in the Secondary Market

SEBI has introduced a wide range of reforms in the secondary market. The important areas are as follows:

1. **Governing board:** Governing Board of the stock exchanges were reconstituted according to the SEBI's directives. Broker and non-broker representations are made 50:50. Sixty per cent of non-broker representation was given in the arbitration, disciplinary and default committees of the exchanges. Regulations regarding the public representatives and the government nominees on the Governing Boards of stock exchanges were issued. As non participants in the market, they have to ensure an impartial and fair governance in the stock exchanges. As per the new guidelines of SEBI, trading members will be given only 40 per cent representation in the governing council of derivatives.
2. **Infrastructure:** To sophisticate the trade on the stock exchange NSE was established with the screen based trading. Then SEBI has allowed the stock exchanges to expand its on-line screen based trading terminals to locations outside their jurisdiction subjected to certain criteria. SEBI decided that recognition to new stock exchanges would be allowed subject to the conditions of On Line Screen Based trading for trade purposes.
3. **Settlement and clearing:** Carry forward transactions were withdrawn and reintroduced by SEBI with modified regulations. SEBI notified all stock exchanges to introduce weekly settlement. Besides, auctions have to be conducted by stock exchanges within eight days of the settlement in case members fail to deliver the

NOTES

shares. The renewal of contract in cash groups from one settlement to another is not permitted.

In the case of short sales, in terms of recommendation of the B. D. Shaw committee, SEBI issued directives to all stock exchanges to submit scrip-wise information (for select scrips) on net short sales position at the end of each trading day for dissemination of information to the public.

SEBI advised all the stock exchanges to set up Clearing Houses, Clearing Corporations or settlement guarantee fund for expediting the process of dematerialization of securities and settlement transaction in the depository mode. SEBI has introduced rolling the settlement trading in the demat segment for all companies with effect from 15 January 1998.

Warehousing facility has been permitted by the SEBI. The 'warehousing' of transaction facilitates execution of a firm client's order for a large quantity in parts during the same trading cycle and the issue of only one contract note at the weighted average price in the end of the trading cycle. SEBI allowed: institutional investors, Flls, stock brokers, stock exchanges to make use of the facility of 'warehousing' of trades, subject to certain conditions.

4. **Debt market segment:** Wholesale debt market segment in the NSE enables the traders to trade in debt instruments. To expedite the settlement SEBI (Depository and participants) Regulations 1996, has been amended. It allowed dematerialization of Government securities. SEBI has allowed the listing of debt instrument on the stock exchange even if the company's equity was not listed earlier. Flls were allowed to invest up to 100 per cent of the funds in debt instruments of Indian companies through 100 per cent dedicated debt fund.

   SEBI has made it mandatory for all the debt instruments to be rated from any one of the authorised credit rating company. Recently, SEBI made dual rating mandatory for the issues of size above `500 million. By doing this the SEBI expects to put an end to credit rating shopping, which allows a company not to disclose an adverse rating and instead seek a favourable rating from other agency. A company will not be allowed to get a rating from an agency which is its associated firm.

5. **Price stabilization:** SEBI has set up a division to monitor the unusual movements in prices, in coordination with the stock exchanges. SEBI has asked stock exchanges to monitor the prices of newly listed permitted scrips from the first day of trading. Necessary circuit breaker system and other market monitoring restrictions could be applied. In case of newly listed scrips, when there is an abnormal price variation, the exchange would impose a special margin of 25 per cent or more on purchase in addition to regular margin. The penal margin is retained by the exchanges for a period of three months or one month after delivery. Intraday margin, gross exposure margin, net exposure margin, mark to market margin, concentration margin and special margins are also imposed on the traders by the stock exchanges to reduce the price volatility.

   To prevent circular trading and price rigging, SEBI has introduced a host of price filters. Intraday price band permits the stock to be traded within a range during a trading session. It aims at preventing intraday price swings. Inter week price band has a wide range within that stocks are permitted to be traded in a week. It prevents wild swings in prices rippling into the next settlement cycle. SEBI imposed

a uniform intra-day price band of 10 per cent for all securities on all stock exchanges, in addition to the 25 per cent weekly price cap being followed by all exchanges.

6. **Delisting:** SEBI tightened the delisting norms by permitting delisting only in accordance with the norms specified by Chandratre committee. The norms are:
   - On voluntary delisting from regional stock exchanges the company would have to make a buy offer to all the shareholders in the particular region
   - The promoters would have to buy or arrange buyers for the security.
   - The listing fees for three years should be taken from the companies at the time of delisting and be kept in Escrow Account with the stock exchange.
7. **Brokers and SEBI:** The regulation of the functioning of the brokers starts with the registration of the brokers. The registration is given on the basis of the eligibility to be a member of any stock exchange, infrastructure facilities like adequate office space, equipment and man power. He should have past experience in the business of buying, selling or dealing in securities. Capital adequacy norms are laid down depending on the stock exchanges and the members' turnover. Registration fees also has to be paid by the brokers.

   Code of conduct is laid down for every stock broker to be registered with SEBI. The code seeks to ensure that a broker should not indulge in malpractice and manipulation. The code of conduct deals with brokers' duty towards: execution of orders, issue of contract note, breach of trust, fairness to clients and investment advice. Contract notes are to show transparency in deals regarding price, brokerage and service tax. The contract notes and bills are to be passed duly on time, failing which brokers are liable to penalties depending on the days of default.

   Brokers shall have to furnish SEBI a copy of the audited balance sheet and profit and loss account within six months of each accounting period. Brokers are expected to preserve the books of account and other records for a minimum period of five years. SEBI has the right to inspect brokers' books of account, other records and documents and can appoint qualified auditors to investigate into the books of accounts.

   Regional offices are set up by the SEBI to attend to the complaints against brokers and stock exchanges. The regional offices are set up in New Delhi, Calcutta, Chennai and Mumbai. Disciplinary actions such as reprimand, warning, suspension and cancellation of the registration are taken against the erring stock brokers. A penalty of suspension of registration of a stock broker may be imposed if:
   - The stock broker violates the provisions of the Act, rules and regulations
   - The stock broker does not follow the code of conduct laid down by SEBI
   - The stock broker:
     - Fails to furnish any information related to his securities as required by the board
     - Furnishes wrong or false information
     - Does not submit periodical returns as required by the Board
     - Does not co-operate in any enquiry conducted by the Board
   - The stock-broker fails to resolve the complaint of the investor or fails to give satisfactory reply to the Board in this behalf

NOTES

- The stock-broker indulges in manipulating price or concerning activities in the market
- The stock-broker is guilty of misconduct
- The financial position of the stock broker deteriorates to such an extent that the Board is of the opinion that his continuance in securities business is not in the interest of investors and other stock brokers
- The stock broker fails to pay the fees
- The stock broker violates the conditions of registration
- The membership of the stock broker is suspended by stock exchange

10.3 COMMERCIAL BANKS, COOPERATIVES AND RRBs

Banks are financial firms and depend on economies of size and gains arising from internalizing certain activities rather than relying on market transactions. Banks provide packages of financial services which individuals find too costly to search out, produce and monitor by themselves.

The evolution of banking, which lasted for centuries until two types of modern banking developed in the industrially-advanced economies in the late 19th century, was an integral part of the expansion of capitalism. The techniques of banking which developed in the 17th century facilitated the industrial and territorial expansion that began about the same time. Banking systems evolved to meet the demands of the constituents, vested interests and regulations governing their establishment. The British system evolved around the central banking system with a central bank and clearing banks with a large network of offices regulated by the central bank; while the German one evolved out of an identification of interests of finance, industry and government to provide multiple services to the constituents. The US system, however, was set apart by the dominance of the unit banks, the role played by an active interbank market in deposits and reserves and the cooperative lending practices. It also featured wholesale banking which was the source of several innovative practices such as rollover credit or flexirate lending. This section will deal with the various types of banks.

10.3.1 Commercial Banks

Among the financial institutions, the role of commercial banks is unique. Firstly, the banks' demand deposit liabilities were ` 2,55,365 crore at the end of March 2005, constituting a large proportion (39.5 per cent) of narrow money, M1 (consisting of currency with the public, demand deposits and other deposits with the RBI) of ` 6,46,263 crore. Of the broader measure of money supply, M3 of ` 22,53,938 crore at the end of March 2005 (which includes M1 + post-office savings banks deposits = M2), time deposits with banks of ` 16,07,675 crore, aggregate deposits of ` 18,91,692 crore with banks constituting 83.9 per cent.

Second, commercial banks are the primary vehicle through which credit and monetary policies are transmitted to the economy. Credit and monetary policies are implemented through action on bank reserves (cash and statutory liquidity ratios), margin requirements and the rate at which scheduled banks can borrow from the RBI. These affect the supply, availability and cost of credit at banks.

Check Your Progress

1. State the powers of the Ministry of Finance according to the SEBI Act.
2. What is badla trade?
3. What is the basis on which the registration of brokers is based?

Third, the nature of lending and investing by commercial banks is multi-functional. They deal in a wide variety of assets and accommodate different types of borrowers. They facilitate the spread of the impact of monetary policy to non-bank lenders and to other sections of the economy. Further, the operations of commercial banks are highly flexible since they provide facilities for financing different types of borrowers which enables them to channel funds according to specified priorities and purposes.

Definition of Banking

The Banking Regulation Act, 1949 defines banking as accepting for the purpose of lending or investment, of deposits of money from the public, repayable on demand or otherwise and withdrawable on demand by cheque, draft or order otherwise.

Special Nature of Banks

Banks are special as they not only accept and deploy large amounts of uncollateralized public funds in a fiduciary capacity, but also leverage such funds through credit creation. Capital represents a very small fraction of total assets of banks especially when compared to non-financial institutions. A minimum percentage of capital of 8 per cent of assets is equivalent to a leverage ratio (debt/equity ratio) of 92/8 = 11.5, which is unsustainable with non-financial institutions. Borrowers would consider it as impairing the repayment ability too much and causing an increase in the bankruptcy risk beyond acceptable levels. The high leverage of banking institutions does not interfere with their functioning because the discipline imposed by borrowers does not apply to depositors who are protected by deposit insurance. Banks require easy and immediate access to financial markets for raising funds as long as the perceived risk by potential lenders remains acceptable. The risks are, however, made visible and explicit by bank ratings.

Special Position of Banks and Financial Institutions

The special nature of banks, creation of liquidity, carries risks unique to management of banks. The basic function of bank management is *risk management*. One would be stretching the point if this is equated with the conferment of special privilege which calls for the imposition of an obligation to provide banking services to all segments of population on equitable basis.

Public sector banks have been, for more than two decades after nationalization in 1969, in the forefront to provide services through branch expansion, like the ` 5 savings account, innovative loan products for small-scale and agriculture sectors. While these endeavours continue, the reform era's emphasis on profitability and stringent accounting standards have cramped the functioning of banks. There is no doubt that extending the reach of banks to poorer sections would ensure that the banking system benefits.

In the developed countries, official policies promote inclusive practices to empower low income groups, going as far as to give them a statutory sanction. In the UK, a dedicated fund has been set up to encourage inclusion. Banks and credit unions have been assigned definite responsibilities. In the US, the Community Reinvestment Act prohibits discrimination by banks against low and moderate-income groups, while in France it is a legal right for anyone to have a bank account.

We can emulate their example if we set up dedicated funds through budgetary allocation. Burdening the fragile Indian banking system to reach out without any regard to the purpose or end use of loan would only contribute to the growth of NPAs. Cash flow generation and augmenting supply of goods and services should guide us to protect

NOTES

the depositors and nation from unviable loans and addition to demand which generate inflationary pressure. Further, budgetary processes have constitutional sanctity and checks and balances exist to ensure that funds voted are spent for the purpose. On the other hand, such mechanisms do not exist in reaching out to the poor.

Functions of Commercial Banks

The functions of commercial banks are:

- To change cash for bank deposits and bank deposits for cash
- To transfer bank deposits between individuals and/or companies
- To exchange deposits for bills of exchange, government bonds, the secured and unsecured promises of trade and industrial units
- To underwrite capital issues.

They are also allowed to invest 5 per cent of their incremental deposit liabilities in shares and debentures in the primary and secondary markets. The commercial banks have set up subsidiaries to provide advice on portfolio management or investment counselling. They also offer their constituents services to pay insurance, advice on tax problems and undertake executive and trustee services.

Payment Systems

Commercial banks are institutions which combine various types of transaction services with financial intermediation. Banks provide three types of transactions: the first is to convert deposits into notes and coins to enable holders of deposits to undertake transactions in cash. Second, bank deposits are used as a means of settling debts. Third, where exchange controls do not exist, banks exchange cash and deposits from one currency into cash and deposits of another currency.

Commercial banks earlier had a monopoly on transaction services. Other financial intermediaries such as savings and loans, savings banks and credit unions in the United States have been authorized to offer transaction accounts. Money market mutual funds, another type of financial service organizations, have developed financial products against which cheques may be written.

Commercial banks are at the very centre of the payment systems. Bank money constitutes 39.5 per cent of the money supply (Ml) of the Indian economy. An efficient payment system is vital for a stable and growing economy and the bank's role is important.

In advanced economies, commercial banks are also at the heart of the electronic payment system which is replacing paper-based payment methods. In the USA, electronic payment between commercial banks is done through Fedwire which is a wholesale wire transfer system operated by the Federal Reserve System. About 3,00,000 transfers per day, amounting to $1 trillion, are made. Large banks in New York operate a private electronic transfer system called CHIPS (The Clearing House Interbank Payments System) which transfers $1 trillion a day involving international movement of funds.

Finally, SWIFT(the Society for Worldwide Interbank Financial Telecommunication) based in Brussels is operated by 2,000 banks, brokerage firms and non-banking financial institutions worldwide.

Intermediation

Commercial banks undertake the important process of financial intermediation whereby the funds or savings of the surplus sectors are channelled to deficit sectors. Commercial

banks, along with other financial institutions, channel the funds of surplus economic units to those wanting to spend on real capital investments. Funds are transferred through lending by banks or by creation of financial liabilities such as bonds and equity shares. Banks intermediate by obtaining the funds of savers in exchange for their own liabilities such as entries in a passbook and then, in turn, make loans to others. Financial intermediaries including banks buy and sell the right to future payments. Banks collect deposits from savers by offering interest. In 2008-09, savings of the households in the form of bank deposits constituted 54.9 per cent of total financial savings. Deposits of commercial banks can be of any denomination which have the characteristics of low risk and high liquidity. The small deposits are put together to make large loans.

Through their intermediary activities, banks provide a package of information and risk-sharing services to their customers. While doing so, they take on part of their risk. Banks have to manage the risks through appropriate structuring of their activities and hedge risks through derivative contracts to maximize their profitability.

Transformation Services

Banks combine various types of transformation services with financial intermediation. They provide three transformation services when they undertake intermediation process. First, liability, asset and size transformation, consisting of mobilization funds and their allocation (provision of large loans on the basis of numerous small deposits). Second, maturity transformation by offering the savers, the relatively short-term claim on liquid deposits they prefer and providing borrowers long-term loans which are better matched to the cash flows generated by their investment. Finally, risk transformation by transforming and reducing the risk involved in direct lending by acquiring more diversified portfolios than individual savers can. Commercial banks, by effectively appraising credit requests, can channel funds into productive uses.

Transformation Services and Risks

Banks incur risks while undertaking transformation services. In the last three decades, banks abroad assumed new roles and accepted new forms of financial intermediation by undertaking currency and interest rate swaps and of dealing in financial futures, options and forward agreements. These new instruments reflect considerable flexibility in responding to market situations and adjusting continually assets and liabilities both on and off-balance-sheet, while enhancing profitability.

Measurement, control and monitoring of risk will help banks to attain the objective. Techniques such as gap duration and value at risk are suggested to analyse risk. Strengthening of information technology in commercial banks is a prerequisite to implement effectively ALM system. The role of a broad-based ALCO in advising boards of banks is of significance.

Financial Services

Commercial banks provide securities-related services. Commercial banks in India have set up subsidiaries to provide capital market related services, advice on portfolio management or investment counselling. In the US, the Glass-Stegall Act of 1933 restricts the nature of services provided by commercial banks. In the US, they may offer discount brokerage services but not general purpose brokerage services. US banks facilitate mergers and acquisitions and trading in currencies and US government securities.

The Glass-Stegall Banking Act prohibits commercial banks from owning a firm dealing in securities. The Act has been challenged by banks offering money market

NOTES

mutual funds and other investment services. The US Federal Reserve Board in January 1997, issued a proposal that would allow banks holding companies and their securities, industry affiliates to offer 'one stop shopping' for their customers. Commercial banks in the US in 1990s have become very active in the management and distribution of mutual funds, managing more than 10 per cent of the assets of all mutual funds. In India, several commercial banks such as Bank of India, Canara Bank, Indian Bank and State Bank of India have set up subsidiaries under the guidelines issued by the Reserve Bank in 1987, followed by guidelines laid down by the Ministry of Finance in 1991.

Fiduciary Services

In the US, banks manage employee pension and profit-sharing programmes that do not show up on banks' balance sheets. In the US, banks operate separate trust departments which manage the funds of others for a fee, under the guidance of a trust agreement. The assets held in trust do not show up on the banks' balance sheets because they do not own these assets.

Off-balance-sheet (OBS) Operations

Banks assume contingent liabilities and sell derivatives which are non-fund based activities to their clients. The off-balance-sheet exposure of banks also extends to balance sheet risk management and generating profits through leveraged positions. Like any on-balance-sheet exposure, an OBS exposure exposes a bank to several risks such as credit risk, liquidity risk, market risk and operational risk. The nature of risks faced by banks in OBS activities is not different from on-balance-sheet items (consisting mainly of foreign exchange contracts 49.3 per cent of total derivatives and single currency interest rate swaps 46.6 per cent).

OBS exposures of ` 18,48,341 crore in total liabilities constituted 119.7 per cent of total liabilities at the end of March 2004. The large amount reflects the impact of deregulation, need for risk management, need for diversified income base due to pressure on margin on conventional on-balance-sheet items and new opportunities thrown up by technological progress.

The risks arising on account of OBS activities of banks are sought to be controlled through a combination of both banks, internal control policies and risk mitigation mechanism imposed by the regulator.

Insurance

Banks abroad have generally been permitted to act as distributors of insurance products. They have set up their own subsidiary or bought shares of insurance companies or swapped shares with insurance companies. Insurance companies, in turn, have acquired stakes in some banks either as investment, diversification or to promote the distribution of insurance products through bank branches.

Banks and insurance companies could combine to mutual benefit. Commercial banks can use their branch network to sell all types of insurance, particularly life insurance to their traditional customers. Insurance companies design complex financial products and offer them for placing savings that private customers find particularly appealing such as retirement funds or single premium insurance policies.

Banks could engage in insurance activities in-house via a department in the bank, a separately capitalized subsidiary of the bank or a separately capitalized affiliate of its

branch-holding company. Expansion into insurance by banks will result in new and complex risks leading to larger claims on deposit insurance or the safety net in general. The combining of banking and insurance activities raises the question of competing jurisdiction of two regulators.

The linkages between banks and insurance companies have to be defined by the government, as an owner of public sector banks, DFIs, insurance companies and as a sovereign that has to assign appropriate regulatory jurisdiction whenever the two activities overlap.

Entry of Banks into Insurance in India

Banks have been permitted to enter into insurance business after the enactment of Insurance Regulatory and Development Authority Act, 1999. Banks having a minimum net worth of ` 500 crore and meeting the criteria for capital adequacy, profitability, NPA level and track record of existing subsidiaries can undertake insurance business through joint venture (50 per cent equity) with risk participation. Banks which are not eligible as joint venture participants can participate without the risk participation basis up to 10 per cent of their net worth or ` 50 crore (whichever is) lower in an insurance company for providing infrastructure and services support without taking on any contingent liability. Insurance products can be distributed by scheduled commercial banks as an agent of insurance company.

10.3.2 Cooperative Banks

The cooperative banking sector in India comprises urban cooperative banks (UCBs) and rural cooperative banks such as state cooperative banks (SCBs) and district central cooperative banks (CCBs). The cooperative banking sector has an extensive branch network and reach in the remote areas.

Cooperative credit institutions have played a catalytic role in mobilizing rural savings and stimulating agricultural investment. Cooperative credit institutions account for the second largest proportion of 28.5 per cent of total institutional credit of ` 1,08,500 crore to agriculture and allied activities in the rural sector in 2004-05. Commercial banks have emerged as the predominant source of institutional credit with agriculture accounting for 61.2 per cent in 2004-05; and regional rural banks provided 9.7 per cent. However, the overall share of institutional credit in the total rural debt market is very small. Institutional credit forms 18 per cent of agriculture and allied activities component of the GDP in 2003-04. The often-quoted statement that share of debt of institutional agencies in the case of rural households has increased from 61.2 per cent to 64 per cent between 1981-91 does not establish that institutional credit to agriculture has gone up. The All-India Debt and Investment Survey, 1991-92 does not depict the true picture since the figures relate to total outstanding debt and not the overall share of institutional credit in the total debt market which is likely to be much smaller. Since the introduction of reforms in 1991-92 there has been very little perceptible improvement in either stability or efficiency of cooperative banks. In particular the asset quality and profitability of scheduled UCBs showed deterioration in the reform period.

Role of RBI

Although cooperatives were established following the enactment of Cooperative Societies Act in 1904, efforts to develop cooperatives were intensified following the recommendations of the All-India Rural Credit Survey Committee. The committee

NOTES

assigned crucial role to the RBI in building up the cooperative credit organization by being:

- An active collaborator in drawing up schemes of development with the Government of India and state governments
- The provider of finance first to the state governments for contribution to the share capital of cooperative credit institutions at various levels and secondly, to the cooperative credit structure itself to meet its requirements of short-term, medium-term and long-term finance

To meet the financial obligations, two funds were set up under the RBI Act—the National Agricultural Credit (long-term operations) Fund and the National Agricultural Credit (stabilization) Fund with contributions from the RBI every year.

The RBI became an active agency for the promotion of an appropriate structure of institutions as well as policies and procedures to enhance the share of cooperative credit in rural credit. Large, short-term and medium-term loans to cooperative banks were given.

The Reserve Bank also undertook the task of building up the cooperative credit structure and rationalization of policies adopted by it. The Committee to Review Arrangements for Institutional Credit to Agriculture recommended, in its report in 1981, that the rural credit structure should be strengthened with a view to evolve an efficient credit delivery system at all levels for effectively implementing the concept of an integrated approach to rural development.

Based on the recommendations of the All-India Rural Credit Review Committee (1969), a multi-agency approach to rural credit was adopted with the commercial banks supplementing the efforts of the cooperative banks. In 1975, to further strengthen the rural credit system, a third channel to supply credit to rural areas was added by setting up regional rural banks.

The cooperative banking system was developed on the premise that both short-term and long-term credit cannot be supplied by the same organization. The short-term structure extending production credit has Primary Agricultural Cooperatives (PACs) at the grass-root level, Cooperative Central Banks (CCBs) at the intermediate level and State Cooperative Banks (SCBs) at the apex level. The long-term structure, extending investment credit have Primary Cooperative Agriculture and Rural Development Banks (PCARDBs) at the base level which are affiliated to the State Cooperative Agriculture and Rural Development Banks (SCARDBs) at the state level. The State Cooperative Bank at the apex level coordinates and regulates the working of central cooperative banks and provides the link between the RBI and the money market on the one hand and the entire cooperative system in the state on the other. Central cooperative banks are the intermediate agency between the state cooperative banks and the primary credit societies at the village level run by agriculturists. The primary credit societies form the base of the cooperative credit system, with most of the members being agriculturists.

Advisory Committee on Flow of Credit to Agriculture and Related Activities (2003)

The Tenth Five-year Plan (2002-07) envisages a substantial jump in credit flows to agriculture to ` 7,36,750 crore in the plan period as compared to the credit flow of ` 2,29,956 crore achieved in the Ninth Five-year Plan period. Credit flow from all formal sources in 2002-03 (` 70,810 crore), in 2003-04 (` 86,981 crore) and in 2004-05 (`

1,08,500 crore) was much below the levels achieved in the Tenth Plan. The budget for 2004-05 proposed the setting up of a Task Force to examine the reforms required in the cooperative banking system. The RBI also constituted an Advisory Committee (Chairman, Prof. V. S. Vyas). Based on the report of the Vyas Committee the RBI announced a number of measures to improve credit delivery to agriculture. These include waiver of margin/security requirements on agricultural loans up to ` 50,000 and in case of agri-business and agri-clinics for loans up to ` 5 lakh, revision of NPA norms for agricultural finance and dispensing with service area approach except for government sponsored schemes.

Organizational Structure

The organizational structure of cooperative credit institutions reveals that cooperative banking has a more widespread and extended network in the rural sector than in the urban sector.

The cooperative banking system consists of rural and urban cooperative banks with the former being characterized by a relatively widespread branch network. The cooperative banking system supplements the effort of commercial banks in mobilizing savings and meeting the credit needs of the population.

The cooperative credit structure consists of 1,00,113 institutions with distinct urban and rural sectors. The urban sector has 1,770 cooperative banks (with 53 scheduled banks) and 98,343 rural cooperative credit institutions. The rural credit institutions are again divided into short-term and long-term credit institutions. The short-term credit institutions consist of state cooperative banks numbering 31, central cooperative banks numbering 371 and primary agricultural credit societies numbering 97,224; and the long-term institutions consist of 20 state cooperative agriculture and rural development banks and 697 primary agriculture and rural development banks.

(i) Urban Cooperative Banks (UCBs)

Urban cooperative banks (primary cooperative banks) have been set up with the objective of promoting banking habits among lower and middle income groups in urban areas and purvey credit to small borrowers including weaker sections of the society. The Reserve Bank is entrusted with the responsibility of regulation and supervision of the banking related activities of urban (primary) cooperative banks under the Banking Regulation (B.R.) Act, 1949 as applicable to Cooperative Societies (AACS). Other aspects such as incorporation, registration, administration, management and winding-up of UCBs are supervised and regulated by the respective state governments through Registrars of Cooperative Societies (RCS) under the Cooperative Societies Acts of the respective states. UCBs with a multi-state presence are registered under the Multi-State Cooperative Societies Act, 2002 and are regulated and supervised jointly by the Central government through the Central Registrar of Cooperative Societies and the Reserve Bank. The number of UCBs stood at 1,770 at the end of March 2008 including 79 salary earners' banks and 105 Mahila banks. The branches number 6,990. Of the urban cooperative, the scheduled banks which are under closer regulatory and supervisory framework of the RBI number 53 at the end of March 2008. Applications for banking licence including licence for opening of new branches are not considered (2005).

Urban cooperative banks display a high degree of heterogeneity in terms of deposits/ asset base, area of operation and nature of business. In view of their importance the sector should emerge as a sound and healthy network of jointly owned, democratically

NOTES

controlled and professionally managed institutions. The sector showed weaknesses out of lack of corporate governance, unethical lending and large loan defaults. A medium-term framework has been evolved on the basis of a vision document to protect depositors' interest and enable UCBs to provide useful service to local communities. The RBI initiated several regulatory measures during 2004-05 to ensure growth of UCBs on sound lines.

Loan accounts of certain UCBs are allowed to be classified as NPA based on 12 months delinquency norm until 2009.

(ii) Rural Credit Institutions

State Cooperative Banks (SCBs): State cooperative banks (39) are at the apex of the three-tier cooperative structure meeting short-term and medium-term credit. Their capital and reserves formed 12.3 per cent of total liabilities and 20.0 per cent loans and advances. Deposits at ` 48,560 crore constituted 56.6 per cent of liabilities. Loans at ` 52,777 crore formed 108.7 per cent of deposits. Loans and investments together formed 83.3 per cent of assets of SCBs. Borrowings at ` 22,256 crore formed 30 per cent of liabilities. Recovery performance was 85.7 per cent.

The aggregate NPAs of SCBs were estimated at ` 8,704 crore or 14.2 per cent of their outstanding loans as at the end of March 2008.

Central Cooperative Banks (CCBs): Central cooperative banks in 2007 numbering 371, form the middle tier in the short-term credit structure of cooperative credit structure. Deposits of ` 94,524 crore constitute the major source of funds accounting for 59.6 per cent of liabilities. Borrowing at ` 20,256 crore accounted for 18.8 per cent of liabilities. Capital and reserve accounted for 16.5 per cent of liabilities and 31.5 per cent of loans and advances. NPAs constituted 18.5 per cent of total loans.

(a) Short-term Rural Credit

Primary Agricultural Credit Societies (PACS)

The short-term cooperative credit system in rural areas rests on the PACS. PACS directly interface with individual farmers, provide short-term and medium-term credit, supply agricultural inputs, distribute consumer articles and arrange for marketing of produce of its members through a cooperative marketing society. They are the grass-root level cooperative credit institutions playing a vital role in the disbursement of short-term credit.

(b) Long-term Credit Structure of Rural Cooperative Banks

The long-term rural cooperative credit structure consists of State Cooperative Agriculture and Rural Development Banks (SCARDBs) at the top level dispensing investment credit and Primary Cooperative Agriculture and Rural Development Banks (PCARDBs) at the bottom level.

Revival of Rural Cooperative Banking Sector (2005)

The Task Force on the revival of rural co-operative credit institutions (Chairman A. Vaidyanathan) recommended the restoration of autonomy to the credit cooperatives by scaling down the control and interference by state governments through amendments to the State Cooperative Societies Act. Amendment to the Banking Regulation Act also was recommended. The Task Force recommended the provision of financial assistance

for recapitalization to fund the accumulated losses of the short-term cooperative credit structure evolving a common accounting system, management information system and computerization and HRD initiatives. The financial assistance is contingent upon introduction of institutional, legal and regulatory reforms. In January 2006, the Central government announced a revival package of ` 13,596 crore to cover assistance for cleansing balance sheets of short-term cooperative credit structure, capital infusion for ensuring CRAR of7 per cent, technical support for capacity building for training, introduction of common accounting and MIS and their computerization.

- B.R. Act, 1949 should be amended to remove dual control and bringing the cooperatives under the regulatory control of the RBI.
- The financial assistance has been estimated at ` 14,839 crore to be shared by the Government of India, state government and the cooperative credit structure. NABARD has been designated as the Total Implementing and Pass Through Agency to coordinate and monitor the programme representing the Government of India.

Rural Infrastructure Development Fund (RIDF)

A Rural Infrastructure Development Fund with a corpus of ` 2,000 crore was constituted by NABARD in April 1995 for advancing loans to state governments and state-owned corporations for quick completion of ongoing projects relating to medium and minor irrigation, soil conservation, watershed management and other forms of rural infrastructure. Eleven branches of allocations have been made towards the fund since then. Scheduled commercial banks/public sector banks contributed to the corpus to make up for their shortfall in their priority. The scope of lending was widened to cover Gram panchayats, self-help groups and projects in social sector covering primary education, health and drinking water. The total corpus of RIDF I to X stood at ` 72,000 crore and assistance sanctioned and disbursed at ` 74,073 crore and ` 45,693 crore at the end of March 2008. Disbursements have remained low on account of difficulties associated with the identification of appropriate projects, lack of budgetary support where only part of the funding was visualized, delays in finalization of formalities for withdrawal of funds and in completing land acquisition and tendering procedures in the case of irrigation projects. No state had utilized RIDF sanctions in full and state wise utilization was uneven. Of the total sanction under RIDF I to X rural roads and bridges accounted for 45 per cent, irrigation projects 34 per cent and others 21 per cent (social sector 9 per cent and power 3 per cent).

STCR Refinance Fund (STCR CF)

The setting up of Short-term Cooperative Rural Credit (STCRC) Refinance Fund with a corpus of ` 5,000 crore was announced by the Union finance minister in the budget for 2008-09.

National Bank of Agriculture and Rural Development (NABARD)

The National Bank for Agriculture and Rural Development was established in July 1982 to take over and decentralize the functions in the sphere of rural credit. It was set-up for providing credit for the promotion of agriculture, small-scale industries, cottage and village industries, handicrafts and other rural crafts and allied economic activities in rural areas with a view to promoting integrated rural development and securing prosperity for rural areas. The new apex bank was envisaged as an organizational device to provide undivided

NOTES

attention, forceful direction and a pointed focus to the credit problems arising out of an integrated approach to rural development. The Committee to Review Arrangement for Institutional Credit for Agriculture and Rural Development (1981) recommended that the new bank will take over from the Reserve Bank the overseeing of the entire rural credit system and the statutory inspection of cooperative banks and regional rural banks on an agency basis; the bank continues to retain its essential controls.

On its establishment, NABARD has taken over the RBI's refinancing functions in relation to state cooperative banks and regional rural banks. The bank is now the coordinating agency in relation to the Central government, Planning Commission, state governments and institutions at all-India and state levels engaged in giving effect to the various policies and programmes relating to rural credit.

NABARD's Capital

The capital of the bank of ` 100 crore was originally subscribed by the Central government and the RBI in equal proportions. It was raised later to ` 500 crore. The Union Budget for 1996-97 had proposed quadrupling of the capital base to ` 2,000 crore in the next five years. The share capital of NABARD was raised to ` 2,500 crore in 2007-08 with an advance of ` 2,500 crore from the Centre (see Table 10.2).

NABARD serves as an apex refinancing agency for the institutions providing investment and production credit in rural areas. NABARD provides refinance to state cooperative agriculture and rural development banks (SCARDBs), state cooperative banks (SCBs), regional rural banks (RRBs) and other financial institutions approved by the Reserve Bank. The ultimate beneficiaries of refinance from NABARD could be individuals, partnership concerns, companies, state-owned corporations or cooperative societies.

Other main activities of NABARD include: (*i*) initiating measures towards institution building for improving absorptive capacity of the credit delivery system, including monitoring, formulation of rehabilitation schemes, restructuring of credit institutions and training of personnel; (*ii*) coordinating the rural financing activities of all institutions engaged in developmental work at the field level; (*iii*) maintaining liaison with the Government of India, the state governments, the Reserve Bank and other national level institutions concerned with policy formulation; and (*iv*) undertaking monitoring and evaluation of projects refinanced by it.

Table 10.2 *National Bank for Agriculture and Rural Development: Liabilities and Assets (Select Years)*

(` *crore*)

Item	*1990-91*	*1995-96*	*1998-99*	*2003-04*	*2007-08*
Liabilities					
Capital	100	500	2,000	2,000	2,000
Reserves	471	1,738	2,820	5,291	8,848
NRC (LTO) Fund	5,687	8,185	10,020	13,070	13,215
NRC (Stab.) Fund	773	840	1,044	1,500	1,534
Borrowings from GOI including IDA/IBRD assistance	1,736	1,294	1,003	563	370
General line of credit from RBI	2,391	4,787	5,649	4,194	0
Borrowings from RBI under ARDR Scheme, 1990	692	204	540	—	0
Open Market Borrowings	633	1,045	1,633	14,383	32,622

NOTES

RIDF Deposits	—	350	3,068	1,208	30,593
Foreign Currency Loan (KFW-Germany)	—	70	268	29	508
Others	268	595	94.1	250	9,163
Total	12,751	19,608	28,986	55,889	98,853
Assets					
Refinance Outstanding (MT & LT)	6,646	11,145	15,752	28,079	32,473
General line of credit (ST)	2,412	4,789	5,711	5,598	17,382
Medium-term loan from NRC (LTO) Fund	260	178	61	9	0
Conversion Loans from NRC (Stab.) Fund	353	32	473	663	2,058
Loans to State Govts. for contribution to share capital of coop. credit societies	226	363	499	450	290
Bills Rediscounted	305	108	21	—	0
Loans under ARDR Scheme, 1990	666	198	—	—	0
Investment in Government Securities	910	1,298	1,245	2,271	1,683
Loans out of RIDF	—	387	3,667	14,004	30,649
Others	973	1,110	1,541	4,792	12,955
Special Deposits with RBI in respect of NRC (Stab.) Fund	—	—	—	—	—
ADFC Equity	—	—	16	17	1,316
Total	12,751	19,608	28,986	55,889	98,853

***Source*:** Reserve Bank of India, *Report on Currency and Finance*, 1995-96, Vol. II, p. 120, *Handbook of Statistics on Indian Economy*, 1999, pp. 110-111, 2003-04, pp.136-137 and 2007-08, pp. 153-154.

Operations

NABARD is empowered to provide short-term refinance assistance for periods not exceeding 18 months to state cooperative banks, RRBs and any financial institutions approved by the RBI for a wide range of purposes including marketing and trading relating to rural economy. NABARD can grant medium-term loans to the state cooperative banks and regional rural banks for periods extending from 18 months to seven years for agriculture and rural development.

Refinance

NABARD is empowered to provide, by way of refinance assistance, long-term loans for 25 years to state loan development banks (now known as state cooperative agricultural and rural development banks), RRBs, scheduled commercial banks and state cooperative banks for the purpose of making investment loans, as well as giving loans to artisans, small-scale industrial units and village and cottage industries. Loans for periods less than 20 years can be made to state governments to enable them to subscribe to the share capital of cooperative credit societies.

Resources

Table 10.1 presents the assets and liabilities of NABARD for the select year. For its short-term operations, NABARD can borrow from the RBI in the form of a line of credit to enable NABARD to extend short-term refinance to state cooperative banks and RRBs. For its term loan operations, NABARD draws funds from the Central government, the World Bank/IDA and other multilateral and bilateral agencies, the market and the National Rural Credit Fund (Long-term Operations) that it has established.

NABARD can raise funds by sale of bonds and debentures, direct borrowing, acceptance of deposits, receipt of gifts and grants. NABARD may borrow foreign

currency from any bank or financial institution in India and abroad with the approval of the Central government which will guarantee such loans.

Supervision of Cooperative Banks

NOTES

NABARD conducts inspection of SCBs and CCBs under the provision of the Banking Regulation Act, 1949. It conducts on a voluntary basis periodic inspections of state level institutions such as SCARDBs. The approach to on-site supervision of cooperative banks is based on CAMELs, supplemented by off-site surveillance through a set of returns, statutory and non-statutory.

Reforms in Cooperative Credits

Apart from prescribing entry point/viability norms for cooperative banks and guidelines relating to their operations, the Reserve Bank also placed stipulations on primary (urban) cooperative banks in respect to income recognition, classification ofassets and provisioning on the lines stipulated for commercial banks.

The financial sector reforms to improve efficiency and productivity of banks were extended in stages to the cooperative banks as well. The measures relating to interest rate deregulation application of prudential norms relating to asset classification, income recognition and provisioning have been extended. Capital adequacy norms are yet to be extended.

Lending and deposit rates of all cooperative banks, excluding PCBs were deregulated in October 1994 subject to an MLR of 12 per cent.

Scheduled PCBs were allowed in April 1996 to undertake equipment leasing and hire-purchase financing activities after meeting certain prudential requirements. In December 1996 PCBs were permitted to extend direct finance for agricultural activities to the entire district of registration and they will be counted as priority sector advances.

The prudential accounting norms were made applicable to state cooperative banks/ central cooperative banks from the year 1996-97 in two phases, 1996-97 and 1997-98.

Supervision

Since 1996, the RBI acquired a statutory basis for inspection of state cooperative banks, central cooperative banks and the primary cooperative banks. The state cooperative banks fulfilling certain conditions became eligible to acquire the status of scheduled banks. The cash reserve was prescribed at 3 per cent of demand and time liabilities to be maintained with the RBI. Other cooperative banks were also required to maintain a 3 per cent cash ratio of their demand and time liabilities with themselves, the RBI, the SBI or the state cooperative bank of the state concerned. Further, liquid assets at 25 per cent of demand and time liabilities were required to be maintained. The control over advances applied to the cooperative credit sector.

Deposit Insurance

To facilitate deposit mobilization by cooperatives, the benefit of insurance to their deposits was extended to cooperatives in 1968.

Review of Working of Cooperative Credit Institutions

The financial health of the cooperatives is affected by poor resource base, high transaction costs with low margins, mounting overdue, lack of professional management and excessive state control. Among the rural credit institutions, the primary agricultural credit societies

are the grass-root level institutions numbering 98,343 with a total membership aggregating two crore. In view of their inability to mobilize deposits they depend mainly on borrowings from higher financing agencies to fund their operation. The decline in their borrowing membership has been affecting the volume of their business.

The organizational structure of the cooperative banking system is mainly responsible for the high transaction costs. For improving the viability of the cooperative banking systems, two alternatives for restructuring the organizational set-up exist: first, merger of long-term and short-term structure and second, delayering of the short-term structure. Merger or integration of the two wings of the credit structure was suggested by the committee under the chairmanship of Dr R. K. Hazari in 1976 to provide credit facilities in an integrated manner. Integration would inure to the benefit of the farmer since it would be one contact point. The institutions would also benefit because it would enlarge their business and consequently their viability. The committee recommended that long-term and short-term wings of the cooperative credit should be integrated at all levels, the primary, the intermediate/district and the apex.

Only Andhra Pradesh implemented the integration of short-term and long-term structure in 1986. The integration resulted in a higher flow, reduction of management costs and reduction of bad debts.

Delayering is suggested to reduce transaction costs. Overdues restrict recycling of funds and render the institutions ineligible for availing of refinance facility from NABARD. CBSR (1998) recommended the delayering of the cooperative credit system to reduce cost of intermediation and provide the benefit of cheaper NABARD credit to the ultimate borrowers.

To reduce transaction costs and improve recovery the linking of self-help groups (SHGs) and non-government organizations has been advocated by NABARD since 1992. NABARD has been providing 100 per cent refinance to banks for financing SHGs. Until March 2005, 16.18 lakh SHGs have been linked covering 5.6 lakh families and involving ` 57.1 crore of bank lending. Intermediation of SHGs and NGOs has helped banks to reduce transaction costs and improvement of recovery. The transaction costs of the borrowers were also reduced due to elimination of cumbersome procedure of documentation and time and cost incurred on repeated visits to banks. Linkage has also shifted loans from non-income generating activities to production activities, improved the income levels of members and social empowerment of members. The linkage programme under the three models discussed above have enlarged the coverage and reduced NPAs to a near-zero level.

The ideal solution would be a consolidation of the cooperative system of credit into one, since the emphasis should be on organization which should be 'local' in nature. All the existing eight types of institutions numbering 1,08,791 could be put under one umbrella and appropriate credit delivery mechanisms could be devised where the loan is small and no collateral exists the help of self-help groups and non-governmental organizations can be enlisted. All proposals generating cash flows could be financed and emphasis should shift from individual to group or cooperative loans. Further, the political and regional element has to be eliminated by disassociating the apex level with the state government. The entire cooperative credit complex can be organized into four or five banks retaining the cooperative form of organization for decision-making, local units electing local unit board members and local unit board members electing the central board members. Members of local and state level institutions should be borrowed from cooperative credit institutions. The cooperative movement should be taken to its logical

conclusion by shedding the political content which is inevitable as long as apex units are equated with state units and political element is allowed free play.

10.3.3 Regional Rural Banks (RRBs)

NOTES

Regional Rural Banks (RRBs) are oriented towards meeting the needs of the weaker sections of the rural population consisting of small and marginal farmers, agricultural labourers, artisans and small entrepreneurs. RRBs were set up after the nationalization of banks in 1969 when the emphasis shifted to providing more credit to weaker sections. After the passage of the Regional Rural Banks Act in 1976 they were set up to make credit available to rural households besides instilling thrift. RRBs as subsidiaries to nationalized banks, were expected not only to provide credit to farmers and village industries but also mobilize deposits from rural households in the long run. They form an integral part of the rural financial architecture in India. In March 2008, there were 91 RRBs catering to the credit requirements of 586 districts in 25 states with a network of 14,790 branches.

RBI Assistance

With a view to facilitate their operations, the RBI gave regional rural banks direct access to refinance assistance at a concessional rate of 3 per cent below the bank rate. They have been allowed to maintain a lower level of statutory liquidity than the scheduled commercial banks. They have been allowed to pay 0.5 per cent more interest on all deposits except those of three years and above. The sponsor banks IDBI, NABARD, SIDBI and other financial institutions are statutorily required under the Regional Rural Bank Act to provide managerial and financial assistance to them.

Growth of RRBs

There were six RRBs in 1975 and in 1981 they grew to 107. The total number of districts covered rose from 12 to 182. The banks which mobilized their own resources (` 5 crore and above) were required to commit their own funds to the extent of 25 per cent of their outstanding loans and advances.

The sponsor banks' involvement varied depending on the regional rural banks' involvement. It was 25 per cent in the case of RRBs with deposits of ` 5 crore and above; and 35 per cent in the case of RRBs with minimum involvement of 15 per cent.

During 1982-83, RRBs were divided into three categories and refinance specified.

	Refinance
Outstanding loan business of ` 8 crore and above	30 per cent
Outstanding loan business of less than ` 8 crore but which are in existence for five years	20 per cent
Not falling in the above two categories	15 per cent

Refinance from sponsor banks for the remaining portion of eligible loans and advances was up to 30 to 35 per cent as the case may be.

During the period 1975-81, the number of branches rose from 17 to 4,795 and deposits from ` 0.20 crore to ` 338 crore. By 1987, the number of RRBs rose to 196, districts covered to 188, branches to 13,533 and deposits to ` 12,305 crore and outstanding advances rose to ` 2,732 crore. On 31 March 2008 to their total deposits were ` 99,095 crore and advances were ` 57,601 crore.

NOTES

Evaluation of RRBs

The committee constituted by the RBI in June 1977 to evaluate the performance of RRBs concluded that with some modifications in their organization and structure, they could become a useful component in the totality of the rural credit structure. The committee also suggested the setting up of RRBs where the cooperative organization was not able to adequately serve the credit needs and even in other areas in order to fill the large gap. This formed the basis for the branch licensing policy during 1979-81.

The Working Group on Multi-agencyApproach in Agricultural Financing, appointed by the RBI, in its report in 1978, underscored the need for the commercial banks and particularly the RRBs to play a supplementary role in providing credit for agriculture and allied activities.

In 1980, the RBI studies of the functioning of RRBs found that they had been able to achieve the main objective of helping the weaker sections in the rural areas in meeting their credit requirements despite their constraint of limited area of operation, unhelpful topography and unenterprising clientele. The study found that RRBs can break-even at ` 8 crore business, through a network of 70 branches in about six years, provided they enjoy a margin of 5 per cent between borrowing and lending rates.

The Committee to Review Arrangements for Institutional Credit for Agriculture and Rural Development, examined the role of RRBs in the rural development work interalia and suggested the following:

- As these banks were more suitable for rural development work, preference should be given to them to open branches in rural areas.
- The eligible business of commercial banks' rural branches may be transferred to RRBs.
- The losses in initial years of RRBs may be met by shareholders and equity capital should also be raised.
- The various facilities provided by sponsor banks should continue for 10 years in each case.
- Concessionary refinance by RBI should be continued.
- The control, regulation and promotional responsibilities relating to RRBs should be transferred from the Government of India to RBI or NABARD.

Restructuring of RRBs

As recommended by the Bhandari Committee, 49 RRBs were taken for restructuring and revival in 1994-95. Action was initiated through developmental action plans by NABARD on the managerial, operational and operational restructuring of RRBs and cleansing of their balance sheets over a five-year time span on the basis of the rolling plan concept. An amount of ` 360 crore made up of 50 per cent from GOI, 35 per cent from sponsor banks and 15 per cent from state governments for restructuring and revival in 1994-95, was provided.

Table 10.3 *Combined Balance Sheet of Regional Rural Banks*

(` crore)

1	31 March 2008
Liabilities	**1,23,541**
Share Capital	196
Reserves	5,687
Share Capital Deposits	2,833

NOTES

Deposits	99,095
Current	5,689
Savings	53,370
Term	40,036
Borrowings	11,649
NABARD	8,350
Sponsor Bank	3,250
Others	49
Other Liabilities	4,081
Assets	**1,23,541**
Cash in Hand	1,412
Balances with RBI	7,164
Other Bank Balances	23,493
Other Investments	25,073
Loans and Advances (net)	57,601
Fixed Assets	214
Other Assets#	8,584
Memorandum Items:	
a. Credit-Deposit Ratio	60.3
b. Investment-Deposit Ratio	45.2
c. (Credit + Investment)-Deposit Ratio	106.4

Source: Reserve Bank of India, *Report on Trend and Progress of Banking in India*, 2007-08.

In December 1995, the Basu Committee set up by NABARD, recommended the selection of 68 RRBs for comprehensive restructuring under Phase II. The initiatives were primarily in the areas of interest rates, relocation of branches, credit allocation, direction of credit and manpower policy in simultaneity with the infusion of capital. The Government of India released a sum of ` 1,867.65 crore between 1994-98 and 1998-99 for the re-capitalization of RRBs.

In addition, additional equity support of ` 305.3 crore was provided in 1998-99. In 1998-99, 175 of the total of 196 RRBs stood fully or partially re-capitalized while two RRBs did not require support. Only 19 RRBs were left outside the ambit of the re-capitalization programme.

Table 10.3 presents the important banking indicators of RRBs at the end of March 2008. The investments of RRBs at ` 25,073 crore were mainly in government securities at 20.2 per cent and while the aggregate investment-deposit ratio was 45.2 per cent, the credit-deposit ratio was 60.3 per cent in 2007-08. The investment and credit to deposit ratio was 106.4 per cent.

Purpose-wise Distribution of Loans

The purpose-wise distribution of loans and advances of RRBs as at the end of March 2008 shows that agricultural loans accounted for 55.4 per cent of total loans and advances of ` 59,751 crore in 2008. Of the agricultural advances, crop loans accounted for 68.8 per cent while term loans accounted for 31.6 per cent (Table 10.4).

NOTES

Table 10.4 Purpose-wise Outstanding Advances by RRBs

(` crore)

Purpose/End-March	2006	2007	2008
1	2	3	4
I. Agriculture (i to iii)	**21,509**	**27,452**	**33,112**
Per cent to total loans outstanding	54.2	56.6	55.4
i. Short-term loans (crop loans)	13,877	18,707	22,644
ii. Term loans (for agriculture and allied activities)	7,632	8,745	10,468
iii. Indirect Advances	—	—	—
II. Non-agriculture (iv to vii)	**18,204**	**21,041**	**26,639**
Per cent to total loans outstanding	45.8	43.4	44.6
i. Rural Artisans, etc.	748	736	671
ii. Other Industries	757	880	1,276
iii. Retail Trade, etc.	3,452	3,677	5,016
iv. Other purposes	13,246	15,748	19,676
Total (I + II)	**39,712**	**48,493**	**59,751**
Memo item:			
a) Priority Sector	32,177	39,852	49,650
b) Non-priority Sector	7,535	8,641	10,101
c) Share of Priority Sector (per cent to total)	81.0	82.2	83.1

Source: Reserve Bank of India, *Report on Trend and Progress of Banking in India, 2007-08.*

Financial Performance of RRBs

The financial performance of RRBs in 2007-08 is presented in Table 10.5. Of 90 RRBs, 82 RRBs were making profits. The ratio of operating profits to total assets was 1.74 per cent in 2007-08. The intermediation cost (operating expenses to total assets) was 2.4 per cent, wage bill 1.79 per cent and the spread was 3.29 per cent. In 2003-04, loss-making units witnessed a decline in interest expenses as well as operating expenses.

Table 10.5 *Financial Performance of Regional Rural Banks (2007-08)*

(` crore)

Particulars	2007-08(P)		
	Loss-making (8)	Profit-making (82)	Total RRBs (90)
1	2	3	4
A. Income (i + ii)	**316**	**8,879**	**9,195**
(i) Interest income	286	8,106	8,392
(ii) Other income	30	773	803
B. Expenditure (i + ii + iii)	**365**	**7,401**	**7,766**
(i) Interest expended	191	3,873	4,064
(ii) Provisions and contingencies	50	675	725
(iii) Operating expenses of which:	130	2,847	2,977
Wage Bill	113	2,102	2,215

NOTES

C. Profit			
(i) Operating Profit/Loss	–5	2,159	2,154
(ii) Net Profit/Loss	–55	1,429	1,374
D. Total Assets	**4,440**	**1,19,101**	**1,23,541**
E. Financial Ratios			
(i) Operating Profit	–0.11	1.81	1.74
(ii) Net Profit	–1.24	1.2	1.11
(iii) Income	7.12	7.46	7.44
(a) Interest Income	6.44	6.81	6.79
(b) Other Income	0.68	0.65	0.65
(iv) Expenditure	8.22	6.21	6.29
(a) Interest expended	4.3	3.25	3.29
(b) Operating expenses of which:	2.93	2.39	2.41
Wage Bill	2.55	1.76	1.79
(v) Provisions and contingencies	1.13	0.57	0.59
(vi) Gross NPAs			5.88
(vii) Net NPAs			3.02

Source: Reserve Bank of India, *P (Provisional) Report on Trend and Progress of Banking in India, 2007-08.*

Regulatory Control

A number of measures were taken since 1995 not only to make RRBs viable but also to enable them to function effectively. Apart from re-capitalization and infusion of equity, the measures include deregulation of interest rates on advances and deposits of above-one-year maturity, rationalization of branches and relaxation of norms relating to investments by RRBs. In 1998-99, NABARD introduced several policy measures for improving its overall performance. These were:

- Quarterly/half-yearly review of RRBs, especially weak ones by the sponsor banks
- Merger of RRBs coming under a sponsor bank and operating in contiguous areas
- Off-site surveillance
- Framing of Appointment and Promotion Rules (1998) for the staff of RRBs
- Introduction of Kisan credit cards for provision of credit to farmers
- Encouraging RRBs to adopt self-help groups for channelling credit to the poor on a sustainable basis.

Rural Credit Delivery System

While the purpose or objectives for which RRBs have been set up have been achieved in money terms, they are total failures in real terms. Credit is helpful only if a viable activity is identified and financed. Two ideas have influenced lending by RRBs: grant of credit at cheap or concessional rates and lending to individuals belonging to weaker sections without checking the viability of the activity proposed to be undertaken. The emphasis or approach to help weaker sections through the banking system can only be based on projects that generate cash flows. The cash flows need not be large for generating profits but should meet interest and repayments to banks and a going wage to the participants. The project should meet costs including wages, interest and repayment of loans. It is only when capital assets are acquired that depreciation provision should be made. At a later stage, equity can be built-up if the activity proves profitable by ploughing profits back.

Micro-enterprise Finance Institutions

In this connection, the three major institutions covered by Mr Y. S. P. Thorat of the Rural Planning and Credit Dept. RBI, in his monograph 'A comparative study of Micro-enterprise Finance Institutions', have been widely discussed to identify fresh initiatives in India. The three major institutions covered are the Bank Rakyat Indonesia (BRI), Indonesia, the Grameen Bank, Bangladesh and the BancoSol, Bolivia.

The Bank Rakyat Indonesia is a bank with a specific mandate to serve rural areas. BRI worked through small outlets, set up in the 1970s, as conduits for subsidized farm credit called 'Bimas'. Along with credit, technical support by the government was provided for the adoption of improved rice varieties and farming practices. The repayments were poor and BRI incurred losses. The benefits were appropriated by the elite. In 1984, the Indonesian system was revamped and character-based loans for productive enterprises for 3-24 months with monthly repayments were offered. BRI serves two million borrowers with outstanding loans at $931 million. BRI treats individual units as profit centres that must maintain their viability.

The Grameen Bank was set up in 1983 by the Bangladesh government. Its distinctive feature is its clientele, 94 per cent of whom are women. There are two million borrowers with outstanding loans of $180 million. Mr Thorat found that the Grameen Bank, in addition to the work of poverty alleviation through credit, pursues a social development agenda which includes sanitation, nutrition, family planning and education. Grameen Bank also promotes informal village schools, distribution of seeds and seedlings and special projects such as fisheries and textile production.

Banco Solidario, S. A. (BancoSol) is a commercial bank set up in 1992 at Lapaz, Bolivia, exclusively for micro-enterprise lending. It offers group loans which promise commercial returns with the objective of promoting socio-economic development of small-scale enterprises and the self-employed. The focus is on the total number of loans, arrear rate and the number of new loans.

These institutions have managed to keep arrears under 5 per cent and long-term losses below 4 per cent; and provisioning has been adopted by building it into the cost structure.

Banks, rural banks and cooperatives have extended microfinance under Self-help Groups: Bank Linkage Programme. Due to low cost, and near-zero NPAs associated with microfinance, financial support was forthcoming from commercial banks (58 per cent of total finance). On cumulative basis the number of self-help groups receiving support from commercial banks increased to ` 40,85,000 and cumulative loans to ` 22,196 crore.

Capital of RRBs

Returning to regional rural banks, we find that their overdue advances constituted 25.6 per cent of total advances. On a share capital of ` 155.64 crore, their advances amounted to ` 7,505 crore or capital assets ratio of 2 per cent. The stipulation that RRBs should adopt income recognition and asset classification from 1995-96 and provisioning norms from 1996-97 onwards are not feasible. Actually, their present position has been propped up by the support of ` 360 crore in 1994-95, ` 223.57 crore in 1995-96, ` 200 crore in 1996-97 and provision' of ` 270 crore in the central budget for 1997-98.

The total support for restructuring and re-capitalization was ` 783.57 crore in the three years 1994-95 to 1996-97; and the provision of ` 270 crore in 1997-98 would take it to ` 1053.57 crore.

NOTES

Despite the emphasis on financing weaker sections by public sector and RRBs since 1995, the access of weaker sections (whose assets were up to ` 1,000) was only 23.3 per cent in respect to borrowing from institutional agencies. For those with assets between ` 1,000 and ` 5,000, it was 26.9 per cent as per the 1981-82 All-India Debt and Investment Survey of the RBI. A more recent study pertaining to 1996 shows that for the poor, formal sources provided only 7 per cent of the credit while the remaining 93 per cent came from informal sources. The low rate of credit access is attributable to the dispersal of the focus of loans as they cater to proposals sponsored by different agencies for different categories with targets and sub-targets such as SC/ST, IRDP, DWCRA, SLARS, special component plan and tribal development plan. Rigidity of procedures and the time needed for banking and credit operations has driven the poor to informal sources.

RRBs and Poverty Alleviation

There is an urgent need to approach the problem of poverty. In practical terms, a large portion of the weaker sections consisting of small and marginal farmers, SC/STs who are beneficiaries of DRI and IRDP schemes fall below the poverty line.

The objective of our plans and the ultimate objective of economic reforms is to promote equitable economic growth and reduce poverty. This has also been the objective of planning in India for the past 60 years and more, encompassing nine five-year plans and several annual plans. There has been economic growth but it has not benefited the poor. The five-year plans and annual plans implemented through the budget make provisions to accelerate the growth process and to benefit the poor, adopt programmes for rural development, generate employment and social goods such as primary education, health and family welfare, and let the measures to work themselves out. The growth process on which billions of rupees have been spent created opportunities for the skilled and those who have property rights. The poor who did not possess the skills could not avail themselves of the opportunities created by the growth process.

The poverty problem in India has to be viewed primarily as one of organizing production to involve the poor directly through a structure of investment to generate income/output in their hands rather than as one of providing opportunities in terms of employment, hoping that the poor would make use of them. The vast number or rural poor have to be brought into the economic framework and into the exchange sector. The western economic process governed by the market would work only if skills exist and are more or less widely and evenly distributed.

Apart from helping the poor through a structure of investment that generates income rather than merely providing job opportunities which may or may not match their unskilled nature, there is another aspect to be considered while making plans for rural development. This relates to provisions for employment promotion in the central budget. For instance, the budget for 1997-98 makes a provision of ` 9,096 crore. There are a lot of other measures enumerated in the budget speech of the finance minister which are not part of the budget. They actually smack of the pre-deregulation era when financial institutions, at the behest of the government, were involved in unviable lending that affected their profitability.

Let us look at the budget provisions for the rural poor. These provisions have always found a place in the budget, but their full benefit has never reached the poor. A former prime minister said on record that only 20 per cent of such provisions reached the beneficiaries. Further, the emphasis should be on output generation.

First, the concept of employment where people work eight hours a day and 40 hours a week within the four walls of a factory is not relevant in India. We just do not have the funds to do that for 60 million people (assuming a worker dependent ratio of 1:5).

Second, it involves movement of people which demands transport and infrastructure along with provision of schools, sanitation and medical care. The nature of activity has to be 'Griha Udyog' type of projects (such as weaving and metal work) within the village.

Finally, the poor, in view of their nutritional deficiency, have neither the stamina nor the enthusiasm for eight-hour work days, five or six days a week. The provisions for the benefit of poor only add to demand without adding to the supply of goods. Even in cases where they are engaged in asset creation, such assets do not endure. Roads and bunds are washed away or become unusable in the next season. The implementation of schemes through state governments and their agencies was never cost-effective. The provisions become a source for personal enrichment and patronage. The government has to integrate the poor below the poverty line into the mainstream of national life irrespective of religion, caste and gender. Divisive policies that have perpetrated through the budget have to be abandoned in the national interest.

As a part of economic reforms, economic integration of the poor into market-oriented activity, with a view to provide an opportunity for all those willing and able to work at the minimum/going wage defined in terms of calories, should be undertaken. We can integrate the public distribution of grains with it. Technology mission at district levels to identify activities which can be organized on a project basis should be constituted. People should be encouraged to come together to organize, execute and run productive activities identified by a technology mission.

Such production activities should have a parallel marketing network to convey information on market requirements as well as lift the production of the project, either directly or indirectly through village dry grocer cooperatives, etc. Infrastructure activities such as biogas, compost, drinking water, construction of recreational facilities, schools which double as a panchayat office, communication facilities, markets for selling village produce and sheds for cooperative dry grocers who may exchange village project production with daily requirements of the project workers more or less on a barter basis organized at the village level.

Apart from infrastructure, on a project basis, technology missions can organize any traditional skills or repetitive activities such as textile printing or assembly of, say, electronics. The project approach not only imparts accountability but eliminates middlemen who have been appropriating 80 per cent of the funds provided by government budgets. In three years' time, all the districts, starting in the first year with the most poverty-stricken 100 districts, can be covered.

The project requirements and marketing network can be manned by educated young men who can be trained to 'serve and share' instead of 'command and patronize' which has become the bane of the Indian society. Further, a countrywide launching of such a programme to enrich the human resources would help fight divisive politics fostered by subsidies and two-rupee schemes. We can conceive of enlisting graduates in commerce, management economics and engineering to serve as a techno-economic team at each of the 14,790 RRB branches. A graduate in arts to man the informal education centre and a graduate in medical science to man the primary health centre can join the team. As an incentive, theymay be provided a stipendof `6,000 per month and preference inemployment after serving six months at the RRB branch level in the village.

NOTES

The expenditure on a techno-economic team of four and a teacher and a doctor would amount to ` 36,000 per month and ` 4,32,000 per annum. It would be desirable to insist on a project report by the fresh graduates to be submitted to the college/university from which they graduated to verify their contribution especially since they would be given preference in employment. The whole scheme would be voluntary and each batch would serve six months. The annual expenditure on a six-member support team for RRB branches around the country would be about ` 630 crore.

Actually, the saving in expenditure by eliminating different agencies for different categories would be adequate to meet the expenditure on graduate volunteers for techno-economic, educational and primary health activities. The first objective of the Ninth Plan (1997-2002) also refers to prioritizing agriculture and rural development to generate adequate productive employment and eradicate poverty. The approach of bank lending during the Ninth Plan needs to shift to financing the poor below the poverty line. Their number is higher and includes weaker sections. Financing BPL families is easier as the definition is clear (` 'X' per person per month). Implementation is simpler and less time-consuming. Further, financing is cost-effective and reduces unproductive expenditure by reducing the number of agencies sponsoring loan proposals.

A self-sustaining economy with a democratic set-up can be built only if India achieves equitable economic growth. The implementation of economic reforms itself is in danger because the benefits of reforms, as was the case under the plans, are not going to the poor. Rather than categorizing them on the basis of caste and religion, the poor have to be looked at as one constituency to channel the benefits of economic growth. What is at stake is not merely the fate of economic reforms but the survival of a stable democratic political system.

Other aspects of economic reform that are likely to step up the overall rate of growth, would naturally improve the economic environment for the poor.

There is an immediate need to render the network of RRBs dynamic by infusing them with funds and supporting them with techno-economic teams, teachers and doctors. RRBs should be allowed to alleviate poverty of those below the poverty line. They should become catalysts for the growth process instead of just sanctioning loans for unviable projects in the name of helping weaker sections.

Check Your Progress

4. How does the Banking Regulation Act of 1949 define banking?
5. What are the two funds that were set up under the RBI Act?
6. Why was the Rural Infrastructure Development Fund constituted?
7. List some of the policy measures introduced by NABARD for improving its overall performance.

10.4 PROBLEMS OF MICROFINANCE

The concept of micro-finance implies informal and flexible approach to the credit needs of the poor. Micro-finance is the provision of thrift, credit and other financial services and products in small amounts to the poor. It helps to generate income in their hands and improve their living standards. Micro-finance has grown at a rapid pace in the past four decades across the world, benefiting millions of poor. The back- drop for micro-finance in India is the international effort in progress since the 1980s.

The Grameen Bank Model in Bangladesh

In 1976, the Grameen Bank (GB) was launched in a village in Bangladesh to assist poor families by providing credit to help them overcome poverty. Similar efforts were made in Indonesia and Latin America. GB was transformed into a bank in 1983.

GB was owned by poor borrowers who were mostly women. It was doorstep banking. The loans did not require any collateral and the system was based on mutual

trust, accountability, participation and creativity. GB offers credit for income-generating activities as opposed to consumption. The borrower has to be part of a group of five which ensures that the borrower behaves in a responsible way but is not responsible for repayment. Loans can be received in a continuous sequence. Since 2000, GB increased emphasis on deposit mobilization.

GB shifted to individual lending after 25 years of operation and its products can be used for everyday money management as well as micro-enterprises.

At the end of March 2008, GB had 7.46 million borrowers, 94 per cent of whom were women. With 2,504 branches GB provides services in 81,574 villages or 97 per cent of villages in Bangladesh.

RBI, NABARD and SIDBI

Recognizing the potential of micro-finance to reduce poverty, Reserve Bank of India, NABARD and Small Industries Development Bank of India (SIDBI) have promoted the micro-finance movement in India. The RBI advised banks in February 2000 to mainstream micro-credit and to extend the outreach of micro-credit providers. Micro-credit extended either directly or through any intermediary is reckoned as part of bank's priority sector lending. Banks are free to prescribe their own lending norms.

10.4.1 Micro-Finance Development Fund

A Micro-finance Development Fund in NABARD with a start-up capital of ` 100 crore from RBI, NABARD, banks and others was announced through Union Budget 2000-01 to provide start-up funds to micro-finance institutions and infrastructure support for training and systems management and data support.

The Union Budget for 2005-06 announced that the Government of India intends to promote micro-finance institutions (MFIs) in a big way. For this purpose, the Micro-finance Development Fund (MFDF) was redesignated as Micro-finance Development and Equity Fund (MFDEF) and the corpus of the fund was increased from ` 100 crore to ` 200 crore. MFDEF is expected to play a vital role in capitalizing the MFIs and thereby improving their access to commercial loans.

The Central government is considering the need to identify and classify the MFIs and rate such institutions to empower them to intermediate between the lending banks and the clients. To facilitate the process of rating of MFIs, NABARD has decided to extend financial assistance to commercial banks and RRBs by way of grant to enable them to avail the services of credit rating agencies for rating of MFIs.

Linkage Models of Micro-credit

Three distinct models of micro-credit are currently being followed under the self-help groups and bank linkage programme called SBLP.

Model I: SHGs promoted, guided and financed by banks

Model II: SHGs promoted by NGOs /government agencies and financed by banks

Model III: SHGs promoted by NGOs and financed by banks using NGOs/formal agencies as financial intermediaries

Model II has emerged as the most popular model under, the SBLP programme. Commercial banks, cooperative banks and the regional rural banks are the active participants.

NOTES

Micro-finance Institution (MFI) Approach

The MFI model in India is characterized by a diversity of institutions and legal firms like trusts, societies, cooperatives, non-banking financial companies—MFIs or NBFCs. MFIs (about 800) are scattered across the country and registered with one of several of registering authorities, under different statutes.

Banks can use MFIs as their agent for handling credit, monitoring, supervision and recovery. Another variation of this model is where the MFI, an NBC holds the individual loans on its books for a while before securitizing them and selling them to the banks, such refinancing through securitization enables the MFIs greater funding access.

The Union Budget for 2008-09 announced that banks would be encouraged to embrace the concept of total financial inclusion. The government advised all scheduled commercial banks to meet the entire credit requirements of SHG members, namely: (*a*) income generation activities, (*b*) social needs like housing, education, marriage and (*c*) debt swapping.

10.4.2 Role of NABARD

NABARD conducts skill development programmes, promotes micro-enterprises, provides marketing support, initiates project to link post offices with SHGs and SHG federations and supports MFIs with grants and loans. NABARD introduces a scheme for commercial banks for rating MFIs.

SIDBI: SIDBI launched its micro-finance programme in 1994 on a pilot basis by providing funds to NGOs to lend. It was the first to recognize the NGO/MF route as an effective delivery channel. It pioneered the concept of capacity assessment rating (CAR) for the MF. SIDBI introduced in 2003 transformation loan to enable MFIs to transform themselves from an informal set-up to more formal entities by setting up SIDBI Growth Fund for MFIs.

Regulation: There is no single regulator for the micro-finance providers. They fall under the Reserve Bank and state governments. Legislation proposed by the Government of India through Micro-financial Sector (Development and Regulation) Bill 2007 is under consideration of the Parliament. The Bill proposes NABARD to be the regulator and micro-finance institutions be registered with it.

Micro-insurance: Micro-insurance schemes cover health care, life's accident expenses, maternity protection and disability. Insurance companies in the public and private sector have tied up with MFIs to offer micro-insurance schemes. The poor, however, have little comprehension of pooling of risk and are not enthusiastic to part with funds with no immediate returns. The Rangarajan committee on micro-finance (2008) emphasized the need to integrate savings, credit and insurance into a holistic framework. Life insurance, health insurance, crop insurance and asset insurance were examined by the committee.

10.4.3 Progress of Micro-Finance: SBLP

Table 10.6 presents the cumulative progress in SBLP programme in select years.

Table 10.6 *Cumulative Progress in SHG-Bank Linkage Programme in Select Years*

(` crore)

Year	*No. of SHGs linked (Cumulative)*	*Bank loan (Cumulative)*	*Refinance/Assis-tance (Cumulative)*
1992-93	225	0.29	0.27
1997-98	14,317	23.76	21.39
2002-03	7,17,360	2,048.67	1,412.79
2007-08	40,85,000	22,196.00	7,062.00

Source: RBI, *Trend and Progress of Banking in India*, 2007-08, p. 211.

The SHG-Bank Linkage Programme made considerable progress since its inception. In March 2008, 40,85,000 SHGs were linked to banks and the cumulative loan assistance was ` 22,196 crore. Commercial banks account for the largest share followed by regional rural banks and cooperative banks. If the total loans disbursed 86.9 per cent of loans were disbursed exclusively for women.

Of the three models under SBLP, Model II viz., SHGs promoted by NGOs/ government agencies and financed by banks has emerged as the dominant model accounting for 80.7 per cent of bank loans of ` 398 crore in 2006-07; Model II, SHGs promoted, guided and financed by banks to 14.4 per cent; and Model III SHGs promoted by NGOs and financed by banks using NGOs/formal agencies to 4.9 per cent loan recovery.

The amount of outstanding loans at the end of March 2007 were ` 12,366 crore. About 37 per cent of banks reported recovery of above 95 per cent under the programme, 36 per cent of banks recovery was in the range of 80-94 per cent and another 20 per cent in the range of 50-79 per cent. Recovery rates of public sector banks were lower than others.

MFIs

Commercial banks were the major source of funds for MFIs. The total loans outstanding were ` 1,584 crore to 550 MFIs on 31 March 2007. The RBI survey found that recovery rate for MFIs was good.

Impact of Micro-finance

SHGs have made a positive impact on the income and employment situation. However, loans were not large enough to take up income-generating activities. As a result, loans were used for consumption purposes or subsistence income-generating activities. SHGs have a weak record of account keeping. However, approximate figures of total savings and total loans outstanding to banks could be provided for information about profits is not known. SHGs do not have a clear policy on how to deal with defaults or dropouts which form 10 per cent of membership.

10.4.4 Outreach and Scale of MFIs

The limited outreach and scale of Indian MFIs, relative to the MFI giants in Indonesia and Bangladesh, reflects, at least in part, the absence of an enabling policy, legal and

NOTES

NOTES

regulatory framework. MFIs suffer from the fact that their regulatory oversight is fragmented across many government agencies. MFIs are not allowed to mobilize deposits (even from their own members) unless they convert themselves into a non-bank finance company (NBFC). And even as NBFCs, an 'investment grade' rating from corporate rating agencies is required for mobilizing deposits. This is difficult for most MFI-NBFCs; based on past examples, on account of the typically geographically concentrated and non-collateralized portfolios that MFIs have, rating agencies, in almost all cases, have not assigned the required credit rating. The minimum start-up capital requirement for registering as an NBFC (` 20 million or US$450,000) is typically beyond the reach of most MFIs. Similarly, the minimum capital requirements for insurance companies (` 1 billion, or US$23 million) are high. MFIs have problems raising equity: NGOs are not allowed to invest in MFI equity, because of the charitable status of NGOs under the Section 11 and 12 of the Income Tax Act. And regulation on equity investment in MFIs dictates that foreign equity must be a minimum of US $500,000, and cannot exceed 51 per cent of total equity; this implies that bringing in US $500,000 foreign equity requires raising an equal amount (almost ` 23 million) from India—an amount that is considered far too high by most Indian MFIs. What's more, since 2002, MFIs are no longer allowed to raise debt from foreign donors and development finance institutions through the 'External Commercial Borrowing' (ECB) route.

Second, the cost of funds for Indian MFIs is relatively high, and unlike in Bangladesh and a number of other countries, the Indian MFI sector has not benefited from grants/subsidized funding. Unlike in, say, Bangladesh, where PKSF lends to MFIs at 4-6 per cent p.a. (less than half the market interest rate), Indian MFIs, right from inception, tend raise debt (from SIDBI, FWWB or commercial banks) at market rates (between 11-13.5 per cent p.a.). While, in many ways, this is a more sustainable way to grow, in practice, the high cost of funds combined with problems in accessing equity, has meant that achieving profitability and growth has been more difficult for Indian MFIs than their counterparts in countries like Bangladesh.

Third, the Indian MFI sector suffers from capacity and skills constraints, and inadequate support systems. As microfinance is a specialized activity and given that many MFIs have evolved from NGOs that have otherwise been focusing on grant based activities, staff tend to have stronger inclination towards social development issues and tend to possess limited skills in finance, accounting and business management. Thus, sensitization to issues like internal controls, importance of credit discipline amongst groups/ members, MIS, financial control and management, financial analysis, business planning, systems development and new product design tend to be of relatively low quality. MFIs need considerable technical assistance to scale up skills in these aspects.

Fourth, most MFIs in India lend to SHGs. This means that MFIs in India are constrained by many of the same factors that have held back the outreach and scale of SHG Bank Linkage. In particular, capacity, time and cost issues related to group formation have posed constraints.

10.4.5 Self-Help Groups

Swarnajayanti Gram Swarojgar Yojana (SGSY) has been launched with the objective of bringing every assisted family above the poverty line within three years, through the provision of micro enterprise.

In view of the above, it was felt that there is a need for a critical examination of the strategies adopted, the interventions sought, funds flow and its utilization, organizational

structure and the mechanism of implementation by the implementing agencies in sampled states to understand the impact, failures and success. Voluntary Operation in Community and Environment (VOICE) has been entrusted to undertake a detailed study of the funds released by the Central and State Governments and utilization of the same by line departments so as not only to assess the extent to which it has been possible to achieve the aims and objectives of the SHG's beneficiaries, but also to review the scheme itself and suggest policy measures to improve the situation.

Self Help Groups (SHGs): Developmental Paradigm

Villages are faced with problems related to poverty illiteracy, lack of skills health care etc. These are problems that cannot be tackled individually but can be better solved through group efforts. Today these groups known as Self-Help Groups have become the vehicle of change for the poor and marginalized.

Self-help group is a method of organizing the poor people and the marginalized to come together to solve their individual problem. The SHG method is used by the government, NGOs and others worldwide. The poor collect their savings and save it in banks. In return they receive easy access to loans with a small rate of interest to start their micro unit enterprise.

A life of dignity is the right of every citizen. Poverty is an obstruction to a dignified life. Self-employment is a significant step to have sustained incomes and remove the shackles of poverty. Programme for self-employment of the poor has been an important component of the antipoverty programmes implemented through government initiatives in the rural areas of India. Government have introduced an effective Self-Employment programme 'Swarnajayanti Gram Swarozgar', or SGSY. This new programme has been launched from April 1999. This is a holistic programme covering all aspects of self-employment such as organization of the poor in to self-help groups, training, credit, technology, infrastructure and marketing. SGSY will be funded by the centre and the states in the ratio of 75:25. With the coming into force of SGSY, the earlier programmes IRDP, TRYSEM, DWCRA, SITRA, GKY and MWS are no longer in operation.

Socio-economic Profile of Swarozgaris

The strata of rural families, handled through self-help groups, are capital scarce, labour surplus, and, by and large, bereft of proper knowledge as well as management skill. At the same time, it also remains a fact that barring certain location limitations, rural areas do throw ample opportunities for installation of micro enterprises by making use of untapped manpower resources, available raw material of various forms, and existing market channels. Given the financial support, together with appropriate knowledge and skill input, the poor people, in general, have the propensity to make better use of labour and capital. Thus, installation of income-generation activities and micro enterprises in the rural areas, in a way, helps promote first-generation micro entrepreneurs with resource mobilization on their own through their SHGs.

About 59 per cent of the sample women SGSY beneficiaries are observed to be women, which is considerably higher that the targeted 40 per cent. Over 69 per cent members belonged to SC, ST groups, about 21 per cent to OBC group and only about 8 per cent to the forward group. The minority community has just registered its presence by about 2 per cent membership. The above distinctions appear to be the product of the SHGs policy of SGSY programme. In this case, the SHG members or Swarozgaris must come from the list of below poverty line (BPL) families and the minimum percentage of SC, ST Swarozgaris should be 50 per cent.

NOTES

A great majority of the members were found to be educated and at least literate in both the groups of SHGs. Over 65 per cent members were capable of reading and writing.

Rural people are engaged in various occupations for earning their livelihood. Information about the occupation of the SHG members under study was collected.

So far as the marital status is concerned, the respondents were found to be either unmarried or married or widowed. A great majority of the respondents in the case of both the groups of SHGs were found to be married. In the case of unmarried category 104 out of 1350 women members in the SGSY supported SHGs were unmarried.

Poverty alleviation bias is well reflected in the data reported, but not without aberration. About 82 per cent members were found to be landless who are most likely to conform to the criterion of being below poverty line (BPL). But 17 per cent of the SHG members who are small landholders may or may not conform to this criterion. Even if it is accepted that these landholders belong to the BPL group, then how it is that no marginal landholder category could be included as members of SHGs.

Impact of SHGs on Swarozgaris

Self- employment is a significant step to have sustained incomes and remove the shackles of poverty. During the study, adequate attention was given to cover various aspects of the programme and understand the impact of the support on the slated objectives. Various economic activities have been pursued by the group in the sampled states.

Activity wise analysis of incremental employment and income suggests that swarozgaris involved in activities like dairy and other farm based activities which require lesser amount of forward and backward linkages have been more successful.

There is no doubt that the SGSY has generated substantial increase in the incomes of swarozgaris in the sample states. Maximum increase was found in Andhra Pradesh, followed by Gujarat, Chhattisgarh, Uttar Pradesh and Bihar in that order. The growth of incremental income in Chhattisgarh and some districts of Bihar was due to relatively low level of pre-project (SGSY) income.

The assistance under SGSY was found to have enabled the swarojgaris enhance their earnings. However, these incremental levels of income are largely due to the low level of earning prior to interventions under SGSY. As such a small increase in income has resulted in a significant rise in the group income.

10.5 NON-BANK FINANCIAL INTERMEDIARIES (NBFIs)

Check Your Progress

8. What is micro-finance?
9. Where was the Grameen Bank launched and why?
10. Which bank was the major source of funds for MFIs?
11. What are self-help groups?

Non-bank financial institutions (NBFIs) consist of all-India financial institutions (AIFIs), non-banking financial companies, primary dealers and non-bank finance companies. Non-bank financial institutions are being increasingly recognized as complementary to the banking system capable of absorbing shocks and spreading risks. Apart from complementing banks in financial intermediation, NBFCs provide depth and resilience to the financial system. Financial intermediaries provide core financial services such as payments and liquidity, maturity transformation, store of value, information processing and pooling of risks. While commercial banks, which have traditionally provided these services, are diversifying into insurance and securities business to enhance their earnings, NBFIs by broadening access to financial services enhance competition and diversification

of the financial sector by offering equity and risk-based products. After, initiation of financial liberalization, some of the foreign banks have established several subsidiaries in the form of either non-bank finance companies or limited companies in the non-financial sector in India that undertake diverse business such as dealing in securities, leasing and finance or information technology. NBFIs constitute a highly complex sector of the financial system owing to their diversified nature. Regulatory safeguards have to be devised for the particular business in addition to the general guidelines.

Nature and Size

Non-banking finance companies which are heterogeneous in nature in terms of activity, organizational structure, portfolio mix, are important financial intermediaries and an integral part of the Indian financial system. The main advantages of NBFCs lie in the lower transaction costs, quick decision-making, customer orientation and prompt provision of services. They have been able to carve out a niche for themselves in meeting the credit needs of unorganized sector and small borrowers at the local level. Their number has gone up from 7,063 in 1981 to 51,929 in 1996. Of these, only 20 per cent used to submit returns to the RBI. In 1996, the regulated deposits of NBFCs amounted to ` 38,110 crore. Aggregate public deposits of 1,547 NBFCs holding public deposits, amounted to ` 20,428.93 crore as on 31 March 1999; and as a proportion to deposits with commercial banks they constituted 2.9 per cent. Non-banking deposits as a proportion of household savings in gross financial assets rose from 3.1 per cent in 1980-81 to 7.4 per cent in 1998-99. NBFCs attracted a large number of small investors since the rate of return on deposits with them was relatively high. NBFCs are quite flexible in meeting the credit needs of specific sectors like equipment leasing, hire-purchase, housing finance and consumer finance, where gaps between the demand and supply of funds have been high and where established financial entities are not easily accessible to borrowers. The growth in the number of NBFCs was facilitated by the ease of entry, limited fixed assets and absence of any need to hold inventories. While their functions and the services they render are different, the common feature is acceptance of deposits from the public, borrowing from banks and in the case of companies organized as public limited companies, accessing the capital market.

The number of NBFCs registered with the RBI was 12,809. NBFCs consist of NBFC–D (deposit taking NBFCs), RNBCs, mutual benefit companies, miscellaneous non-banking companies (MNBCs) and Nidhi companies. The deposit taking NBFCD are 364 and RNBC two (former IFCI and TFCI).

10.5.1 Definition of Non-banking Finance Company

According to the Reserve Bank (Amendment Act), 1997, 'A Non-banking Finance Company' (NBFC) means:

(i) A financial institution which is a company

(ii) A non-banking institution which is a company and which has, as its principal business, the receiving of deposits under any scheme or arrangement or in any other manner or lending in any manner

(iii) Such other non-banking institution or class of such institutions as the bank may with the previous approval of the Central government specify

The definition excludes financial institutions besides institutions which carry on agricultural operations as their principal business.

Categories of Non-finance Companies

Non-banking finance companies consist mainly of finance companies which carry on hire-purchase finance, housing finance, investment, loan, equipment leasing or mutual benefit financial companies but do not include insurance or stock exchanges or stockbroking companies.

The non-bank finance companies are categorized into:

- An equipment leasing company (EL)
- A hire-purchase company (HP)
- A housing finance company (HFC)
- An investment company (IC)
- A loan company (LC)
- A mutual benefit financial company (MBFC), *i.e.*, Nidhi companies
- Amiscellaneous non-banking company, *i.e.*, chit fund companies.

RNBCs: Residuary Non-banking Company (RNBC) is a company which receives any deposit under any scheme or arrangement, by whatever name called, in one lump-sum or in installments or in any other manner and which does not fall into any of the above categories.

Finally, a non-banking, non-financial company is defined as an industrial concern or a company whose principal activity is agricultural operations or trading in goods and services or real estate and which is not classified as a financial, miscellaneous or a residuary non-banking company.

10.5.2 Mutual Benefit Finance Companies (MBFCs)

Mutual benefit finance companies (Nidhis) were exempt from most of the provisions of the Reserve Bank's NBFC's directions. However, the RBI imposed on, 8 July 1996, a ceiling of 15 per cent interest rate on deposits and prohibited them from issuing advertisements in any form and paying any brokerage for soliciting deposits. NBFCs' deposit interest rates were freed on 24 August 1996 along with the rationalization measures for registered NBFCs. The ceiling, however, does not apply unless MBFCs have positive net-owned funds (NOFs) as on 31 March 1996, are able to repay the amount of their liabilities including the interest payable to their depositors and have a ratio of NOF to deposits not exceeding 1 : 20 as on the date of application.

Again, on 15 January 1997, the prescribed ratio of NOF to deposits not exceeding 1:20 was made applicable only on the incremental deposit liabilities after 15 January 1997. However, MBFCs with NOF to deposit ratio of 1:20 or less on 15 January 1997, should not exceed the prescribed ratio of 1:20 on the aggregate deposit liabilities.

Regulation of Non-banking Companies

Four categories of non-bank finance companies (EL, HP, IC and LC) submit statutory annual schedules and returns to the Reserve Bank. The Reserve Bank has issued a separate set of directions of financial, miscellaneous and residuary non-banking companies governing their deposit acceptance activity. The activity of deposit acceptance by non-banking, non-financial companies (manufacturing companies) is being regulated by the

Government of India, under (Acceptance of Deposits) Rules, 1975, framed under Section 58A of the Companies Act.

The regulation of the deposit acceptance activities was undertaken initially to effectively supervise, control and regulate them. In order to moderate the deposit mobilization of NBFCs and protect depositors, the quantum of deposits was linked to Net Owned Fund (NOF), which is the aggregate of the paid-up capital and free reserves reduced by balance of loss, deferred revenue expenditure and other intangible assets. Investments in same group or other NBFCs, beyond 10 per cent of owned funds were also excluded from NOF. The regulations did not extend to the assets side of NBFCs.

Financial Sector Reform of NBFCs

In view of the important role of NBFCs in the financial system, the need for subjecting them to financial reform was felt. The Committee on the Financial System recognized that NBFCs should have prudential norms and guidelines as in the case of commercial banks. For this purpose, a Working Group on Financial Companies was constituted in May 1992 under the Chairmanship of Dr A. C. Shah. The major recommendations of the Working Group are:

- Category-wise classification of financial companies may be abolished and uniform regulations applied to all companies.
- Regulatory attention by the RBI may be confined to companies with net owned funds of ` 50 lakh and above.
- As regards new financial companies' entry norms, minimum net owned funds of ` 50 lakh and a cooling period before accepting deposits have been suggested.
- The regulations should be on the asset side, such as limit on credit concentration.
- Capital adequacy standards may be laid down based on risk assessment of assets and credit conversion factors for off-balance sheet items.
- The exempted category of deposits should be removed and all deposits should be brought under the regulatory framework. A clear distinction should be made between deposits and borrowing from banks/institutions.
- Non-banking financial companies may be allowed to accept deposits for periods ranging from 12 months to 84 months.
- Prudential norms for income recognition, transparency of accounts and provisions for bad and doubtful debts may be prescribed.

The bank implemented some of the recommendations with modifications and introduced a number of changes from 12 April 1993.

In the first phase, the minimum period of deposits of NBFCs has been brought down from 'over 24 months' to '12 months' and maximum period to 84 months in the case of residuary non-banking companies. The definition of regulated deposits was widened to include intercorporate deposits, borrowing through issue of debentures and monies received from directors or shareholders of private limited companies which were earlier under the exempt category.

Maintenance of liquid assets: Hire-purchase finance and equipment leasing constitute 10 per cent of deposits out of which 5 per cent in government securities; and residuary non-banking companies constitute 10 per cent of deposits in government securities within the limit of 70 per cent investment in approved securities.

NOTES

Registration

Companies with net owned funds of ` 50 lakh and above have been advised to register with the RBI.

Prudential Guidelines

Detailed prudential guidelines to NBFCs were issued in June 1994. The registered NBFCs were required to achieve a minimum capital adequacy norm of 6 per cent by 31 March 1995 and 8 per cent by 31 March 1996 and also obtain a minimum credit rating. CRAR requirement ranged between 12-15 per cent depending on the principal line of business activities of an NBFC (2002).

Liberalization Measures for NBFCs (1996)

The RBI announced liberalization measures for NBFCs on 24 July 1996. They are aimed at encouraging disciplined NBFCs which are run on sound business principles. The two major classifications of NBFCs are:

I. Equipment leasing and hire-purchase companies (financial companies)

II. Loan and investment companies: Loan and investment companies were further divided into four major categories by the RBI circular. They are:

(i) Registered finance companies complying with credit rating requirements and prudential norms.

(ii) Registered finance companies complying with either credit rating requirements or prudential norms.

(iii) Registered finance companies complying with neither the credit rating requirements nor prudential norms.

(iv) All other finance companies.

Category (i) finance companies will now enjoy the following benefits:

(a) No ceiling on deposits

(b) Freedom to determine the rate of interest on deposits

(c) SLR reduced from 15 per cent to 12.5 per cent.

For category (ii) companies, only the maximum ceiling on deposits will be removed and other benefits available to category (i) will not be available.

Category (iii) companies will not enjoy the above benefits and will have to comply with the following requirements:

(a) The overall ceiling on deposits reduced with immediate effect from 10 times the net owned funds (NOF) to seven.

(b) Rate of interest and SLR requirements remain unchanged.

Category (iv) companies have to reduce their deposits with immediate effect from 10 times of NOF to five. The benefits listed under categories (i) and (ii) will be available only on obtaining a certificate of compliance from the RBI. There is no ceiling on the maximum amount of deposits which can be accepted by the finance companies and the rate of interest offered on them. The rate of interest may vary from company to company depending on their resource mix and advances portfolio.

Category (i) finance companies may have to face severe competition from (iii) and (iv) category companies in their efforts to mobilize deposits since these companies

are likely to offer higher incentives through the broker network on their deposits. As a result, the interest rate will be fixed based on marketability of fixed deposit schemes by each of these companies.

There is a greater risk of interest mismatch from this source of funding since the interest rate in a free market is related to the competitiveness of the company. However, in case of bank funding, even though such a risk is present, the maturity mismatch is almost ruled out. Even though cash credit limits are sanctioned for a period of one year, the limits are renewed by the banks every year except in rare cases. Again, bank funding is cheaper compared to the cost of sourcing fixed deposits. Hence, finance companies which rely more on bank funding stand to gain in the process.

10.5.3 Regulations for NBFCs Accepting Public Deposits

The term 'public deposit' has been defined on the lines of the definition provided in the Companies (Acceptance of Deposits) Rules, 1975 framed under the Companies Act, 1956 and the Reserve Bank of India Act. Public deposits will include fixed, recurring etc. deposit received from relatives and friends, shareholders by a public limited company and the money raised by issue of unsecured debentures/bonds. Public deposits would, however, exclude money raised by NBFCs by way of issue of secured debentures/ bonds, borrowings from banks/financial institutions, deposits from directors, intercorporate deposit (ICD), deposit received from foreign citizens and those received by the private limited companies from their shareholders.

Prohibitions from accepting public deposits for NBFCs having net owned fund (NOF) of less than ₹25 lakh: The NBFCs having net owned fund of less than ` 25 lakh will not be entitled to accept deposits from the public. However, they can raise borrowings from other resources.

Credit rating requirement and quantum of public deposit: The ceiling on quantum of public deposit for an NBFC has been linked to its level of credit rating, given by approved credit rating agencies. Higher credit rating will entitle NBFCs to raise large amount of public deposit as per details given here under:

Table 10.7 *Credit Rating Level*

Levels of Credit Rating	*Equipment Leasing Hire-purchase Companies*	*Loan and Investment Companies Multiple of NOF*
A	3 times	2 times
AA	2 times	1 time
AAA	1 time	0.5 time

Interest rate ceiling: All the NBFCs have now been subjected to the interest rate ceiling of 16 per cent per annum. NBFCs which are presently offering interest rates in excess of the prescribed ceiling are required to reduce their interest rate to bring it within the ceiling with immediate effect.

Brokerage: Brokerage payable by NBFCs on deposit of one year to five years has now been uniformly fixed at 2 per cent as against the varying rates earlier. In addition, they may also pay to the brokers by way of reimbursement on the basis of vouchers/bills produced, an amount not exceeding 0.5 per cent of the deposits collected.

NOTES

Submission of returns: NBFCs accepting public deposits only will be required to submit to RBI annual statutory returns and financial statements. NBFCs not accepting public deposits are exempted from this requirement.

Classification of NBFCs

(i) The Reserve Bank has also made certain refinements in the norms for classification of NBFCs into various sub-groups based on their principal activity as evidenced from the asset/income pattern. With regard to the special regulatory dispensation accorded to equipment leasing and hire-purchase, the criteria for classification of the NBFCs has been tightened. Thus, an NBFC to be eligible for being classified as equipment leasing company or a hire-purchase finance company shall have not less than 60 per cent of its assets and shall derive income from these activities taken together.

(ii) All new NBFCs incorporated after 9 January 1997 will be provisionally classified for a period of one year and reviewed thereafter on the basis of their asset/ income pattern as disclosed in their balance sheet/profit and loss account and other related aspects. Existing NBFCs, as well as those unclassified will be classified on the basis of their principal activity as evidenced from their financial statements into various categories such as Equipment Leasing Companies, Hire-purchase Finance Companies, Loan Companies, Investment Companies, Miscellaneous Non-banking Companies or Residuary Non-banking Companies as the case may be. Only such of the NBFCs as have been specifically notified under Section 62A of the Companies Act, 1956 by the Government of India, will be classified as Nidhi Companies. NBFCs which have been incorporated with the intent to function as Nidhis will be classified as Loan Companies and the directions as applicable to 'Loan Companies' will be made applicable to them till such notification.

Depositors cautioned: The Reserve Bank has cautioned the depositors on repayment of deposit, in spite of the rigorous regulation.

Depositors should, therefore, at their own risk and responsibility, be circumspect and satisfy themselves about the financial soundness and health of the companies before placing their deposits.

RBI modified regulations for NBFCs (31 January 1998): The detailed regulations issued on 2 January 1998 to ensure orderly functioning of the NBFCs and for protection of the interest of the depositors, were modified partly to remove problems in their implementation.

Check Your Progress

12. Apart from complementing banks in financial intermediation what do the non-bank financial institutions (NBFIs) provide?
13. What are the main advantages of NBFCs?
14. What are the two major classifications of NBFCs?

10.6 INSURANCE INSTITUTIONS

The insurance industry is an integral part of the financial services' industry and profound changes have been taking place in this sector in India. This is because a well-developed and properly evolved insurance sector is needed for economic development. It not only provides long-term funds needed for infrastructural development but also strengthens the risk taking ability of individuals and institutions.

The Concept of Insurance

Insurance is a form of risk management primarily used to hedge against a risk of a contingent loss. It is defined as the transfer of the risk of a potential loss, from one entity

to another, in exchange for a premium. Insurance applies to situations where a loss may or may not occur. It cannot apply to situations where the loss is expected to happen. While an insurer is a company that sells the insurance services, an entity seeking to transfer the risk (an individual, corporation or association of any type and the like) becomes an 'insured' party. Insurance policy is a contract between an insured and an insurer where the latter assumes a risk. Generally, an insurance contract includes the following elements:

- Parties (insurer, insured and the beneficiaries)
- Premium (the amount charged for insurance coverage)
- Period of coverage (the duration for which a particular loss is covered)
- Amount of coverage (the amount to be paid to the insured or beneficiary in the event of a loss)
- Exclusions (the events not covered)

Thus, an insured person is indemnified against the event of loss covered in the policy. If the insured parties incur a loss on account of a specified peril, the coverage gives a right to the policyholder to 'claim' against the insurer for the covered amount of loss as specified in the policy. The fee paid by the insured to the insurer for assuming the risk is called 'premium'. Insurance premiums from the many insured are used for the later payment of the claims. An insurer maintains adequate funds for anticipated losses and a certain margin towards his profit. In a nutshell, insurance provides for an indemnification against the loss or liability from specified events and circumstances that may occur or be discovered during a specified period (Statement of Financial Accounting Standards No. 113).

Principles of Insurance

Insurance is based on the principle of probability and cooperation. This can be explained as:

- **Principle of probability**

The degree of loss depends upon various factors. All the affecting factors are analysed and the probability of loss is calculated. Probability helps to find the chance of occurrence of loss. The law of large numbers is applied in the principle of probability. It is assumed that the past events will occur in the same pattern. The premium is fixed on the basis of past events, present conditions and future prospects.

- **Principle of co-operation**

In insurance, the loss is shared by a group of persons who are willing to co-operate. The insurer collects the shares of the insured members in advance and accumulates a fund. A member is paid from this fund at the time of occurrence of the insured risk. At present the insured pay the premium to join the insurance schemes. Thus, the insured cooperates to share a loss of an individual by advance payments of premium amount.

Features of Insurable Risks

The features of insurable risks are as follows:

- **Large number of homogeneous exposure units**

Insurance is based on the operation of the law of large numbers. There must be sufficient number of risks of a similar nature, being insured, so that the probability of loss could be

NOTES

estimated. A majority of the insurance policies are provided for individual members of a very large group. However, large commercial property policies may insure exceptional properties, for which, there are no 'homogeneous' exposure units.

- **Loss is definite**

Pure risk has an element of loss or a break-even but not gains. At least in principle, the event which is insured and results in a loss, should take place at a known time, in a known place and from a known cause. The death of an insured on a life insurance policy can be cited as an example. Fire, automobile accidents and workers' injuries also meet this criterion.

- **Loss is accidental**

The event that leads to a claim should be unexpected or should at least be outside the control of the insured. The loss should be pure and there is only an opportunity cost. Events that contain speculative events like ordinary business risks are not considered for insurance.

- **Premium should be affordable**

The premium is related to the amount of protection offered. If the estimated loss due to the insured event is high, naturally the premium fixed would be large. However, the premium should not be so large that there is no reasonable chance of a significant loss to occur for the insured.

- **Loss is calculable**

The probability of the occurrence of the loss event and the attendant cost related to the event should be calculable. Probability of loss is generally empirical and can be calculated.

- **Risk exposure is limited**

The risk is often aggregated. The ability of the insurer to issue policies becomes constrained, if the same event can cause losses to many policyholders of the same insurer. The sum of all policyholders so exposed is the constraint. Normally, the insurers prefer to limit their exposures to a loss from a single event to some small portion of their capital base.

Advantages of Insurance

Some of the advantages of insurance are as follows:

- It involves proper planning and administration to reduce the loss due to uncertainties.
- It ensures certainty of payment at an uncertain event of loss and thereby provides protection.
- All the persons who are exposed to the risk share the loss.
- The insurer employs the funds in productive channels.
- It improves the efficiency of the insured person because it liberates him from the worry of the loss.

Categories of Insurances

Three broad categories of insurances are offered. These are given below.

(a) **Life insurance:** It deals with various plans connected with the life of a person.

(b) **General insurance:** All kinds of insurance policies that are not related to life are known as general insurance.

(c) **Reinsurance:** It is insurer's insurance.

NOTES

Life insurance in its existing form came in India from UK in 1818 with Oriental Life Insurance Company. The Indian Life Assurance Companies Act, 1912 was the first measure to regulate life insurance business. Later, in 1928 the Indian Insurance Companies Act was enacted, which was amended in 1938. Finally, this Act was amended by the Government of India in 1950. So long as insurance remained the monopoly of the government, the need for an independent regulatory authority was not felt that strongly. However, with the acceptance of the entry of private insurance entities, the need for a regulatory authority became essential. The regulatory framework in relation to the insurance companies seeks to take care of three major concerns:

(a) Protection of consumers' interest

(b) To ensure the financial soundness of the insurance industry

(c) To help the healthy growth of the insurance market

Originally, the insurance industry was controlled by the Insurance Act 1938, the Life Insurance Corporation Act 1956 and General Insurance Business (Nationalization) Act, 1972. The Insurance Regulatory and Development Authority Act, was enacted in 1999. The IRDAAct of 1999 had paved the way for the entry of private players into the insurance market, which had hitherto been the exclusive privilege of public sector insurance companies/corporations.

10.6.1 Insurance Regulatory and Development Authority (IRDA)

As per the provisions of the IRDAAct, Insurance Regulatory and Development Authority (IRDA) was established on 19 April 2000, to protect the interests of insurance policyholders and to regulate, promote and ensure the orderly growth of the insurance industry. The authority has its headquarter at Hyderabad. Insurance Regulatory and Development Authority (Protection of Policyholders' Interests) Regulations was passed in the year 2002.

The authority has notified 27 regulations on various issues which include registration of insurers, regulation on insurance agents, solvency margin, reinsurance, obligation of insurers to rural and social sector, investment and accounting procedure, protection of policy holders' interest etc.

Registration of Private Players

- The company should be formed and registered under the Companies Act, 1956.
- The aggregate holdings of equity shares by a foreign company, either by itself or through its subsidiary companies or its nominees, should not exceed 26 per cent paid up equity capital of the Indian insurance company.
- The company's sole purpose is to carry on life insurance business, general insurance business or reinsurance business.
- The minimum paid up equity capital for life or general insurance business is ` 100 crore.
- The minimum paid up equity capital for carrying on reinsurance business has been prescribed as ` 200 crore.

NOTES

Protection of the Interests of the Policyholders

IRDA has the responsibility of protecting the interests of the insurance policyholders. IRDA has notified The Protection of Policyholders' Interests Regulations, 2001, to provide for the following:

- Policy proposal documents in easily understandable language
- Claims procedure in both life and non-life
- Setting up of grievance redressal machinery
- Speedy settlement of claims and policyholders' servicing

The regulation also provides for the payment of interest by the insurers for the delay in the settlement of claims.

- The insurers are required to maintain solvency margins so that they are in a position to meet their obligations towards the policyholders with regard to the payment of claims.
- It is obligatory on the part of the insurance companies to clearly disclose the benefits and the terms and conditions under the policy. The advertisements issued by the insurers should not mislead the insuring public.
- All the insurers are required to set up proper grievance redressal machinery in their head office and at their other offices.
- The authority takes the matter with the insurers, if any complaint is received from the policyholders in connection with the services provided by the insurers under the insurance contract.

Policyholders' Servicing

(1) An insurer carrying on life or general business, as the case may be, shall at all times, respond within 10 days of the receipt of any communication from its policyholders, in all matters, such as:

- Recording change of address
- Noting a new nomination or change of a nomination under the policy
- Noting an assignment on the policy
- Providing information on the current status of a policy, indicating matters such as, accrued bonus, surrender value and entitlement to a loan
- Processing papers and disbursal of a loan on the security of a policy
- Issuance of a duplicate policy
- Issuance of an endorsement under the policy; noting a change of interest or sum insured or the perils insured, financial interest of a bank and other interests
- Guidance on the procedure for registering a claim and the early settlement thereof

Insurance Ombudsmen

Efficient customer services in the insurance sector has been one of the major areas of concern for the government. With a view to ensure expeditious redressal of public grievances relating to the settlement of the claims, the government has introduced a system of ombudsman in insurance sector with effect from 11 November 1998. Insurance ombudsmen are currently located in twelve cities. Each ombudsman is empowered to redress customers on personal lives where the insured amount is less than ` 20 lakh.

10.6.2 Life Insurance

Life insurance different from other types of insurances. The subject matter of insurance is the life of human beings. Life insurance is a contract, which provides risk coverage to the insurer. The purchaser of insurance pays a fixed premium in exchange of a promise of compensation in the event of some specified loss. Thus, life insurance is a contract for payment of money to the person assured (or to the person entitled to receive the same) on the occurrence of the event insured against. Usually the contract provides for:

- The payment of an amount on the date of maturity or at specified periodic intervals, or at death, if it occurs earlier.
- Periodical payments of insurance premium by the insured, to the corporation which provides the insurance.

Any person above 18 years of age is eligible to enter into a valid contract. Subject to certain conditions, a policy can be taken on the life of a spouse or children.

Payment of Premiums

The policyholders can pay premiums other than single premiums, to the insurer in yearly, half-yearly, quarterly or monthly instalments or through a salary savings scheme. If the mode of payment of premium is yearly or half-yearly, some insurers give a rebate of 3 per cent and 1.5 per cent respectively on the premium. If the mode of payment is monthly, some insurers charge an additional 5 per cent for the facility offered. This additional charge is waived for the salary saving scheme, which provides for the payment of premium through monthly deductions by the employers from the salary of their employees.

Bonus

The insurer distributes profits among the policyholders every year in the form of a bonus/ profit share. An insurance policy can be 'with' or 'without' profit. In a 'with' profit plan, any bonus declared is allotted to the policy and is paid at the time of maturity or death along with the contracted amount. In a 'without' profit plan, the contracted amount is paid without any profit share. The premium rate charged for a 'with' profit policy is therefore, higher than that of a 'without' profit policy.

Additions

These may be guaranteed additions or loyalty additions.

- Guaranteed addition means that the insurer guarantees the bonus/profit declared as a certain amount per thousand rupees of the sum assured of the policy. Sum assured is the amount that an insurer agrees to pay an insured or his nominees on occurrence of contingencies, e.g., death or on maturity. This assured bonus is credited to the policyholder irrespective of the performance of the insurance company and is known as guaranteed addition. Guaranteed additions are payable at the end of the term of the policy or on the early death of the policyholders.
- Loyalty addition is the additional amount paid over and above the guaranteed addition. The insurer declares and credits to the policyholder, an additional amount per thousand of the sum assured every five years, depending on its performance.

Benefits

Insurance provides several benefits. These are:

- ***Survival benefits*:** This means that a part of the sum assured is paid to the policyholder at fixed intervals of time before the actual maturity date. The risk

NOTES

cover for life continues for the full sum assured even after the payment of the survival benefits and bonus is calculated on the full amount of the sum assured. If the policyholder survives till the end of the term, the survival benefits would be deducted from the maturity value.

- ***Accident benefits:*** These are given on payment of an additional premium of Re 1 per ` 1000 of sum assured per year and the insured is entitled to the following benefits:

  In the case of accidental death, the nominee shall receive double the sum assured. In the case of total and permanent disability due to an accident, risk coverage continues without further payment of premium. In addition, an amount equal to the sum assured is paid to the insured in monthly instalments spread over a period of 10 years. However, subsequent accidental death will not entitle the nominee for double the sum assured.

- **Disability benefits:** These are provided, if the insured becomes totally and permanently disabled due to any accident. He need not pay future premiums and his policy shall remain in force for the full sum assured.

Life Insurance Policies

Some of the common life insurance policies have been discussed here:

(i) Endowment policy

An endowment policy covers the risk for a period specified by the insurer. At the end of the specified period, the sum assured is paid back to the policyholder, along with the bonus accumulated during the term of the policy. In an endowment policy, capital is accumulated for a specific purpose and it is a protection against the saver's premature death. Premium for an endowment life policy is much higher than that of a whole life policy.

Many investors use endowment life insurance to fund anticipated financial needs, such as college education of their sons and daughters or their own retirement. Unlike whole life, an endowment life insurance policy is designed primarily to provide for a living benefit during the life time of the individual and only secondarily to provide for life insurance protection. Therefore, it is more of an investment than a whole life policy.

(ii) Whole life policy

A typical whole life policy remains as long as the policyholder is alive. The risk is covered for the entire life of the policyholder. Hence, it is known as a whole life policy. The whole life policy amount and bonus are payable only to the nominees upon the death of the policyholder. The policyholder is not entitled to receive any money during his or her own lifetime, i.e., there is no survival benefit.

(iii) Term life policy

It is a policy for a chosen period. The risk is covered only for that period. A term plan meets the needs of people who are initially unable to pay a larger premium required for a whole life or an endowment assurance policy, but would be able to pay for a policy in the near future. Surrender, loan or paid-up values are not granted under this policy because reserves are not accumulated. If the premium is not paid within the grace period, a policy will lapse without acquiring any paid-up value. The policyholder may survive the term but, the risk cover comes to an end.

However, a lapsed policy may be revived during the lifetime of the life insured but before the expiry of the period of two years from the due date of the first unpaid premium, on the usual terms. Accident and/or disability benefits are not granted on policies under the term plan.

Money Back Policy

This is basically an endowment policy for which a part of the sum assured is paid to the policyholder in the form of survival benefits, at fixed intervals before the maturity date. The risk cover on life continues for the full sum assured even after the payment of survival benefits and bonus is calculated on the full sum assured. If the policyholder survives till the end of the policy term, the survival benefits would be deducted from the maturity value.

An important feature of this type of policy is that in the event of death at any time within the policy term, the death claim comprises the full sum assured, without deduction of any of the survival benefit amounts, which might have been paid already as money-back components. Similarly, bonus is also calculated on the full sum assured.

Joint life policy

These are similar to endowment policies. They too offer maturity benefits to the policyholders, apart form covering the risks just like all the other life insurance policies. However, joint life policies are categorized separately as they cover two lives simultaneously. They offer a unique advantage for a married couple or for the partners in a business firm.

Children's Insurance Policy

Children's insurance policies include those through which parents or legal guardians provide for life insurance for their child from birth. The risk cover commences when the child attains the age of 12, 17, 18 or 21, as per the policy document.

Group Policy

Life insurance protection under the group policies is provided to various groups such as employers-employees, professionals, cooperatives, weaker sections of the society, etc. It also provides insurance coverage at the lowest possible premium cost for people in certain approved occupations. Besides providing insurance coverage, it also offers group schemes to employers who allow the funding of gratuity and pension liabilities of the employers.

Matters to be Stated in a Life Insurance Policy

1. According to Insurance Regulatory and Development Authority (IRDA) Act, the following have to be stated in a life insurance policy:
    - Name of the plan governing the policy, its terms and conditions
    - Whether it is participating in the profits or not
    - Basis of participation in profits such as cash bonus, deferred bonus, simple or compound reversionary bonus
    - Benefits payable, the contingencies upon which these are payable and the other terms and conditions of the insurance contract
    - Details of the riders attached to the main policy

NOTES

- Date of commencement of risk and the date of maturity or date/dates on which the benefits are payable
- Premiums payable, the periodicity of payment, the grace period allowed for payment of the premium, the date of the last instalment of the premium, the implications of discontinuing the payment of an instalment/instalments of the premium and also the provisions of a guaranteed surrender value.
- Age at entry and whether the same has been admitted
- The policy requirements for conversion of the policy into a (i) paid up policy, (ii) surrender (iii) non-forfeiture and (iv) revival of lapsed policies
- Contingencies excluded from the scope of the cover, both in respect of the main policy and the riders
- The provisions for nomination, assignment and loans on security of the policy and a statement that the rate of interest payable on such loan amount shall be as prescribed by the insurer at the time of taking the loan
- Any special clauses or conditions, such as, first pregnancy clause, suicide clause etc.
- The address of the insurer to which all communication in respect of the policy shall be sent
- The documents that are normally required to be submitted by a claimant in support of a claim under the policy

2. While acting under Regulation 6(1) in forwarding the policy to the insured, the insurer shall inform by a letter forwarding the policy, that he has a period of fifteen days from the date of receipt of the policy document to review the terms and conditions of the policy. If the insured disagrees to any of these terms or conditions, he has the option to return the policy stating the reasons for his objection. He shall then be entitled to a refund of the premium paid. This is subject to a deduction of a proportionate risk premium for the period on cover and the expenses incurred by the insurer on medical examination of the proposer and stamp duty charges.
3. In respect of a unit linked policy, in addition to the deductions under sub-regulation (2) of this Regulation, the insurer shall also be entitled to repurchase the unit at the price of the units on the date of cancellation.
4. In respect of a cover, where the premium charged is dependent on age, the insurer shall ensure that the age is admitted as far as possible before the issuance of the policy document. In case where age has not been admitted by the time the policy is issued, the insurer shall make efforts to obtain proof of age and admit the same as soon as possible

Claims Procedure of a Life Insurance Policy

The claims procedure of a life insurance policy is as follows:

- A life insurance policy shall state the primary documents which are normally required to be submitted by a claimant in support of a claim.
- Alife insurance company, upon receiving a claim, shall process the claim without delay. Any queries or requirements of additional documents, shall be raised all at once and not in a piece-meal manner, within a period of fifteen days of the receipt of the claim.

- A claim under a life policy shall be paid or be disputed, giving all the relevant reasons within thirty days from the date of receipt of all relevant papers and the clarifications required. However, where the circumstances of a claim warrant an investigation in the opinion of the insurance company, it shall initiate and complete such an investigation at the earliest and in any case, not later than six months from the time of lodging the claim.
- Subject to the provisions of Section 47 of the Act, where a claim is ready for payment but the payment cannot be made due to any reason such as that of a proper identification of the payee, the life insurer shall hold the amount for the benefit of the payee and such an amount shall earn interest at the rate applicable to a savings bank account with a scheduled bank (effective from thirty days after the submission of all papers and information).
- Where there is a delay on the part of the insurer in processing a claim for a reason other than the one covered by sub-regulation (4), the life insurance company shall pay interest on the claim amount at a rate which is 2 per cent above the bank rate prevalent at the beginning of the financial year in which the claim is reviewed by it.

Life Insurance Providers

Life insurance was the monopoly of the Life Insurance Corporation (LIC) for a long time. An Act of Parliament, viz., Life Insurance Corporation Act, formed the Life Insurance Corporation of India in September 1956, with capital contribution from the Government of India. Since nationalization, LIC has built up a vast network of 2048 branches, 100 divisions and 7 zonal offices spread over the country. The Life Insurance Corporation of India also transacts business abroad and has offices in Fiji, Mauritius and United Kingdom. LIC is associated with joint ventures abroad in the field of insurance, namely, Ken-India Assurance Company Limited, Nairobi; United Oriental Assurance Company Limited, Kuala Lumpur and Life Insurance Corporation (International) E. C. Bahrain.

The government had opened up the insurance sector for private participation in 1999 and had also allowed the private companies to have foreign equity up to the level of 25 per cent. Following the opening up of the insurance sector, twelve private sector companies have entered into the life insurance sector.

Public Sector Providers

1. Life Insurance Corporation of India

   www.licindia.com

Private Sector Providers

1. Allianz Bajaj Life Insurance Company Limited

   www.allianzbajaj.co.in
2. Birla Sun-Life Insurance Company Limited

   www.birlasunlife.com
3. HDFC Standard Life Insurance Company Limited

   www.hdfcinsurance.com

4. ICICI Prudential Life Insurance Company Limited
   www.iciciprulife.com
5. ING Vysya Life Insurance Company Limited
   www.ingvysayalife.com
6. Max New York Life Insurance Company Limited
   www.maxnewyorklife.com
7. MetLife Insurance Company Limited
   www.metlife.com
8. Om Kotak Mahindra Life Insurance Company Limited
   www.omkotakmahnidra.com
9. SBI Life Insurance Company Limited
   www.sbilife.co.in
10. TATA AIG Life Insurance Company Limited
    www.tata-aig.com
11. AMP Sanmar Assurance Company Limited
    www.ampsanmar.com
12. Dabur CGU Life Insurance Company Private Limited
    www.avivaindia.com

Factors to be Considered

The need for insurance differs from family to family. The insurance policy and the amount taken often depend on the following factors:

- One's own income level
- Tax planning
- The number of dependents
- The wealth, income and expense levels of the dependents
- Their significant foreseeable expenses
- The inheritance and the lifestyle to be provided for them

10.6.3 General Insurance

General insurance covers the loss due to unforeseen events such as accidents, illness, fire and burglary. Unlike life insurance, general insurance is not meant to offer any return but is a protection against contingencies. General insurance policy may be termed as a contract of indemnity as the insurer normally makes good the actual amount of the loss suffered.

General insurance includes property insurance, liability insurance and various other forms of insurance. Fire and marine insurances are strictly called property insurances. Liability insurance includes loss due to vehicle theft, fidelity and machine loss insurance. The strict form of liability insurance is fidelity insurance, where the insurer compensates the loss to the insured when he is under the liability of a payment to a third party.

There is no certainty in the loss of the asset, which is insured against. Hence, the premium is decided by the value of the asset and the probability of such a loss. Under certain Acts of Parliament, some types of insurances like motor insurance and public liability insurance have been made compulsory.

General Insurance Products

The broad categories of general insurance are:

- Fire Insurance
- Marine Insurance
- Motor Insurance
- Health Insurance
- Miscellaneous Insurance

Fire Insurance

It is a comprehensive policy. It covers the risk of loss due to fire. Besides covering loss on account of fire, it also covers loss on account of earthquakes, riots, strikes, malicious acts and floods. Any movable and immovable property having a monetary value is covered under a fire insurance policy. This is a material damage policy and it is also called a standard fire policy. Fire in a factory may result in a total or partial stoppage of production leading to a financial loss. Fire insurance policy covers such a loss. It is a blanket policy which covers risks related to both fixed as well as current assets. However, these policies can be differentiated on the basis of the agreed sum.

Valued policy

The insurance company pays the agreed value of the property.

Average policy

Fire claims are paid to the insured as a proportion of the actual value of the property, at the time of loss.

Specific policy

Risk on account of fire is insured for a specific sum. The maximum coverage under this policy shall be up to the total amount of the insurance policy.

Floating policy

The risks related to one or two different kinds of goods are covered for a single sum and for a single premium in a floating policy.

Excess policy

The risk coverage is to the extent of the maximum additional amount by which the stock may sometimes increase. Such additional risk coverage is provided for in an excess policy.

Marine Insurance

This has developed over a period of many centuries. Marine insurance covers the risk arising from and incidental to marine operations related to cargo, hull, freight and the like. The Marine Insurance Act, 1963 is the basis for the transactions in marine insurance in India. Marine insurance policies are broadly classified into:

- Marine hull insurance
- Marine cargo insurance

NOTES

Marine hull insurance is related to the insurance of hulls. Marine cargo insurance is related to the insurance of goods in transit from one place to another, by any single mode or combined modes of transit through sea, rail, road, air and inland waterways.

A marine cargo insurance policy is an important document in international trade and is provided as a collateral security to the banks.

Motor Insurance

It is mandatory for all the vehicles in India to have a third party insurance. There are two types of motor insurance: third party insurance, which only insures the party/parties other than the owner in an incident and comprehensive insurance, which insures the owner as well as the third party involved, in addition to loss or damage caused to the vehicle due to accident, fire, explosion, self-ignition, lighting, burglary, house breaking, theft, riots and strikes, earthquakes, flood, hurricanes, storms, cyclones, malicious acts and the like.

The premium for motor vehicles is decided on the value of the vehicle and on the location where it has been registered. The premium for a heavy commercial vehicle is decided on the value of the vehicle and its gross laden weight.

Health Insurance

The ever-increasing cost of medical treatment today is beyond the reach of the common man. In the case of a medical emergency, the cost of hospital, room rent, the doctor's fees, medicines and related health services could work out to an enormous sum. In such a situation, health insurance provides the much needed financial assistance for medical treatment.

Health insurance policy is a legal, and binding contract between an insurance company and a customer. The major difference between health insurance and life insurance is that in the latter, a person may purchase guaranteed renewable insurance for the whole life of the insured at a constant premium rate, whereas the former is generally purchased year by year with no assurance of renewability. Generally, if a policy is renewable, there is no guarantee that premium rates will not increase.

These policies are regulated by the General Insurance Corporation (GIC) and marketed by the four big insurance companies: United India Insurance Company Ltd., New Delhi Assurance Company Ltd., Oriental Insurance Company Ltd. and National Insurance Company Ltd.

Mediclaim Policy

Mediclaim (in India for individuals and groups) and overseas mediclaim policies (abroad) are available. These policies provide for reimbursement of hospitalization, domiciliary hospitalization expenses for illness and diseases suffered or accidental injuries sustained during the policy period. The expenses that would fall under different heads are mentioned below. The policy generally covers the following expenses which should not exceed the total sum insured.

(a) Nursing expenses

(b) Fees of surgeon, anaesthetist, medical practitioner, consultants, specialists etc.

(c) Anaesthesia, blood, oxygen, operation theatre charges, surgical appliances, medicines and drugs, diagnostic materials and x-rays, dialysis, chemotherapy,

radiotherapy, cost of pace maker, artificial limbs and cost of organs and other similar expenses.

This insurance is available to persons between the ages of 5 and 80 years. Children between the ages of 3 months and 5 years can be covered provided one or both the parents are covered concurrently. Premium up to a sum of ` 10,000 qualifies for tax benefits under section 80-D of the Income Tax Act.

Group mediclaim policy

This is available to any homogeneous group of individuals/associations/institutions/ corporate bodies provided it has a central administration and is subjected to a minimum of 100 persons.

In health insurance, unhealthy people are more likely to purchase health insurance policies because they anticipate large medical bills. On the other hand, people who consider themselves to be reasonably healthy, may decide that medical insurance is an unnecessary expense for them. Hence, the financial burden would be heavy for the service providers. Therefore, insurance companies use the term 'adverse selection' to describe the tendency of only those people who will benefit from insurance to buy it.

Miscellaneous Insurance Policies

The miscellaneous insurance policies are as follows:

Agricultural insurance

Agricultural pump set insurance: In this, the property covered is pump sets whether operated by electricity, diesel or oil. The perils covered are fire, lightning, mechanical and/or electrical breakdowns.

Cattle insurance: It provides for cover against death of animals within the specified geographical area mentioned in the policy. Death may arise from contraction of any disease or occurrence of an accident after the commencement of the insurer's liability under the policy.

Crop insurance: It provides a cover against unavoidable loss of production as a result of one or more of these causes: climatic reasons such as drought, flood, frost and cyclone, pest infestation, plant diseases, as well as riots and strikes. The General Insurance Corporation has introduced crop insurance in certain states with the participation of the states as co-insurers.

Travel policy

While traveling, any tourist may lose his baggage or passport. He may also meet with an accident. Travel policies are designed to take care of all the problems that generally occur while traveling. The various travel policies have been given below:

- Videsh Yatra Mitra
- Personal accident — family
- Baggage insurance policy
- Executive travel insurance
- Suhana Safar

NOTES

Business policy

It covers the risks of loss of business goods, plant, machinery and the like. The common types of business policies are:

- Product liability insurance
- Shopkeeper's insurance policy
- Burglary insurance for business premises
- Partnership insurance
- Workmen's compensation insurance
- Professional indemnity insurance
- Fidelity guarantee insurance
- Machinery breakdown policy
- Contractor's all risk policy
- Loss of stock in cold storage insurance policy
- Air transit insurance policy

Matters to be Stated in the General Insurance Policy

1. A general insurance policy shall clearly state:
    - The name/names and address/addresses of the insured and of any bank/ banks or any other person having financial interest in the subject matter of insurance
    - Full description of the property or interest insured
    - The location or locations of the property or interest insured under the policy and wherever appropriate, with respective insured values
    - Period of insurance
    - Sums insured
    - Perils covered and not covered
    - Any franchise or deduction applicable
    - Premium payable and the basis of adjustment of the premium should be stated, where the premium is provisional subject to adjustment
    - Policy terms, conditions and warranties
    - Action to be taken by the insured upon the occurrence of a contingency which is likely to give rise to a claim under the policy
    - The obligations of the insured in relation to the subject matter of insurance upon occurrence of an event giving rise to a claim and the rights of the insurer under the circumstances
    - Any special conditions attached to the policy
    - Provision for cancellation of the policy on grounds of misrepresentation, fraud, non-disclosure of material facts or non-cooperation on the part of the insured
    - The address of the insurer to which all the communications in respect of the insurance contract should be sent
    - The details of the riders attached to the main policy
    - The pro forma of any communication that the insurer may seek from the policyholder to service the policy

2. Everyinsurer shall periodically ask the insured about the requirements to be fulfilled by the insured regarding the lodging of a claim arising in terms of the policy and the procedures to be followed by him to enable the insurer to settle a claim early.

NOTES

Claim Procedure of a General Insurance Policy

- An insured or the claimant shall give notice to the insurer of any loss arising under the contract of insurance at the earliest or within such extended time as may be allowed by the insurer. On receipt of such communication, a general insurer shall respond immediately and give a clear indication to the insured on the procedures that he should follow. In cases where a surveyor has to be appointed for assessing a loss or a claim, it shall be done within seventy-two hours of the receipt of the intimation from the insured.
- Where the insured is unable to furnish all the particulars required by the surveyor or where the surveyor does not get the full cooperation of the insured, the insurer or the surveyor as the case may be, shall inform in writing to the insured about the delay that may result in the assessment of the claim. The surveyor shall be subjected to the code of conduct laid down by the authority while assessing the loss and shall communicate his findings to the insurer within thirty days of his appointment with a copy of the report being furnished to the insured, if he so desires. In case of some special circumstances, either due to its special and complicated nature, the surveyor shall under intimation to the insured, seek an extension from the insurer for submission of his report. However, in no case shall a surveyor take more than six months from the date of his appointment to submit his report.
- If an insurer, on the receipt of a survey report, finds that it is incomplete in any respect, he shall require the surveyor under intimation to the insured, to furnish an additional report on certain specific issues as may be required by the insurer. Such a request may be made by the insurer within fifteen days of the receipt of the original survey report. However, the facility of requesting for an additional report by the insurer shall not be resorted to, for more than once in the case of a particular claim.
- The surveyor on receipt of this communication shall furnish an additional report within three weeks of the date of receipt of the communication from the insurer.
- On receipt of the survey report or the additional survey report, as the case may be, an insurer shall offer a settlement of the claim to the insured within a period of thirty days. If the insurer decides to reject a claim under the policy for any reasons, it has to be recorded in writing and communicated to the insured. This shall be done within a period of thirty days from date of the receipt of the survey report or the additional survey report, as the case may be.
- Upon acceptance of an offer of settlement as stated in sub-regulation (5) by the insured, the payment of the amount due shall be made within seven days from the date of acceptance of the offer by the insured. In cases of delay in the payment, the insurer shall be liable to pay interest at a rate which is 2 per cent above that of the bank rate prevalent at the beginning of the financial year in which the claim is reviewed by it.

General Insurance Providers

The Mercantile Insurance Ltd. was the first company set up to transact all types of general insurance business in 1707. The first general insurance company, Triton Insurance

NOTES

Company Ltd. was established in Calcutta in 1850. General Insurance Council, a wing of the Insurance Association of India, framed a code of conduct for ensuring fair and sound business practices in 1957. The general insurance business in the country was nationalized with effect from 1 January 1973 by the General Insurance Business (Nationalization) Act, 1972.

More than 100 non-life insurance companies including branches of foreign companies operating from within the country, were amalgamated and grouped into four companies namely, The National Insurance Company Ltd. The New India Assurance Company Ltd., the Oriental Insurance Company Ltd. and the United India Insurance Company Ltd., with head offices at Calcutta, Bombay, New Delhi and Madras respectively.

General Insurance Corporation which was the holding company of the four public sector general insurance companies was delinked from the latter and approved as the 'Indian reinsurer' from 3 November 2000. The share capital of the GIC and that of the four companies are held by the Government of India. All the five entities are government companies registered under the Companies Act.

Public Sector

1. National Insurance Company Limited
   www.nationalinsuranceindia.com
2. New India Assurance Company Limited
   www.niacl.com
3. Oriental Insurance Company Limited
   www.orientalinsurance.nic.in
4. United India Insurance Company Limited
   www.uiic.co.in

Private Sector

1. Bajaj Allianz General Insurance Company Limited
   www.bajajallianz.co.in
2. ICICI Lombard General Insurance Company Limited
   www.icicilombard.com
3. IFFCO-Tokio General Insurance Company Limited
   www.itgi.co.in
4. Reliance General Insurance Company Limited
   www.ril.com
5. Royal Sundaram Alliance Insurance Company Limited
   www.royalsun.com
6. TATAAIG General Insurance Company Limited
   www.tata-aig.com
7. Cholamandalam General Insurance Company Limited
   www.cholainsurance.com

8. Export Credit Guarantee Corporation
   www.ecgcindia.com
9. HDFC Chubb General Insurance Company Limited
   www.hdfcergo.com

NOTES

10.7 MUTUAL BENEFIT FUNDS

'Put your money in trust, not trust in money' attracts the small investors who are risk neutral or risk averse. Small investors prefer some kind of collective investment vehicle which can pool their marginal resources, invest it in securities and distribute returns among them on cooperative principles. This led to the growth of mutual fund industry both in the developed and developing capital market.

Concept of Mutual Fund

Mutual fund is a mechanism for pooling resources by issuing units in securities to the investors and investing funds in accordance with the objectives disclosed in the offer document. Amutual fund is a corporation, trust or partnership, which manages the collected money with the help of professional expertise. Different persons have defined mutual funds different ways. 'A mutual fund is almost like a cooperative society of investors. That is why the word 'mutual' is used. It collects money from investors by issuing mutual fund units, invests it in securities, and divides whatever dividend or interest is received among its members' (A. John Halin).

The SEBI Mutual Fund Regulations, 1993 defines mutual fund as 'a fund established in the form of a trust by a sponsor, to raise money by the trustees through sale of units to the public, under one or more schemes, for investing in securities in accordance with the regulations'. Mutual funds are financial intermediaries which bring a wide variety of securities within the reach of the most modest investors. The financial intermediary is known as 'investment company' in the US and most other countries. They are called 'investment trusts' in the United Kingdom. In India, they are known by the term 'mutual funds'.

Origin of Mutual Fund

The history of mutual fund can be traced back to Europe where William I established a society in Belgium for such a purpose. The foreign and colonial Government Trust of Lund in 1868 is considered to be the forerunner of the concept of mutual fund. Massachusetts Investor's trust was the first mutual fund set up in the US in the year 1929. The mutual fund industry witnessed a boon in the US market after the 1990's and became a popular source of investment. In India, the Unit Trust of India set up the first mutual fund in 1964.

10.7.1 Mutual Funds in India

Until 1987, Unit Trust of India was the sole mutual fund in the country. This was due to the restrictive policies adopted by the Government of India with regard to the financial services industry. The growth of the mutual fund industry has been divided into different phases.

Check Your Progress

15. What is a premium?
16. When and why was the Insurance Regulatory and Development Authority (IRDA) established?
17. What are the factors to be considered while taking an insurance policy?
18. How are marine insurance policies classified?

Phase I

The mutual fund in India came into existence in 1964 when Unit Trust of India was incorporated as a statutory corporation. The maiden scheme launched by the Unit Trust of India was the unit scheme of 1964, an open-ended scheme, which is still in operation. At that time, public awareness about mutual fund was limited. There was no disclosure norm. The institution was modelled along the lines of mutual funds in the UK. The name 'Unit Trust' itself has been borrowed from the UK where mutual funds are called 'investment trusts'.

The Unit Trust of India played a commendable role by launching a number of open as well as close ended schemes, keeping in view the varied needs of the different groups of investors. The schemes targeted everyone from a new-born child to a retired individual.

Phase II

Unit Trust of India's monopoly came to an end in 1987. The Government of India amended the Banking Regulations Act, permitting commercial banks in the public sector to set up mutual funds. The first non UTI mutual fund was launched by the State Bank of India in November 1987 by the name 'SBI Mutual fund'. Its first scheme, Magnum Regular Income Scheme launched in 1987, was well received by the investors.

Canara Bank established its subsidiary, Canbank Mutual Fund in Dec 1987. It launched two schemes, i.e., Canstock (income scheme) and Canshare (growth scheme), which were both close ended. They were also followed by two open-ended schemes—Cancigo and Cangilt in the succeeding year.

Indian Bank, Bank of India and Punjab National Bank introduced mutual funds during the year 1989–90. The government permitted insurance corporations in the public sector to establish mutual funds. Life Insurance Corporation of India set up LIC mutual fund in June 1989. It targeted small investors particularly from rural and semi-urban areas. Unlike the other mutual funds, LIC offered insurance protection to the investors. This was in addition to the benefits of liquidity, safety and return. Shortly the General Insurance Corporation of India also entered into the mutual fund industry.

The Government of India issued comprehensive guidelines in June 1990 covering all mutual funds. Registration of mutual funds with the SEBI was made compulsory. The guidelines covered the norms for registration, management, investment objectives, disclosure and pricing. The Securities Exchange Board of India (mutual funds) Regulations, 1993 came into effect on 20 January 1993. The establishment of Asset Management Company (AMC) and the listing of close-ended schemes became mandatory. Disclosure norms were tightened to protect the small investors.

Phase III

The innovative promotional campaigns launched by different mutual funds created investor awareness. Exclusion of the private sector was widely criticized. The liberalization policy and new economic policy advocated by Doctor Manmohan Singh paved way for the entry of private sector into the mutual fund industry. The SEBI accorded approval to a number of players in the private sector to launch mutual funds in October 1993. Kothari group in collaboration with the Pioneer fund, the oldest fund in U.S, launched Prima fund in November 1993. The other private sector mutual funds include Twentieth Century Mutual Fund, Taurus Mutual Fund, Morgan Stanley Mutual Fund, HDFC Mutual Fund

and Zurich Mutual Fund etc. After the entry of the private sector, the declaration of Net Asset Value (NAV) of the schemes became regular. At present NAV's are declared weekly. The portfolios are also disclosed periodically.

NOTES

Phase IV

After 1996, the mutual fund industry witnessed a healthy growth. This is shown in Table 10.7. With the growth of investors' interest in mutual funds, the number of players operating in the industry reached new heights. SEBI (Mutual Funds) Regulations, 1996 was introduced by SEBI to set uniform standards for all mutual funds in India and safeguard the interest of the investors. The Union Budget in 1999 exempted all dividend incomes of the mutual funds in the hands of investors from income tax. The SEBI and the Association of Mutual Funds in India (AMFI) launched various Investor Awareness Programmes to educate investors and inform them about the mutual fund industry.

Table 10.8 *Fund Mobilization* *(` in crore)*

Years	*UTI*	*Public Sector*	*Private sector*	*Total*
1998–99	11,679	1732	7966	21377
1999–2000	13,536	4039	42,173	59,748
2000–01	12,413	6192	74,352	92,957
2001–02	4643	13,613	1,46,267	1,64,523
2002–03	5505	22,923	2,20,551	2,48,979

Phase V

The Unit Trust of India Act 1963 was repealed in 2003, and Unit Trust of India was bifurcated into two separate entities. The US 64 scheme which assured return and certain other schemes were brought under the Specified Undertaking of the Unit Trust of India with ` 29,835 crore of assets under the management as on January 2003. This Specified Undertaking of Unit Trust of India does not come under the purview of the Mutual Fund Regulations, but under the rules framed by the Government of India.

The second is the UTI Mutual Fund Ltd., sponsored by SBI, PNB, BOB and LIC. It is registered with SEBI and functions under the Mutual Fund Regulations. It was in this phase that the mutual fund industry witnessed a consolidation phase. Mergers and acquisitions became common in the mutual fund industry. Some examples of these are Birla Sun Life mutual fund's acquisition of schemes ofAlliance Mutual Fund and Principal Mutual Fund's acquisition of Sun F&C Mutual Fund and PNB Mutual Fund. Many international mutual fund players like Fidelity, Franklin Templeton Mutual Fund etc. have entered India. There were twenty-nine funds in the end of March 2006. The growth phase is still continuing in spite of the temporary oscillations in the performance.

Table 10.9 *Fund Mobilization*

Source: AMFI Reports

NOTES

10.7.2 Structure of a Mutual Fund

The formation and operations of the mutual funds are governed by the Securities Exchange Board of India Mutual Funds Regulation 1993. Later, it was replaced by the SEBI Mutual Fund Regulations 1996. The mutual funds comprise four separate entities, namely, sponsor, mutual fund trustee, asset management company and the custodian. They are assisted by independent entities such as banks, registrars and transfer agents.

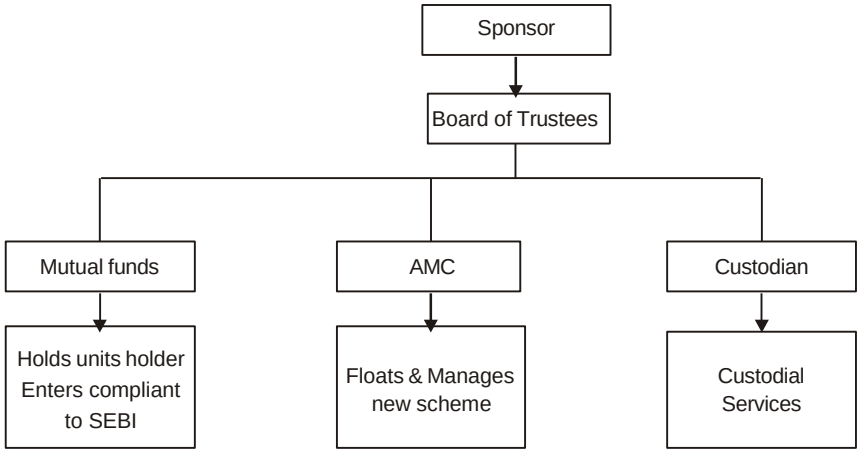


Fig. 10.1 *Structure of a Mutual Fund*

Sponsor

A mutual fund is to be established by a sponsor and registered with the SEBI. A sponsor can be any person acting alone or in combination with a corporate body. However, a sponsor should have the following requirements:

- Should have a sound track record
- Should have been carrying on business in financial services for a period of not less than five years
- Should have a positive net worth in all the preceding five years
- Should have profits after providing for depreciation, interest and tax in three out of the immediately preceding five years including the fifth year
- Should contribute at least 40 per cent of the net worth of the Asset Management Company
- Should not have been guilty of fraud or convicted for any offence

Trustee

Mutual funds are established in the form of a trust. The trustees should be persons of ability, integrity and standing. Two-third of them should be independent persons (1998). A trustee should not be an associate or a subsidiary or be associated with a sponsor in any manner. He should be appointed with the prior approval of the SEBI. The meeting of the trustees shall be held at least once in two calendar months and at least six such meetings shall be held in every year. An AMC or any of its officers/employees are not eligible to act as trustees to any mutual funds.

The trustees have the responsibility to safeguard the interest of the investors. They have wide powers to overview, supervise and monitor the activities of an asset

management company. If the conduct of the business is not in compliance with SEBI's regulations, they can take remedial measures against an AMC as required. They have powers to dismiss an AMC. Nevertheless, it should be approved by SEBI.

NOTES

Asset Management Company

A sponsor or the trustees appoint an AMC, and it should be approved by the board. An appointee can be terminated by a majority of the trustees or 75 per cent of the unit holders of the scheme. An AMC should have a sound track record, general reputation and fairness in transactions. The directors of AMC should possess adequate professional experience in finance and financial services. An AMC should have a net worth of not less than ` 10 crore. Each director of an asset management company is required to give the details of his dealings in securities with the trustees, on a quarterly basis.

An AMC manages the various schemes of mutual funds with the help of a team of professionals with adequate experience. They carry out market research for building the portfolio of a particular mutual fund. An AMC takes all reasonable steps, exercises and due diligence to ensure that investment of funds pertaining to any scheme is not contrary to SEBI's regulations and trust deed.

Custodian

Mutual funds have a custodian. Custodians carry out custodial services for the various schemes of a fund. It is their duty to send intimation of the custodial services rendered to the board. A custodian, or its directors, or partners will not be directly or indirectly associated with any AMC in any way. A custodian shall not be appointed in case a sponsor or its associates hold 50 per cent or more of the voting rights of the share capital of a custodian, or where 50 per cent or more directors of the custodian represent the interest of the sponsors or its associates.

10.7.3 Operation of the Mutual Fund

Mutual fund offers units or shares to the public by issuing an offer document or prospectus and collecting the funds. The money collected is invested as per the investment objectives stated in the offer documents/prospectus. Mutual funds generally invest in a wide range of securities in different industries with the aim of reducing the risk exposure. The investments should be within the norms prescribed by SEBI.

An expert fund manager is employed to manage each scheme. He carries out the specific task of purchasing and selling shares and debentures at the appropriate time in the market. Mutual fund managers are under the control of the board of trustees of the fund. They guide the operation of the fund.

All the mutual fund websites allow investors to download the application forms and offer documents of their products. Some of the mutual funds permit downloading other transaction forms such as redemption slips and transfer forms as well.

Mutual funds generally publish their net asset value every Friday. After the annual accounts are audited, the mutual funds ascertain the income earned by them. They distribute at least 90 per cent of their income by the way of dividends to the unit holders. After the duration of the scheme is over, mutual funds sell the securities pertaining to the scheme. It redeems the units by paying the investors their capital and also pays capital gains according to the number of units held by the investors.

Net Asset Value (NAV) of a Mutual Fund

The NAV of a mutual fund is like a company's book value. NAV is the market value of the assets of a fund scheme for every outstanding unit as on the date of mutual fund valuation. Net asset is the asset under a mutual fund as on the date of NAV computation. Net asset value is calculated on the basis of total asset value of a company, minus the administrative expenses. Repurchase price for units of a mutual fund is computed after considering the net asset value. NAV is universally used because it is a single number and is easily related to the face value of the unit.

The formula for calculating the NAV is to divide the net assets by the total number of units. In other words, it is obtained by dividing the difference between the total assets and liabilities by the total number of units.

NAV = (Total assets – liabilities) ¸ total number of units or

= Market value of all investments + other assets – liabilities ¸ total number of units

Consider a mutual fund with the following investments:

X Ltd. 500 shares of ` 10 each (current market price ` 40)

Y Ltd. 1000 shares of ` 10 each (current market price ` 110)

Other assets ` 10000

Accrued expenses ` 15000

The market value of investment

$= 500 \times 40 + 1000 \times 110 =$ ` $20000 + 110000$

Total assets = 130000 + 10000 = 140000

The fund has issued 10000 units

$$\text{NAV} = \frac{140000 - 15000}{10000} = \frac{125000}{10000} = 12.5\%$$

The NAV is received as a barometer of performance of schemes. When the NAV is lower than the face value of the unit, it signifies a poor performance and vice-versa. However, if the NAV appears simple, it may not be completely relied upon. The way in which the liabilities and the expenses of running the scheme are accounted for and apportioned, affect the NAV. The net asset value of the mutual funds is provided either in a table or as a running ticker on the home page. All the mutual funds update their NAVs on a daily basis. The historical values of NAVs are available only for a few funds. Prudential ICICI Mutual fund, Kotak Mahindra Mutual fund and Sundaram Newton fund provide historical NAVs.

Debt-equity mix

The debt equity mix also affects the NAV. In a rising interest rate situation, if the debt component is high, the NAV will be low and vice- versa. Thus, in the case of balanced fund or debt fund, an investor should look into the market interest rate as well.

10.7.4 Types of Mutual Funds

A mutual fund scheme can be classified into close ended or open ended depending on its maturity period.

NOTES

Close-Ended Scheme

It has a prefixed maturity period, e.g., five to seven years. Both the corpus amount and the number of units are prefixed. The fund is open for subscription only for a specified period after the launch of the scheme. Mutual funds are required to despatch certificates or statements of accounts within six weeks from the date of closure of the initial subscriptions of the schemes. The investors can invest in the scheme during this period. After the closure of the subscription period, investors can buy and sell the units of the scheme at the stock exchanges where the units are listed. They would either get a demat account statement or unit certificates as traded in the stock exchanges.

According to SEBI regulations, one or two exit routes should be provided to the investors. It may either be in the form of regular repurchase or by listing them in stock exchanges. Some of the close-ended mutual funds provide the option of selling back of the units to the mutual funds. The prices are fixed on the basis of net asset value. The NAV of the schemes is disclosed on a weekly basis.

The entire corpus is disinvested after the maturity period and the proceeds are distributed among the investors in proportion to their unit holdings.

Open-Ended Schemes

These are available for subscription and repurchase on a continuous basis. These schemes do not have a maturity period. Investors can buy and sell units at prices fixed by a mutual fund. Prices are fixed on the basis of NAV. The NAVs of these schemes are declared daily. Liquidity is the main advantage of the open-ended scheme. The main difference between the open-ended and the close-ended schemes is that the latter is traded on stock exchanges, whereas the former is not. Also, open ended schemes are available at all times, whereas the close-ended schemes are available only for a prescribed period.

10.7.5 Investments by Mutual Funds

SEBI has laid down rules and regulations regarding the investments of mutual funds to protect an investor. Some of these are listed below:

- The funds collected under any scheme of the mutual fund shall be invested only in transferable securities in the money, capital and debts markets.
- Money market scheme of the mutual fund shall be invested in the money market instruments in accordance with the directions issued by the RBI.
- The mutual fund shall not advance any loans for any purpose.
- Every mutual fund shall compute and carry out valuation of its investments in its portfolio and publish the same in accordance with the valuation norms specified in the eighth schedule.
- Every mutual fund shall compute the NAV of each scheme by dividing the net assets of the scheme by the number of units outstanding on the date of valuation.
- The NAV of a scheme shall be calculated and published at least in two daily newspapers at intervals of not exceeding one week.
- The price at which the units are sold and repurchased should be made available to an investor.
- A mutual fund scheme can invest up to 10 per cent of its NAV in the listed and unlisted securities or units of venture capital funds.

NOTES

Restrictions on Investment

Mutual funds should not invest more than 15 per cent of its NAV in a single debt instrument rated below the investment grade by a credit rating agency. It can be extended to 20 per cent of the NAV of a scheme with prior approval by the board of trustees and the board of AMC. The above condition is also applicable to the unrated debt instruments. No mutual fund under all its schemes should own more than 10 per cent of any company's paid up capital carrying voting rights. Transfer of investments from one scheme to another scheme in the same mutual fund shall only be allowed if:

- Such transfers are done at the prevailing market price for quoted instruments on spot basis.
- The securities so transferred should be in conformity with the investment objective of a scheme to which such transfer has been made.
- The aggregate of the inter scheme management made by all schemes under the same management or in schemes under the management of any other scheme of AMC should not exceed 5 per cent of net asset value of the fund.
- Every mutual fund shall get the securities purchased or transferred in the name of the mutual fund on account of the concerned scheme, wherever investments are intended to be of long-term nature.
- If there is a pending deployment of funds in securities in terms of investment objectives of the scheme, a mutual fund can invest the funds of the scheme in short term deposits of scheduled commercial banks.
- No mutual fund scheme shall make an investment in any unlisted security of an associate or group company of the sponsor; or any security issued by way of private placement by an associate or group company of the sponsor; or the listed securities of group companies of a sponsor which are in excess of 30 per cent of the net assets of all the schemes of a mutual fund.
- No mutual fund scheme shall invest more than 10 per cent of its NAV in the equity shares or equity related instruments of any company. However, the limit of 10 per cent shall not be applicable for investments in index funds or sector or industry specific schemes.
- A mutual fund scheme shall not invest more than 5 per cent of its NAV in equity shares or equity related instruments of any company in case of open-ended schemes and 10 per cent of its NAV in case of close-ended schemes.

10.7.6 Mutual Funds and Taxation

The government policy on taxing mutual funds changes year to year. The following details give the taxation status of dividend given to the investors.

- Pre-1999: Dividends from MFs were taxable in the hand of the investors.
- 1999: Dividends from MFs were exempted for investors for three years. Equity schemes, US 64, were exempted from distribution tax. Debt MFs were subject to dividend distribution tax of 10 per cent. Long-term capital gains tax was cut from 20 to 10 per cent.
- 2000: Dividend distribution tax on debt schemes rose from 10 to 20 per cent. Equity schemes continued to be exempted.
- 2001: Dividend distribution tax on MFs was reduced from 20 to 10 per cent.

- 2002: MFs were exempt from dividend distribution tax. Dividends were again taxable in investor's hands.
- 2003-04: Equity MFs were to pay tax at 10 per cent during this period.

NOTES

The Present Scenario

The mutual fund industry has grown enormously in a short span of time. Its swelling corpus is an example for it. The private sector funds are growing faster than the public sector funds and there is more product diversification in the former.

Investors

The majority of the investors in mutual funds are corporate and institutional sectors. Nearly 60–70 per cent of the investment comes from corporate/institutional sectors. The investors are mainly concentrated in metropolitan cities like Mumbai, Delhi, Chennai, Bangalore and Kolkata.

Wide variety of funds

A number of funds are available under the broad category of equity related and debt related funds. Often the market is flooded with new varieties of funds. In spite of these, the mutual fund industry is suffering. The problems faced by it are discussed below.

The Problems

The various problems are as follows:

Poor performance of mutual funds

Many of the mutual funds are not performing well. Therefore, many investors are disappointed with the returns that they get from the mutual funds. The reasons for poor performance are given below:

- Excessive diversification of portfolio, i.e., buying more number of equities results in the addition of poor performers in the portfolio. The management of a widely diversified portfolio is also difficult.
- Frequent turnover of portfolios maylead to huge payment of brokerage/commission.
- Improper investment planning results in poor performance. If the entry and exit timing of an investment is not properly calculated, it results in a huge loss. Wrong selection of stocks also leads to less return.
- Purchase of blue chip securities during the bull run results in a heavy loss during the bear period. The scope for further risk in price is also limited.
- A steady decline in the interest rates has affected the bond funds. The scope for trading gains is limited by the falling interest regime.

More concentration on the metro markets

Mutual funds prefer metro markets because of the presence of a distribution mechanism and availability of an informed base in the cities. Even though metro markets have huge savings potential, people of the metros are more likely to change their investments from one avenue to another. To maintain a sustainable rate of growth, mutual funds must concentrate more on the urban and semi urban areas as they have a high savings potential.

Lack of informed selling

To capture distribution, mutual funds have up-scaled the commissions. The funds rely on distributor's feedback to change or make additions to products. The focus is mainly on

NOTES

short-term performance rather than the long-term one. The sellers should understand the products fully and it should be 'informed selling'.

Investors' lack of knowledge

There is a wide range of mutual fund products. The presence of these varieties confuses the investors as a majority of them are not able to understand the risk reward relationship. The complexity of the products offered by the mutual funds confuses the investors. Therefore, designing simpler products and using intelligible tools for communication become essential.

Lack of attracting retail saving

By definition and purpose, a mutual fund is about retail savings. However, at present it has got skewed towards the corporate segment. The retail investors have been traditionally investing in fixed deposits of banks. This mind-set has to change and the mutual fund industry essentially needs to focus on the corporate segment because it has a huge base with a potential to grow.

Lacunae in Regulations

The mutual fund regulations were first issued in 1992. Later they were altered and amended several times and still need to be fine tuned.

The working arrangement in practice in all mutual funds is in the form of a tripartite agreement among the trustees, the AMC and the custodian. However, the SEBI guidelines do not envisage such an agreement. It is necessary to define the role and relationship among the three in definite terms. The trustees consist of men of eminence but they are not expected to work in the mutual fund office on a full-time basis. They mostly depend on the AMC, more particularly the MD or CFO of the AMC for guidance and information. There is no structured arrangement for the board of trustees to meet the board of directors of AMC. However, a regular interaction would benefit the mutual fund. The right to call for information is not sufficient. The system should facilitate a flow of information and encourage active participation from all its members.

Several quantitative and qualitative restrictions are placed in the area of equity investment. A debt instrument should be rated but rating does not eliminate duration risk and credit risk. For example, a commercial paper is of short-term in nature and though rated, carries a high credit risk because it is an unsecured paper. Restrictions on debt instruments are also necessary.

The offer documents are informative but are too elaborate and confusing to an ordinary investor. Therefore, it would be helpful to the investors if offer documents give a summary sheet of the information in easy-to-understand language.

Check Your Progress

19. How does the SEBI Mutual Fund Regulations define mutual fund?
20. When did mutual fund in India come into existence?
21. Who governs the formation and operations of the mutual funds?
22. How is the net asset value calculated?

10.8 SUMMARY

In this unit, you have learnt that,

- The capital market is susceptible to fraudulent and unfair practices. It is important to protect the trader/investor from such practices.
- A three tier regulatory structure comprising Ministry of Finance, Securities and Exchange Board of India and Governing Boards of the stock exchanges regulate the functioning of stock exchanges.

NOTES

- Even though established in 1988, the Securities and Exchange Board of India received statutory powers only on 30 January 1992. Under the SEBI Act, wide powers are vested in SEBI.
- Promoters' contribution means contribution by those described in the prospectus as promoters, directors, friends, relatives and associates.
- SEBI has set up a division to monitor the unusual movements in prices, in coordination with the stock exchanges. SEBI has asked stock exchanges to monitor the prices of newly listed permitted scrips from the first day of trading.
- The regulation of the functioning of the brokers starts with the registration of the brokers. The registration is given on the basis of the eligibility to be a member of any stock exchange, infrastructure facilities like adequate office space, equipment and man power.
- Banks are financial firms and depend on economies of size and gains arising from internalizing certain activities rather than relying on market transactions.
- Commercial banks are the primary vehicle through which credit and monetary policies are transmitted to the economy.
- The special nature of banks, creation of liquidity, carries risks unique to management of banks. The basic function of bank management is risk management.
- Banks and insurance companies could combine to mutual benefit. Commercial banks can use their branch network to sell all types of insurance, particularly life insurance to their traditional customers. Insurance companies design complex financial products and offer them for placing savings that private customers find particularly appealing such as retirement funds or single premium insurance policies.
- The cooperative banking sector in India comprises urban cooperative banks (UCBs) and rural cooperative banks such as state cooperative banks (SCBs) and district central cooperative banks (CCBs). The cooperative banking sector has an extensive branch network and reach in the remote areas.
- A Rural Infrastructure Development Fund with a corpus of ` 2,000 crore was constituted by NABARD in April 1995 for advancing loans to state governments and state-owned corporations for quick completion of ongoing projects relating to medium and minor irrigation, soil conservation, watershed management and other forms of rural infrastructure.
- The National Bank for Agriculture and Rural Development was established in July 1982 to take over and decentralize the functions in the sphere of rural credit.
- Regional Rural Banks (RRBs) are oriented towards meeting the needs of the weaker sections of the rural population consisting of small and marginal farmers, agricultural labourers, artisans and small entrepreneurs.
- The concept of micro-finance implies informal and flexible approach to the credit needs of the poor. Micro-finance is the provision of thrift, credit and other financial services and products in small amounts to the poor.
- In 1976, the Grameen Bank (GB) was launched in a village in Bangladesh to assist poor families by providing credit to help them overcome poverty.
- Recognizing the potential of micro-finance to reduce poverty, Reserve Bank of India, NABARD and Small Industries Development Bank of India (SIDBI) have promoted the micro-finance movement in India.

NOTES

- The limited outreach and scale of Indian MFIs, relative to the MFI giants in Indonesia and Bangladesh, reflects, at least in part, the absence of an enabling policy, legal and regulatory framework.
- Swarnajayanti Gram Swarojgar Yojana (SGSY) has been launched with the objective of bringing every assisted family above the poverty line within three years, through the provision of micro enterprise.
- Villages are faced with problems related to poverty illiteracy, lack of skills health care etc. These are problems that cannot be tackled individually but can be better solved through group efforts. Today these groups known as Self-Help Groups have become the vehicle of change for the poor and marginalized.
- Self-help group is a method of organizing the poor people and the marginalized to come together to solve their individual problem. The SHG method is used by the government, NGOs and others worldwide. The poor collect their savings and save it in banks. In return they receive easy access to loans with a small rate of interest to start their micro unit enterprise.
- There is no doubt that the SGSY has generated substantial increase in the incomes of swarozgaris in the sample states. Maximum increase was found in Andhra Pradesh, followed by Gujarat, Chhattisgarh, Uttar Pradesh and Bihar in that order.
- Non-bank financial institutions (NBFIs) consist of all-India financial institutions (AIFIs), non-banking financial companies, primary dealers and non-bank finance companies.
- A non-banking, non-financial company is defined as an industrial concern or a company whose principal activity is agricultural operations or trading in goods and services or real estate and which is not classified as a financial, miscellaneous or a residuary non-banking company.
- The Reserve Bank has also made certain refinements in the norms for classification of NBFCs into various sub-groups based on their principal activity as evidenced from the asset/income pattern.
- The insurance industry is an integral part of the financial services' industry and profound changes have been taking place in this sector in India.
- Insurance is a form of risk management primarily used to hedge against a risk of a contingent loss. It is defined as the transfer of the risk of a potential loss, from one entity to another, in exchange for a premium.
- Life insurance in its existing form came in India from UK in 1818 with Oriental Life Insurance Company. The Indian Life Assurance Companies Act, 1912 was the first measure to regulate life insurance business.
- Life insurance is different from other types of insurances. The subject matter of insurance is the life of human beings. Life insurance is a contract, which provides risk coverage to the insurer.
- A typical whole life policy remains as long as the policyholder is alive. The risk is covered for the entire life of the policyholder. Hence, it is known as a whole life policy.
- General insurance covers the loss due to unforeseen events such as accidents, illness, fire and burglary. Unlike life insurance, general insurance is not meant to offer any return but is a protection against contingencies.

- General insurance includes property insurance, liability insurance and various other forms of insurance. Fire and marine insurances are strictly called property insurances.
- The Mercantile Insurance Ltd. was the first company set up to transact all types of general insurance business in 1707. The first general insurance company, Triton Insurance Company Ltd. was established in Calcutta in 1850.
- Mutual fund is a mechanism for pooling resources by issuing units in securities to the investors and investing funds in accordance with the objectives disclosed in the offer document. A mutual fund is a corporation, trust or partnership, which manages the collected money with the help of professional expertise.
- Mutual funds have a custodian. Custodians carry out custodial services for the various schemes of a fund. It is their duty to send intimation of the custodial services rendered to the board.
- Many of the mutual funds are not performing well. Therefore, many investors are disappointed with the returns that they get from the mutual funds.

10.9 KEY TERMS

- **Promoters' contribution:** It means contribution by those described in the prospectus as promoters, directors, friends, relatives and associates.
- **Micro-finance:** It is the provision of thrift, credit and other financial services and products in small amounts to the poor.
- **Residuary Non-banking Company (RNBC):** It is a company which receives any deposit under any scheme or arrangement, by whatever name called, in one lump-sum or in installments or in any other manner and which does not fall into any of the above categories.
- **Non-banking, non-financial company:** It is defined as an industrial concern or a company whose principal activity is agricultural operations or trading in goods and services or real estate and which is not classified as a financial, miscellaneous or a residuary non-banking company.
- **Insurance:** It is a form of risk management primarily used to hedge against a risk of a contingent loss. It is defined as the transfer of the risk of a potential loss, from one entity to another, in exchange for a premium.
- **Premium:** The fee paid by the insured to the insurer for assuming the risk is called 'premium'.
- **General insurance policy:** It may be termed as a contract of indemnity as the insurer normally makes good the actual amount of the loss suffered.
- **Mutual fund:** It is a mechanism for pooling resources by issuing units in securities to the investors and investing funds in accordance with the objectives disclosed in the offer document.

10.10 ANSWERS TO 'CHECK YOUR PROGRESS'

1. According to the SEBIAct, the Ministry of Finance has appellate and supervisory powers over SEBI.

NOTES

2. The carry forward system is mainly used by speculative brokers and large traders. Speculative deals may lead to abnormal price rise and payment crises and affect the investors' confidence. Badla should be considered as a lending and borrowing system with suitable checks and balances. Index futures offers an opportunity for hedgers and speculators.
3. The regulation of the functioning of the brokers starts with the registration of the brokers. The registration is given on the basis of the eligibility to be a member of any stock exchange, infrastructure facilities like adequate office space, equipment and man power.
4. The Banking Regulation Act, 1949 defines banking as accepting for the purpose of lending or investment, of deposits of money from the public, repayable on demand or otherwise and withdrawable on demand by cheque, draft or order otherwise.
5. To meet the financial obligations, two funds were set up under the RBI Act—the National Agricultural Credit (long-term operations) Fund and the National Agricultural Credit (stabilization) Fund with contributions from the RBI every year.
6. A Rural Infrastructure Development Fund with a corpus of ` 2,000 crore was constituted by NABARD in April 1995 for advancing loans to state governments and state-owned corporations for quick completion of ongoing projects relating to medium and minor irrigation, soil conservation, watershed management and other forms of rural infrastructure.
7. NABARD introduced several policy measures for improving its overall performance. These were:
   - Quarterly/half-yearly review of RRBs, especially weak ones by the sponsor banks
   - Merger of RRBs coming under a sponsor bank and operating in contiguous areas
   - Off-site surveillance
8. The concept of micro-finance implies informal and flexible approach to the credit needs of the poor. Micro-finance is the provision of thrift, credit and other financial services and products in small amounts to the poor.
9. In 1976, the Grameen Bank (GB) was launched in a village in Bangladesh to assist poor families by providing credit to help them overcome poverty.
10. Commercial banks were the major source of funds for MFIs.
11. Self-help group is a method of organizing the poor people and the marginalized to come together to solve their individual problem. The SHG method is used by the government, NGOs and others worldwide. The poor collect their savings and save it in banks. In return they receive easy access to loans with a small rate of interest to start their micro unit enterprise.
12. Apart from complementing banks in financial intermediation, NBFIs provide depth and resilience to the financial system.
13. The main advantages of NBFCs lie in the lower transaction costs, quick decision-making, customer orientation and prompt provision of services.

14. The two major classifications of NBFCs are:
    - Equipment leasing and hire-purchase companies (financial companies)
    - Loan and investment companies
15. The fee paid by the insured to the insurer for assuming the risk is called 'premium'.
16. As per the provisions of the IRDAAct, Insurance Regulatory and Development Authority (IRDA) was established on 19 April 2000, to protect the interests of insurance policyholders and to regulate, promote and ensure the orderly growth of the insurance industry.
17. The need for insurance differs from family to family. The insurance policy and the amount taken often depend on the following factors:
    - One's own income level
    - Tax planning
    - The number of dependents
    - The wealth, income and expense levels of the dependents
    - Their significant foreseeable expenses
    - The inheritance and the lifestyle to be provided for them
18. Marine insurance policies are broadly classified into:
    - Marine hull insurance
    - Marine cargo insurance
19. The SEBI Mutual Fund Regulations, 1993 defines mutual fund as 'a fund established in the form of a trust by a sponsor, to raise money by the trustees through sale of units to the public, under one or more schemes, for investing in securities in accordance with the regulations'.
20. The mutual fund in India came into existence in 1964 when Unit Trust of India was incorporated as a statutory corporation.
21. The formation and operations of the mutual funds are governed by the Securities Exchange Board of India Mutual Funds Regulation 1993.
22. Net asset value is calculated on the basis of total asset value of a company, minus the administrative expenses.

10.11 QUESTIONS AND EXERCISES

Short-Answer Questions

1. What are the powers vested in the Securities and Exchange Board of India (SEBI)?
2. State the objectives of SEBI.
3. What is the role of SEBI in the primary market?
4. State the functions of commercial banks.
5. Write a note on the cooperative credit structure.
6. What do you understand by micro-credit? How can regional rural banks help in financing micro-enterprise institutions?
7. What are the problems related to micro-finance?
8. How do the SHGs help the poor section of the society? What is the impact of

9. What do you understand by non-bank finance companies? How many categories are there and what is the basis of classification?
10. What is insurance? What are the features of risk associated with insurance?
11. What is general insurance? What are the matters to be stated in it?
12. Write short notes on (a) marine insurance (b) fire insurance (c) health insurance.
13. Define mutual funds and explain their operation. How is a mutual fund managed?
14. Briefly state the various stages of growth of mutual fund industry in India?

Long-Answer Questions

1. Evaluate the working of Capital and Securities and Exchange Board of India (SEBI).
2. Discuss the powers and functions of SEBI.
3. Explain the role of commercial banks among the financial intuitions.
4. Evaluate the role and operation of National Bank for Agriculture and Rural Development.
5. Discuss the role of regional rural banks in rural credit.
6. Discuss the contribution of micro-finance to eliminate rural poverty.
7. 'Self-help group is a method of organizing the poor people and the marginalized to come together to solve their individual problem.' With regard to this statement, describe the role of self-help groups.
8. Evaluate the non-banks financial intermediaries (NBFIs).
9. Discuss the different categories of insurance.
10. Distinguish between life insurance and general insurance.
11. Critically evaluate the functioning of mutual funds in India.
12. What is the present scenario of the mutual fund industry in India and what are the problems faced by them?

10.12 FURTHER READING

Mankiw, N Gregory. 2010. *Macroeconomics.* Worth Publishers, New York.

Shapiro, Edward. 1996. *Macroeconomic Analysis.* Galgotia Publication, New Delhi.

Branson, WH. 1979. *Macroeconomic Theory and Policy*. Universal Book Stall, New Delhi.

Jha, R. 1999. *Contemporary Macroeconomic Theory and Policy*. New Age International, New Delhi.

Gupta, SB. 1995. *Monetary Economics: Instruments and Policy*. S Chand & Co. New Delhi.