CHAPTER 3

THEORETICAL BASE FOR MACROECONOMETRIC MODELLING

3.1 INTRODUCTION

This chapter reviews the literature related to macroeconometric modelling in preparation for the presentation of the empirical model for the economy of Lesotho. The review seeks to explore the origins, developments and salient features of macroeconometric modelling from a theoretical point of view. The approach adopted in this exercise is that of dwelling on the role of the individual sectors in a macroeconomic setting and their interlinkages with other sectors.

3.2 BACKGROUND TO MACROECONOMETRIC MODELLING

The primary concern of macroeconomic modelling has been to provide a framework that is a representation of the real economy. Its development has undergone phases that have made it more practical and usable for policy analysis. Early works on macroeconometric modelling laid robust foundations for the development of modelling although they were heavily criticised for presenting a static representation of the economy and for being abstract.[16] Later works represented a major breakthrough in terms of theoretical foundations and improved estimation techniques.[17]

Major developments in the construction of macroeconometric models can be traced from the work of Tinbergen (1936) and that of the Cowles Commission under Klein (1950).[18] In the following periods, macroeconomic modelling became the leading basis for the formulation of macroeconomic policy and forecasting. However, with time and as models failed to deliver desired explanations to policy makers, their popularity waned. This was

[16] See for example, Walras, Pareto, Frisch (1933) and Kalecki (1935).
[17] For instance, the Klein-Goldberger (1969) model and the Tinbergen (1936) model.
[18] See Wallis (1994) and Allen and Hall (1997).

followed by scepticism on the part of policy makers and heavy criticism by authorities and academics alike.

The most influential criticisms were those that were based on identification restrictions. Because of its requirement of a complex system of simultaneous equations, the identification process would depend to a large extent on the presence of dynamics in the model. Moreover, the interaction of policy regimes and the importance of the role of expectations rendered the process of identification a difficult one.

These criticisms and subsequent debates not only fuelled interest in the development of other modelling techniques, notably, VAR and CGE modelling techniques, but also led to intensive research in developing techniques that overcame the weaknesses inherent in macroeconometric modelling. With the initial works of Davidson and Hall (1991) and others, came more powerful modelling strategies. In particular, the techniques provided ways to estimate and test the structural relationships and over-identifying restrictions. Because of these and other developments, macroeconometric modelling has since regained its status in the policymaking sphere and is still used today to tackle policy-making in challenging situations.

3.3 PRINCIPLES OF STRUCTURAL MODELLING

A considerable amount of work has gone into developing macroeconometric models. The IS-LM framework can be recalled as one of the major reference frameworks and has served as a benchmark for constructing macroeconomic models for a long time. It gained much popularity prior to the 1970s partly because of its simplistic principle of determining only two variables. However, persistent inflation in the 1970s introduced scepticism with regard to its applicability because of its inability to determine an equilibrium price level. Following this, the AS-AD framework became dominant as an analytical framework. In this framework, income and price level are determined endogenously. These frameworks are summarised as the Keynes-Klein (KK) type of

models. These models used to occupy a dominant position in macroeconometric modelling as they were used extensively to model developed and under-developed market economies alike (Challen and Hagger 1983).

One of the frameworks that enjoyed considerable success is the Mundell-Flemming framework, developed in the 1960s and used extensively afterwards. This framework however had a fatal flaw of not taking into account the aggregate supply or exchange rate influence on domestic prices. Two-sector open economy models based on the tradable and non-tradable sectors were then custom designed to bridge this gap. In these models the purchasing power parity was restricted to tradable markets while prices of non-tradables were influenced to a large extent by domestic factors. In line with the assumption of a small open economy, the price taker assumption was applied to both goods and asset markets.

In the complex economic environments of recent times, a common problem in modelling is the choice of a suitable theoretical framework, the extent of institutional and structural detail and the use of the model. In principle, it is the varying perceptions of the economic system that are formulated into theories of economic behaviour that constitute the different analytical frameworks for economic analysis. These theoretical frameworks in turn determine the specifications of the equations in the system and implicitly the ultimate use of the model. Lee (1997) outlines three aspects that are usually considered in developing a model. These are: (i) the relevance of the model to the key attributes of the economy on which policy should be directed; (ii) the coherence of the model to the analytical framework of how the economy operates; and (iii) the model's ability to represent historical data sufficiently.

Although macroeconometric modelling can be outlined in steps as done by, among others, Hall and O'Sullivan (1997), Greenslade and Hall (1996) and Jacobs and Sterken (1995), there is still a lack of consensus on analytical macroeconomic models for developing countries, particularly at the empirical level.[19] One of the contentious subjects

[19] See Haque *et al*. (1990).

in macroeconometric modelling is the level of disaggregation in the model. The process of disaggregation of variables is said to refine the model and linkages between the variables with the aim of avoiding the 'black box' phenomenon. However, it is also argued that even though a model that is highly disaggregated can be interesting in terms of its capabilities, there is a possibility that it might contain details that obscure the main features of the model and therefore reduce its relevance to the problem that it is supposed to address.

Varying assumptions about the degree of substitutability between foreign and domestic assets and goods is yet another issue of contention in developing models. These assumptions include issues such as the degree of capital mobility and purchasing power parity. For instance, the issue of the substitutability of assets in the fixed and flexible exchange rate regimes and the way in which foreign bonds are viewed by domestic residents are at the centre of the contention.

The use of two- or more-sector models has also been discussed and researched at length in the sphere of macroeconometric modelling. A major argument for these kinds of models is that there are large relative price movements between tradables and non-tradables as well as a way of explaining purchasing power parity (Prachowny 1984:3). However, in some cases this segregation has been seen as impractical for several reasons. Single sector models are preferred in such cases.

3.4 MACROECONOMETRIC MODELLING IN THE CONTEXT OF DEVELOPING ECONOMIES

Within a broad category of developing countries, African countries collectively have particularly unique economic structural features. A general observation is that the effects of policy and structural changes are rarely reflected in most models used by policy makers in African countries (Soludo 1995:2).

Among their common characteristics when considering macroeconometric modelling, is the problem related to the inadequacy of time series data in terms of both quantity and

quality. Other problems related with data include discontinuities, changes in definitions and missing observations (Jacobs and Sterken 1995:112). It is important to incorporate these aspects explicitly in constructing models for such economies. According to Hall and Pauly (2001:2), econometric modelling can still be useful despite structural change if it takes into account the form of change that has taken place.

On the practical front, the development of economy-wide macroeconometric models for purposes of dynamic policy analysis is still at a very rudimentary phase for many developing countries. Even though a number of models have been developed for a large number of African developing countries, they are usually constructed for academic purposes and are hardly maintained for purposes of policy analysis. Emphasis is often put on the economic and structural history of the economic system, rather than on the analytical transparency of model design and results. Because of this, many models apply to short-run policy issues and ignore the long-run sustainability of economic behaviour. These designs are usually justified on the basis of structural rigidities and institutional features such as the dual economic systems, the underdeveloped financial systems and excessive government intervention. This often leads to lack of consistency in policy formulation and introduces some scepticism in the relevance of conventional analytical frameworks in explaining economic behaviour (Soludo 1995:2).

Another important issue in modelling developing countries, in general, concerns the correct stance of theory and structural properties of the economy in policy analysis. In some cases more emphasis is placed on economic theory while in others the structural features of the economy dominate the design of the model. Some analysts argue that economic theory is economy specific and that structural and institutional features should be taken as exogenous. Such models tend to be robust, given that the policy environment and structures are stable. A plausible alternative is to assume that economic theory is given. Under these circumstances, it is easy to attribute failure of theory to work to distortions in the institutional and structural framework. A halfway approach that seems practical, while it accommodates as much information as possible, is to entertain both aspects in order to preserve analytical intuition.

The treatment of expectations still poses a major hurdle in modelling developing economies. It seems difficult to justify the use of any type of expectations. The argument is that both adaptive and forward-looking expectations are extreme characterisations (Soludo 1995:15). In recent attempts, learning models of expectations take precedence, though with some criticism.

Hall and Pauly (2001) provide a framework for modelling individual African economies. Two key factors are highlighted. Firstly, that the framework should assume that in the long run the economy would conform to the predictions of economic theory. This effectively gives way to calibrating long-run parameters in line with economic theory as a point of departure. Secondly, it is important to bear in mind that structural changes may render some coefficients unstable and lead to measurement problems. In many cases the Kalman filter is used to deal with these two problems while it also introduces the learning type of expectations in the system.

The use of some variant of the ECM is appropriate to represent the basic structure of the model. Because of the limited span of data and measurement problems, it recommended that the model should be kept small and simple while it incorporates as much prior information and economic theory as possible. In order to preserve the degrees of freedom it is also important to limit the amount of lags that are used in the model. Since structural change affect the parameters in the model, the use of time-varying parameter estimation techniques is appropriate to capture these changes.

3.5 MACROECONOMIC MODELLING IN THE PRESENCE OF STRUCTURAL CHANGES

Conventional econometric models rest to a large extent on the assumption that the structure of the economy is stable in the short and long run (Hendry and Clements 2000). It is noteworthy, however, that structural breaks are the most common features in macroeconomic data sets of developing and transitional economies alike. The presence of structural breaks in the data generating mechanism implies that it is impossible to

establish the primacy for forecasting of causal information over non-causal variables since the economy is basically moving between two different regimes with different structural characteristics. This then raises scepticism towards the usefulness of causal information for forecasting, the role of parsimony and the impact of collinearity (Hendry and Clements 2000). The two sets of coefficients will essentially follow a non-stationary process during the change because the underlying model parameters will have changed. Because structural changes introduce volatility in the long-run parameters of the model and affects the forecasting performance of the model, it is crucial that it is incorporated in the model explicitly (Greenslade and Hall 1996). In accommodating structural change, one of the options that modellers have in such cases is to specify the equations to fit only the latest data set using *a priori* assumptions on how they believe the economy has changed. Greenslade and Hall (1996) however warn that although this approach may work for models whose purpose is forecasting, it is not suitable for models that are to be used for policy analysis and understanding past trends.

Hendry and Clements (2000:4) outline a number of approaches that can be used to avoid systematic forecasting errors due to deterministic shifts caused by structural breaks. These include intercept corrections, differencing, co-breaking and regime-switching models. A more plausible remedy often used for this condition is time-varying parameters and the Kalman filter (Hall and O'Sullivan 1997:267).[20] Intercept corrections are described by Hendry and Clements (2000:15) as non-zero values for a model's error terms added over the forecast period to adjust a model-generated forecast to prior beliefs thus allowing for expected future events that are not explicitly catered for in the model. It is also crucial to incorporate as much information as possible concerning the form and timing of the structural change. This may include issues such as the type of the change, the time at which the change began and when it ended, the speed of adjustment and the rate of change of parameters.

Alternatively, the use of switching models can be instrumental in portraying structural change. This approach was traditionally used to model disequilibrium markets that switch

[20] See also Basdevant (2000).

between demand and supply constraint conditions. The basic assumption of this approach is that the model switches in a discrete pattern between two sets of parameters representing different states of nature with only a random error. The role of the Kalman filter in this and other techniques is then to introduce a learning type of expectations that evolve with the model. Since it is built into the model, it automatically changes the values of the perceived parameters in line with observed expectation errors. If the Kalman filter can be used for these purposes it can obviously be used to allow parameters of the model to change in the face of structural change. The implication of this approach is that through the Kalman filter, the simulation properties of the model may be preserved while the dynamics of the model will change in response to structural change to produce forecasts that are consistent with the new regime (Hall and O'Sullivan 1997:268-270).

3.6 THEORETICAL GUIDELINES TO SECTORAL MODELLING IN THE CONTEXT OF MACROECONOMIC MODELS

3.6.1 Theoretical models of aggregate supply

Increasingly detailed modelling of the supply has come under the spotlight in macroeconomic modelling than was formerly the case (Challen and Hagger 1983). In contrast to the Keynesian view that expanding aggregate demand will raise income and reduce unemployment, proponents of the supply side focus on the view that increased saving is necessary for increasing capital formation that can raise productivity and hence economic growth. Instead of demand management, a key feature of the supply side view is therefore to create capacity through capital formation. According to Meiselman (1982:42), and Roberts (1982:49), supply side economics acknowledges that fiscal policy, in particular the tax component of fiscal policy, affects incentives to invest, economic efficiency and economic growth by way of changing relative prices. This suggests that in order to evaluate policies designed to redress the improvement of capital formation, it is important to analyse factor supply decisions in the economy. Following the neo-classical flexible accelerator theory pioneered by Jogernson (1969), the decision to invest is influenced by, among others, the cost of capital which, in turn depends on tax

policy and other investment incentives such as low rates of interest and inflation, a stable foreign exchange market and general macroeconomic stability. A reduction in the erosion of investment incentives can encourage both private and public capital accumulation substantially. However, the increase in demand for either type on investment needs to be accompanied by an increase in the supply of financial capital available to finance investment. According to Boskin (1982:22), if supply is not increased simultaneously, the increased demand will merely hike interest rates.

Conventional models of the supply side consist of two approaches namely, the production function approach and the cost structure approach. While many recent works tend to use the cost structure to model the supply side, the production functions approach is also widely used and suited for cases in which there are data limitations. The former approach places emphasis on wages and prices. The latter approach involves the presentation and estimation of the functional relationship between inputs and output in the production process. Much emphasis is placed on the treatment of technology and the demand and supply of factors. In principle, the two approaches can be seen as identical in many ways. It is the technical issues that often lead to divergences in treatment (Whitley 1994:111). While the choice of variables in each case depends on individual analyses, a number of basic functional forms have been developed over time.

3.6.1.1 The production function approach

Assuming a general production function of the following form

$$Y = f(L, K, T) \tag{3.1}$$

Where Y is output, L is employment, K is capital stock and T is technology, factor demands can be derived from the marginal productivity conditions obtained by profit maximization. Under conditions of perfect competition, the marginal productivity conditions imply that firms equate the real wage to the marginal product of labour such that

$$\frac{w}{p} = f(L, K, T) \tag{3.2}$$

The demand for labour can then be derived from the above relation as

$$L^d = f(\frac{w}{p}, Y, T) \tag{3.3}$$

The Cobb-Douglas production function is one of the widely exploited functional forms that allow substitution between the factors of production. The function takes the following form

$$Y = L^{\alpha} K^{\beta} \qquad \alpha, \beta > 0 \tag{3.4}$$

Among the desirable properties of this function is the relative ease with which it can be manipulated. More importantly, the distributive factor shares of are given by the exponents of the factors and are constant. In addition, the Hicks and Allen elasticities of substitution are identical for all factors and equal to unity. These properties originate from the additivity and homotheticity assumptions imposed on the Cobb-Douglas production function and represent some of its major limitations.

A major theoretical improvement with regard to the flexibility of the unitary elasticity of the Cobb-Douglas function by independent efforts of Arrow and Solow in 1961 led to the introduction of the constant elasticity of substitution (CES) production function given by:

$$Y = \gamma[\delta L^{-\rho} + (1-\delta)K^{-\rho}]^{-\nu/\rho} \tag{3.5}$$

In this function γ is the efficiency parameter, δ is the distribution parameter and ρ is the substitution parameter. The returns to scale are given by ν while the Hicks and Allen

elasticities of substitution are identical and given by $\frac{1}{(1+\rho)}$. A major breakthrough of the CES is that it allows for an arbitrary degree of substitution between factors. However, because it is homogeneous and quasi additive, its factor shares are independent of total output and its elasticity of substitution is the same for all input pairs. According to Chung (1994:111), although the constant value of the elasticity of substitution can assume a wide range of values, it is not as flexible as a variable. Thus, the range of variation of the elasticity is to a large extent still limited. In addition, as a corollary to the homotheticity and quasi-additivity characteristics, the factor shares of output are also quite rigid.

The transcedental logarighmic (translog) function originated from Kmenta's 1967 log-linearization of the CES function and has been widely used in recent empirical analysis. It arose as an attempt to develop a more flexible technology and therefore removes the assumptions of homotheticity and separability. Although these two restrictions can be imposed as testable hypotheses, they are not prerequisites. The translog function is written as follows for an n-input case

$$\ln Y = \ln a + \sum_{i=1} \alpha_i \ln x_i + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \beta_{ij} \ln x_i \ln x_j \qquad (i \neq j, \quad i, j = 1,...,n) \tag{3.6}$$

This function provides a greater variety of substitutions of transformation patterns than those restricted by the constant elasticity of substitution. The form of the functions determining output shares of inputs and the Allen-Uzawa partial elasticities of substitution for the production and cost frontiers are identical. Because of its flexibility, the translog function has proven highly useful in bridging the gap between theoretical and empirical research and is considered more desirable.

Whitley (1994:112) points out several practical problems of using the production function approach. One of them is that real wage elasticities with respect to employment are often below unity. This tends to conflict with the built-in restrictions of the Cobb-

Douglas function. Moreover, output should be treated as an endogenous variable in the estimation. It also becomes difficult to ascertain consistency between factor demands and price setting decisions under the production function approach.[21]

3.6.1.2 The cost function approach

The use of the cost structure addresses the problem of consistency between factor demands and price setting decisions in particular. In principle, the approach involves minimisation of the cost function of the following form with respect to the production function

$$C = wL + rK \tag{3.7}$$

Where w and r are the wage rate and the cost of capital respectively. As with the production function approach, this approach yields factor demand as functions of relative factor prices, an exogenous output level and technology. Thus the labour demand function takes the following general form

$$L^d = f(\frac{w}{r}, Y, T) \tag{3.8}$$

An important feature of this approach is that is possible for the price equation to be derived within a consistent framework together with factor demands. While the approach seems to have desirable properties in technical terms, it is also noteworthy that the explicit estimation of the production function facilitates the derivation of a more direct measure of capacity utilisation, a feature that the cost function approach lacks, yet a primary objective of many macroeconometric models.

[21] See Allen (1997:169) and Whitley (1994:112).

3.6.2 Theoretical models of aggregate demand

The modelling of the demand side involves the determination of the behaviour of aggregate demand in the economy by modelling the individual expenditure behaviour of all economic agents based on theoretical and practical experiences. Viewed from the demand side, actual economic activity is assumed to be influenced by demand factors and is therefore represented by the national income identity.

$$Y^{ACT} = C_P + C_G + I_P + I_G + (X - M) + INV \qquad (3.9)$$

Where C_P is private consumption expenditure, C_G is government consumption expenditure, I_P is private investment expenditure, I_G is government investment expenditure, X is exports of goods and services, M is imports of goods and services and INV represents the change in inventories.

3.6.2.1 Consumption behaviour

Four basic theories of consumption behaviour, dating back to the 1930s, form the core of the theoretical literature on consumption and its extensions. Collectively, these theories attempt to provide understanding of the determinants of optimal levels of consumption. With their foundations in the microeconomic analysis of consumer behaviour, they explain the consumption behaviour of larger units in the economy.

The pioneering absolute income hypothesis (AIH) developed by Keynes in 1936, predicts a positive relationship between the individual's consumption at any time, and the current level of disposable income. On the other hand, the relative income hypothesis (RIH) developed independently by Duesenberry (1949) and Modigliani (1949) relates consumption to the income of the individual relative to that of the rest of society. Drawing from optimization behaviour, it contends that an individual maximizes utility subject to a weighted average of the population's consumption. Solution to this problem suggests that consumption is dependent on the individual's percentile position in the

income stream as well as past levels of this variable. This model has been generalised in the habit persistence model of Brown (1952).

The permanent income hypothesis (PIH) developed by Friedman (1957) and the life–cycle hypothesis (LCH) developed by Modigliani and Brumberg (1954) and Ando and Modigliani (1963) have gained greater popularity in empirical analyses. The LCH relates consumption to lifetime income as opposed to contemporaneous income. The underlying problem in this model is the maximization of intertemporal utility subject to a lifetime budget constraint. The theory presupposes that an individual maintains a constant or steadily increasing consumption pattern throughout a lifetime given low or negative levels of income at the beginning and end of life and higher income levels at mid-life. The basic idea is that the individual maximises utility in such a way that it does not exceed the present value of lifetime income. The outcome is a consumption function that is dependent on the unobservable present value of lifetime income. To overcome this measurement problem, the theory defines total income as consisting of human and non-human income. This ultimately yields consumption as an increasing function of labour income and non-labour income or household net worth.

Starting from similar foundations, the permanent income hypothesis (PIH) developed by Friedman (1957) contends that consumption is a positive function of permanent income which is defined in simple terms as the average of present and future income. Thus the slope of the consumption function is determined by the relative variation in permanent and transitory incomes. An important implication of this hypothesis is that the rise in current income is associated with an increase in consumption only to the extent that it reflects an increase in permanent income. In turn, permanent income is defined as the product of real rate of return on total wealth and the present value of lifetime income. It is the measurement of permanent income, which introduced the role of expectations in the analysis of consumption behaviour. Early works on PIH-based consumption functions, including that of Friedman (1957), made extensive use of adaptive expectations and a distributed lag formulation with a Koyck transformation in the measurement of permanent income. With time and theoretical developments, the use of rational and

learning models of expectations has taken the centre stage.[22] This was also a result of the instability of the consumption function stressed by Lucas (1976) in the critique on econometric policy assessment. This has made forward-looking models of consumption more popular.

Different approaches have been used to build on the basic theoretical frameworks in the preceding discussions. These include analysis of consumption paths under conditions of uncertainty, extending analysis to account for multiple assets and risk, liquidity constraints and buffer stock models. A host of other variables that are hypothesised to influence consumption patterns and which have surfaced implicitly in the original frameworks have been given more attention. These include the interest rate, different types of wealth, taxation, financial intermediation and demographic factors.

Consideration of uncertainty follows the analysis analogous to that of PIH and yields the proposition of the unpredictable consumption patterns and Hall's (1978) hypothesis that consumption follows a random walk. Counter to predictions of the basic frameworks, Hall's extension posits that unexpected fall in income led to decline in consumption only by the amount of the fall in permanent income. Thus, changes in consumption are unpredictable in the face of uncertainty.[23]

The rate of interest is introduced into the model by relaxing the assumption of zero interest rates and thus allowing for a non-zero discount rate. The analysis under these conditions suggests that consumption need not be a random walk. Consumption rises over time if the rate of interest exceeds the discount rate and falls otherwise. Variations in the real rate of interest then suggest variation in the predictable component of consumption growth.[24]

---

[22] Among others, Hall (1978) reformulated the LCH and the PIH by adding the assumption of rational expectations to the first order conditions of the intertemporal optimizing model of consumer behaviour.

[23] It is noteworthy however that most empirical works fail to confirm that lagged values of income and consumption fail to predict changes in consumption. See for example, Hall (1978), Campbell and Mankiw (1989a) and Shea (1995).

[24] Early empirical works however suggest that growth of consumption is not very responsive to changes in the rate of interest, thus implying a low intertemporal elasticity of substitution. See for example, Mankiw (1981), Hansen and Singleton (1983), Hall (1998b) and Campbell and Mankiw (1989a).

The relation between consumption and taxation is implicit in the relation of the former to disposable income and provides a nexus between the real sector and the government sector. This link stresses the role of taxes paid by the households to government. The obvious effect of taxations is traced to the budget constraint of the household. According to Sachs and Larrain (1993:98), higher taxes reduce consumption for a given path of output by reducing the present value of disposable income. In turn, whether changes in taxes are transitory or permanent may potentially have important policy implications.

Accounting for multiple assets provides a link between the household behaviour and asset markets. Within an intertemporal setting, and given expected returns to assets, the model shows that the aspect of riskiness, that matters to the decision of whether to increase holdings of a particular asset, turns out to govern the relation between the assets' payoff and consumption. As the individual increases investments in assets, consumption tends to depend on the assets' payoff. This suggests that hedging risks is important to optimal portfolio choices. The model further shows that if the expected returns on assets are determined by their demands, using the consumption capital-assets pricing model brings us to the conclusion that the premiums that assets offer are proportional to the covariance between an asset's return and consumption (see Romer 1996:330).

The PIH has also been extended to incorporate the effects of liquidity constraints. This effectively relaxes the assumption that consumers borrow and save at the same rate of interest in intertemporal models of consumption. If individuals face higher interest rates for borrowing, there is a likelihood of lower borrowing for purposes of smoothing out consumption given lower current income. Thus liquidity constraints tend to put more weight on current income as a determinant of consumption than predicted by the PIH.[25]

The rate of inflation or prices in the consumption function is used to capture the inflation loss on liquid assets (Whitley and Bai 1997:69). According to Thomas (1993), a rise in

[25] See Hayashi (1982), Hayashi (1985) and Flavin (1985) for empirical evidence on the effect of liquidity constraints on consumption.

the price level, with real income and real wealth held constant, should produce an equiproportionate rise in nominal income and wealth and hence should not affect consumption expenditure. A positive coefficient of the price level would suggest money illusion. Besides the inflation loss on liquid assets, the inflation variable in the consumption has been rationalised by Deaton (1978) as a proxy for uncertainty particularly in cases of accelerating inflation.[26]

3.6.2.2 Investment demand

Empirical analysis of investment behaviour follows form a number of theoretical models proposed within different schools of thought. While the Keynesian present value and marginal efficiency of investment criteria trace investment behaviour from the basic profit maximisation behaviour of a firm and have been widely used in empirical analysis, later models such as the Jorgenson's (1971) accelerator principle linked the theory of investment to growth models.

A major argument of these later developments lies in the introduction of dynamics to the static present value model. Restrictive assumptions behind the accelerator theory prompted further developments in the theory of investment by, among other, Jorgenson and Hall (1971). In this approach, the level of output and the user cost of capital influence optimal/desired capital stock. In turn, the user cost of capital depends on the price of capital, the real rate of interest and the rate of depreciation. Although it uses a different approach of introducing adjustment costs, the Tobin Q principle by Tobin and Brainard (1968) and Tobin (1969) complements the accelerator theory (Sundararajan and Thakur 1980:825; Hayashi 1982:214).

Of these frameworks, the neoclassical flexible accelerator theory and the Tobin Q theory are considered more suitable for modelling investment behaviour in cases where the economy is supply-constrained. Their appropriateness stem from the ease of identifying

[26] While empirical works by Deaton (1978) and Juster and Watchel (1972) using US data suggested that consumption levels fall as the price level rises, Branson and Klevorick (1969) found contrasting results for the same data.

policy instruments that government has at its disposal to influence or enhance aggregate supply from the resulting investment function. Although Q models of investment have quite appealing features that provide firm theoretical and empirical foundations for the analysis of investment, in that they start from the micro foundations of optimisation behaviour and rationalise the importance of expectations in investment decisions (Blundell *et al*. 1992:234), the flexible accelerator theory is usually chosen over the Tobin Q theory for modelling investment for several reasons. Firstly, the measurement problems associated with the unobservable marginal Q (usually measured by the average Q which is the ratio of the market value of the existing capital stock to its replacement cost) in the case of the Tobin Q principle pose threat to the performance of a model if used in such a context. Hayashi (1982:218) shows that the marginal and average Q can only be the same in the special case where the firm is a price taker and the production function and the installation function are the same. Measurement problems are at the core of the poor performance of the Q models and have cast scepticism on the adequacy of the forward-looking neoclassical investment theory itself (Allen 1997:42).[27] The problem is further compounded by the realisation that it is highly likely that in many instances financial markets measure the financial firm inaccurately (Blundell *et al*. 1992).[28] The application of the Tobin Q model within the context of developing economies is limited by rudimentary stage of capital markets (Geda *et al*. 2001).

On the other hand, the accelerator principle also has some shortcomings. The first shortcoming concerns the impact of changes in exogenous variables and the second deals with the role of expectations in investment decisions (see Romer 1996:347). Firstly, by implication, the model suggests that discreet changes in exogenous variables require infinite rates of investment. However, it is common knowledge that investment in the economy is constrained by the level of output and therefore cannot be infinite. Secondly, it is assumed that firms equate the current revenue product of capital to its current user cost. However, in practice, expectations about demand and costs are central to investment

[27] A vast amount of empirical analysis provides comprehensive of this deficiency. These include von Furstenberg (1977), Summers (1981) and Blundell *et al*. (1992). See Jorgenson (1971) for a detailed survey of empirical works on investment behaviour,
[28] See also Blanchard *et al*. (1993) and LeRoy (1989).

decisions. Firms expand their capital stocks when they expect their sales to be growing and the user cost of capital to be low. The introduction of adjustment costs in this case caters for the difference introduced by expectations. These consist of internal and external adjustment costs. Internal adjustment costs are associated with changing or installation of new capital stock while external costs are associated with changes in the price of capital relative to other goods so that firms have to make decisions of whether to invest or disinvest. To account for these arguments, liquidity and credit allocations in the economy may be included to capture adjustment costs and to reflect the fact that in many developing countries financial constraints and limited availability of foreign exchange affect investment decisions.

A variant of the accelerator principle permits the exploration of the contention that public investment may be a stimulus to or hinder private investment. A number of potential channels through which government investment is related to private investment are outlined in the literature. Firstly, government may compete with the private sector for scarce physical and financial resources, hence exerting a negative influence on private investment at least in the short run. Secondly, since public investment essentially complements private investment by contributing to increasing the productivity of private capital stock, private investment requirements per unit of output may be eroded. Thirdly, an increase in public investment may raise the demand for the output of the private sector thereby affecting output expectations and investment requirements of the private sector. Finally, by raising output and savings, thereby supplementing economic resources, public investment may work to offset part of the crowding out effects.

3.6.3 The government sector in a macroeconomic context

Though it is not usually accorded the attention that it deserves in macroeconomic models, the modelling of the governments sector should occupy a central position and made as comprehensive as possible. This argument stems from the fact that the behaviour of government almost invariably is the major driver of many key macroeconomic variables. This is especially true for developing economies in which the private sector is at an infant

phase and government plays a major part in economic decisions. According to Schmidt-Hebbel (1994:15), overindebtedness and the resulting debt crisis, high inflation and poor investment and growth performance in developing countries can all be traced back to imbalances in the public sector. This implies a strong linkage between the public sector and other sectors of the economy. From the monetarist point of view, large deficits are translated into inflationary pressures. The linking factor in this analysis is the way in which government chooses to finance its deficit (Easterly and Schmidt-Hebbel 1994:15; Biggs 1998:238). Domestic borrowing is highly linked with credit squeeze and rising interest rates, while foreign borrowing and money creation are linked with external imbalances and inflationary pressures respectively. According to Easterly and Schmidt-Hebbel (1994:17), external borrowing may lead to a current account deficit though either higher interest rates or credit allocation practices and real exchange rate appreciation and sometimes a balance of payments crisis or debt crisis depending on the level of reserves. Another likely outcome in all these cases is the fall in investment and economic growth induced by the erosion in the profitability of investments. The real sector may also be affected through the response of consumption and investment expenditures to budget imbalances. With regard to consumption expenditures, the Ricardian equivalence hypothesis by Barro (1974) is an attempt to establish the relationship between private consumption and changes in taxation. The degree of substitutability or complementarity of public investment has important policy implications.[29]

The debate of the possible mechanisms though which the behaviour of government influences economic activity dates as far back as the times of Mercantilism and the early classical economists such as Hume, Smith, Ricardo and Mills.[30] A vast amount of literature attests to the fact that the main ways in which the behaviour of government affects macroeconomic activity are taxation and its expenditure patterns. Budget imbalances in turn affect the performance of the economy in two main ways, namely, their effects on resource allocation and through their effects on the rate of capital accumulation (Biggs 1998:235). Within the neoclassical school, even though an increase

[29] See Buiter (1997) for a taxonomy on crowding out theories.
[30] See Mundell (1993).

in deficits may seem to increase output in the short run, the effect is very temporary, for an increase in deficits has detrimental effects on economic growth in the long run. The capital accumulation channel works regardless of the budget financing choice of government. The rate of interest is the main conduit in this analysis and works in the general case to stifle private investment. In principle, a deficit is a reflection of increased consumption levels and by implication, lower levels of saving. Given that demand for loanable funds is in excess of supply at a given rate of interest, this creates an upward pressure on interest rates and hence discourages investment. In the context of developing economies, deficits have been known to crowd out exports as well. The rise in interest rates further induces capital inflows and an appreciation of the real rate of interest. In turn, the competitiveness of local commodities in international markets is hampered.

The Keynesian mechanism operates by way of changes in resource utilisation and produces results that are directly opposite to that of the neoclassical view. In the Keynesian view, private and public investments are complementary to each other and work to produce a higher overall level of capital accumulation. A deficit is perceived as a reflection of an increase in aggregate demand and ensures that previously unemployed resources are utilised to increase output. In addition, investors view the increased aggregate demand as an inducement to expect increased profitability from investments. Although the resource allocation mechanism seems to be a theoretically sound argument, its practical use for policy purposes may however be limited. It is argued that spare capacity that is needed for output to rise may not be available in all sectors of the economy. In such situations, prices may be pushed up or a surge in imports may be realised depending on the openness of the economy. Ultimately, this leads to a worsening of the current account translating into a reduction in reserves or balance of payments crisis, which can easily lead the economy into a debt trap. On the other hand, this reduces investor confidence while it also increases the risk of investments and consequently a fall in investment.

3.6.4 Theoretical models of the external sector

International trade in goods and services is the primary channel that links the domestic economy with other economies. While some proponents argue that the relationship may have been overstated over the years[31], the relationship between the openness of the economy through trade and economic development is the basis for the importance of the external sector in any economy. Economic theory postulates that through specialisation, increased efficiency and skills and technology transfer, international trade increases output and welfare across nations. This follows from the pioneering works of Adam Smith (1776) through the theory of absolute advantage, Robert Torrens (1808) and David Ricardo (1821) through the theory of comparative advantage and Gottfried Haberler (1936) through the theory of opportunity costs, among others. Although refinements and modifications have been developed over time to accommodate practical trade relations, the basic notion that underlies the relation between the external sector and the rest of the economy is still that unrestricted trade provides more benefits overall.

While ideas regarding the linkage between trade policy and economic development seem to be changing rapidly in the wake of liberalisation and globalisation in recent years, the notion that trade policy is central to the overall design of policies to enhance economic development remains at the fore. While the idea of inner oriented policies such as import substitution and high levels of protection enjoyed much favour in their times, an outer oriented trade regime coupled with the removal of restrictions has become dominant in recent years (Krueger 1997:1). With the emphasis of protectionist strategies and the invariable need for foreign exchange, the former led invariably to insurmountable BOP crises across developing countries. The subsequent intervention by the World Bank and the IMF by way of the stabilisation programs whose main recommendation entailed measures that restrained domestic demand (Crockett 1981:61; Khan and Knight 1982:713). On the other hand the stabilisation programs have tended to loose their popularity in many developing countries owing to the deterioration of country economic performances. The present idea of outward oriented trade policies supports supply-side

[31] See Rodriguez and Rodrik (1999).

policies intended to increase economic efficiency. These measures have the potential to increase the overall level of capacity output given the existing stock of resources without lowering current levels of consumption. They basically involve the overall removal of restrictions and other distortions with the objective of improving efficiency and resource allocation.[32] While these strategies have their own drawbacks, the move is clearly towards the revival of the neoclassical growth and trade theory encouraging free trade and more integration of international markets.

Because of the growth of international trade and developments in international market relations and policy, as well as the sceptical view that empirical evidence in support of the globalisation is very weak,[33] the development of international trade policy remains at stake for developing countries at large. While the current wave of globalisation seems to have overwhelmed policy makers, there is still need for an in-depth analysis of domestic characteristics that will help to design trade policies that best suit the domestic economy, while at the same time fits well with changes in the global economy.

Earlier works on the balance of payments include the elasticities approach due to Robinson (1937) and Machlup (1955), and the absorption approach to the balance of payments developed by Alexander (1952). Inclusive among the theories of the balance of payments are the highly exploited Mundell-Fleming approach developed by Fleming (1962) and Mundell (1962, 1963). An addition to these is the monetary approach to the balance of payments following Dornbusch (1973), Whitman (1975) and Frenkel and Johnson (1976). The portfolio balance approach is a later and expanded development stemming from the monetary approach.

The main focus of the elasticities approach to the balance of payments is on the demand conditions and on the current account of the balance of payments with heavy reliance on the Marshall-Lerner condition. This approach puts more emphasis on the price effects of a devaluation. Its core message centres on the effect of a devaluation on the current

[32] See for example, Khan and Knight (1982).
[33] See for example, Rowthorn and Kozul-Wright (1998).

account balance and hence the overall balance of payments. Depending of the price and volume effects, a devaluation will improve the position of the current account only if the sum of the foreign elasticity of demand for exports and home country elasticity of demand for imports exceeds unity, that is, $\eta_x + \eta_m > 1$. If the sum of the elasticities is less that unity then devaluation will lead to a deterioration of the current account.

The absorption approach complements the elasticities approach by extending the analysis to encompass the income effects that result from a devaluation and hence also relies on the Marshall-Lerner condition. In this method a current account imbalance is perceived as the difference between domestic output and absorption, in which case a surplus in the current account imply an excess of domestic output over absorption and a deficit will reflect a shortfall of absorption over output. It predicts that if a devaluation raises domestic income relative to domestic absorption, the current account improves, provided that the marginal propensity to absorb is less that unity. On the other hand, if a devaluation raises domestic absorption relative to domestic income, the current account will deteriorate. The overall effect of a devaluation depends on the employment and terms of trade effects of a devaluation on income. In turn, direct absorption is affected by a devaluation through various channels including the real balance effect, the income distribution, the money illusion effect, the expectational effect and the Laursen-Metzler effect. Despite their simplistic assumptions, ambiguous conclusions, and deficiencies, the two approaches have remained influential as they contain useful messages for policy makers and have been widely used in empirical analyses.[34]

The Mundell-Flemming model is one of the frequently used Keynesian approaches to modelling the balance of payments. Its main contribution is that it integrates the capital account of the balance of payments in the analysis. By an application of the portfolio theory pioneered by Tobin (1958) and Markowitz (1959) to international markets, the theory relates interest rate differentials to capital flows. The model rests on the law of one price and is less dependent on the Marshall-Lerner condition than other models

[34] For a review of empirical works see for example Gylfason (1987), Knight (1984), Goldstein and Kahn (1985) and Krugman (1991).

(Stevenson *et al.* 1988:213). Although the model enjoyed considerable success, its fatal flaw is that it allowed neither aggregate supply considerations nor the exchange rate to influence domestic prices (Prachowny 1984:3). Besides the capital account, the model also links the balance of payments to the monetary sector with implications that a balance of payments surplus will increase domestic money supply depending on the sterilisation policy being followed in the economy.

The monetary approach to the balance of payments developed by Dornbusch (1973) makes use of the law of one price in both goods and assets markets and full or natural level of employment, a common feature of the monerarist models. The fundamental idea of this approach is that the balance of payments is the outcome of a flow divergence between the growth of the demand for money and the growth of credit with the money stock consequences of the balance of payments bringing the money market into equilibrium. In the words of Frenkel and Johnson (1976:21), the model "stresses the budget constraint imposed on the country's international spending, and views the various accounts of the balance of payments as the windows to the outside world, through which the excesses of domestic flow demands over domestic flow supplies, and of excess domestic flow supplies over domestic flow demands are cleared". This elucidates the monetarist view that the balance of payments is entirely a monetary phenomenon. One of the notorious policy implications dictated by the monetary approach is the impotence of monetary policy in its effects on domestic variables, while it becomes fully potent in its effects on the balance of payments. The Mundell-Flemming model under conditions of no sterilisation also dictates this result.[35] In addition, the monetary approach predicts an automatic adjustment mechanism for the balance of payments with no permanent effects, provided that a shock to the balance of payments is not underwritten by a change in domestic credit (Stevenson *et al*. 1988:243).

Early versions of the portfolio models of the balance of payments can be found in Oates (1966), McKinnon (1969) and Branson (1974). The main contribution of these models is the integration of the stock theory of the capital account and the role of wealth effects.

[35] For a synthesis of the two approaches, see Frenkel *et al.*, (1980).

The model is an extension of the monetary approach designed to embrace more liquid assets and hence draws heavily on the diversified portfolio theory pioneered by Tobin (1958), Markowitz (1959) and Sharpe (1964), which in turn is superimposed on the real sector. The basic element of the model is to trace the process of portfolio adjustment to the balance of payments following a monetary disturbance. According to the portfolio theory, changes in the money stock are translated into changes in wealth, and hence wealth effects, that change the demand for varying types of assets in the economy depending on a number of factors, including the sizes of different portfolio multipliers and the degree of substitutability between assets (Stevenson *et al*. 1998:162). Within the context of the balance of payments any disturbance in the current account gives rise to wealth and substitution effects that counteract each other so that the long-run equilibrium income level is given by the current account, unless they are sterilised. In essence the portfolio balance approach suggests changes in domestic and foreign interest rates, changes in wealth resulting from changes in savings and expectations as important determinants of the balance of payments. Although this approach seems to take on a generic form, it is often criticised on the grounds that it does not appropriately cover the real sector. In addition, because it assumes that markets clear, it is hardly suitable for application in developing countries where financial markets are characteristically narrow and inefficient.

In practice, issues such as the level of disaggregation and the distinction between the tradable and non-tradable sectors of the economy seem to play a major role in modelling the external sector. Most importantly, considerations of the size of the economy, the degree of capital mobility and the exchange rate regime take precedence in individual country models.

3.6.5 The labour market, prices and wage determination

The control of inflation and unemployment are some of the major objectives of stabilization macroeconomic policy. While the labour market is directly linked to the supply side of the economy, the price sector links the supply side of the real sector to

other sectors of the economy, in particular the final goods sector, the external sector and the monetary sector. This is an integral part of the model as it deals with factors that are responsible for the generation of price instability in the economy. As with the monetary sector, it is difficult to disentangle price developments from the labour market. Theoretical foundations of inflation almost invariably have their roots in the monetary impulses, and more so, in the labour markets. In the view of Whitley (1994:121), price equations tend to be a mark-up on costs of production in partial wages and imports costs with the mark-up being sensitive to demand influences. However, in supply-constrained economies, this represents a narrow view of the price mechanism since prices may be quite responsive to changes in supply conditions as well. In addition, because of the reliance of these economies on imports, import prices have to be taken into account. It may prove crucial to distinguish between imports costs related to cost of raw materials and other inputs from the competing prices of final goods.

Two types of analytical frameworks, namely, the monetary models of inflation and the wage price mechanisms have dominated advancements in macroeconomic price theory. Monetary models stress the role of monetary variables in the generation of price instability. In these models, increases in money supply are transmitted into the real sector via changes in the price level. On the other hand, the wage-price mechanism moves along with the Keynesian view by emphasising the role of excess demand in changing prices. Lately, structural models of the inflation have also taken their stand, relating inflationary pressures to limitations of the market mechanism.

In another view, price changes can be traced back to both supply and demand shocks that originate at home and from abroad. This brought forth the distinction between demand-pull and cost-push theories. Either a monetary or fiscal stimulus or a change in private spending behaviour that causes aggregate demand to increase can initiate demand-pull inflation. The monetarist version of demand-pull theories emphasised the causative role of monetary changes while the Keynesian view stressed non-monetary impulses. In contrast, cost-push inflation is initiated by supply factors. Supply shocks take a number of forms. A reduction in harvest and therefore output can lead to a rise in prices for

agricultural based economies in particular. Similar increases in wages can be interpreted as a supply shock as it in effect reflects a rise in the costs of production. Wage-push from unions as well as profit-push generated by administered pricing constitute cost-push factors. On the other hand, an increase in the price of oil is a famous example of an external supply shock that influence price changes especially in oil importing economies. Within the monetarist doctrines, the distinction between cost-push and demand-pull is spurious because an increase in wages or profits can only raise the level of prices and not their rate of change unless accompanied by a faster monetary growth. Thus, any sustained inflation, whether Keynesian fiscal induced money-accommodated inflation or whether quantity theory money initiated inflation, amounts to the same thing in monetarist principles.

Traditional Keynesian analysis puts more emphasis on quantity changes and the multiplier mechanisms within the context of closed economies and does not dwell much on price and wage formation in open economies. One result of this has been an underdevelopment of the mechanisms that explain international price linkages. Later improvements of the Keynesian model such as the Mundell-Fleming model, the Dornbusch sticky price model, the Solow-Swan model and others incorporated international relations but emphasized monetary transmission mechanisms and stabilization policy in the context of exogenous wages and domestic prices. These models still did not explicitly and adequately address the question of price and wage determination.

The development of the Phillips curve relation in 1958 introduced a formalised way of examining the relationship between inflation and unemployment having been explored as early as 1926 by Irvin Fischer (Santomero and Seater 1978:500). According to the Phillips curve relation, as in other markets, excess demand in the labour market indicated by the (un)employment rate is the source of wage inflation. However, the implied trade-off between unemployment and inflation faced serious criticisms founded in both its theoretical and methodological procedures. The most detrimental flaw of the Phillips curve arose from its lack of stability, which made it difficult for it to produce any

conclusive results (Frisch 1983:30). Later insights into the relation were introduced by Friedman (1968) and Phelps (1968) with the notion of the natural rate of unemployment. Attempts to explain changes in the closed-economy scenario centred on what came to be known as the expectations augmented Phillips curve. This body of theory explains variations in inflation and unemployment as caused by inflationary expectations. The central message of the augmented Phillips curve analysis is that wages and prices are rigid and tend to adjust slowly to shocks because of the existence of long-term contracts and costs of wage and price changes.

The case for international price linkages in open economies is usually tackled with the classical assumptions of perfect price and wage flexibility and full employment in contrast to the Keynesian sticky price models. Like in the case of closed economy models, the wage and price formation mechanisms are embodied in the sphere of monetary factors. These models have tended to sacrifice the analysis of inflation-unemployment trade-offs at the expense of models that elaborate the monetary mechanism. A typical example of these is the Hume's price specie flow mechanism. In these models long-run equilibria are the primary focus while the short- to medium-run analysis of wage-price-output interactions, which are essential for policy designs come secondary.

3.6.5.1 Models of inflation based on the Phillips curve

Macro models of inflation

According to the early works on the wage price mechanism, prices are mark-ups over unit costs at standard rates of output and capacity utilization and did not respond to demand movements (Blanchard 1990:785) it was believed that prices only played a passive role and that their adjustment to changes in nominal money only reflected wage increases. The relation between prices and nominal money was considered quite stable such that the direct effect of demand on prices was negligible. A major shortcoming of this approach has been cited as its lack of theoretical foundations and in the view of

Nordhaus (1972), although mark-up pricing was designed to address non-competitive markets, it became optimal only under conditions of perfect competition and constant returns to scale. According to Gordon (1976), prices should not be set simply as a mark-up over labour costs but as a weighted average of all costs of production.

Though attempts were made to provide theoretical bases for the Phillips curve relation by, among others, Lipsey (1960) and Samuelson and Solow (1960), their analysis was later perceived to be inadequate in explaining changes in inflation.[36] Constructive theoretical developments, especially with regard to measurement issues subsequently emerged.[37] Santamero and Seater (1978:513) suggest a number of measures of inflationary pressures. A number of variables are suggested as measures of unemployment.[38] Of these, the one that received most recognition is that of the concept of the natural rate of unemployment by Friedman (1968) and Phelps (1967). With regard to the rate of change of the price level, others such as France (1962:67) vaguely hinted the expectations hypothesis. (see also Archibald 1965:5 and Archibald *et al*. 1974:7). Arguments for and against productivity as an explanatory variable have been advanced in Kuh (1967:118) and Vanderkamp (1972:220), among others. Profit is also considered in many works on the grounds that unions may be more hostile and firms more relaxed when profits are high. Early woks that consider profit as an explanatory variable include Ecksterin and Wilson (1962), Bhatia (1962), Schultze and Tryon (1965) and Hamermesh (1970). The hypothesis that union power may cause an increase in the wage rate has been tested using varying measures by, among others, Hines (1964, 1968 and 1971), Throop (1968), Thomas (1974), Ashenfelter and Johnson (1972), Schmidt and Strauss (1976), Tobin (1972), Ashenfelter and Johnson (1969), Godfrey (1971) and Ashenfelter *et al*. (1972).[39] The influence of socio-political factors on wage inflation has also been explored

[36] A salient outcome of Lipsey (1960) is that wage inflation originates from unemployment represented by excess demand in the labour market, which in turn is derived from excess demand in the goods market. The Samuelson and Solow (1960) analogy implies that the rate of inflation is determined by demand pressure in the labour market, the expected rate of inflation and the proportion of growth in labour productivity that is not transferred to the workers in the form of money wages (see for example, Klein, 1978).
[37] See Santamero and Seater (1978) for a critique of the inflation-unemployment trade-off.
[38] For these see Lipsey (1960:127), Dow and Dicks-Mireaux (1958:50), Hansen (1970:97), Simler and Tella (1968:197), Taylor (1970:202), Mackay and Hart (1974:137), Gordon (1977:83).
[39] See also Johnston (1972), Pencavel (1970) and Zis (1974).

by many studies such as Beveridge (1945), Turvey (1951), Pitchford (1957, 1961 and 1968) and Tobin (1972).

The expectations augmented Phillips curve

The contributions of Friedman (1968) and Phelps (1967) can be summarised by the following relation

$$\frac{D_w}{w} = f(U) + \left(\frac{D_p}{p}\right)^* \tag{3.10}$$

Where $\frac{D_w}{w}$ is the rate of wage inflation, U is the natural rate of unemployment and $\left(\frac{D_p}{p}\right)^*$ is the expected rate of inflation.

By this relation, a rise in inflation triggers revisions of inflationary expectations causing the Phillips curve to shift and unemployment to return to its natural rate and introducing the concepts of the short-run and long-run Phillips curves. Thus, if expectations are allowed to vary, the long-run Phillips curve turns out to be vertical at the natural rate of unemployment, which is consistent with any rate of inflation that is anticipated. In this framework, expectations are assumed to adjust through adaptive behaviour. In practical applications, this version of the Phillips curve implies that lagged inflation rate, the level and the change in the level of excess demand in the labour market explain actual inflation.

The expectations augmented Phillips curve was challenged by among others, Lucas (1972, 1973) and Sargent and Wallace (1975), on the basis of the use of adaptive expectations. Arguing that expectations are formed by intelligent agents who take into account all available information, they introduced the notion of rational expectations in the inflationary process. In this variant, the natural rate hypothesis is fitted with rational

expectations and an excess demand relation that is derived from the Keynesian IS-LM system and Okun's law. The degree of excess demand in the labour market is a function of the rate of growth of real money supply so that if the rate of monetary expansion exceeds the rate of inflation, the positive real balance effects generates excess demand and reduces unemployment. Thus the rate of inflation is explained by the expected rate of inflation, the rate of growth of money supply and some random component. An important implication of this formulation is that the actual rate of unemployment tends to oscillate randomly around the natural rate. Thus, the Phillips curve is unstable even in the short run and deviations of unemployment from its natural level are attributed entirely to random factors. This has been termed 'the persistence dilemma' by Gordon (1981).

Micro models of the Phillips curve

In search of alternative explanations, attention has since been focused on the microeconomic foundations of inflation. Most notable of these theoretical advancements are the search and matching models of unemployment pioneered by Stigler (1962). The crux of these theories is the existence of imperfections such as incomplete information in the labour and commodity markets, so that allowance is made for some normal level of search unemployment. In these models workers and jobs are heterogeneous, so that they engage in a costly process of matching up idiosyncratic preferences, skills and needs, misperceptions and misinformation. On the other hand, the contribution of Phelps (1969) to this body of models stresses the firms' relative wage setting behaviour. Firms are viewed as prices setters and the desired differential between the firms' wage offer and the average market wage depends negatively on aggregate unemployment and positively on the aggregate vacancy rate relative to the firm's vacancies (Santamero and Seater 1978:518). In principle, three categories of search models of unemployment are identified. The first group of models has the basic feature that individuals speculate on the distribution of nominal wages over current vacancies. In the second group individuals misperceive their real wage while in the third category individuals speculate over time on the normal level of nominal wages. The first group consists of models of among others Alchain (1970), McCall (1970), Mortensen (1970), Lucas and Prescott (1974) and Siven

(1974).[40] Starting from the basic problem of maximising the present value of lifetime income, the basic ideology of these models is that an individual will change the market in which he offers his labour services if he perceives a change in his wage relative to the wages associated with current vacancies. This translates into choosing the optimal reservation or acceptance wage that is given by the wage that equates the marginal cost of searching to its marginal gains. The Phillips relation arises from these models by imposing the assumption that people's inflationary expectations are slow to adjust to changes in the actual aggregate rate of inflation. An important implication of these models is that unanticipated increases in inflation will reduce unemployment only temporarily. Another body of search theories stresses the notion that labour supply depends positively on the expected real wage. While the individual has full information regarding changes in the nominal wage, he misperceives the changes in the price level. This is the path followed by, among others, Friedman (1968), Almonacid (1971), Lucas (1973, 1975) and Sargent (1973). In these models individuals maximise income, and supply labour, but do not engage in search activities. As in the first category of models, people's perceptions of inflation are slow to adjust to changes in the actual rate. In addition their perception of the increase in prices is slower than in wages. In this way, they continuously have a mistaken belief that real wages have increased. The basic outcome is that while the short-run Phillips curve is negatively sloped, there is no trade-off in the long run.

In the third category of models due to Lucas and Rapping (1970), Lucas (1972) and Weiss (1972), the individual maximises lifetime utility and has to decide on the amount of time spent on work and leisure. The intertemporal labour supply depends on the time path of nominal wages. The current labour supply depends on the current and expected future real wages, the rate of interest and the current real assets. The Phillips curve is derived from these models by assuming that the substitution effect of a change in the perceived real interest rate outweighs the income effect, the perceived real interest rate is inversely related to the perceived inflation rate and that the perceived rate of inflation is inversely related to the current price level (Santamero and Seater 1978:520). The general

[40] See also Gronau (1971), Parsons (1973) and Salop (1973).

outcome in this case is a negatively sloped short-run Phillips curve and a positively sloped long-run Phillips curve. Criticisms of the search theories have brought into consideration resignation and layoff decisions and have led to major debates and yet another body of literature.[41]

Non-competitive models of inflation and unemployment

A shift from search models introduced the view that unemployment is a direct result of market failure. Among others, Hansen (1970) and Grossman (1974) share this view. In line with these developments, Friedman (1977) introduces a positively sloped variant of the Phillips curve characteristic of a transitional period in which government intervention, the existence of contracts, and uncertainty and rigidities in the labour market are dominant. In this case, markets do not clear instantaneously because of structural imbalances and frictional forces. It is assumed that an increase in aggregate demand will reduce unemployment in all markets, producing a standard Phillips curve. An expected change in the wage rate produces a wedge between actual and market clearing wages so that it makes sense to include both wage expectations and excess demand in the determination of nominal wage adjustment. Thus a cycle in aggregate demand can produce either clockwise or anti-clockwise loops. In this framework the tendency for clockwise looping is greater, the larger the adjustment parameter for price expectations and employment. Ultimately there is no trade-off between inflation and unemployment even in the long run. Other models that yield similar results are that of Ross and Wachter (1973) and Brunner and Meltzer (1976).

[41] See for example Tobin (1972:210), Gordon (1976: 205), Santamero and Seater (1978:523), Grossman (1973:85), Lucas (1975), Barro and Grossman (1971, 1976), Okun (1975), Barro (1977:23), Azariadis (1975), Baily (1974) and Gordon (1974, 1976).

3.6.5.2 Monetary models of inflation

Closed economy monetary models

The explanation of inflation has also played a pivotal role in monetary models. The basis of these models is that inflation is entirely a monetary phenomenon. The rate of growth of money supply solely explains the rate of inflation. These models are almost invariably founded on four basic propositions. Firstly, it is assumed that the economic system and markets are inherently stable so that labour markets in particular tend to equilibrate. Secondly, these models invoke the long-run quantity theory and perfect foresight so that the rate of growth of money supply is translated to the rate of growth of prices. Thirdly, the short-run dynamics are such that the growth of money induces only a temporary change in real output that lasts only during the adjustment process. Lastly, demand management policy, whether monetary or fiscal, is considered to be a source of instability in the economy. As in the wage-price mechanisms, subdivisions according to ideologies have developed in the monetarist theories. There is a group that believes in the distinction between the short-run and the long-run Phillips curve, commonly known and monetarist mark I, and the rational expectations school, known as monetarist mark II.

The former advocates for changes in real variables at least during the adjustment period. In this model, the rate of growth of money supply determines the rate of inflation with the Phillips curve as a nexus between the monetary and real sectors and facilitation of the adjustment process. A presentation of this model takes the following form.

$m = x + \pi$	-	The quantity theory	(3.11)
$\pi = \pi^{*} - b(u - u^{*})$	-	The Phillips curve	(3.12)
$u - u_{-1} = -a(x - x^{*})$	-	Okun's law	(3.13)

where m represents the rate of growth of money supply, x is the rate of growth of output and π is the inflation rate, π^* is the inflationary expectations, u is the unemployment rate, u^* is natural rate of unemployment. In this model there are no inflationary expectations in

the short run and hence real variables are affected. In addition, the Phillips curve remains stable and short-run equilibrium prevails. Thus, a monetary impulse leads to a real as well as an inflationary effect. However, in the long run expectations shift according to adaptive behaviour so that the Phillips curve and Okun's law are adjusted so that the real effects vanish. It is noteworthy that the long-run adjustment process in this model is subject to under- or overshooting because of system oscillations.

In contrast, the rational expectations school assumes a continuous sequence of long-run equilibria. A distinction is made between systematic and unanticipated changes in money supply. The former triggers a change in expected and actual inflation and remains neutral even in the short-run. It is only the latter which cannot be influenced by systematic economic policy that can produce sustained changes in real output and employment. The model takes the following form.

$$m = \pi + x^* + g + e^1 \quad - \quad \text{The quantity theory} \quad (3.14)$$

$$\pi = \pi^* - b(u - u^*) + e^2 \quad - \quad \text{The Phillips curve} \quad (3.15)$$

$$u = u^* - ag + e^3 \quad - \quad \text{Okun's law} \quad (3.16)$$

The major contention of this model is that an acceleration of real growth pushes the rate of unemployment below its natural rate. In addition, the generation of monetary impulses follows a particular rule given as follows:

$$m = \mu_0 + \mu_1 g_{-1} + \mu_2 \pi_{-1} + e \quad (3.17)$$

By this rule, the rate of monetary growth depends on an autonomous component and the deviation of its growth from its natural rate and the rate of inflation both lagged by one period. In addition, it is assumed that agents know this systematic component of money supply growth and take it into account in forming expectations. This then implies that the actual rate of inflation will be explained by the rationally expected component and some random component given by the deviation of the actual rate of growth from its expected

rate. This implies that the rates of unemployment and growth of real output will depend on the expectational error. Thus, the real sector is totally independent of anticipated monetary shocks.[42]

In an attempt to divide monetary shocks into inflation and output effects, Friedman (1970) identified both over-utilisation of capacity and an unanticipated acceleration of the rate of real growth as determinants of inflation. The incorporation of Okun's law and the Phillips curve then suggest that inflationary expectations, a surprise acceleration of real growth and the level of excess demand explain inflation. This can be contrasted to the traditional quantity theory in which the money market plays a significant role in the determination of price and the Keynesian system in which the price level is determined exogenously. Friedman (1971) later developed an intermediary between these two with the rate of interest assumed to be constant. In this framework, the price level increases proportionately with money supply. As in the Keynesian system, fiscal policy plays an important role while monetary policy is completely impotent. A dynamic implication of this framework is that unanticipated changes in the rate of growth of money supply produce a deviation in the rate of growth of nominal income from its trend growth. The rate of inflation rises only in the long run. When a higher rate of inflation is anticipated, the system reaches a new steady state characterised by a higher money supply with an unchanged real growth. This has been interpreted to represent a long-run theory of inflation.

The inflation tax model is one of the important mechanisms of the inflation process in developing countries (Blejer and Cheasty 1988:869). The existence of fiscal deficits, regardless of the method by which they are financed, are viewed as a major driver of growth of money supply. This mechanism is restricted to cases in which deficits are financed by money creation. Because of the link between fiscal deficits and the rate of domestic credit expansion, a distinction between monetary and fiscal policies becomes

[42] This result has triggered criticism based on the implicit assumption that markets clear. (see for example, Tobin 1980; Buiter 1980; Friedman 1981 and Gordon 1981). Models that eliminated this assumption show the potency of discretionary policy. These make use of assumptions such as multiperiod contracts (Fischer 1977), staggered wage setting (Taylor 1979), predetermination of prices (Phelps and Taylor 1977) and slow wage and price adjustments (Buiter 1980).

vague, especially if the exchange rate is fixed. Thus, changes in money supply are by definition equivalent to changes in credit to government and private sector and international reserves (Khan and Knight 1982:715). The basic assumption of the model is that the maintenance of deficits requires the generation of inflation tax, which in turn requires continuous increases in money supply leading to inflation (Aghevli and Khan 1978:384). Thus, a reduction in fiscal deficits is expected to be deflationary.[43]

Open economy monetary models of inflation

The question of how inflationary impulses are transmitted from one economy to another is an important one for open economies. Apart from the traditional routes in which increased foreign demand and increased input costs reach the domestic price level through the effect of higher exports on aggregate demand and directly through the Keynesian multiplier, as well as through the inclusion of import prices in the aggregate mark-up price equation, open economy inflation has been analysed by means of two predominant frameworks, namely, the monetary approach to the balance of payments, and the Aukrust- Edgren, Faxen and Odhner (EFO) approach. The crux of the former model is that changes in the supply of domestic money do not only influence the domestic rate of inflation but the overall balance of payments as well. Thus the money and capital markets play a major role in the transmission of inflation across national borders and a disequilibrium in the balance of payments is a reflection of a imbalance in the money market. This model rests, to a large extent, on the assumption that there exists a stable and well-defined demand for money function. The monetary approach introduced two additional assumptions. The first is the assumption of purchasing power parity, that all goods are tradable with prices quoted in international markets. Secondly, the domestic holdings of foreign reserves are allowed to increase as a direct result of the export surplus and also because the higher price level raises the demand for money relative to the original supply. In practice, the rate of interest plays a crucial role here. High rates of interest in developed countries lower the rate of inflation there, while they also attract

[43] It is noteworthy however, that feedback relationships may complicate the determination of optimal fiscal policy in situations of stable prices and high inflation. In particular, the size of the deficit itself may be a function of the inflation rate (See Crockett 1981 and Khan and Knight 1981).

foreign capital, thereby increasing the exchange rate. This, in turn, implies that imports from developed countries to developing countries are more expensive and therefore directly influence the increase in prices in the latter. Thus, the transmission runs from high interest rates, increased demand for foreign currencies, high exchange rates, and increased cost of production and ultimately increased domestic prices. In another dimension, high interest rates in developed countries trigger interest rate hikes in international money and capital markets. This raises the cost of credit, resulting in a fall in output and an increase in prices worldwide. In addition, the increase in financial liquidity internationally contributes directly to demand-pull inflation.

Another variant of the model starts from a consolidated balance sheet of the banking system and contends that a change in foreign exchange reserves is given by the difference between that change in the existing money supply and the change in the volume of domestic credit (Johnson 1972 and 1976). In turn, this difference gives the overall balance of payments. The model assumes a small open economy, producing one internationally traded commodity, a fixed exchange rate and hence equality between domestic and world inflation. Instantaneous adjustment is assumed in the money markets because of open international markets of commodities and securities. Thus, the central monetary authorities do not have control over money supply. For example, excess money supply is translated instantaneously into a fall in reserves and a deterioration of the balance of payments. It is only the composition of money supply that changes. Given that money demand depends on the price level, the level of output, and for simplicity, a constant rate of interest, the monetary approach comes to a conclusion that an increase in foreign reserves or an improvement in the balance of payments depends positively on the rate of growth of domestic real income and world inflation. Thus, for a given rate of world inflation, real growth in the domestic economy leads to accumulation of foreign exchange reserves or balance of payments surplus if a restrictive monetary policy is pursued. The monetary authorities can only control inflation by controlling the balance of payments or foreign reserves.

This is in contrast to the Keynesian model in which a growth of real income in a small open economy would lead to an increase in imports and a balance of payments deficit, so that high levels of employment come at the expense of external stability. This is also in contrast with a closed economy version in which the monetary authorities can control money supply and the rate of inflation. The adjustment process is taken to be instantaneous and incorporates the Alexander (1952) absorption mechanism.[44] If real cash balances increase because of an expansion in domestic credit at a given price level, domestic absorption increases relative to output. This induces a deficit in the balance of payments and a reduction in foreign exchange reserves, money supply and subsequently real absorption. This process depends to a large extent on a positive relationship between absorption and real cash balances and a positive relationship between the balance of payments and a change in the nominal money supply.

Another conduit of international inflation is provided by the system of floating exchange rates. Changes in the exchange rate tend to determine the pattern of real economic activity domestically and abroad. In principle, the stronger the currency of a country, the lower the inflation rate. The importance of the exchange rate in the determination of inflation depends to a large extent on the share of foreign trade in the country's economy. The greater the share of foreign trade in economic activity, the higher the stake that exchange rates take in the determination of inflation. A depreciation results in higher costs of inputs. This causes the costs of final goods to increase and leads to a mechanism of cost-push inflation. While domestic prices rise, so do prices of exported goods. This in turn leads to a devaluation undertaken as an attempt to increase competitiveness of exports in foreign markets. The costs of inputs then rise again and so does the domestic price level. In the context of developing countries, given that export and import demand elasticities are usually low, a devaluation hardly improves the balance of payments. The only sure outcome is accelerating inflation, chronic balance of payments problems and large amounts of foreign debt.

[44] See also Swoboda (1976 and 1977).

The view of Kydland and Prescott (1977) that lack of prior commitment in monetary policy can lead to inefficiently high inflation has given rise to a body a literature that also provides a linkage between domestic inflation and openness of the economy. In their view, the existence of imperfect competition and distortions produces sub-optimal levels of output which gives policy makers an incentive to create inflation in an attempt to affect real output through monetary policy. In essence, an inverse relationship between openness and inflation is predicted in the absence of precommitment. Subsequent works of among others, Barro and Gordon (1983), Taylor (1983), Rogoff (1985) and Fischer (1990) have suggested mechanisms by which policy makers can avoid the tendency towards inefficiently high inflation rates. In particular, Rogoff (1985) points out that an unanticipated monetary expansion results in a depreciation of the real exchange rate, which in turn reduces the incentives for expansion of output. Policy co-ordination in this case yields better results. If countries co-ordinate their monetary policies, they form a coalition of a single larger and less open economy that reduces the impact of depreciation and raises the equilibrium rate of inflation.

The basic analytical framework used in these models is that of a closed economy model of the dynamic inconsistency of optimal monetary policy (Romer 1993:872). It is assumed that unexpected monetary shocks affect both prices and real output. This implies that the difference between actual output from its equilibrium value is positively related to the difference of actual inflation from expected inflation as a result of either imperfect information about prices or incomplete adjustment of prices. The models also assume that policy makers prefer higher output levels to inflation. The sub-optimality of the levels of output arises from imperfectly competitive forces or distortionary taxes in the economy. The policy maker's problem involves choosing the optimal rate of money growth or equivalently the rate of inflation. This problem yields a sub-optimal outcome of positive inflation and output that is at a natural rate. Thus it would be desirable for the policy maker to precommit to a zero inflation policy.

However, the basic implications of this framework are that an increase in imports affects inflation in two ways. Firstly, the increase in the degree of openness reduces the benefits

of increased output. Domestic expansion increases domestic output relative to foreign output. This means that the relative price of domestic goods falls unless domestic and foreign goods are perfect substitutes. Secondly, while it increases output, the increase in the degree of openness also increases the rate of inflation. The reason for this is the real depreciation associated with an increase in domestic output. The depreciation affects inflation in two ways. Firstly, the impact of inflation is higher for imported commodities than for local commodities, and secondly, the costs of domestic firms are increased through the input costs. Thus, increased openness means that a monetary expansion will lead to a larger increase in domestic prices for a given increase in output. Because openness affects the main determinants of inflation, the incentives of policy makers to expand are eroded in more open economies so that equilibrium inflation is lower.

3.6.5.3 Structural models of inflation

The long-run tendency of inflation is often attributed to failure of the market mechanism resulting from a number of factors. These include differences in productivity and price and income elaticities in the industrial and services sector while growth of money wages in these sectors is usually uniform and there are rigidities of prices and wages particularly in a downward direction (Frisch 1983:153). In many structural models, the economy is sub-divided into the industrial or progressive sector and the service or conservative sector. Changes on relative prices in different sectors are linked to changes in the general price level (Canavese 1982:523). The Baumol (1976) model of unbalanced growth is usually considered as a prototype of structural models. In this and other models, the industrial sector is assumed to exhibit an increase in labour productivity while the service sector is characterised by constant labour productivity. Given that money wages grow at the same rate in both sectors, and that firms in the services sector use the mark-up pricing rule, the supply price in this sector rises relative to that of the industrial sector. This proposition assumes a small price elasticity and a large income elasticity of demand for the output of the services sector. The practical implication for inflation is that while prices in the industrial sector remain unchanged, prices in the services sector increase at the same rate as money wages because money wages rise at the same rate as labour

productivity. With time, the general price level rises at a rate proportional to the rate of increase in the price of the services sector.

A variant of this theory is the Hicks-Tobin theory of labour supply due to Tobin (1972) and Hicks (1974) whose core argument is the existence of contracts in the labour market. With similar basic assumptions this model assumes downward rigidity of money wages because of the contractual agreements between employers and workers. The basic implication is that the contracts eliminate or undermine the power of the market forces to enforce adjustments. Even if markets would respond, the principle of continuity, which depends to a large extent on the principle of fairness, leads to a lagged response of money wages to market forces. It is however possible that in the long run market forces and changes in productivity will determine wages. The question then is what factors determine real wages or relative money wages in the short run? In the view of Tobin (1972:3), it is only rational that workers judge the fairness of their compensations on the basis of wage differentials or relative wages. Thus, if real wages fall, the Tobin-Hicks hypothesis suggests that workers will attempt to defend their relative wage as opposed to their absolute wage (Frisch, 1983:162). This is in harmony with the Keynesian theory of labour supply (Keynes 1936:14), in which if money wages of one group are raised, other groups will strive for the same change in order to stabilise its real wage. This maintenance of wage differentials results in cost-push inflation.

The monetarist version of world inflation due to Johnson (1976), Swoboda (1977) and Classen (1976 and 1978) applies the classical version of the quantity theory. In its simplest form the model assumes there is only one world money. The world price level is then determined by the relationship between world real money supply and world real money demand. An increase in foreign money supply is met by an increase in foreign money demand and is translated into an increase in world price. Thus world inflation is endogenous. This model has raised concern over the monetary effects on output and employment, at least in the short run.

Dornbusch (1973) extended this model to allow for both traded and non-traded goods. The movement of goods, in particular raw materials such as oil, also plays a major role in the international transmission of inflation. This movement affects inflation through the changes in the terms of trade. If prices of exports grow more relative to import prices, a country essentially becomes a net exporter of inflation. In effect, the prices of domestic non-traded goods are flexible and the labour market clears instantaneously stimulating foreign demand. In the Dornbusch model, this in turn will require an increase in the nominal wage rate and a fall in the relative price of non-traded goods. The inflow of reserves increases the domestic money supply by a sufficient amount to finance a rise in the relative price of non-traded goods to the initial level. In the final analysis all nominal variables will have increased by the same proportion as the increase in the world price.

The Scandanavian model of inflation formulated by Edgren, Fax'en and Odhner (EFO) in 1973 and Aukrust (1977) supports the structural foundations of inflation and uses the same analogy as the Dornbusch model. In contrast to the above, the model is formulated in a manner that emphasises a small open economy hence the subdivision of the economy into the tradable and non-tradable sectors. Nevertheless, the basis principle still lies in the differences in productivity between the two sectors. With the exchange rate, labour productivity and money wages fixed, the basic prediction of the model is that while the domestic rate of inflation is given by weighted inflation rates of the two sectors, its deviation from world inflation is founded in structural factors. Thus, the rate of inflation is independent of excess demand. Instead, it depends on the difference between the rates of growth in labour productivity in the two sectors. In addition, the share of non-tradables in total expenditures determines the magnitude of the difference between domestic and world inflation. In other words, domestic inflation is determined solely by cost factors and results from imported inflation and productivity differences. As in the Dornbusch model, wage determination does not feature in this model so that differences in demand and supply of labour do not play any role in the adjustment process. This is in direct contrast to the Phillips-Lipsey model in which inflation is determined solely by demand factors and results from excess demand. Attempts to reconcile these contrasting views have been provided by, among others, Branson and Myhrman (1975). In this

interpretation, the Scandanavian model is considered as a long-run supply model with implicit demand conditions. By assuming that the rate of unemployment is exogenous, demand management policy is made to be strong in achieving a given level of unemployment, thereby implicitly choosing a given level of inflation with the help of the Phillips curve. In another reconciliation attempt by Aukrust (1977) and Calmfors (1977), the long-run Phillips relation becomes handy in determining the degree of excess demand compatible with the Aukrust-EFO model.

3.6.6 Theoretical models of the monetary sector

There exists a massive body of literature on the transmission of monetary impulses to the macro economy.[45]As in other fields, the debates on this issue are plagued with controversies and hence the varying schools of thoughts. While classical economic theory holds that in a framework of flexible prices, money should be neutral so that changes in nominal money is reflected only in nominal prices and not in output, strong linkages have since been identified in the literature between the monetary sector and other sectors of the economy. For example, some of the imperfect information models suggest that changes in the price level are triggered by changes in output, originating from changes in the monetary sector (Blanchard 1990:798).[46]

A common view that changes in money supply are translated into changes in aggregate demand dominated the perceptions of researchers until the 1970s. It was believed that changes in nominal money led to sustained changes in real money and in output. This in principle reflected the stickiness of wages and prices to changes in employment and output. Models of the transmission mechanism of money to aggregated demand and those of the wage-price mechanism provided the basis for analysis.[47] In general terms, the models suggested that changes in money lead to a slow change in prices and wages and therefore a sustained effect on output and real wages. However, the introduction of rational expectations later brought scepticism into the existence of a long-run Phillips

[45] See for example, Laidler (1999), Christiano *et al.* (1999), Fares and Srour (2001), Dib (2003) and Klaeffling (2003).
[46] See also Fry (1995).
[47] See, for example, Tobin (1972) for a generalization of the wage price mechanism.

curve on which the wage-price mechanism was based. Major theoretical shifts in the theory or money and economic growth include the elimination of the long-run trade-off and the introduction of exchange rates. These in turn introduced the price behaviour of economic sectors that are related to foreign competition and the distinction between producer and consumer prices (Blanchard 1990:787). Early evidence on these theoretical developments is found in among others, Friedman and Schwarts (1963), Sims (1972), Barro (1977), Mishkin (1983) and King and Plosser (1984).

Models of imperfect competition in which even changes in nominal money balances could affect output followed this body of theoretical advancements. Lucas (1972) developed macroeconomic models in line with this development. The core message of these models was that anticipated money changes had no effect on output and how unanticipated money could influence output under conditions of market clearing and imperfect information. However, as a policy implication, the models revealed that if the monetary authority had no more information than the public, the effect was unlikely to improve welfare.[48]

A third body of literature accommodates the aspect of imperfect information by incorporating rational expectations within the wage price mechanism. Fischer (1977) and Taylor (1980) developed models of this nature. Within Fischer's framework, demand for money shocks affect output only to the extent that they are unanticipated. In line with Keynesian analysis, given flexible prices and fixed nominal wages, demand shocks cause an increase in prices, a fall in wages and an increase in employment and output. In the context of Taylor's framework, wages are fixed at some nominal level for two periods. An increase in unanticipated money therefore has no effect on wages or prices in the current period implying a full-scale effect on output. However, with time wages and prices adjust so that output returns to the equilibrium level.

[48] See for example, Lucas (1973), Sargent (1973) and Sargent and Wallace (1975).

3.6.6.1 Models of perfect competition

Mishkin (1995:4) concurs with Tobin (1968) that while monetary policy is a powerful tool, it has the potential to produce unexpected and sometimes dreadful results. In another dimension, while others (for example Laidler 1999:9) believe that a transitory monetary shock cannot have lasting effects on the economy while a permanent shock will, depending on among others the nature of the monetary system, others believe that whether transitory or permanent, a monetary shock will affect the real economy. On the other hand, early evidence such as Friedman and Schwartz (1963) suggest that changes in money supply create movements in real variables.

While the Keynesian belief is that money is non-neutral in the economy, mainly because of the existence of wage and price rigidities, the classical belief is that money is neutral with changes in nominal money being reflected only in nominal prices and not in output and employment. Beginning from the development of the quantity theory of money by Fischer (1920) and the cash balance equation by the Cambridge school, the classical doctrine has maintained the neutrality of money. Within this framework, changes in money balances are reflected in changes in the price level and are regarded as a source of instability in the economy. This mechanism implicitly assumes perfect foresight.

Since the 1930s to the mid-1970s, economists shared a common view that changes in nominal money balances brought about changes in aggregate demand, thereby influencing real variables. The adjustment was believed to take place by way of the rate of interest. Given the Keynesian fixed price system, nominal wages and prices were slow to adjust to changes in employment and output so that changes in nominal money balances could be translated into changes in real money and ultimately output. The transmission from money to aggregate demand and the wage price mechanism dominated research at that time. The fixed price model of Barro and Grossman (1971, 1976) provides the basis for exploring the effectiveness of policy regimes in a non-Walrasian

setting.[49] Blanchard (1990:782) presents the Keynesian framework that supports the transmission from money to aggregate demand as follows.

$$y = a(m - p) \qquad a > 0 \tag{3.18}$$

$$y = b(w - p) \qquad b < 0 \tag{3.19}$$

$$w = w^{*} \tag{3.20}$$

where y is real output, m, w and p are real money, real wages and prices respectively. The first equation represents aggregate demand. Equation (3.19) represents aggregate supply while equation (3.20) represents fixed real wages. This framework is based on the assumption of a Walrasian *tatonnement* process that real wages provide an adjustment mechanism in the labour market. However, workers suffered from money illusion so that they observed nominal wages, which are rather sticky compared to prices (Gordon, 1976:188). For this reason, and because the process of adjustment would take some time, an increase in nominal balances would lead to a rise in prices, a reduction in real wages and ultimately an increase in employment and output.

3.6.6.2 Wage-price mechanisms

As already mentioned, according to the wage price mechanism, prices were regarded as mark-ups over unit costs and were not believed to be following movements in aggregate demand (Blanchard 1990:784; Gordon 1976:212).[50] Thus prices played a passive role in the adjustment process so that real wages changed instantaneously and fully. In this mechanism, the augmented Phillips curve played a major role in the determination of wages. Tobin (1972) and Blanchard (1990) present the wage price mechanism by the following system.

$$y = a(m - p) \tag{3.21}$$

$$p = w \tag{3.22}$$

[49] See also Benassy (1990).
[50] See also Tobin (1972) and Santomero and Seater (1978).

$$w - w_{-1} = b(p_{-1} - p_{-2}) + cy \qquad 0 < b < 1; c > 0 \qquad (3.23)$$

As in the previous system, the first equation represents aggregate demand. Equations (3.22) and (3.23) show the instantaneous pass through of prices into real wages and the wages equation that gives wages as a function of lagged prices and output as a proxy for unemployment respectively. This model shares a common view with the Keynesian system that aggregate demand determines the level of output, at least in the short-run. A rise in nominal money leads to an increase in output and aggregate demand initially, leading to a rise in prices. However, in this system there is no money illusion so that real wages are changed instantaneously. In the final analysis, output returns to its original level.

3.6.6.3 Models of imperfect competition

Imperfect information and expectations

Towards the mid-1970s the wage price mechanism faced a serious crisis based on weaknesses in its theoretical foundations. Phelps (1967) and Friedman (1968) advanced one of the major criticisms by introducing the natural rate hypothesis and the rational expectations. This implied that unexpected demand movements only and not systematic components, could explain unemployment. By this, the trade off between inflation and unemployment was very much doubted. This swerved further developments into two directions. One involves a move back to the classical based models of perfect competition and price flexibility but featuring imperfect information. The other takes the imperfect information route. The essence of this line of thinking is that with agents having limited information, there is potential that nominal monetary shocks could influence output. One of these models is that of Lucas (1972), which suggested that with imperfect information, money could influence real output. The development of wage price mechanisms with imperfect information has to a large extent followed the theoretical foundations of the search models of unemployment. Within a framework of imperfect competition, the

Lucas model maintains the assumption of perfect competition and is represented by the following system.

$$y = a(m - p) \tag{3.24}$$

$$p_i = p + e_i \qquad i = 1,...,n \tag{3.25}$$

$$y_i = b(p_i - E_i p) \tag{3.26}$$

Equation (3.24) is the aggregate demand while equation (3.25) gives the price that is faced by each firm p_i, which is different from the price level p, by the random component e_i. Equation (3.26) is the supply of each firm, which depends on price expectations $E_i p$. In this model the fact that there is imperfect competition implies that there is a component of money supply increase that is not anticipated. It is this component that leads to a change in output. Thus, firms misperceive monetary shocks for relative price shocks. A major shortcoming of this model is failure to explain the mechanism through which monetary shocks are translated into permanent changes in output. Further investigation by Lucas and Rapping (1969) suggested that while it is unlikely for perceived permanent changes in real wages to trigger changes in supply because of their conflicting income and substitution effect, temporary changes could have the potential to trigger larger responses because they depend only on substitution effects. Thus if firms realised a temporary increase in the price of output, they would increase output.[51]

Two channels through which misperceptions of monetary shocks could lead to permanent changes in output were in turn identified. Lucas (1975) pointed out that if misperceptions led agents to change variables that affected their decisions in the future, for example, capital, then the initial shock would have lasting effects.[52] The second channel identified by Lucas (1975) and Taylor (1975) reveals that if there still is a lack of information, *ex post*, large permanent shocks could be misperceived for changes in the relative price for a long time during which they will change output. The model developed by Weiss (1980) to assess the role of monetary policy under conditions of perfect and imperfect

[51] See also Hahn (1978), Hart (1982), Snower (1983) and Weitzman (1985).
[52] See also Howitt (1986), Blinder and Fischer (1981) and Sargent (1979).

information reaches similar conclusions. The model of Benassy (1987) also provides theoretical understanding of monetary policy under conditions of imperfect competition and unemployment and can be used to explain conditions of monetary neutrality and non-neutrality.

The general policy implication of these models is that while anticipated money cannot influence output, unanticipated monetary shocks could only affect output. This, in turn, implies that the role of monetary policy is reduced significantly as compared to its role under standard wage price mechanisms. Notwithstanding, a word of caution is provided that if monetary authorities had no more information than the public, the outcome would reduce the allocation of efficiency of the price system, as it would provide incorrect signals (Blanchard 1990:797).[53]

Research along the lines of imperfect information later took the route of real business cycle models. In many of these models, it is not the role of monetary shocks per se that takes the centre stage, but fiscal shocks and changes in technology. A salient feature however is the reverse causality between money and output. An exception in which money precedes output is when money is regarded as a factor or production.

Nominal rigidities and rational expectations

One of the models that introduced nominal rigidities and rational expectations within the context of the wage price mechanisms is that developed by Fischer (1977). The framework is represented by the following system.

$$y = (m - p) + u \tag{3.27}$$

$$y = -(w - p) \tag{3.28}$$

$$w = E(P|-1) \tag{3.29}$$

[53] Also see Lucas (1975), Sargent and Wallace (1975) and Sargent (1973).

Equation (3.27) is aggregate demand. Equation (3.28) is output supply and equation (3.29) indicates that nominal wages are preset at the beginning of the period on the basis of available information. Solving for real wages in this system gives the outcome that both demand and monetary shocks influence output only when they are unanticipated. The transmission channel is similar to that of the Keynesian models. Given that prices are flexible and money wages are rigid, an increase in demand leads to a rise in prices, a fall in real wages and a rise in output. In this model, policy makers need to have more information than the wage setters to be able to maintain stable output levels. In the staggered wage setting model in which wages are set for periods longer that the time between policy decisions, the models suggest that activist monetary policy can be used effectively to offset the effects of multi-period predetermination of wages.

In Taylor's (1980) model, wages are not only predetermined but are fixed to two consecutive periods. The model as presented in Taylor (1979) is as follows.

$$y = (m - p) \tag{3.30}$$

$$p = \frac{1}{2}(w - w_{-1}) \tag{3.31}$$

$$w = \frac{1}{2}[p + E(p(+1)|-1)] + \frac{1}{2}a[E(y|-1) + E(Y(+1)|-1)] \tag{3.32}$$

Equation (3.30) is aggregate demand, equation (3.31) gives the price level as a weighted average of the wages and equation (3.32) is the wage chosen in the current period for the current and next period. The solution of this model under rational expectations reveals that money has no effect on wages or prices in the current period so that it has a full effect on output. However, with time, nominal wages and prices adjust and output returns to its original path. In this model, the effects of money last for a longer period than in the previous model of Fischer. While the two models predict non-neutrality of money, it is important to note that this outcome requires the existence of real rigidities.

3.6.6.4 Standard monetary policy channels

A digression from the wage price mechanisms is the standard monetary policy channels. The mechanism of monetary impulses through the rate of interest represents one of the important elements of the Keynesian system. According to this mechanism, contractionary monetary policy leads to a rise in the rate of interest, which in turn raise the cost of capital so that investment spending falls. Falling investment is then transmitted into a fall in aggregate demand and output. Within this framework the interest rate effect has not only come to be associated with investment spending but also with durable consumer spending as well (Taylor 1995).[54] However, Laidler (1999:7) points out that the account of this mechanism is incomplete, given the complex interactions of open economies.[55] He emphasises the credit channel arguing that the willingness to buy or borrow following a fall in the rate of interest does not depend so much on the interest elasticity of demand for money as it depends on the interest elasticity of supply of indebtedness in the banking system.

The exchange rate represents one of the important conduits of monetary impulses in the advent of flexible exchange rates and intensified integration of economies (Taylor 1995; Obstfeld and Rogoff 1995). This channel is however dependent on the effect of the interest rate on the exchange rate. By this mechanism expansionary monetary policy leads to a fall in the rate of interest, which in turn leads to outflows of portfolio investment and a depreciation of the domestic currency. This makes domestic commodities less expensive than foreign commodities, hence a rise in net exports and output.

The effect of money supply on prices of other assets takes a rather universal approach to the assessment of the effects of relative prices of assets and real wealth. Two main channels are usually associated the effect of relative prices. These are the Tobin Q

[54] This proposition is however not supported by empirical evidence in many cases. In addition, evidence of the interest elasticity of investment is hard to come by, particularly in developing countries. See for example, Fry (1980), Molho (1986), Oshikoya (1992) and Hadjimichael and Ghura (1995).
[55] Also, this mechanism has been heavily criticized on the use of only one relative asset price, the rate of interest (see Tobin 1968:29).

investment channel and the wealth effects channel that operates through consumption spending. The former works itself though the valuation of assets. The argument is that a monetary expansion increases expenditures on investment through a higher demand for stocks. In turn, this increases the prices of equities. This increases the Q value of stocks, increasing investment and output. This channel closely resembles the Keynesian interest rate channel though it uses a narrower definition of assets. The consumption channel works via the relative asset prices through the life cycle hypothesis. By this mechanism, when stock prices rise, the value of financial wealth rises, increasing the lifetime resources of consumers. Thus, lifetime consumption rises. According to Meltzer (1995), this analysis can be extended to its effect on land and property values.

Bernanke and Gertler (1995) discuss two basic channels of monetary transmission namely, the bank lending channel and the balance sheet channel. The bank-lending channel works through the role of the banking institutions in the financial sector. According to this route, monetary expansion that increases bank reserves and deposits is expected to cause a rise in bank loans, investment expenditures and income.[56] On the other hand, the balance sheet channel operates through the net worth of firms. Several ways is which this channel works have been identified. One way is through the change in equity prices, which in turn leads to two different changes namely, changes in the behaviour of the firm and changes in its portfolio composition. Another view is through the interest rate channel. Starting with the former, an increase in stock prices originating from monetary expansion may lower adverse selection and moral hazard behaviour so that the net worth of the firm rises. This in turn raises investment and aggregate demand. Similarly, a rise in stock prices increases the value of financial assets, reducing the likelihood of financial distress. This in turn leads to a rise in expenditures on consumer durables and therefore income. The interest rate channel works such that a fall in the rate of interest following a monetary expansion improves the balance sheet of the firm, so cash flow rises. This in turn reduces adverse selection and moral hazard behaviour. In that way lending increases, and so does investment and output.

[56] This effect is however believed to have eroded considerably with developments in the financial sectors over the years (Edwards and Mishkin 1995; Meltzer 1995 and Bernanke and Gertler 1995).

3.7 CONCLUSION

The literature reviewed in this chapter has laid the foundations for the interrelationships that exist among the different sectors of the economy. The mechanisms by which the variables tend to affect each other provide a basis for modelling the individual sectors with the view of integrating them in a macroeconomic system. This review brings out clearly the importance of the government sector in the context of a macroeconomic model and how imbalances in that sector affects key macroeconomic variables. One of the important features that are highlighted by the review is the interdependence of the monetary sector, the prices sector and the labour market at a theoretical level. This review is used as a basis for the specification of the model in chapter five.