



CHAPTER 1

INTRODUCTION AND BACKGROUND

1.1 INTRODUCTION

During the two decades following the second world war, the objective of full employment was the first priority in the market economies and the Keynesian tenet set the (then) undisputed foundations for conducting economic policy. The nature of the unemployment problem was mostly cyclical and limited attention was devoted to the role and functioning of labour markets in the economy and specifically to their allocative, distributional and informational functions and capacities. Demand-management policies, coupled with relatively high economic growth rates, were deemed adequate to deal with unemployment. This, however, also meant an uninterrupted period of rising prices, as a result of which inflation came to be seen as a typical condition of the economy.

As a result of the inflationary pressure exerted by the international oil crisis at the time, unemployment once more became a serious problem during the 1970s. Unemployment in most countries became, to a greater or lesser degree, a non-cyclical phenomenon, i.e. unemployment rates tended to increase steadily despite periods of relatively high economic growth. Unemployment increasingly displayed structural characteristics and labour markets showed larger imbalances and rigidities. Rising unemployment, especially since the early 1980s, caused the labour market and its relationship with the rest of the economy to become the primary focus of policymakers and researchers.

This study shows that the South African labour market has lost its capacity to perform its allocative, distributional and informational functions efficiently and thus also became relatively inflexible in adapting to internal and external shocks to the economy. South Africa as a developing country is also experiencing a high and growing level of unemployment and since the late 1980s the capacity of the formal economy to provide sufficient employment opportunities for its growing labour force, has steadily been eroded to the present position where it is virtually non-existent. Apart from the immense economic cost to society, this structural problem also has a direct impact on the socio-economic conditions and political developments in the country.

Structural unemployment will only be permanently alleviated by addressing the structural impediments of the labour market and by accomplishing sustainable long-term economic growth and stability. Long-run economic growth will in turn be enhanced by supply-side policies in general and labour market policies in particular focusing on the improvement and quality of human capital. This notion is according to the growth theory pioneered by Paul Romer (Romer 1986: 1002-37), stating that knowledge, resultant from investment in human capital, is just as important a component of economic growth as raw labour and physical capital.

1.2 OBJECTIVE AND RESEARCH METHODOLOGY

Supply-side theory, policy and modelling have become imperative in economic analysis. This is due to the deficiencies of demand-oriented theory, policy and models to satisfactorily address unemployment and inflation. For many decades the Keynesian foundation for conducting economic policy was undisputed, but its failure to explain and solve the problems of stagnation, lagging productivity, double-digit inflation, high interest rates and depreciating currencies, led to the emergence of supply-side economics and the formulation of supply-side propositions

Macroeconomic modellers have consequently taken up the challenge to coincide with the supply-side notion of long-run growth in an economy with stable employment, productivity, wages and prices and thereby acknowledged the cost-minimising or profit-maximising decision-making processes of the firms which are responsible for production activities in the economy. For a macro-econometric model to be instrumental in policy analysis that goes beyond short-term forecasting requirements, distinct consideration must be given to its long-run equilibrium properties and stability with respect to output, employment and inflation. These, in turn, depend crucially on the consistency and structure of its supply-side specification.

The purpose of the study is to develop a theoretically consistent supply-side model of the South African economy. Ultimately, the objective is to integrate the supply-side model with a full-sector macroeconomic model of South Africa. The macroeconomic model developed by the econometric research team of the University of Pretoria (SAMEM) and which is primarily demand-driven, is used for this purpose.

The methodology of this study comprises:

- (i) A thorough investigation of the theoretical principles of supply-side economics and the developments in macroeconomic modelling to ensure compliance with the requirements for theoretical consistency, forecasting and policy analysis in particular. These results are used to evaluate and restructure the demand-driven SAMEM.
- (ii) The specification, estimation and validation of a neoclassical supply-side model of the South African economy, encompassing the recent, leading developments in the field such as cost-minimising behaviour, market imperfections and collective bargaining. This involves the derivation and estimation of single equations for production (actual and potential), capacity utilisation, fixed investment, corporate savings, demand for labour (skilled and unskilled), labour supply (total and skilled), wage rates (skilled and unskilled), unemployment (NAWRU) and prices (production and consumption prices).
- (iii) Finally, the system of supply-side equations is subjected to a series of policy-scenario simulations for the purpose of proposing an optimal set of policy measures that will resolve or at least alleviate the severe labour market inefficiencies and related unemployment problem of the South African economy



1.3 OUTLINE OF THE STUDY

Based on the above-mentioned methodology, the study is divided into three sections: (1) the theory and background of supply-side modelling; (2) estimation of a neoclassical supply-side model of the South African economy and (3) critical policy implications. The first section is covered in chapters 2 and 3. Chapters 4 to 9 deal with the estimation and validation of the model and chapter 10 presents the policy analysis and critical policy implications for South Africa.

In chapter 2 the theoretical principles of supply-side economics are explored, as they need to form the basis of a supply-side macro-econometric model of the South African economy. Supply-side effects and particularly the role of taxation, which is the main supply-side instrument, are modelled within a neoclassical framework of profit-maximising or cost-minimising behaviour of firms. These neoclassical principles are adopted in several, different ways by macro-economic modellers. The structural properties of the existing (mainstream) supply-side models are investigated to establish how they compare in modelling the neoclassical and associated supply-side principles.

Chapter 3 presents a brief outline of the current structure and shortcomings of SAMEM, along with a proposed restructured framework, incorporating a neoclassical supply-side model of the South African economy. The properties and objectives of the supply-side model are identified and an outline of the proposed methodology for the development of the supply-side model is presented.

The purpose of chapter 4, is to estimate an aggregate neoclassical production function for the South African economy as the key component of the supply-side model. The production function is used to derive a measure for capacity utilisation in the model, a property which a cost function does not exhibit. The resulting production function identifies the technology and properties of the South African production structure. Two aspects are considered in the estimation of a production function: (1) the functional form; and (2) whether to estimate the production function directly or to estimate a cost function and derive a consistent production function based on duality principles. The cost-function approach has the advantage of allowing for consistency between costs, prices and factor demands in a neoclassical framework.

Duality principles, however, according to which unique relationships exist between the coefficients of cost and production functions, only apply to restricted functional forms such as the Cobb-Douglas and the Constant Elasticity of Substitution (CES) functions. The Transcendental logarithmic (Translog) form on the other hand, for which duality principles do not apply, is more flexible and may therefore be considered preferable. As a result, two different approaches are followed: (1) the direct estimation of a Cobb-Douglas cost function, used for the derivation of a consistent production function based on Shephard's duality principles; and (2) the estimation of a non-homothetic Translog cost function, tested for the validity of imposed restrictions that will result in a Cobb-Douglas cost function. If the Translog cost function cannot reduce to a Cobb-Douglas cost function, an equivalent Translog production function has to be estimated by



imposing restrictions, similar to those validly imposed when estimating a cost function. This, however, will imply theoretical inconsistency in the derivation and estimation of the price and factor demand equations.

An aspect that deserves considerable attention in supply-side modelling, is the development of some measure for potential output. Measuring productive potential and the deviation between actual and potential output (i.e. the output gap), provides a number of key insights into macroeconomic performance. In macro-econometric context, capacity utilisation serves as a determinant of the behaviour of prices and wages and influences all key macroeconomic variables through a well-developed supply system.

However, modelling the output gap or capacity utilisation is a complicated matter for a number of reasons. First, different concepts of potential output have been proposed in the literature and are used in different models. Second, a wide variety of empirical methods are used to measure potential output, ranging from time-series and trend-type analyses to production function-based methodologies, with the precise answers sensitive to the method selected. Finally, actual output could be determined directly from Keynesian demand or from a production-function (supply) approach may be used. In chapter 5, (1) a brief review is presented of the different concepts and methodologies of potential output and output gaps; (2) two measures for potential output are identified and explained; and (3) the analytical framework, methodology and estimates for potential output and capacity utilisation for the South African economy are presented.

A model of aggregate fixed investment is proposed, derived and estimated in chapter 6. Investment plays an important role in the gross domestic product of the economy for a number of reasons. First, investment increases a country's productive capacity to the extent that investment outlays (plant and equipment) are long-lived, durable goods, increasing potential capacity output supply and to the extent that new investment goods embody the most recent technical advances. Second, investment expenditure induces shifts in the aggregate levels of employment and personal income by affecting the demand for capital goods. Third, the sensitivity of aggregate supply and demand to changes in investment is important empirically, since investment is a severely volatile component of the GDP. This volatility and the fact that investment movements have important consequences for productive capacity, employment demand, personal income and the balance of payments, make it important to understand the fundamental causes of variations in aggregate investment.

Gross domestic fixed investment in South Africa has deteriorated significantly with the imposition of economic sanctions and resulting disinvestment since 1985. The situation has not improved with the abolition of sanctions and the end of disinvestment in South Africa in 1994. Instead, the greater degree of openness of the economy serve to expose South Africa's vulnerability to international financial market instability, as was only too apparent during the East Asian and Russian crises of 1998.

The neoclassical (Jorgenson) approach is the most suitable in estimating a domestic fixed investment function as it has to be consistent with a supply-side model for the South African economy, incorporating all cost-minimising and profit-maximising decision-making processes by firms. A further advantage of the Jorgenson approach, is the fact that supply-side policy instruments such as taxes, interest rates and funding play an integral role.

However, based on earlier reasoning, it is necessary to model the significant role of financial constraints (internal and external) on investment in South Africa. An attempt is therefore made to extend the neoclassical specification by incorporating the financial constraints as specified by cash-flow models.

For the purpose of accommodating the principles of the cash-flow model, an aggregate financial constraint variable is constructed, incorporating both internal (domestic) and external (foreign) sources of funding. Domestic financial constraints consist, in accordance with the exposition of the national accounts, of savings by households, corporate enterprises and the government, as well as replacement investment or depreciation in real capital stock. External financial constraints consist of net foreign capital flow and the value of the change in gold and other foreign reserves. Assuming the behaviour of all role-players, except business corporations (firms) as exogenous, it is necessary to estimate an empirical equation for corporate savings – an important source of internal funding.

Gross domestic investment in South Africa is therefore modelled by a system of equations: a stochastic function for gross domestic fixed investment, identities for the real capital stock and aggregate financial constraints in nominal terms and a stochastic function for nominal corporate savings.

The purpose of chapter 7 is to develop a neoclassical labour model of the South African economy. The resulting wage and employment levels will influence economic activity through the supply side of the macroeconomic model. For empirical purposes, the South African labour market is divided into two parts: a skilled and an unskilled labour market. The reason for this is assumed differences in both the wage determination processes and firms' employment behaviour concerning differences in the levels of productivity and the role of labour unions applicable to skilled and unskilled workers.

Although an attempt is made to model the labour participants in the informal sector separately from the formal labour market activities, it is done without any contemporaneous feedback from the informal sector to the supply-side model in general and the formal labour market in particular. The reason for this is that informal activities are inadequately recorded in the production sector time series over the sample period under consideration. The contribution of the informal sector has been presented more comprehensively in the system of national accounting after 1995. The value of an equation for the informal labour activities, is that it gives an indication of the magnitude of the informal sector and the unexplored potential for the formal economy.



The modelling of wages and employment, essentially according to a systems approach to ensure consistency in a neoclassical framework, is based on the work of Layard and Nickell (1985, 1986; Nickell 1988). They use a framework of wage bargaining under imperfect competition, emphasising labour market interactions. Their approach also includes the role of labour unions and labour taxes on employers. This study deviates slightly from this approach by ultimately including a production function and not a cost function in the complete supply-side system, as this approach allows the derivation of an estimate for capacity utilisation – a key component in the price mechanism (structure) of the economy.

Although a production function is included in the model, it wasn't estimated directly, but derived from an estimated cost function for the South African economy on the basis of Shephard's duality principles. The direct estimation of a cost function and subsequent derivation of factor demand and price functions, ensure consistency with the profit-maximising or cost-minimising decision-making processes of firms.

In chapter 8 a pricing structure for the neoclassical supply-side model is developed. The estimated price-setting equations may also be used as tools to explain the high inflation levels the South African economy has been plagued with since 1970. For purposes of consistency, the Layard-Nickell approach of cost minimisation, utilised in chapter 7 to model wages and employment, is again employed.

The estimations for production, fixed investment, corporate savings, the labour model and the pricing system are combined into a neoclassical supply-side model of the South African economy in chapter 9. The system is closed by introducing a number of identities and definitions, linking every endogenous variable in the model and thereby ensuring a fully dynamic system. The model is subsequently evaluated along the full ideal principles of model selection.

First, the theoretical structure of the estimated model (empirical specification) is evaluated to determine if the model complies with (1) the *a priori* objectives of neoclassical supply-side modelling, (2) economic theory – for the model in general and the individual equations in particular, (3) rival models, i.e. the extent to which the model encompasses the characteristics of rival models and (4) policy analysis, i.e. the relevancy of the specified model for policy analysis.

Second, a simulation (i.e. *ex-post* forecast) of the full system of equations is conducted. These dynamic simulation properties of the model are evaluated in terms of (1) the statistical significance (*ex-post* forecast ability) and (2) the stability of the model over the simulated sample range. The statistical significance (goodness-of-fit) of the full system is measured in terms of simulation/forecast error statistics and confirmed by the graphical representations of the simulations.

Third, a series of dynamic, *ex-post* simulations are conducted by shocking every stochastic variable in the system. The simulation results are used to (1) determine the long-run (steady-state) multipliers and elasticities of the system and once again, (2) to evaluate the statistical



significance and sensitivity of the model in terms of the degree, speed and stability of convergence. The robust (stable) nature of the model serves as an indication of the forecasting ability of the model.

The level of unemployment, associated with a consistent level of output, can be explained in terms of the structural, long-run or supply-side properties of the economy. The essence of a neoclassical supply-side model is to capture and explain the underlying production structure of the economy, associated with consistent factor demand and price relationships. The resulting levels of production and employment are forthcoming from firms' decision-making processes which, in turn, are driven by profit-maximising or cost-minimising goals.

In chapter 10, (1) a brief description of the labour conditions and unemployment problem in South Africa is given; (2) a set of policy rules (proposals) which may increase the labour absorption capacity of the economy and subsequently reduce the unemployment problem is identified; and (3) the suggested policy measures are empirically validated through a series of dynamic simulations of the estimated supply-side model.



CHAPTER 2

A REVIEW OF SUPPLY-SIDE ECONOMICS

2.1 INTRODUCTION

Given the range and nature of uses of macroeconomic models, i.e. structural analysis, forecasting and policy analysis, it is necessary to recognise the important balance between statistical goodness-of-fit, structural simplicity and theoretically plausible behaviour. For a macro-econometric model to be useful for policy analysis that goes beyond short-term forecasting requirements, particular attention must be paid to its long-run equilibrium properties and stability with respect to output, employment and inflation. These, in turn, depend crucially on the consistency and form of its supply-side specification. At the same time, appropriate econometric methods are needed to ensure that short-term dynamic properties and underlying estimated properties are data consistent and well determined.

Against this background, the theoretical principles of supply-side economics are explored, as they need to form the basis in the development of a supply-side macro-econometric model of the South African economy. Supply-side effects and particularly the role of taxes, which are the main supply-side instruments, are modelled within a neoclassical framework of profit maximising or cost minimising behaviour of firms. These neoclassical principles are adopted in several, different ways by macro-economic modellers. The structural properties of the existing (mainstream) supply-side models are investigated to see how they compare in modelling the neoclassical and associated supply-side principles.

The purpose of this chapter is therefore three-fold:

- (i) to present an overview of supply-side economic theory;
- (ii) to discuss the modelling of supply-side effects in a neoclassical framework and
- (iii) to compare the structural properties of existing supply-side models.

2.2 THE THEORY OF SUPPLY-SIDE ECONOMICS

Although some economists believe supply-side economics to be a modern concept, its origin can be traced back to the classical doctrines of the 1800s and 1900s. Widespread unemployment and the worldwide depression of the 1930s led to demand management policies of the Keynesian Revolution, with its rejection of the classical assumptions. Keynesian analysis became the widely accepted foundation of economic policies, but its failure to address and explain problems such as stagnation, lagging productivity, double-digit inflation, high interest rates and depreciating currencies, led to the emergence of supply-side economics and the development of supply-side propositions.

This section therefore commences by discussing the classical roots of supply-side economics and contrasts it with the Keynesian approach. A discussion of the emergence of supply-side economics and the development of supply-side propositions then follows and the section is



concluded by addressing different concepts associated with the theoretical principles of supply-side economics.

2.2.1 The supply-side notion

Divergent ideas and controversy reign with regard to the basic concepts and theory of supply-side economics. It is therefore important to investigate the different aspects of what is actually meant by supply-side economics. According to Hailstones, supply-side economics can be defined as a study of policies designed to stimulate economic growth and promote price stability through various measures that affect the supply of goods and services. These measures include lower taxation, increased savings, greater investment and stronger work motivation (Hailstones 1982: 3).

Lyle E. Gramley, a member of the Board of Governors of the Federal Reserve System and former member of the USA President's Council of Economic Advisers, said the following about supply-side economics:

What do we mean by supply-side economics? Conceivably, a wide range of things could be included - energy policy, manpower training, Federal support for higher education, and other programs that might increase the growth of supply or enhance productivity (Gramley 1980: 2).

Professor Arthur B. Laffer from the University of California refers to supply-side economics as the

new, new economics of individual incentives (Business Week 1979: 116).

While Michael Evans who developed the well-known Chase Econometric model of the USA economy states that

Keynesian models cannot deal with current economic ills because they concentrate on questions of demand.

According to Evans, we need models that stress the supply-side, focusing on the stimulation of productivity (Evans 1980: 3).

Martin Feldstein, professor of Economics at Harvard and President of the National Bureau of Economic Research comments that *demand-side* economists emphasise federal budgetary and monetary policies as a means of manipulating demand for goods and services, thereby spurring on the production side of the economy. Supply-side economists emphasise the need for new tax incentives to encourage people to save and business to invest in new and more efficient factories and machinery (Golden 1980: 33).



In referring to supply-side economics, an International Monetary Fund (IMF) aide noted that the supply-side school stresses tax incentives for business and upper-bracket individuals in the hope of spurring capital investment to modernise ageing plants and equipment (Janssen 1980: 26).

According to **Norman B. Turé**, president of his own consultant firm and a pioneer in supply-side theory, this theory is based on principles developed by the classical economists:

Supply-side economics is shorthand for a way of analysing the effects of government policies and actions on the economy... its basic concepts predate Keynesian by a century and a half. What's new about "supply-side" economics is its application to public economic policy problems, particularly tax and fiscal policies (Raboy 1980: 18).

Professor **Robert Mundell** of Columbia University, an ardent supply-sider, believes that high tax rates have had such a large inflationary impact on the cost of many goods and services that people have been induced to forego their purchases and perform services for themselves that could be more efficiently performed by specialists. According to Mundell, this reduces productivity and the level of economic activity. This may not be true, however, if the work was performed in the spare time of the potential purchaser.

However, Mundell's point is verified to some extent by the rapid growth in the so-called subterranean or underground economy. This non-monetary, or sometimes called invisible economy, produces increasingly larger percentages of the GNP of many economic systems. Stressing that changes in tax rates are the heart of supply-side economics, **Dr. Paul Craig Roberts**, Associate Editor of the Wall Street Journal, states that:

The essence of supply-side economics is to regard tax-rate changes as relative price changes affecting the supply and form of labour, savings, investment and visible economic activity (Business Week 1980).

According to Roberts there are basically two important relative prices governing production on the supply-side. One determines the choice between additional current income and leisure; the other determines the choice between additional future income (investment) and current consumption. Both prices are affected by the marginal tax rates (Roberts 1982: 1-13). Roberts and some of his colleagues claim that higher after-tax incomes will encourage work through increased overtime, less absenteeism and shorter periods of unemployment. On the basis of their thinking, supply-siders opt for the following:

- (i) large and sustained personal and corporate tax cuts to induce more work and capital investment;
- (ii) keeping monetary growth in line with the long-run growth potential of the economy, including perhaps a return to some form of a gold standard and

- (iii) a slowdown in the growth of government spending and a lessening of the nations' tax burden relative to the GNP, thereby freeing more financial resources for private investment.

Hailstones and many others believe that although supply-side economics is thought of as a new concept, its origin can be traced back to the classical doctrines (economics) of the 1800s and 1900s. Widespread unemployment and worldwide depression of the 1930s led to demand management policies of the Keynesian revolution, with its rejection of the classical assumptions. Keynesian analysis became the widely accepted foundations of economic policies, but its failure to address and explain new problems and issues, especially stagnation, lagging productivity, double-digit inflation, volatile interest rates and a declining external value of the national currency, led to the emergence of supply-side economics and the development of supply-side propositions.

The classical roots of supply-side economics is investigated next, followed by a look at the basic principles of and differences found in the Keynesian approach, also known as the income-expenditure analysis. A discussion of the emergence of the supply-side economics and the development of supply-side propositions follows, after which the section is concluded by highlighting key concepts associated with the theoretical principles of supply-side economics.

2.2.2 The classical roots of the supply-side

From the classical doctrine which is based on the concepts of economic freedom, self interest, competition and *laissez faire*, one of the foremost and most relevant propositions for the supply-side theory, is the fact that savings provide the funding for growth in the economy. This is based on Say's Law, which in broad context, states that production of a given level of output generates sufficient income to purchase that amount of output, and that savings are directly or indirectly converted into investment. The Keynesians, of course, take a different view and state that aggregate demand may fall short of aggregate supply.

Say's law implies that supply and demand will always be equal and any amount of goods and services can be cleared from the market. Therefore, the economy will automatically move towards a full-employment level. Furthermore, the restriction of government spending to necessities or emergencies such as war is advocated. Most classical economists advocate a balanced budget and early repayment of debt incurred. These are the most distinctive propositions of the classical doctrine, which differ significantly from those of the Keynesian approach.

2.2.3 The Keynesian view

With the global depression and sustained unemployment of the 1930s, a change in economic thinking emerged, led by John Maynard Keynes. He departed from the conditions of Say's Law, presented a thorough critique of the classical doctrine and proposed a new set of macroeconomic principles regarding production, employment and income.

These principles and relationships of the various steps of the income-expenditure analysis can be summarised as follows:

- (i) Production (GNP), employment, income and prices depend on effective demand.
- (ii) Effective demand is measured by the total of consumption, investment and government spending.
- (iii) Consumption depends on the size of income and the average propensity to consume at that level of income.
- (iv) Since the consumption function is relatively stable, and assuming government spending has a neutral effect, changes in employment and income will result primarily from changes in investment.
- (v) Investment is determined by the marginal efficiency of capital relative to the interest rate.
- (vi) Marginal efficiency of capital is dependent on profit expectations compared to the cost of capital assets.
- (vii) The interest rate depends largely on liquidity preference compared to the quantity of money.
- (viii) Liquidity preference is dependent on the strength of the precautionary, transactions and speculative motives for holding money.
- (ix) Government spending is the total of national, provincial and local expenditures, which may be either or emergency spending.
- (x) If consumption, investment and regular government spending are insufficient to provide a high level of economic activity, emergency government spending may be used to raise the level of GNP, income and employment.

Ceteris paribus, an increase in the size of income will, for example, bring about higher consumption, increase effective demand and raise the level of production, employment and income. If the economy is in a state of full employment, however, such a change will result in higher prices. A strengthening of liquidity preference will raise the interest rate, decrease investment, lower effective demand and result in a decrease of production, employment and income. A lowering of the interest rate will have the opposite effect. It can be seen, further, that deficit government financing will raise effective demand, while a surplus budget will have a dampening effect on the economy unless used to combat inflation. All savings are not automatically converted into private, voluntary investment.

An aspect which also needs to be highlighted as it constitutes a significant divergence between the Keynesian approach on the one side and the classical and supply-side approaches on the other, is the fact that savings, according to the Keynesians, is considered to be a leakage and therefore, depending on the propensities to consume and save, an increase in savings will lead to a decline in economic activity and growth.

2.2.4 The emergence of the supply-side economics

Although the term supply-side economics had been known as early as the mid-1960s, the concept was widely used only by the late 1970s, when it received widespread attention and became



subject to more analytical scrutiny. Many different views and interpretations, and therefore also

macroeconomic problems, such as increasing inflation, lower productivity, increasing unemployment, lower potential growth rates and weakening exchange rates, was a shift in resource allocation from investment to consumption, both private and public. Fiscal and monetary policies tilted in the direction of subsidising consumption and penalising investment. When these policies lead to excess demand, monetary policy was invoked to reduce investment, causing a recession. This vicious cycle led to an ever-increasing rate of inflation. To an extent, this cycle has been fuelled by political considerations. Tax cuts for lower income individuals are easier to defend than tax cuts for businesses (Evans 1982: 253 and Hailstones 1982: 112).

These features are embedded in the large-scale econometric models, based on the principle that an increase in demand will automatically trickle down to increase aggregate supply, thus ensuring balanced non-inflationary growth. However, there is nothing magical about the balance between aggregate demand and supply. If incentives for investment are lacking, capital formation will stagnate. If incentives are lacking for employment, labour force participation will be reduced and productivity will diminish. As a result, the total productive capacity of an economy will grow more slowly than total demand, and bottlenecks, shortages and higher inflation must then be fought by causing a recession and reducing aggregate demand. It is true that the gap between aggregate demand and supply must be closed in order to diminish inflationary pressures. However, there are two ways to accomplish the same aim. One is indeed to diminish demand, thereby causing higher unemployment. The other is to increase aggregate supply, thereby raising the production possibility curve of the economy and increasing jobs and output at the same time that inflation is being lowered. This is the fundamental hypothesis underlying Evans's supply-side modelling (Evans 1982: 254).

Evans, as well as the other supply-side advocates, accuse Keynesian policies as being one-dimensional insofar as they concentrate on changing effective demand and ignoring supply. They also challenge the Keynesian tenet that a redistribution of income *via* taxation in favour of lower income groups will raise spending, output and employment because poor people spend a larger share of their income than rich people. They claim that this could cause a decline in savings and investment and have an adverse effect on production, employment and income. It is this latter possibility that is lacking in the Keynesian models (Evans 1982: 254 and Hailstones 1982: 112).

They also question the Keynesian tenet that spending stimulates demand while savings retard demand. Savings actually can be useful in providing funds for investment. Moreover, supply-siders claim that personal saving is affected by the after-tax return earned on savings. Of particular importance is the work of another supply-sider, Michael Boskin of Stanford University, whose findings reveal that a 10 percent reduction in taxes generates a 2 percent increase in personal saving. It is this saving that is not converted into investment (a part of effective demand) by the Keynesian models (Evans 1982: 254 and Hailstones 1982: 112).

Finally, the Keynesian tenet that government spending will result in a larger increase in demand and output than an equivalent reduction in taxes, is also challenged. The Keynesian reasoning states that the entire amount of government spending is reflected by additional demand. If taxes are cut, however, some of the tax remission may be channelled into saving, which do not

contribute to increased demand (consumption or investment). Thus, according to Keynesian policy, the effect of a tax cut on the economy is less than that resulting from an equivalent increase in government spending. This implies, also, that a simultaneous cut in taxes and government spending will lead to a decline in economic activity (Evans 1982: 254 and Hailstones 1982: 112).

Associated with this is the Keynesian belief that a personal income tax cut has a greater impact on the economy than an equivalent corporate tax cut because individuals spend a larger portion of their income than corporations (Evans 1980). The supply-siders do not agree, since savings are not lost, but converted into investment.

From the vast body of theoretical exposition, the principles of supply-side economics may be summarised as follows:

- (i) A reduction in tax rates increases the incentive of individuals to save by raising the rate of return on assets held by individuals. The higher saving leads to lower interest rates and higher investment (Evans 1982: 254; Evans 1980; Evans 1982: 108; Klein 1982: 247; Roberts 1982: 4; Roberts 1982: 49, 51; Keleher 1982: 265; Turé 1982: 35; Boskin 1982: 18,20 and Kemp 1982: 31).
- (ii) Corporate tax rate cuts or similar measures, such as increasing the investment tax credit or liberalising depreciation allowances, improve investment directly by increasing the average after-tax rate of return (Evans 1982: 254-255; Evans 1980; Evans 1982: 108; Klein 1982: 247; Roberts 1982: 4; Roberts 1982: 49; Keleher 1982: 266; Boskin 1982: 21 and Kemp 1982: 31).
- (iii) Increases in both personal and corporate savings lead to greater liquidity and less loan demand, thereby lowering interest rates. These effects help both capital spending and residential investment (Evans 1982: 255; Evans 1980; Roberts 1982: 49-50; Keleher 1982: 266; Turé 1982: 37, 42-47; Boskin 1982: 22 and Weidenbaum 1982: 10).
- (iv) Higher investment leads to an increase in productivity, which means that more goods and services can be produced per unit of input. As a result, unit costs do not rise as fast and inflation grows more slowly (Evans 1982: 255; Evans 1980; Klein 1982: 247, 249; Roberts 1982: 4, 7; Keleher 1982: 266; Turé 1982: 37, 38, 47, 49 and Boskin 1982: 22).
- (v) A reduction in personal income tax rates leads to a rise in labour force participation and work effort, increasing the quality and quantity of work and therefore the supply of labour necessary to produce more goods and services. It raises the overall growth rate in productivity and productive capacity, thereby contributing to the slowdown in the rate of inflation (Evans 1982: 255; Evans 1980; Evans 1982: 109, 189; Roberts 1982: 4; Keleher 1982: 265-256; Turé 1982: 37, 40-42, 48, 52-56 and Boskin 1982: 19-20). This proposition is, however, debated by Lawrence Klein. The sensitivity between tax changes and labour changes is based on the theoretical principles of the tax labour wedge model

(1982: 247, 250). Klein argues that according to the Wharton model, a reduction in indirect taxes lowers the real wage incentive (1982: 247, 249).

- (vi) Labour supply, capital stock and productivity are all increased by lower tax rates, thereby expanding the maximum productive capacity of the economy (Evans 1982: 255; Evans 1980; Keleher 1982: 267; Klein 1982: 247 and Turé 1982: 37).
- (vii) As a result of higher maximum capacity, the inflationary pressures of shortages and bottlenecks diminish, thereby reducing the rate of inflation (Evans 1982: 255; Evans 1980 and Turé 1982: 37, 38, 40).
- (viii) An increase in maximum capacity also permits the production of more goods and services for export markets as well as domestic consumption. This improves the net foreign balance and strengthens the exchange unit, thus leading to lower inflation because imported goods decline rather than increase in price (Evans 1982: 255; Evans 1980; Keleher 1982: 267 and Turé 1982: 39, 47, 56-60).
- (ix) Lower tax rates result in more modest demands for wage increases, since real income has risen by virtue of the tax cut and workers do not suffer a loss of real income by moving into higher tax brackets. This in turn reduces inflation further (Evans 1982: 255; Evans 1980; Evans 1982: 109; Klein 1982: 247; Roberts 1982: 49; Turé 1982: 40-42, 47-49, 52-56 and Boskin 1982: 19).
- (x) Lower tax rates therefore cause a reduction in inflation through several channels. Inflationary pressures decline as the gap between actual and maximum potential GNP rises; productivity increases, thereby lowering unit labour costs; the exchange unit (currency) strengthens, causing less imported inflation and wage rates rise more slowly (Evans 1982: 255; Evans 1980; Klein 1982: 247 and Turé 1982: 37, 39, 40-49).
- (xi) Lower inflation leads to an increase in real disposable income (bracket inflation is mitigated) and hence a rise in consumption, output and employment (Evans 1982: 255; Evans 1980 and Turé 1982: 46-57).
- (xii) Lower inflation leads to lower interest rates, stimulating investment in both plant and equipment and in housing (Evans 1982: 255; Evans 1980 and Turé 1982: 46-52).
- (xiii) The lower rate of inflation causes an increase in net exports, which strengthens the value of the monetary unit (currency). This leads to further reductions in the rate of inflation because imported goods decline rather than increase in price (Evans 1982: 255 and Evans 1980).
- (xiv) The increased demand for goods and services stemming from lower inflation is matched by the rise in the maximum potential capacity of the economy to produce these goods and services, thereby resulting in balanced, non-inflationary growth (Evans 1982: 255; Evans



1980, Evans 1982: 109; Keleher 1982: 265; Turé 1982: 36, 37, 49, 50; Boskin 1982: 22 and Weidenbaum 1982: 9).

The above theoretical principles are supported by the new developments in macro-econometric modelling by economists such as Stephen Nickell (1988: 202-221), Chris Allen, Stephen Hall and James Nixon (1994; Hall 1995: 974-988), John Helliwell (1995), Robert Coen and Bert Hickman (1995), Thomas Thomsen (1995) and others.

A main contention is that the above tax effects take place in the first instance on account of changes in relative prices and not as a result of changes in disposable income as the Keynesians suggest. According to Roberts (1982: 2, 49), Turé (1982: 36) and others, there are two important relative prices governing production. One price determines the choice between additional future income (investment) and current consumption and the other the choice between current income and leisure. Both prices are affected by the marginal tax rates. Boskin is of the opinion that supply-side economics is better described by the term incentive-oriented economics, as the basis of supply-side economics lies in the incentives to produce income and wealth. According to Keleher (1982: 264-276), the tax changes which are especially relevant to aggregate supply in the sense that they influence behaviour and incentives, are changes in marginal tax rates (the rate at which the additional increment of activity is taxed) and not tax revenues. Weidenbaum (1982: 9-15) stresses the fact that government legislation should not oppose these tax incentives.

Feldstein (1982: 146-157) stresses that supply-side economics focuses on capacity creation through capital formation and research. Supply-siders also recognise that saving is a prerequisite for increased capital formation that can raise productivity and the standard of living.

Concerning the role of fiscal and monetary policy, Kemp (1982: 27-39) and Feldstein (1982: 146-157) say that monetary policy should be used to stabilise the value of the currency and fiscal incentives to encourage individual and business production, saving and risk taking. Meiselman (1982), however, feels that the lags involved in the effect of changes in the stock of money and the problems caused by uncertainty about future changes in the stock of money, necessitate not only slow and stable money growth, but slow and stable money growth mandated by law.

In modelling a supply-side, Stephen Nickell (1988: 202-221) focuses on factors describing the behaviour of firms and determining wages. A clear distinction is made between supply and demand and also the role of surprises. According to his model, supporting the basic theoretical principles discussed above, a rise in demand will typically lead to a rise in inflation and to an increase in output and employment - causing firms to apply upward pressure on prices relative to wages and workers to apply upward pressure on wages relative to prices. The resultant real wage outcome depends on price setting, which in turn depends on the technology of the firm and competitiveness of the model. If strict normal cost pricing pertains, there is no effect from demand to prices and the rise in inflation will ensure that the price-wage mark-up falls. Real wages, as well as employment, will therefore rise as a consequence of a positive demand shock. If firms, however, are capital constrained, the demand shock will have no effect on output and employment and will typically reduce real wages through the inflationary pressure on prices.

The response of the model to supply shocks will originate from wage changes. A typical supply shock, such as a rise in the real price of imports, will exert upward pressure on wages relative to value-added prices. This pressure will induce a rise in both real wages and inflation and, even for given real demand, will produce a fall in output as each firm individually feels that it is becoming uncompetitive. The result is unemployment and stagflation. If there is no reduction in real demand, inflation will continue to rise for as long as the supply shock causes wage pressure. The specification of the wage equation is crucial in this regard. If real demand falls, inflation will be stabilised. Unemployment will increase still further and the effect on real wages will again be determined by the pricing behaviour of firms.

John F. Helliwell (1995) focuses on the role of measured factor inputs and technical progress contained in the underlying production functions, instead of merely focusing on the equations for prices and wages, supported by explanations of labour supply and demand. He argues that, in spite of the effects of globalisation, there are international differences in the rates of growth of aggregate efficiency and potential output and that these differences are related to national economic structures and policies. Supply-side modelling should recognise this endogeneity and make models appropriately responsive to national policies and institutions. However, these influences are in general slow moving, so that for the purposes of short-term forecasting, more traditional methods of forecasting the growth of potential output as some combination of the growth rates of factor inputs are not likely to be too misleading.

In developing a new version of the London Business School model, Chris Allen, Stephen Hall and James Nixon (1994; Hall 1995: 974-988) based this model around an aggregate restricted cost function. Non-energy production is modelled as a function of four factor inputs: capital, labour, fuels and raw materials. They also modelled technical innovation by the inclusion of a time trend.

In developing a stochastic model of potential or full-employment output, Robert Coen and Bert Hickman (1995) treat full-employment output as an endogenous variable. Their concept incorporates a piecewise exponential trend of technical progress, but it also allows for other factors affecting the path of full-employment productivity and output. In the deterministic context, productivity is affected by capital deepening induced by changes in the wage-rental ratio due to price shocks and changes in taxation, and it additionally reflects shocks to labour demand in the stochastic variant. The labour supply is also affected by real consumption wage shocks, and, in the stochastic variant, by shocks to the average hours and participation rate.

To the extent that the shocks to labour demand stem from the production technology, their stochastic full-employment path resembles that of real business cycle theory, except for the crucial difference that they do not assume perfect competition and continuous market clearing, and hence do allow for departures of actual GDP and factor inputs from full-employment levels.

2.2.6 Key concepts associated with supply-side economics

There are a few interesting, but still contested issues in the supply-side debate. The first is the Laffer effect, regarding the relationship between tax rates and tax revenues. It is based on the assumption that people will work more when their after-tax wages rise and that businesses will invest more than the increase in their after-tax profit or rate of return on investment. This, in turn, will result in greater tax revenue.

The Laffer effect depends on the size of the tax change, the point from which the tax change takes place, as well as the incentives people get to change their work effort, productivity and investment spending due to tax changes. The assumption that people will work more, leading to greater productivity, when their after-tax wages rise and that businesses will invest more than the increase in their after-tax profit or rate of return on investment, is questioned. The debate regarding the elasticity or sensitivity between changes in tax rates and the labour supply or the work-leisure ratio continues. Assuming that this relationship exists, Laffer and other supply-siders utilise the labour and capital wedge models to indicate that tax rates affect the quantity of labour and capital demanded as well. For example, a tax increase will increase the cost of hiring a worker and decrease the real wage received by the worker and therefore decrease both the demand and supply of labour. The tax wedge, which is the difference between the hiring cost and actual wage paid to the worker, consequently increases. The same applies in the case of capital.

Another issue is the effect of tax changes on saving. Although there is a lack of substantial empirical evidence, indications are that tax cuts increase saving and, according to recent studies, decrease interest rates.

Whereas the accepted, conventional doctrine holds that tax cuts have relatively small effects on the supply of savings, as well as on the supply of labour, consensus indicates that tax changes can significantly affect investment. The evidence suggests that tax cuts directed at investment may be the most potent area to stimulate aggregate supply *via* their expansionary effect on the capital stock.

The conventional view holds that tax cuts do, to some extent, increase the supply of labour, saving, investment, and hence, aggregate supply. However, in view of the conventional perception of the elasticities of various factor supplies (with respect to taxes), the conventional view holds that the effect of tax cuts on aggregate supply will not be very large. This view however has been disproved by supply-side models built by *inter alia* Laffer, Evans and Turé.

Supply-siders claim that their suggested policies and measures will increase productivity. They maintain further, as do most Keynesians, that increases in productivity will reduce inflation.

Some like Turé, consider supply-side economics as the application of price theory to government fiscal measures. An income tax reduction, for example, increases the price of leisure, raises the price of current consumption compared to future consumption, increases the value of market



work compared to self-work (underground economy) and increases the value of taxable investment compared to tax shelters.

In studying supply-side economics, one has to consider the effect of tax cuts on the supply of labour, on saving, investment, aggregate supply and tax revenues. After analysing the available empirical data, Robert Keleher (1982) concludes that:

- (i) a supply-side cut in income and business taxes will probably result in some increase in the supply of labour, saving, investment and aggregate supply;
- (ii) due to additional real economic growth, the tax base will increase and revenues will not fall in proportion to tax rates – in short, because of feedback effects, the deficit will not be as large as some predict;
- (iii) despite the increase in aggregate supply, tax cuts will produce an increase in budget deficit in the short run, but in the long run, the supply-side effects of tax cuts should be more potent and the deficit less of a concern.

The above theoretical principles form the basis of a supply-side macro-econometric model.

2.3 DEVELOPMENTS IN SUPPLY-SIDE MODELLING

Supply-side economics deal with issues such as the growth in production and supply, productivity, the role of technology, tax incentives for saving and investment, tax effects on employment and wages, relative prices governing these supply-side decisions and the tax revenue (or Laffer) effect. Keynesian economics disregards the incentive effects of tax changes on supply in the economy and ignores the fact that savings are converted into investment. According to the supply-side approach, however, fiscal policy first of all changes relative prices, incentives for leisure or work (income), consumption and investment. These changes have a definite effect on the production and aggregate supply of the economy. But, fiscal policy changes also produce resource reallocations with adverse demand-side implications for employment and the rate of economic growth. All these features need to be addressed when modelling a supply side for purposes of extended policy analysis.

2.3.1 Earlier approaches in supply-side modelling

Early supply-side econometric models are mostly variations or extensions of traditional macro-economic models based on Keynesian or demand-side economics, based on the income-expenditure approach with the level of output and employment principally determined by the level of demand. The following framework or equations represent the demand-side:

$$Y = C + I + G + \Delta IV + X - M \quad (2.1)$$

$$C = c(Y - HP) \quad (2.2)$$

$$I = i(Y, r) \quad (2.3)$$

$$\Delta IV = v(Y) \quad (2.4)$$

$$M = m(Y, ep; p^*) \quad (2.5)$$

$$X = x(Y^*, ep / p^*) \quad (2.6)$$

Equation 2.1 is the standard national income accounting identity, which sums the expenditure components of GDP (private consumption, fixed investment, government spending, changes in inventories, exports and imports). Equation 2.2 is the consumption equation, which explains personal consumption by real income (Y) and credit restrictions (HP). It is of a Keynesian form with consumers assumed to be constrained by real incomes and credit conditions. Equations 2.3 and 2.4 explain fixed investment and changes in inventories by simple accelerator models (changes in output) with some allowance for interest rate effects on fixed investment. Finally, equations 3.5 and 3.6 explain trade, exports and imports, by domestic and foreign income (y, y^*) and by competitiveness (ep / p^*) where e is the exchange rate measured as the foreign price of domestic currency, p the domestic price level and p^* the foreign price level. A rise in this term is associated with a worsening of competitiveness. Most of the early models did not have an endogenous explanation of exchange rates (e) so that most of the variation in competitiveness in model simulations came from changes in relative prices (p / p^*). Monetary influences were largely absent, their main impacts running through the role of credit conditions in personal consumption and a small interest rate influence on fixed investment.

The determination of aggregate supply was rudimentary. It came not from an explicit output supply relationship or constraint, but from increasing inflation through the mechanism of the Phillips-curve treatment of wages, implying a long-run trade-off between unemployment and inflation (Whitley 1994: 43):

$$\dot{w} = w(u, \dot{P}^e) \quad (2.7)$$

with price expectations (\dot{P}^e) given by past inflation:

$$\dot{P}^e = \dot{P}_{t-1} \quad (2.8)$$

and the price level in turn given by a constant mark-up on costs:

$$P = \beta(w - \pi) + (1 - \beta)P_m \quad (2.9)$$

where $(w - \pi)$ represents unit labour costs (wages less productivity) and P_m is the price of imports.

An important practical problem was that empirical estimates of the wage equation (2.7) proved unstable. It was particularly difficult to find a statistically significant and robust role for excess demand (unemployment) on wages. Unemployment itself was determined by a simple identity between employment and the labour force:

$$U = L - N \quad (2.10)$$

The labour force (L) was treated as predetermined apart from demographic trend, and employment (N) was based on an inverted production function where employment lagged behind output:

$$N = f(Y) \quad (2.11)$$

Given the fact that the main source of supply-side influence was through the wage equation, the policy implications which emerge are that demand policies could have powerful effects on output and employment with little role or need for supply-side policies (Whitley 1994: 44).

Supply-side economics, however, stresses the need to take into account the effects of economic policies on incentives to save, invest and work. Therefore, some of the macro models have been modified to take into consideration the effects of fiscal policies on savings, investment, labour supplied, productivity and inflation (Hailstones 1982: 103-104).

Two such models are the Wharton and Evans models. In a version of the Evans model (1980), for example, *the real after-tax interest rate* was added to factors such as disposable income and the price of non-durable consumer goods relative to the price of other consumer goods as a determinant of the consumption of non-durable goods. Thus, a higher after-tax interest rate yield resulting from a tax reduction would encourage saving and lessen consumption. In another equation in the model *the real after-tax wage rate* was added to lagged inflation and lagged unemployment as a determinant of the labour participation rate.

In traditional macro models, a tax cut was always followed by increasing demand and inflation, whereas the supply-side extensions endeavour to show tax cuts raising productivity, increasing supply and reducing inflation as well.

Thus, a cut in taxes would show a higher labour participation rate as the real after-tax wage increased.

Neither the Keynesian nor earlier supply-side models adequately explain stagflation, the simultaneous occurrence of unemployment and inflation. This is because both types of models treat inflation as a full employment phenomenon, which calls for a decrease in effective demand on the Keynesian income-output model (consumption function model) or an increase in the quantity supplied on a supply-side model. Closer inspection of the models, however, would indicate that if inflation occurs at less than full employment, a lowering of the supply price on the Keynesian aggregate demand-aggregate supply model would grant some relief. In a supply-side model an increase in supply (as opposed to an increase in the quantity supplied) will ease inflationary pressure.

Supply-side econometric models need to provide a framework for designing anti-inflationary measures that do not aggravate unemployment during periods of stagflation. In fact, it appears

that the answer during such periods is to find ways to reduce the cost of supply rather than increase supply. If the economy is suffering from stagflation, it implies that there is ample supply but insufficient demand to clear the market. Increasing the supply, without a reduction in cost may merely result in inventory accumulation. But, reducing the cost of supply would permit a reduction in price that would generate an increase in the quantity demanded.

2.3.2 Adapted neoclassical approach

Macroeconomic models came under heavy attack in the late 1970s. They had long been criticised for inadequate theoretical foundations, but began to exhibit serious forecasting failures. Their main failure was the inability to predict stagflation, the simultaneous occurrence of unemployment and inflation.

Changes in macroeconomic modelling came not only from the monetarist and rational expectations¹ challenges, but from the frequency of supply shocks during the 1970s, the need to relate to the new economic policies of the 1980s, and the desire for greater theoretical consistency.

The macroeconomic models started to take up the above-mentioned challenges in the 1980s. They utilised a neoclassical framework, albeit with several critical differences. The changes in the models were evolutionary, rather than revolutionary.

The neoclassical approach can be characterised as a combination of market clearing and rational expectations, and it emphasises the role of stocks rather than flows. The neoclassical approach stresses the supply-side of the economy, not through the inclusion of an explicit production function, but rather through a representation of the labour market which is responsive to changes in benefits and taxes, but not to the level of demand. The approach adopted by the mainstream models has not been to follow the competitive paradigm of the classical school; instead, development was centred around a framework of imperfect competition in goods and labour markets, and the adoption of a bargaining approach to wage determination, following the work of Layard and Nickell² (1985).

The general acceptance of this framework has meant that wage equations are now specified in terms of the level of real wages. Real wages are in principle affected by any factors that influence the bargaining strengths of employers and employees (such as trade unions), including supply-side variables such as tax rates, unemployment benefits and labour market mismatch. Here the need to model the *incentive* effects of the new macroeconomic policy is evident, although the tax coefficients are typically ill defined in empirical work.

An additional feature of the Layard-Nickell approach is the incorporation of demand pressure

¹ Based on the Lucas critique.

² See Nickell (1988) for a detailed exposition of the derivation and structural properties of neoclassical supply-side models.

variables in price-setting, although the specification remains largely cost based. Layard and Nickell emphasise that the long-run solution to the wage and price equations delivers the non-accelerating inflation rate of unemployment (NAIRU). Although both the true neoclassical and the Layard-Nickell approaches support the principle of no long-run trade-off between inflation and unemployment, the concept of the unemployment equilibrium differs. The natural rate of the neoclassical school relates to a competitive solution given the existence of market imperfections, whereas the Layard-Nickell (LN) framework and its NAIRU is based on a bargaining process under imperfect competition (Whitley 1994: 48).

The Layard-Nickell approach is generally accepted as the neoclassical framework for modelling supply-side behaviour and is adopted by many mainstream macroeconomic models, albeit with several structural differences.

The neoclassical approach views inflation as determined wholly by excess money, but the LN supply-side approach holds the view that inflation is generated by excess demand in goods and labour markets and by the inconsistency of wage claims by the unionised sector with the wage that employers are prepared to concede (the *affordable wage*). A necessary feature of the supply-side approach is not only that inflation is generated by additional demand, but that higher inflation itself reduces demand so that in the long-run output is supply determined. Many models have incorporated a measure for capacity utilisation to capture the inflationary effects of excess demand. Although investment expenditures are now more sensitive to inflation than before, often through liquidity effects, the key demand elements which are sensitive to inflation are net trade and consumption.

The developments in macroeconomic models can broadly be summarised by the following set of stylised equations:

$$\text{Income-expenditure identity: } Y = C + I + G + \Delta IV + X - M \quad (2.12)$$

$$\text{Consumption: } C = c(Y, r, W) \quad (2.13)$$

$$\text{Investment: } I = i(Y, w / ck) \quad (2.14)$$

$$\text{Changes in inventories: } \Delta IV = v(Y, cv) \quad (2.15)$$

$$\text{Imports: } M = m(Y, cu, ep / p^*) \quad (2.16)$$

$$\text{Exports: } X = x(Y^*, cu, ep / p^*) \quad (2.17)$$

The consumption equation now includes a role for interest rates (r) and real wealth (W). Investment is influenced by relative factor prices (w / ck) or simply the real cost of capital depending on the precise derivation used, and changes in inventories is influenced by the cost of inventory changes (cv). Exports and imports are additionally determined by capacity utilisation (cu). The implications are that aggregate demand is more interest elastic than before and monetary factors are more important through their direct influence on consumption, investment and inventory changes.

Aggregate supply, based on cost minimising behaviour in the LN-framework, can be described by

the following equations:

$$\text{Wage:} \quad w - p = w(U, t, z, ep / p^*) \quad (2.18)$$

$$\text{Price:} \quad p = p(cu, (w - \pi), ep / p^*) \quad (2.19)$$

$$\text{Exchange rate:} \quad e = e(e^e, r - r^*, \gamma) \quad (2.20)$$

Wages (w) now respond to excess demand (represented by unemployment, U), a vector of tax rates (t), the real exchange rate and wage-push variables (z). The price mark-up (p) is sensitive to demand pressure (cu) and prices charged abroad (ep / p^*). The exchange rate is dependent on future expectations of the exchange rate (e^e), relative interest differentials ($r - r^*$) and a risk premium (γ) is included. Expectations may appear additionally in wage equations (expected prices) or in the demand equations (expected output).

2.4 COMPARISON OF THE STRUCTURAL PROPERTIES OF THE MAINSTREAM SUPPLY-SIDE MODELS

The modelling of the supply-side and particularly the role of taxes have been made possible by the use of the LN framework. Although the LN framework is generally accepted as the leading neoclassical supply-side modelling approach, many macroeconomic models have not yet adopted this framework, but still follow the standard Phillips curve approach. The mainstream models, which have revised their structures to incorporate the neoclassical approach, have however applied the LN principles in several different ways. For purposes of this study it is useful to investigate the main structural properties and differences amongst those supply-side models which have adjusted to the LN framework.

The main differences between the so-called mainstream supply-side models are based on the following conceptual issues:

- (i) A cost or a production function approach, i.e. an explicit or implicit incorporation of the production technology;
- (ii) The functional form and underlying production technology for the cost/production function;
- (iii) Measurement or estimation of technical progress;
- (iv) The incorporation and/or measurement of capacity utilisation (expenditure/demand *versus* production/supply as measure for actual output);
- (v) The role of potential output;
- (vi) The role of the NAIRU/NAWRU;
- (vii) Investment (neoclassical/Jorgenson v. Cash flow v. Tobin's q);
- (viii) A measure for user-cost-of-capital;
- (ix) A market-clearing *versus* a non-market clearing approach in labour modelling;
- (x) Wage-setting model (framework of market imperfections: union-firm wage-bargaining);

- (xi) Price-setting model (framework of market imperfections: mark-up on unit cost of production);

The main differences are made explicit by the modelling research conducted in the UK on the one hand and the models constructed by the OECD on the other hand. The model of the London Business School (LBS) may be regarded as representative of the main properties of the UK models.

A summary of the structural properties of the LBS and OECD models are reported in table 2.1.

2.5 CONCLUSION

Supply-side modelling has become imperative if a macro-econometric model is to be useful for policy analyses that go beyond short-term forecasting requirements. However, the structure and specification of supply-side models should be such that they are consistent with theoretical principles and that they are successful in forecasting stagflation, the simultaneous occurrence of unemployment and inflation.

Against this background, the theoretical principles of supply-side economics were explored, as they need to form the basis in the development of a supply-side macro-econometric model of the South African economy. Supply-side effects and particularly the role of taxes, which are the main

Table 2.1 Structural properties of neoclassical supply-side models based on the Layard-Nickell framework

Structural properties		LBS	OECD
Production	<i>Cost/production approach</i>	Cost function	Production function
	<i>Functional form</i>	Translog	Cobb-Douglas and CES
	<i>Technical progress</i>	Labour augmenting (Harrod neutral)	Labour augmenting (Harrod neutral)
Capacity utilisation (cu)	<i>Actual output</i>	n.a.	Actual output estimated by production function
	<i>Potential output</i>	n.a.	Normal (trending) output, a structural production function approach
Investment	<i>Model</i>	Tobin's q	Neoclassical (Jorgenson)
	<i>User-cost-of-capital (r)</i>	Rental cost	User-cost-of-capital (r) = $f(\text{price of capital, depreciation, rates of return})$

Labour: demand	<i>Market- or non-market clearing</i>	Non-market clearing approach, assuming NAIRU	Non-market clearing approach, assuming NAWRU
Labour: supply	<i>Exogenous/LFP</i>	Labour force participation function	Labour force participation function
Wage-setting	<i>Model</i>	Wage-bargaining & wage productivity	Wage-bargaining & wage productivity
Price-setting	<i>Model</i>	Mark-up on unit costs	Mark-up on unit costs, cu, r
Expectations		Learning in exchange rate function	none

supply-side instruments, are modelled within a neoclassical framework of profit maximising or cost minimising behaviour of firms.

The neoclassical approach can be characterised as a combination of market clearing and rational expectations and it emphasises the role of stocks rather than flows. The neoclassical approach emphasises the supply-side of the economy, not through the inclusion of an explicit production function, but rather through a representation of the labour market which is responsive to changes in benefits and taxes, but not to the level of demand. The approach adopted by the mainstream models has not been to follow the competitive paradigm of the classical school; instead development centred around a framework of imperfect competition in goods and labour markets and the adoption of a bargaining approach to wage determination, following the work of Layard and Nickell (1985).

These neoclassical principles are adopted in several, but different ways by macro-economic modellers. The structural properties of the existing (mainstream) supply-side models were investigated to see how they compare in modelling the neoclassical and associated supply-side principles. It can be concluded that the Layard-Nickell framework for neoclassical supply-side modelling is consistent with both the theoretical and policy principles of supply-side economics and need to form the basis in the development of a consistent supply-side model of the South African economy.



CHAPTER 3

THE SOUTH AFRICAN MACROECONOMIC MODEL AND SUPPLY-SIDE ECONOMICS

3.1 INTRODUCTION

The purpose of the study is to develop a theoretically consistent supply-side model of the South African economy. The ultimate objective is to integrate the supply-side model with a full-sector macroeconomic model of South Africa. The macroeconomic model developed by the econometric research team of the University of Pretoria (SAMEM) is used for this purpose.

The structure of the SAMEM, currently used for forecasting and policy analysis purposes, is demand-driven and therefore vulnerable to valid theoretical and empirical points of critique (see chapter 2). The appropriate solution for maintained theoretical consistency and a consistent framework for policy analysis, is to develop a neoclassical supply-side model based on the principles of the Layard-Nickell framework discussed in chapter 2.

The purpose of this chapter is four-fold:

- (i) to present a brief outline of the current structure and shortcomings of SAMEM;
- (ii) to restructure SAMEM for the incorporation of a neoclassical supply-side model;
- (iii) to identify the properties and objectives of the supply-side model; and
- (iv) to present an outline of the proposed methodology for the development of the supply-side model.

3.2 THE STRUCTURE OF THE SOUTH AFRICAN MACROECONOMIC MODEL (SAMEM)

The macroeconomic model of the Department of Economics of the University of Pretoria was originally developed in 1974 with the purpose of short and medium-term forecasts of the South African economy. The original model was developed by De Wet and Dreyer (1976, 1978). Further expansion and developments on the monetary sector were undertaken by De Wet and Herbst (1981) and later by De Wet and Jonkergouw (1995). De Wet and Van der Walt (1994) developed the trade balance as satellite model of the balance of payments sector.

SAMEM consists of three sectors: the real sector, the balance of payments (of which the trade balance is a satellite model) and the monetary sector. The real sector is concerned with the specification of production, expenditure and income and is based on the income-expenditure approach with the level of output and employment principally determined by the level of demand. The current structure of the model is represented by the set of equations in table 3.1 with an explanatory list of variables in table 3.3.

Table 3.1 Structural equations of SAMEM*

Aggregate Demand		
Income-expenditure identities	$yb = bbb - zgnfd + xgnfd + residu$	(3.1)
	$BBB = CDS + CND + IPNL + IV + IPL + CG + IG + RESIDU$	(3.2)
	$bbb = BBB / Pbbb$	(3.3)
	$ybnl = yb - ybl$	(3.4)
	$YB = yb * Pbbp$	(3.5)
Disposable income	$PC = YB - TI - WNLLNL - WLLL$	(3.6)
	$SC = PC - DE - NEBC - TC - OC$	(3.7)
	$YD = YB - SC - FB - DE - TC - TP - TI - YE + OGB - RESIDU$	(3.8)
	$yd = YD / Pv$	(3.9)
Consumption	$cds = f(yd, ESKOM)$	(3.10)
	$cnd = f(M3)$	(3.11)
	$CDS = cds * Pcds$	(3.12)
	$CND = cnd * Pcnd$	(3.13)
Investment	$ipnl = f(hbb, ESKOM)$	(3.14)
	$kvn1 = 0.995 * kvn1_{t-1} + ipnl + ig$	(3.15)
	$IPNL = ipnl * Pi$	(3.16)
	$IPL = exogenous$	(3.17)
Government	$CG = exogenous$	(3.18)
	$IG = exogenous$	(3.19)
Change in inventories	$IV = exogenous$	(3.20)
Imports	$zgnfd = ZGNFD / Pz$	(3.21)
Exports	$xgnfd = XGNFD / Px$	(3.22)
Excess demand	$SV = hbb / yb$	(3.23)
Aggregate supply		
Wage	$WNL = exogenous$	(3.24)
Prices	$Pz = f(Pxi, rS, Pz_{t-1})$	(3.25)
	$Px = f(Pwkm)$	(3.26)
	$Pbbp = f(LPEU, SV, Px)$	(3.27)
	$Pbbb = f(LPEU, SV, Pz)$	(3.28)
	$Pcnd = f(Pbbb)$	(3.29)
	$Pcds = f(Pbbb)$	(3.30)
	$Pv = f(Pcnd, Pcds)$	(3.31)
	$Pi = f(Pbbb)$	(3.32)
Employment	$LNL = f(ybnl)$	(3.33)
	$LPEU = WNLLNL / ybnl$	(3.34)
	$WNLLNL = WNL * LNL$	(3.35)
Unemployment	$None$	(3.36)

* Lower case denotes real values (constant prices) and upper case nominal values.

Equation 3.1 is the standard national income accounting identity in constant prices, divided into two components, namely domestic expenditure (consumer expenditure, fixed investment, government spending and investment and changes in inventories) and net exports (exports and imports). Disposable income in nominal terms (equation 3.8) is computed by subtracting net payments and transfers by households, as well as business and foreign enterprises' shares in income from GDP. Households' expenditure is again divided into two components, i.e. consumption of durable and semi-durable goods (equation 3.10) and consumption of non-durable goods and services (equation 3.11). Personal consumption of durable and semi-durable goods is explained by real disposable income and credit restrictions, i.e. the bond or long-term interest rate (ESKOM). Consumption of non-durable goods and services depends on the cash flow situation of households and is therefore explained by the broad money supply (M3). The change in inventories is an exogenous variable and fixed investment is explained by a simple accelerator model (change in output) with some allowance for interest rate effects. Although a distinction is made between government consumption and investment, both components are exogenous due to the nature of government spending in South Africa. Exports and imports (equations 3.21 and 3.22) enter the real sector after being endogenously determined in the balance of payments sector of the model. They are dependent on domestic and foreign income, relative or competitive prices (domestic prices relative to foreign prices) and the export and import prices of goods and services respectively. Excess demand (equation 3.23) is defined and utilised as the demand pressure variable in the price-setting equations.

The determination and treatment of aggregate supply is very rudimentary and almost non-existent. The underlying production structure, technology and supply constraints of the economy are totally ignored. Wages (equation 3.24) are exogenous (not even explained in terms of the already out-dated Phillips curve mechanism) and influence domestic prices through the wage per unit of output (equation 3.34). Employment (equation 3.33) is explained in terms of the level of output and no recognition is given to an equilibrium level of unemployment, market imperfections and the role of labour unions in the economy. Since wages are exogenous, there is no need to derive the level or rate of unemployment for the economy. Unemployment therefore has no contemporaneous feedback in the model and its role is totally ignored.

The policy implications are that demand policies could have powerful effects on output and employment with no role for supply-side policies.

Based on the above analysis of the structural properties and limitations of SAMEM, it is necessary to improve the theoretical foundations, as well as the forecasting and policy analysis ability of the existing model by incorporating a consistent neoclassical supply-side model of the South African economy. This process is conducted in two steps. First, restructure the existing demand-driven model to create the necessary links between demand (expenditure) and supply. Second, develop (specify, derive and estimate) a supply-side model that is consistent with the neoclassical theory of profit maximisation or cost minimisation of firms and consistent with a supply-side policy framework.

3.3 RESTRUCTURING THE SOUTH AFRICAN MACROECONOMIC MODEL (SAMEM) TO INCORPORATE A CONSISTENT SUPPLY-SIDE MODEL

Two distinct approaches can be identified for modelling the supply-side of an economy: first, the use of an explicit production function and, second the cost function approach. The advantage of the cost function approach is that it enables the consistent derivation of price, wage and demand functions within a neoclassical supply-side framework, a feature which the production function approach does not include. Where a production function is explicitly estimated, it can be used to derive estimates for capacity utilisation, but the cost function approach typically has to resort to indirect measures or independent equations to determine capacity utilisation.

Before any restructuring of SAMEM can be done, it is necessary to decide on either a production or cost function approach. The production function approach is opted for in this study, since it allows the estimation of capacity utilisation – a key variable in linking production constraints with demand behaviour via costs and prices. However, based on the theoretical consistency of the cost function approach, an attempt will be made to estimate a cost function directly for the South African economy, and derive consistent functions for production (based on Shephard's duality principles), prices and factor demands (see chapter 4 for a detailed discussion).

Restructuring SAMEM to incorporate a supply-side model, therefore necessitates the following adjustments and additions to the existing structural relationships:

- (i) Replace the income-expenditure identity (equation 3.1) with a stochastic production function (equation 3.50). Introduce an identity to match gross domestic production at factor cost and market prices (equation 3.51).
- (ii) Include an estimate of potential output (equation 3.52).
- (iii) Included an estimate for capacity utilisation (equation 3.53).
- (iv) Substitute the employment function (equation 3.33) with a labour market model (equations 3.60, 3.62, 3.63 and 3.54), assuming non-market clearing behaviour within a framework of market imperfections.
- (v) Substitute exogenous wages (equation 3.24) with a wage-setting equation consistently derived from cost-minimising behaviour within a framework of collective bargaining, thus allowing for the role of labour unions (equation 3.54).
- (vi) Substitute the accelerator model for fixed investment (equation 3.14) with a neoclassical model (Jorgenson) for investment, allowing for taxes, depreciation rates, asset value prices and inflation to contribute to the unit-cost-of-capital. Extend the investment model to incorporate the financial constraints on investment, which are very real in the South African scenario (equation 3.45).
- (vii) Include an identity for the financing of investment, consistent with the structure of the national accounts (equation 3.41).
- (viii) Include a stochastic equation for corporate savings as part of the identity for financial constraints (equation 3.43).

- (ix) Replace the domestic price equations (3.25 to 3.28) with equations consistently derived within a framework of cost minimising and market imperfections (equations 3.55, 3.56, 3.65 and 3.66).
- (x) Link demand and supply by substituting exogenous inventory changes (equation 3.20) with an identity where inventory changes drop out as the difference between supply-side production and expenditure on gross domestic product (equation 3.49).
- (xi) Adjust the identities for gross domestic expenditure to be consistent with the new framework (equations 3.37 and 3.38).

Without specifying the detail of the different supply-side equations, the restructured SAMEM are presented in table 3.2.

Table 3.2 Structural equations of amended SAMEM to incorporate a consistent supply-side model of the South African economy*

Aggregate Demand		
Income-expenditure identities	$bbb = cnd + cds + cg + i + residu$	(3.37)
	$BBB = bbb * Pbbb$	(3.38)
	$ybnl = yb_{@m}^A - ybl$	(3.39)
	$YB_{@m}^A = yb_{@m}^A * vpi$	(3.40)
Disposable income	$FINCON = SP + SC + SG + DE + NETCAPFL + GOLD_RESERV$	(3.41)
	$GOSTC = GOS - TC$	(3.42)
	$SC = f(GOSTC)$	(3.43)
	$fincon = FINCON / ppi$	(3.44)
	$YD = YB - SC - FB - DE - TC - TP - TI - YE + OGB - RESIDU$	(3.8)
	$yd = YD / Pv$	(3.9)
Consumption	$cds = f(yd, ESKOM)$	(3.10)
	$cnd = f(M3)$	(3.11)
	$CDS = cds * Pc ds$	(3.12)
	$CND = cnd * P cnd$	(3.13)
Investment	$i_{fix} = f(yb_{@f}^A, ucc, fincon, k)$	(3.45)
	$k = 0.95 * k_{t-1} + i_{fix}$	(3.46)
	$i = i_{fix} + iv$	(3.47)
	$l = i * Pi$	(3.48)
Government	$CG = exogenous$	(3.18)
	$IG = exogenous$	(3.19)
Change in inventories	$iv = yb_{@f}^A - bbb - xgnfd + zgnfd - residu$	(3.49)
Imports	$zgnfd = ZGNFD / Pz$	(3.21)
Exports	$xgnfd = XGNFD / Px$	(3.22)
Excess demand	$SV = bbb / yb_{@m}^A$	(3.23)
Aggregate supply		
Production	$yb_{@f}^A = f(k, N, technical\ progress)$	(3.50)
	$yb_{@m}^A = yb_{@f}^A + ti$	(3.51)

	$yb_{@f}^P = f(k, N^{pot}, technical\ progress^{smooth})$	(3.52)
Capacity utilisation	$cu = yb_{@f}^A / yb_{@f}^P$	(3.53)
Wages	$w = f(product, vpi^e)$	(3.54)
Prices	$Pz = f(Pxi, r\$, Pz_{t-1})$	(3.25)
	$Px = f(Pwkm)$	(3.26)
	$Pbbp = f(ppi)$	(3.65)
	$Pbbh = f(vpi)$	(3.66)
	$ppi = f(w / product, ucc)$	(3.55)
	$vpi = f(ppi, pz\$, r\$, SV)$	(3.56)
	$Pend = f(vpi)$	(3.57)
	$Pcds = f(vpi)$	(3.58)
	$Pv = f(Pend, Pcds)$	(3.31)
	$Pi = f(ppi)$	(3.59)
Employment	$N = f(yb_{@m}^A, w / ucc)$	(3.60)
	$N^{pot} = f(L^{smooth}, NAWRU)$	(3.61)
	$product = yb_{@m}^A / N$	(3.62)
Labour force	$L = f(total_pop, w)$	(3.63)
Unemployment	$U = L - N$	(3.64)

* Lower case denotes real values (constant prices) and upper case nominal values.

Table 3.3 Explanatory variable list for SAMEM*

<i>bbb (BBB)</i>	Gross domestic expenditure
<i>cds (CDS)</i>	Consumption, durable and semi-durable goods
<i>CG</i>	Public consumption
<i>end (CND)</i>	Consumption, non-durable goods and services
<i>eu</i>	Capacity utilisation
<i>DE</i>	Depreciation allowances
<i>ESKOM</i>	Long-term interest rate
<i>FB</i>	Factor payments (net)
<i>fincon (FINCON)</i>	Financing of gross domestic investment (financial constraint)
<i>GOLD_RESERV</i>	Change in gold and other foreign reserves
<i>GOS</i>	Gross operating surplus
<i>GOSTC</i>	Gross operating surplus, adjusted for direct company taxes
<i>i</i>	Gross domestic investment
<i>i_{fix}</i>	Gross domestic fixed investment
<i>ig (IG)</i>	Public investment
<i>IPL</i>	Investment, private agricultural sector
<i>ipnl (IPNL)</i>	Investment, private non-agricultural sector
<i>IV</i>	Change in inventories
<i>k</i>	Capital stock
<i>kvnl</i>	Capital stock, non-agricultural sector
<i>L</i>	Labour force: economic active population
<i>L^{smooth}</i>	Smoothed labour force (time-series technique such as Hodrick-Prescott filter)
<i>LNL</i>	Employment, non-agricultural sector
<i>LPEU</i>	Wage per unit output
<i>M3</i>	M3 money supply
<i>N</i>	Employment
<i>N^{pot}</i>	Potential employment
<i>NAWRU</i>	Non-accelerating wage rate of unemployment
<i>NEBC</i>	Net income from property
<i>NETCAPFL</i>	Net capital flow

<i>OC</i>	Net transfers made to households, general government and rest of the world
<i>OGB</i>	Transfers from government and foreign sector
<i>Pbbb</i>	Deflator: gross domestic expenditure
<i>Pbbp</i>	Deflator: gross domestic product
<i>PC</i>	Profits of companies
<i>Pcda</i>	Deflator: consumption, durable and semi-durable goods
<i>Pcnd</i>	Deflator: consumption, non-durable goods and services
<i>Pi</i>	Deflator: investment
<i>ppi</i>	Production price index
<i>product</i>	Labour productivity
<i>Pv</i>	Deflator: total consumption
<i>Pwkm</i>	World commodity price index
<i>Px</i>	Deflator: total exports
<i>Pxi values</i>	Deflator: industrial countries export unit values
<i>Pz</i>	Deflator: total imports
<i>Pz\$</i>	Deflator: total imports in US dollar
<i>r\$</i>	Rand/dollar exchange rate
<i>RESIDU (residu)</i>	Residual
<i>SC</i>	Corporate saving
<i>SG</i>	Saving of general government
<i>SP</i>	Personal saving
<i>SV</i>	Excess demand
<i>TC</i>	Taxes on companies
<i>TI (ti)</i>	Net indirect taxes
<i>total_pop</i>	Total population
<i>TP</i>	Personal taxes
<i>U</i>	Unemployment
<i>ucc (UCC)</i>	User-cost-of-capital
<i>vpi</i>	Consumer price index
<i>vpi*</i>	Expected consumer prices
<i>w</i>	Total wage rate
<i>WLLL</i>	Wage income, agricultural sector
<i>WNL</i>	Wage rate, non-agricultural sector
<i>WNLLNL</i>	Wage income, non-agricultural sector
<i>yb (YB)</i>	Gross domestic product at market prices
<i>yb^A_{@f}</i>	Actual gross domestic product at factor cost
<i>yb^A_{@m}</i>	Actual gross domestic product at market prices
<i>yb^P_{@f}</i>	Potential gross domestic product at factor cost
<i>ybl</i>	Gross domestic product, agricultural sector
<i>ybnl</i>	Gross domestic product, non-agricultural sector at market prices
<i>yd (YD)</i>	Personal disposable income
<i>YE</i>	Private income from property minus interest on public debt
<i>xgnfd (XGNFD)</i>	Exports of goods and non-factor services
<i>zgnfd (ZGNFD)</i>	Imports of goods and non-factor services

* Lower case denotes real values (constant prices) and upper case nominal values.

3.4 THE PROPERTIES AND OBJECTIVES OF THE PROPOSED SUPPLY-SIDE MODEL

Given the deficiencies of demand-side models identified in chapter 2, the analysis and proposed new structure of SAMEM, as well as the proposed Layard-Nickell framework for a consistent neoclassical supply side, the following objectives are identified for the development of a supply-side model of the South African economy:

The *primary objectives* are:

- (i) To develop an aggregate neoclassical supply-side model of the South African economy, pertaining to the structural and long-run properties of the economy, as well as the profit maximising and cost minimising decision-making processes of firms;
- (ii) To specify, derive and estimate every structural relationship (equation) of the model jointly for the purpose to ensure consistency between costs, demand factors and prices throughout the analytical framework;
- (iii) To incorporate an estimate for capacity utilisation which serves as a significant variable in explaining price and wage-setting behaviour and influences every key macroeconomic variable in a well-developed supply-system;
- (iv) To endogenise technical progress in the cost/production relationship;
- (v) To incorporate price expectations;
- (vi) To allow for the specific and rather unique characteristics of the South African economy;
- (vii) To incorporate a set of target or policy variables allowing for policy proposals with specific reference to labour problems in South Africa; and
- (viii) To maintain a balance between the detail required for policy analysis and the stability of the model to ensure reliable forecasts.

A number of *conceptual issues* have to be decided on and dealt with to achieve the above-mentioned objectives:

- (i) A cost or a production function approach in estimating the structural production properties of the South African economy;
- (ii) An appropriate functional form and underlying production technology for the cost/production function;
- (iii) An appropriate measure or estimate for technical progress;
- (iv) Assumptions about the factor intensity of production and the returns to scale property of the South African production structure;
- (v) An appropriate measure for capacity utilisation (expenditure/demand *versus* production/supply as measure for actual output);
- (vi) An appropriate measure for potential output;
- (vii) An appropriate measure for the NAIRU/NAWRU;
- (viii) An appropriate model for investment (neoclassical/Jorgenson *versus* cash flow *versus* Tobin's q);
- (ix) A measure for user-cost-of-capital;
- (x) A market-clearing *versus* a non-market clearing approach in modelling labour;
- (xi) A distinction between skilled and unskilled labour;
- (xii) An appropriate wage-setting model (framework of market imperfections: union-firm wage-bargaining);
- (xiii) An appropriate price-setting model (framework of market imperfections: mark-up on unit cost of production);
- (xiv) Exclusion (standard approach) *versus* inclusion of unit cost of capital;

Developing a supply-side model consistent with the above-mentioned objectives entails the estimation of equations for the following structural relationships:

- (i) production (following a cost function approach),
- (ii) potential production;
- (iii) capacity utilisation;
- (iv) fixed investment;
- (v) corporate saving;
- (vi) demand for labour (skilled, unskilled and labour participants in the informal sector),
- (vii) labour supply (total and skilled with unskilled as the residual);
- (viii) wage-setting (skilled and unskilled);
- (ix) unemployment (total, skilled and unskilled); and
- (x) price-formation (production and consumption prices).

3.5 ESTIMATION METHODOLOGY

Having stated the objectives and specified the framework for the development of a neoclassical supply side of the South African economy, the next step would be to estimate single equations for each of the above-mentioned relationships within a consistent framework of cost minimising, assuming market imperfections.

The estimation technique used is the Engle and Yoo (1989) three-step cointegration procedure, which is an extension of the Engle-Granger (1987) two-step procedure (see Appendix 1 for an exposition of the technique).

The estimation of every single equation consists of:

- (i) an investigation of the theoretical foundations;
- (ii) specification of a consistent theoretical/empirical model;
- (iii) the data generating process and subsequent identification of the univariate characteristics of the data;
- (iv) cointegration (long-run or equilibrium) estimation (Engle-Granger first step);
- (v) estimation of the short-run dynamics, i.e. the error correction model (Engle-Granger second step);
- (vi) diagnostic testing;
- (vii) cointegration correction and adjustment of long-run coefficients (Engle-Yoo third step);
- (viii) ex-post dynamic simulation of the model;
- (ix) computation and evaluation of the simulation error statistics (Appendix 2); and
- (x) investigation and evaluation of the dynamic response properties and robust nature of the model.

The estimated single equations are then combined into a neoclassical supply-side model of the South African economy. The system is closed by introducing a number of identities and

definitions, linking every endogenous variable in the model and thereby ensuring a fully dynamic system.

The model is subsequently dynamically simulated and evaluated along the criteria (full ideal principles of model selection) set out in Appendix 3.

First, the theoretical structure of the estimated model (empirical specification) is evaluated to determine the extent to which the model complies with (1) the *a priori* objectives of neoclassical supply-side modelling, (2) economic theory – for the model in general and the individual equations in particular, (3) rival models, i.e. to what extent the model encompasses the characteristics of rival models, and (4) policy analysis, i.e. the relevancy of the specified model for policy analysis.

Second, a simulation (i.e. *ex post* forecast) of the full system of equations is conducted. These dynamic simulation properties of the model are evaluated in terms of (1) the statistical significance (*ex post* forecast ability) and (2) the stability of the model over the simulated sample range. The statistical significance (goodness-of-fit) of the full system is measured in terms of simulation/forecast error statistics and confirmed by the graphical representations of the simulations.

Third, a series of dynamic, ex-post simulations are conducted by shocking every stochastic variable in the system. The simulation results are used to (1) determine the long-run (steady-state) multipliers and elasticities of the system and, once again, (2) to evaluate the statistical significance and sensitivity of the model in terms of the degree, speed and stability of convergence. The robust (stable) nature of the model serves as an indication of the forecasting ability of the model.

Finally, the supply-side model of South Africa will be utilised in a policy analysis of the unemployment problem in South Africa. The purpose is to identify and validate a set of policy proposals that will increase the labour absorption capacity of the economy and subsequently reduce the unemployment problem.

3.6 CONCLUSION

In this chapter the structural properties (demand-driven) and inadequacies of SAMEM are identified given the theoretical and empirical points of critique raised against demand-driven models in Chapter 2. A new structure is proposed, incorporating a consistent neoclassical supply side and allowing for the principles of the Layard-Nickell framework. The subsequent properties and objectives of the neoclassical model are identified. The chapter concludes with an outline of the proposed methodology to be followed in the development and application of the supply-side model.



CHAPTER 4

A NEOCLASSICAL PRODUCTION FUNCTION FOR THE SOUTH AFRICAN ECONOMY: THEORY AND EVIDENCE

4.1 INTRODUCTION

In the past two decades supply-side theory, policy and the modelling thereof, have become more popular in the field of economics. This is due to the inadequacy of demand-oriented theory and policy to explain and deal with unemployment and inflation. It is increasingly recognised that the cost-minimising or profit-maximising decision-making processes of the firms responsible for production activities in the economy, need to be examined and modelled. Supply-side economics stresses the necessity of understanding the structure of the production process and the effect of each of the production factors on the level of output. A further aspect to be taken into account when modelling supply-side behavioural equations, is the incorporation of supply-side policy instruments and their effect on the economy (Nickell 1988: 202).

The purpose of this chapter is to estimate an aggregate neoclassical production function for the South African economy as the key component of a supply-side model. The production function may then be used to analyse the production structure and properties of the South African economy. At the macro-level the production function may be used to explain economic growth, the prices of various factors of production and the extent to which these factors are utilised. At a micro-level, a production function is useful to analyse the degree of substitution between the various factors of production and also the extent to which firms experience decreasing or increasing returns to scale as output expands. On both macro- and micro-level, the production function may be used as a tool to assess the proportion of any increase (decrease) in output over time which may respectively be attributed to, firstly, increases (decreases) in the inputs of factors of production; secondly, to the existence of increasing (decreasing) returns to scale; and thirdly, to the technical progress (or lack of it) taking place in the economy.

Two aspects need consideration when estimating a production function:

- (i) the functional form; and
- (ii) whether to estimate the production function directly or to estimate a cost function and derive a consistent production function based on duality principles.

Although there are certain advantages to the latter option, duality principles - according to which unique relationships exist between the coefficients of cost and production functions - only apply to restricted functional forms. Examples of these restricted forms are the Cobb-Douglas and the Constant Elasticity of Substitution (CES) functional forms. The Transcendental logarithmic (Translog) form is an example of a more flexible functional form for which the duality principles do not apply. Two different avenues are therefore explored:

- (i) A Cobb-Douglas cost function is estimated and used to derive and develop a consistent production function based on duality principles.



- (ii) Given the advantages of the more flexible Transcendental Logarithmic (Translog) functional form, the non-homothetic Translog cost function is estimated and tested for the validity of the imposed restrictions, so that it can reduce to the Cobb-Douglas cost function. If the Translog cost function cannot collapse into a Cobb-Douglas cost function, an equivalent Translog production function has to be estimated by imposing restrictions, similar to the ones validly imposed when estimating a cost function.

Once a production function has been estimated based on the above principles, the properties of the South African production structure can be evaluated and interpreted.

4.2 THE THEORETICAL FRAMEWORK: A SURVEY OF THE FUNCTIONAL FORMS FOR PRODUCTION

Several functional forms¹, each representative of the technology of production, evolved from attempts to estimate production behaviour by means of mathematical and statistical modelling. Of the many functional forms, the most commonly used are the Cobb-Douglas, Constant Elasticity of Substitution (CES), Variable Elasticity of Substitution (VES), Transcendental logarithmic (Translog) and the generalised Leontief functions. In order to select the appropriate functional form, the properties of the Cobb-Douglas, CES and the Translog functional forms should be analysed.

The Cobb-Douglas and CES functional forms represent a constant elasticity of substitution production structure. The first has the unique property of a unitary elasticity of substitution, while the latter allows for a less restrictive, not necessarily unitary, but still constant elasticity of substitution. The Transcendental logarithmic (Translog) functional form is representative of a more flexible, variable elasticity of substitution production structure.

4.2.1 The Cobb-Douglas functional form

The best known production function is the Cobb-Douglas function. The Cobb-Douglas production function, first published by C.W. Cobb and P.H. Douglas in 1934, has the following functional form²:

$$Q = AV_1^\alpha V_2^\beta \tag{4.1}$$

with

¹ See Appendix 5 for a review on the theory of production functions.

² Cobb and Douglas originally constrained the exponents of V_1 and V_2 so that they sum to unity, i.e. to embody constant returns to scale and therefore reducing the functional form to:

$$Q = AV_1^\alpha V_2^{1-\alpha}$$

The production function can also be generalised to incorporate more than two inputs. The form then becomes:

$$Q = AV_1^\alpha V_2^\beta V_3^\delta \dots V_n^\Omega$$

Q = output or production (value added);

A = efficiency parameter³ ($A > 0$);

V_i = production factors (inputs) with $i = 1, 2$; and

α and β = input elasticities of production with respect to V_1 and V_2 ($0 < \alpha, \beta < 1$).

The Cobb-Douglas production function has a number of convenient properties.

The marginal products (MP) of the inputs are given by:

$$MP_1 = \frac{\partial Q}{\partial V_1} = \alpha A V_1^{\alpha-1} V_2^\beta = \alpha \frac{Q}{V_1}, \quad MP_2 = \frac{\partial Q}{\partial V_2} = \beta A V_1^\alpha V_2^{\beta-1} = \beta \frac{Q}{V_2} \quad (4.2)$$

Both are positive ($\alpha, \beta > 0$) and diminishing ($\alpha, \beta < 1$). Assuming the firm is a price-taker and a profit maximiser, equations 4.2 imply that the marginal productivity conditions for this production function are:

$$\alpha \frac{Q}{V_1} = \frac{p_1}{p}, \quad \beta \frac{Q}{V_2} = \frac{p_2}{p} \quad (4.3)$$

that can be rewritten as

$$\alpha = \frac{p_1 V_1}{p Q} \quad \text{and} \quad \beta = \frac{p_2 V_2}{p Q} \quad (4.4)$$

Thus, if the marginal productivity conditions hold, the exponents α and β in the Cobb-Douglas function are equal to the respective shares of the inputs in the value of total output.⁴

The cost-minimising condition for the Cobb-Douglas function is given by:

$$MRS = \frac{\partial Q}{\partial V_2} / \frac{\partial Q}{\partial V_1} = \frac{\beta V_1}{\alpha V_2} = \frac{p_2}{p_1} \quad (4.5)$$

The optimising (profit maximising) conditions imply that:

$$\frac{V_1}{V_2} = \left(\frac{\alpha}{\beta} \right) \left(\frac{p_2}{p_1} \right) \quad (4.6)$$

For a given factor price ratio, the greater α/β , the greater the optimal input (V_1/V_2) ratio. Thus, the magnitude of the exponent α , relative to that of β , determines the factor-intensity of a production process represented by the Cobb-Douglas function.

³ For fixed inputs V_1 and V_2 , the larger A , the greater is the maximum output Q obtainable from the inputs.

⁴ These shares will only be equal to the value of total output in the case of constant returns to scale.

The Cobb-Douglas function is homogeneous of degree $\alpha + \beta$ since

$$Q(\lambda V_1, \lambda V_2) = A(\lambda V_1)^\alpha (\lambda V_2)^\beta = \lambda^{\alpha+\beta} A V_1^\alpha V_2^\beta = \lambda^{\alpha+\beta} Q(V_1, V_2) \quad (4.7)$$

For $\alpha + \beta > 1$, the production structure is characterised by increasing returns to scale. For $\alpha + \beta < 1$, the production structure exhibits decreasing returns to scale and if $\alpha + \beta = 1$, constant returns to scale. A significant constraint of the Cobb-Douglas production structure is, however, that the returns-to-scale property remains the same at all levels of output.

The Cobb-Douglas function is also restricted in that the elasticity of substitution is always constant and equal to unity (for all levels of output and for any factor combinations). It therefore represents a technology where a one percent change in the factor price ratio leads to a 1 percent change in the factor input ratio in the opposite direction, that is a one percent increase in p_1 relative to p_2 will lead to a one percent reduction in V_1 relative to V_2 .

The Cobb-Douglas functional form has the advantage of being stable and robust and it is relatively easy to estimate. It does, however, have the disadvantage that it is very restrictive and fails to allow for flexibility in the technology of the production sector being analysed. Restricting the Cobb-Douglas production function to a constant and specifically a unitary elasticity of substitution of the inputs, eliminates an investigation of the extent to which factor substitution is possible and how such substitution may vary between firms and industries. The more flexible CES functional form was therefore developed where σ (the elasticity of substitution), although still constant, can take on values other than unity (Fuss and McFadden 1979: 311 - 363).

4.2.2 The Constant Elasticity of Substitution (CES) functional form

Cobb and Douglas pioneered the development of a production function in the 1920s and 1930s. The next meaningful step came forty years later when the economists, Arrow, Chenery, Minhas and Solow published the CES function (Arrow *et. al.*, 1961). This was an attempt to derive a production function with the properties of (i) homogeneity, (ii) constant elasticity of substitution between the inputs, and (iii) the possibility of different elasticities for different industries.

The following form specifies the CES function⁵:

$$Q = \gamma (\delta V_1^{-\theta} + (1-\delta)V_2^{-\theta})^{-\frac{1}{\theta}} \quad (4.8)$$

with

Q = output or production (value added);

⁵ The CES production function can also be generalised to include more than two inputs:

$$Q = \gamma (\delta_1 V_1^{-\theta} + \delta_2 V_2^{-\theta} + \dots + \delta_n V_n^{-\theta})^{-\frac{1}{\theta}}$$

where $\sum \delta = 1$.

- γ = efficiency parameter⁶ ($\gamma > 0$);
- V_i = production factors (inputs) with $i = 1, 2$;
- θ = substitution parameter ($-1 < \theta < \infty$); and
- δ = distribution parameter ($0 < \delta < 1$)

The marginal products (MP) of the inputs are given by:

$$MP_1 = \frac{\partial Q}{\partial V_1} = \frac{\delta \gamma}{V_1^{1+\theta}} \left[\delta V_1^{-\theta} + (1-\delta) V_2^{-\theta} \right]^{(1+\theta)/\theta} = \frac{\delta}{\gamma^\theta} \left(\frac{Q}{V_1} \right)^{1+\theta},$$

$$MP_2 = \frac{\partial Q}{\partial V_2} = \frac{(1-\delta) \gamma}{V_2^{1+\theta}} \left[\delta V_1^{-\theta} + (1-\delta) V_2^{-\theta} \right]^{(1+\theta)/\theta} = \frac{(1-\delta)}{\gamma^\theta} \left(\frac{Q}{V_2} \right)^{1+\theta} \quad (4.9)$$

Assuming profit maximisation under perfect competition, the marginal productivity conditions corresponding to equation 4.3 are:

$$\frac{\delta}{\gamma^\theta} \left(\frac{Q}{V_1} \right)^{1+\theta} = \frac{p_1}{p}, \quad \frac{(1-\delta)}{\gamma^\theta} \left(\frac{Q}{V_2} \right)^{1+\theta} = \frac{p_2}{p} \quad (4.10)$$

The cost-minimising condition for the CES function, subject to a predetermined output, is given by:

$$MRS = \frac{\partial Q}{\partial V_2} / \frac{\partial Q}{\partial V_1} = \left(\frac{1-\delta}{\delta} \right) \left(\frac{V_1}{V_2} \right)^{1+\theta} = \frac{p_2}{p_1} \quad (4.11)$$

The optimising (profit maximising) conditions imply that

$$\frac{V_1}{V_2} = \left(\frac{\delta}{1-\delta} \right)^{1/(1+\theta)} \left(\frac{p_2}{p_1} \right)^{1/(1+\theta)} \quad (4.12)$$

Since the quantity $(\delta/(1-\delta))^{1/(1+\theta)}$ is a constant, it follows that a one percent rise in the factor price ratio p_1/p_2 leads to a $(1/(1+\theta))$ percent rise in the factor input ratio (V_1/V_2) .

The elasticity of substitution of the CES function is given by⁷:

$$\sigma = - \frac{d(V_1/V_2)}{V_1/V_2} / \frac{d(MRS)}{MRS} = \frac{1}{\theta+1} \quad (4.13)$$

Possible values for θ range from $\theta = \infty$ (when $\sigma = 0$ and substitution is impossible) to $\theta = -1$ (when $\sigma = \infty$, the isoquants are straight lines and substitution possibilities are greatest). When $\theta = 0$, $\sigma =$

⁶ For given values of δ and θ and for fixed inputs V_1 and V_2 , the larger γ , the greater the maximum output Q obtainable from the inputs becomes.

⁷ Because of its relationship with σ , θ is known as the substitution parameter: $\theta = \frac{1}{\sigma} - 1$.

1 as the CES function reduces to the Cobb-Douglas function. The Cobb-Douglas is therefore a special case of the CES function.⁸

The cost minimisation condition (equation 4.10) may also be rewritten as:

$$\frac{p_2 V_2}{p_1 V_1} = \left(\frac{1-\delta}{\delta} \right) \left(\frac{V_1}{V_2} \right)^\theta \quad (4.14)$$

so that, for a given factor input ratio (V_1/V_2) and a given value of θ , δ increases the ratio of input 2's share in total output relative to the share of input 1.⁹ In contrast, the ratio of factor shares in a Cobb-Douglas function is a constant.

The CES function (equation 4.8) implies constant returns to scale and may be generalised to

$$Q = \gamma (\delta V_1^{-\theta} + (1-\delta)V_2^{-\theta})^{-\frac{\nu}{\theta}} \quad (4.15)$$

The CES function is homogeneous of degree ν since

$$Q(\lambda V_1, \lambda V_2) = \gamma [\delta (\lambda V_1)^{-\theta} + (1-\delta)(\lambda V_2)^{-\theta}]^{-\frac{\nu}{\theta}} = \lambda \gamma [\delta V_1^{-\theta} + (1-\delta)V_2^{-\theta}]^{-\frac{\nu}{\theta}} = \lambda^\nu Q \quad (4.16)$$

Thus, for $\nu > 1$, $\nu = 1$ and $\nu < 1$ the function exhibits increasing, constant or decreasing returns to scale respectively.¹⁰

The CES functional form appears to offer sufficient flexibility in modelling the technical alternatives for a firm producing a single output using two inputs. In the context of several inputs, however, the CES function implies that the partial elasticities of substitution between all pairs of production factors are equal. This rules out the possibility of complementarities between any pairs of factors. The strongly separable CES and Cobb-Douglas functional forms possess the distinctive feature of self-duality. In the context of producer theory, this means that both the production and cost functions are elements of the same set of functional forms.¹¹ Therefore, preference will dictate

⁸ For the CES function to become a Cobb-Douglas function, it is necessary to restrict θ to zero so that σ (elasticity of substitution) is equal to one. Unfortunately, if $\theta = 0$ is substituted into the CES function the equation collapses. A solution is therefore obtained by using L'Hospital's rule. Making use of this rule and differentiating with regard to θ and then substituting $\theta = 0$ into the function yields $Q = \gamma V_1^\alpha V_2^\beta$, which is a Cobb-Douglas function. $Q = \gamma V_1^\alpha V_2^{\beta(1-\alpha)}$ is a Cobb-Douglas function with constant returns to scale. (For the complete mathematical derivation of the Cobb-Douglas function using the CES function, see Heathfield (1987)).

⁹ For this reason δ is known as the distribution (or capital intensity) parameter.

¹⁰ For this reason ν is known as the returns-to-scale parameter.

¹¹ Shephard's duality exists between the production function and the cost function of functional forms exhibiting convex and monotonic production technology properties. The production technology is therefore identically represented by either the production function or the corresponding cost function, which allows for direct transformation between the two. Given the separability and homothetic nature of both the Cobb-Douglas and CES functional forms, it is possible to derive the dual cost function given the production function.

whether one chooses to describe the technology by means of, for example, a CES production function or a CES cost function, since both incorporate the same hypothesis.

Recently, several sets of functional forms, which are more flexible than the functions discussed above, have been proposed. These functional forms allow for an arbitrary set of partial elasticities of substitution between pairs of inputs at a given input price or quantity. They offer substantial gains in terms of flexibility, and enable the investigator to test important hypotheses remaining from earlier work. However, because these functional forms are in general not self-dual, the decision of whether to adopt a production function or a cost function as a point of departure, is of grave importance. The type of analysis envisaged will have to be the deciding factor. One of the most widely used and flexible functional forms to be introduced into econometric analysis is the Translog functional form.

4.2.3 The Transcendental logarithmic functional form

The CES production and cost functions were natural extensions of the Cobb-Douglas functional form in that they permit the elasticity of substitution to deviate from unity, although still constant. Subsequently, a function which allows the elasticity of substitution to vary with output and/or factor proportions was generated and led to the development of variable elasticity of substitution (VES) production functions which also allow for variation in the elasticity of scale¹² (returns to scale).¹³ The elasticity of substitution (σ) can vary due to (i) variations in the factor input ratio

The Cobb-Douglas cost function is specified by:

$$C = \left(\frac{q}{A}\right)^{1/\varepsilon} p_1^{\alpha/\varepsilon} p_2^{\beta/\varepsilon} \left(\frac{\alpha}{\beta}\right)^{\beta/\varepsilon} \left(\frac{\beta}{\alpha}\right)^{\alpha/\varepsilon} \text{ where } \varepsilon = \alpha + \beta \text{ (elasticity of scale) or}$$

$$C = \left(\frac{q}{A}\right) \left(\frac{p_1}{\alpha}\right)^\alpha \left(\frac{p_2}{\beta}\right)^\beta \text{ if the production function itself is constrained to have constant returns to scale.}$$

The CES cost function is specified by:

$$C = q^\nu \gamma^{-1} p_1 p_2 [(1-\delta)^\sigma p_1^{-\sigma\theta} + \delta^\sigma p_2^{-\sigma\theta}]^{1/\sigma\theta}$$

When the underlying production function is of constant returns to scale ($\nu = 1$), costs do not vary with output. When $\nu > 1$, average cost decreases with q (increasing returns to scale) and when $\nu < 1$, average cost increases with q (decreasing returns to scale).

Both the cost functions are homogeneous of degree one in prices.

¹² The elasticity of scale is the ratio of the proportionate increase in output to the proportionate increase in inputs. If it is assumed that both inputs increase by the same percentage so that $dV_1/V_1 = dV_2/V_2 = dV/V$, the elasticity of scale can be defined by the equation:

$$\varepsilon = (dQ/Q)/(dV/V)$$

If $\varepsilon < 1$, then doubling inputs will lead to a less than doubling of output (decreasing returns to scale).

If $\varepsilon = 1$, then doubling inputs will lead to a doubling of output (constant returns to scale).

If $\varepsilon > 1$, then doubling inputs will lead to a more than doubling of output (increasing returns to scale).

¹³ In the case of both the Cobb-Douglas and CES functional forms, the returns to scale were fixed. If the returns to scale were "n" for one level of output and one factor combination then it will be "n" for all levels of output and all factor combinations. This results in the long-run average cost curve, which is either continuously rising (decreasing returns to scale), a horizontal line (constant returns to scale) or continuously

$(V_1/V_2)^{14}$ or (ii) technical progress which affects the ease with which the factors can be substituted for each other, even in the case of a constant V_1/V_2 ratio. One such general function is the Translog production function. The Translog functional form proposed by Christensen, Jorgenson and Lau (1973) approximates the logarithm of output by a quadratic (second order) in the logarithms of the inputs.

For the two-input case the function is:¹⁵

$$\ln Q = \ln \alpha_0 + \alpha_1 \ln V_1 + \alpha_2 \ln V_2 + \beta_{11} (\ln V_1)^2 + \beta_{22} (\ln V_2)^2 + \beta_{12} \ln V_1 \ln V_2 \quad (4.17)$$

The Translog functional form is a second-order Taylor expansion or approximation of the quantities (in the case of the production function) or prices (in the case of the cost function) of any number of inputs (Allen 1997: 26). The functions (production and cost) are quadratic in the logarithms. Unlike the case with the Cobb-Douglas and CES functions, the Translog function is not restricted by the assumptions of homotheticity and additivity (or separability) and is therefore more flexible. The assumption of homotheticity (in the case of a production function) implies that the factor shares of output are independent of total output, while the additional assumption of separability (additivity) implies that the elasticities of substitution are constant and equal for any pair of inputs. The assumptions of homotheticity¹⁶ and separability can, however, be imposed on the Translog function as testable parameter restrictions. See Appendix 4 for a discussion of the implications of separability and additivity on functional structures.

The Translog form reduces to the Cobb-Douglas form when all second-order terms vanish.¹⁷ It also provides a second order approximation of the CES form with non-linear restrictions on the

falling (increasing returns to scale). Thus the long-run average cost curve cannot take the "U"-shape often assumed in the theory of the firm.

¹⁴ The greater the ratio, i.e. the greater the V_1 intensity of production, the less likely it is to substitute further V_2 for V_1 and the lower σ is likely to be.

¹⁵ The Translog production function can be generalised to include more than two inputs:

$$\ln Q = \ln \alpha_0 + \sum_{i=1}^n \alpha_i \ln V_i + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \beta_{ij} \ln V_i \ln V_j$$

where $\beta_{ij} = \beta_{ji}$ and $i \neq j$; $i, j = 1, \dots, n$.

¹⁶ The assumption of homotheticity on a Translog production function of the form:

$$\ln Q = \ln \alpha_0 + \sum_{i=1}^n \alpha_i \ln V_i + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \beta_{ij} \ln V_i \ln V_j$$

implies the following parameter restrictions:

$$\sum \beta_{ij} = \sum \beta_{ji} = 0, \forall i, j$$

(Chung 1994: 142-143)

¹⁷ The Translog production function of the form:

$$\ln Q = \ln \alpha_0 + \sum_{i=1}^n \alpha_i \ln V_i + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \beta_{ij} \ln V_i \ln V_j$$

parameters.¹⁸ However, based on the fact that the Translog form is non-homothetic and non-separable, it is not self-dual.¹⁹

The elasticity of scale (ε) for the Translog function is defined by:

$$\varepsilon = \alpha_1 + \beta_1 + (2\alpha_2 + \gamma_1) \ln V_1 + (2\beta_2 + \gamma_1) \ln V_2 \quad (4.18)$$

From this it is clear that in general the elasticity of scale changes with factor proportions and with the level of production. In the absence of any *a priori* restrictions imposed on the elasticity of scale, it is therefore necessary to calculate the elasticity of scale for each point (V_1, V_2) on the production function.

The elasticity of substitution (σ) for the Translog function is defined by:

$$\sigma = -\frac{(A+B)}{Q} \left\{ A+B - 2\alpha_2 \left(\frac{A}{B} \right) - 2\beta_2 \left(\frac{B}{A} \right) - 2\gamma_1 \right\}^{-1} \quad (4.19)$$

where $A = \beta_1 + 2\beta_2 \ln V_2 + \gamma_1 \ln V_1$ and $B = \alpha_1 + 2\alpha_2 \ln V_1 + \gamma_1 \ln V_2$.

Thus the elasticity of substitution of the Translog function depends on the level of output and on the level of V_1 and V_2 .

The Translog cost function is not derived by optimising the Translog production function. The cost function is simply set up in the same way as the production function using costs and prices rather than quantities of inputs (Heathfield 1987: 110-111).

Thus, the non-homothetic and non-separable Translog cost function is defined by:

$$\ln C = \ln \alpha_0 + \alpha_q \ln q + \sum_{i=1}^n \alpha_i \ln p_i + \frac{1}{2} \beta_{qq} (\ln q)^2 + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \beta_{ij} \ln p_i \ln p_j + \sum_{i=1}^n \beta_{iq} \ln q \ln p_i \quad (4.20)$$

$(i \neq j; i, j = 1, \dots, n)$

The cost function must, however, satisfy the restriction of being homogenous of degree one in prices, since doubling all prices and leaving all quantities unchanged must double costs.

reduces to the Cobb-Douglas production function if the following parameter restrictions apply:

$$\beta_{ij} = \beta_{ji} = 0, \forall i, j$$

(Thomas 1993: 330; Nicholson 1995: 342)

¹⁸ See Appendix 11.

¹⁹ See Appendix 6.

²⁰ The elasticity of scale is rendered independent of the level of production (but dependent on factor proportions) if $\alpha_1 = -\beta_2$ in which case: $\varepsilon = \alpha_1 + \beta_1 + (2\alpha_2 + \gamma_1) \left[\ln \left(\frac{V_1}{V_2} \right) \right]$.

The elasticity of scale is rendered independent of V_1 by constraining $2\alpha_2 = -\gamma_1$ and independent of V_2 by constraining $2\beta_2 = -\gamma_1$.

The elasticity of scale is rendered independent of V_1 and V_2 altogether by constraining $\gamma_1 = -2\alpha_2 = -2\beta_2$. For this to be further constrained to constant returns to scale then $\alpha_1 + \beta_1 = 1$.

Constraining the cost function to be homogenous of degree 1 in prices requires parameter restrictions of:

$$\sum_{i=1}^n \alpha_i = 1 \text{ and } \sum_{i=1}^n \beta_{iq} = 0, \sum_{i=1}^n \beta_{ij} = \sum_{j=1}^n \beta_{ij} = 0, \text{ and } \beta_{ij} = \beta_{ji}$$

The underlying production technology associated with the above cost function will be homothetic if $\beta_{iq} = 0$ for all i , linearly homogeneous if $\alpha_q = 1$ and $\beta_{iq} = \beta_{iq} = 0$, and separable (linearly) if $\beta_{ij} = 0$ for all i and j ($i \neq j$) so that all of the cross-product terms of inputs are zero.

An interesting, but important relationship exists between the CES and the Translog functional forms, in that the CES functional form can be expanded into a Translog form, by making use of a Taylor series. The following section proves how this can be done. The necessity of this relationship will become apparent in section 4.5 where the relationship is used to prove that the functional form used to represent the South African economy, is indeed a true representation of the production technology in the South African economy (Fuss and McFadden 1979: 239).

4.3 AN EMPIRICAL PRODUCTION FUNCTION FOR THE SOUTH AFRICAN PRODUCTION SECTOR: A COBB-DOUGLAS APPROACH

In this section an aggregate production function, based on Cobb-Douglas technology, is estimated for the South African economy. The Engle and Yoo (1991) three-step cointegration estimation technique is employed. The resulting production function is subjected to comprehensive evaluation and testing to ensure that the function complies with the "full ideal principles" of model selection (Appendix 3).

4.3.1 The theoretical model

A cost function is estimated and utilised in the derivation of a production function. This derivation is based on the duality theory (Appendix 6). The form initially used to estimate the cost function is a Cobb-Douglas functional form. Therefore, in addition to the assumptions of homotheticity and homogeneity, the stringent assumption of unitary elasticity of substitution applies.

The method used to estimate this Cobb-Douglas cost function is the Engle and Yoo three-step cointegration procedure (Appendix 1). Duality principles are used to derive the production function from the resulting long-run cost function.

The economic theory of cost states that a firm's costs are a function of the input costs and output. It follows that the costs of both production factors (capital and labour) should have a positive impact on the cost of production. These two costs are therefore included in a cost function for the South African economy. In estimating the function, a positive coefficient for the cost of labour and the cost of capital may be expected. Another factor which should have an influence on the cost of production is the technology used to produce a certain output. The expected effect of technology on the cost of production is a negative one; technical advance should cause firms' costs to decline. The expected coefficient for technical progress should be negative in the estimated cost function.

Turning to the level of production, an increase in output should lead to an increase in the costs of firms and, once again, a positive coefficient is expected.

The empirical function for the cost of production (c) in South Africa therefore is:

$$c = f(\underset{+}{gdp@}, \underset{+}{factor\ cost}, \underset{+}{user\ cost\ of\ capital}, \underset{+}{labour\ cost}, \underset{-}{technical\ progress})$$

4.3.2 The data

An exposition on the data and related processes utilised in the estimation of the cost/production function is presented in Appendix 7. A variable list and graphical illustration of all the variables encountered in both the long-run cointegration and short-run error correction model are presented in Appendices 8 and 9 respectively.

4.3.3 The estimation results of the cointegration equation

4.3.3.1 A cointegration equation for the cost of production

The first step of the Engle and Yoo (1991) three-step estimation technique (Appendix 1) was employed to test whether the set of variables specified in the empirical model is cointegrated, i.e. whether the particular combination of variables is consistent with the long-run equilibrium relationship.

The long-run coefficients of the Cobb-Douglas cost function were estimated by restricting the sum of the coefficients of input prices to one. The restriction follows from economic theory, defining a cost function as homogeneous of degree 1 in input prices, that is, a doubling in input prices should lead to a doubling in cost. The cointegration results are reported in table 4.1.

Table 4.1 Cointegration equation: Cost of production

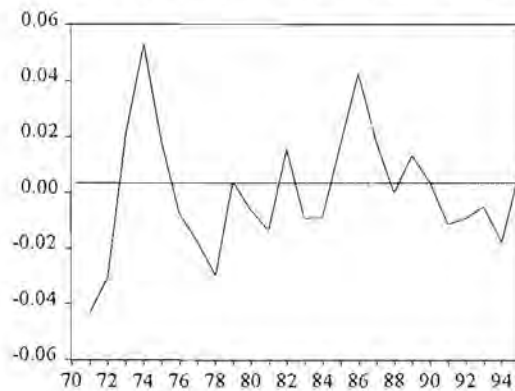
Dependent Variable: ln_c90p_lrem5				
Method: Least Squares				
Sample(adjusted): 1971 1995				
Included observations: 25 after adjusting endpoints				
ln_c90p_lrem5 = c(1)*ln_bbpfact_90p + c(2)*ln_ucc2_90p + (1-c(2))*ln_wtot_ppi_rat + c(3)*tecno_index				
	<i>Coefficient</i>	<i>Std. Error</i>	<i>t-Statistic</i>	<i>Prob.</i>
c(1)	0.300673	0.015849	18.97057	0.0000
c(2)	0.115779	0.011901	9.728241	0.0000

Table 4.1 (cont.)

c(3)	-0.079248	0.040007	-1.980867	0.0602
R-squared	0.988108	F-statistic	913.9758	
Adjusted R-squared	0.987027	Prob(F-statistic)	0.000000	

Comparing the Engle-Granger test statistic of -4.77 with the computed MacKinnon²¹ and the specified cointegration augmented Dickey-Fuller critical values respectively, resulted in the rejection of the null of no-cointegration in favour of stationary residuals and cointegrated variables. Figure 4.1 represents a plot of the stationary residuals.

Figure 4.1 Residuals: Cost of production (ln_c90p_lrem5)



4.3.3.2 Derivation of the long-run (cointegration) production function

By applying duality theory, the estimated Engle and Granger cointegration coefficients of the Cobb-Douglas cost function can be used to derive a consistent long-run Cobb-Douglas production function.

Before the production function can be derived, it is necessary to compute whether the economies of scale (r) are increasing, decreasing or constant. From the estimated equation it follows that average cost for the function always increases and therefore, technology in the South African production sector displays decreasing returns to scale. This result is confirmed by using the calculation of economies of scale, given a cost function with technology integrated into the equation, as described by Berndt (1991):

$$r = \frac{1}{1 + c(1)} = \frac{1}{1 + 0.300673} = 0.7688327$$

with $c(1)$ the estimated coefficient of gross domestic product (output) in the cost function.

²¹ Critical values for the relevant response surfaces can be found in MacKinnon (1991). The response surface for any number of regressors, excluding any constant and trend components, $1 \leq n \leq 6$, can be calculated as $C(p) = \phi_\infty + \phi_1 T^{-1} + \phi_2 T^{-2}$, where $C(p)$ is the p percent critical value.

By multiplying each of the input price coefficients and that of technology by the returns to scale (r), the coefficients of the inputs in the production function²² are obtained (Berndt 1991: 62 - 75).

These calculations are based on the principles of duality and are as follows:

$$\alpha = c(2) * r = 0.115779 * 0.300673 = 0.0890146$$

$$\beta = (1 - c(2)) * r = (1 - 0.115779) * 0.300673 = 0.7003954$$

$$\delta = c(3) * r = 0.079248 * 0.300673 = 0.0609284$$

With $c(2)$ the coefficient of user-cost-of-capital in the estimated long-run cointegrated cost function and $c(3)$ the coefficient of the cost of labour in the estimated long-run cointegrated cost function. From these results it is already apparent that production in South Africa is extremely labour-intensive, which has significant consequences for economic policy.

The Engle and Granger long-run cointegration Cobb-Douglas production function, in non-logarithmic form, can therefore be written as:

$$Q = A_0 e^{(0.0609284 * T)} K^{0.0890146} L^{0.7003954}$$

Although the function gives a good indication of the trend in the dependent variable, there is a difference in the level between the actual and estimated values. It is, therefore, necessary to estimate an associated constant (A_0 , i.e. the Hicks-neutral component of technical progress) for the production function. The ordinary least squares estimation results are reported in table 4.2.

Table 4.2 Cointegration equation: Real gross domestic product at factor cost

Dependent Variable: ln_bbpfact_90p				
Method: Least Squares				
Sample: 1970 1995				
Included observations: 26				
ln_bbpfact_90p=0.0890146*ln_kap_r+0.7003954*ln_n + 0.0609284*tecno_index + c(1)				
	<i>Coefficient</i>	<i>Std. Error</i>	<i>t-Statistic</i>	<i>Prob.</i>
c(1)	9.488127	0.005413	1757.760	0.0000
R-squared	0.969263			
Adjusted R-squared	0.969263			

²² The coefficients of the Cobb-Douglas production function $Q = A_0 e^{\delta T} K^\alpha L^\beta$, with K = real fixed capital; L = demand for labour; T = the technical index; α, β the coefficients of the elasticities of the inputs, and δ the marginal influence of technical progress on production, are derived by making use of the calculations illustrated by Berndt (1990).

4.3.4 The short-run dynamics: error correction model (ECM)

After the long-run cointegration relationship has been determined, the second stage of the Engle and Yoo procedure consists of an estimation of the error correction mechanism (ECM) (see Appendix 1) in order to capture the short-run or dynamic adjustment process to the long-run equilibrium. It incorporates the equilibrium error (residual terms) estimated from the long-run equilibrium relationship. The estimation results of the ECM are reported in table 4.3.

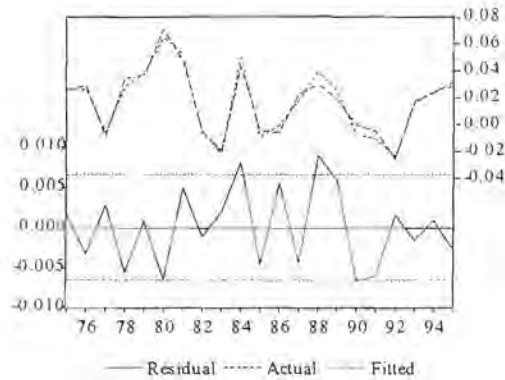
Table 4.3 Error correction model: Real gross domestic product at factor cost

Dependent Variable: $\Delta(\ln_bbpfact_90p)$				
Method: Least Squares				
Sample(adjusted): 1975 1995				
Included observations: 21 after adjusting endpoints				
<i>Variable</i>	<i>Coefficient</i>	<i>Std. Error</i>	<i>t-Statistic</i>	<i>Prob.</i>
residual(-1)	-0.445426	0.083627	-5.326362	0.0002
$\Delta(\ln_kap_r)$	1.565503	0.342552	4.570115	0.0008
$\Delta(\ln_kap_r(-1))$	-1.467576	0.294881	-4.976840	0.0004
$\Delta(\ln_n)$	0.606497	0.200341	3.027321	0.0115
$\Delta(\text{tecno_index})$	0.147462	0.050808	2.902366	0.0144
$\Delta(\text{tecno_index}(-4))$	-0.077070	0.032019	2.407010	0.0348
drought_dum	-0.036744	0.005245	-7.005830	0.0000
sanction_dum	-0.029635	0.005917	-5.008796	0.0004
imf_dum	-0.025683	0.005586	-4.597443	0.0008
c	0.025426	0.009005	2.823603	0.0166
R-squared	0.962824	F-statistic	31.65443	
Adjusted R-squared	0.932407	Prob(F-statistic)	0.000001	
S.E. of regression	0.006535			

Appendix 7 gives an explanation of the dummy variables included in the error correction model.

A data plot of the actual and fitted values of gross domestic production is provided below (figure 4.2).

Figure 4.2 Actual, fitted and residual values of $\ln_bbpfact_90p$



4.3.5 Diagnostic testing

The production function was submitted to rigorous diagnostic testing. Once again it must be noted that since all the variables in the ECM are stationary, the assumptions of classical regression analysis are fulfilled. Standard diagnostic tests can therefore be used to determine which variables should be included in the final specification of the ECM (Harris 1995: 24). The diagnostic test results reported in table 4.4 indicate that the function passes all these tests.

Table 4.4 Diagnostic tests: Real gross domestic product at factor cost ($\ln_bbpfact_90p$)

<i>Purpose of test</i>	<i>Test</i>	<i>d.f.</i>	<i>Test statistic</i>	<i>Probability</i>
Normality	Jarque-Bera	JB(2)	1.151718	[0.562222]
Homoscedasticity	ARCH LM	$nR^2(1)$	0.159922	[0.689229]
Homoscedasticity	White	$nR^2(15)$	17.88064	[0.268992]
Serial correlation	Breusch-Godfrey	$nR^2(2)$	3.286585	[0.193342]
Serial correlation	Lung Box Q	Q(12)	8.370700	[0.756000]
Misspecification	Ramsey Reset	LR(2)	2.478033	[0.138886]
Parameter stability	Recursive estimates	Indicative of stability		

4.3.5 Cointegration correction and adjusted coefficients

In this step, the Engle and Yoo technique (Appendix 1) is applied to adjust the coefficients and t -statistics so that they are closer to their true values. The variables included in the long-term regression can then be evaluated statistically, and the variables which are not statistically meaningful, can be discarded or adjusted. This approach is similar to the one used in classical regression analysis. Tables 4.5 and 4.6 summarise the third-step estimation results and the adjusted coefficients.

Table 4.5 Engle-Yoo third-step estimation: Real gross domestic product at factor cost (ln_bbpfact_90p)

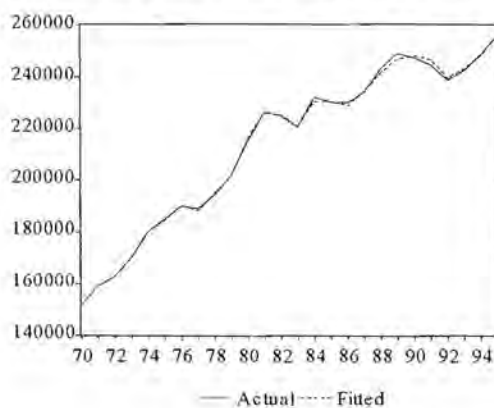
Dependent Variable: residual_ecm			
Method: Least Squares			
Sample(adjusted): 1975 1995			
Included observations: 21 after adjusting endpoints			
<i>Variable</i>	<i>Coefficient</i>	<i>Std. Error</i>	<i>t-Statistic</i>
(0.445426)*ln_kap_r	-0.010152	0.012699	-0.799432
(0.445426)*ln_n	0.072389	0.089060	0.812812
(0.445426)*tecno_index	-0.016105	0.019560	-0.823368

Table 4.6 Cointegration correction: Real gross domestic product at factor cost (ln_bbpfact_90p)

<i>Variable</i>	<i>Adjusted Coefficient</i>	<i>Adjusted t-Statistic</i>
ln_kap_r	0.0788626	6.2101425
ln_n	0.7727844	8.6771210
tecno_index	0.0448234	2.2915849

The Engle and Yoo adjusted coefficients are used to dynamically simulate the final version of the model, combining the long and short-run characteristics. The overall fit of the model is depicted in figure 4.3

Figure 4.3 Actual and fitted values of ln_bbpfact_90p



The fit of the estimated equation is evaluated in an *ex-post* simulation context by means of a number of quantitative measures. From the simulation error statistics (Appendix 2) reported in table 4.7 it can be concluded that the estimated equation represents a good fit of the corresponding actual data series.

Table 4.7 Simulation error statistics of real gross domestic product at factor cost

<i>Actual v Fitted : ln_bbpfact_90p</i>	
Root Mean Squared Error	0.005138
Mean Absolute Error	0.004348
Mean Absolute Percentage Error	0.035276
Theil Inequality Coefficient	0.000208
Bias Proportion	0.000000
Variance Proportion	0.004090
Covariance Proportion	0.995910

Economic evaluation of the estimation results renders what was expected *ex ante*. Although the South African production technology is in the process of improving its utilisation of capital and technology, the elasticity of labour is expected to be higher than that of capital since labour is still the most utilised production factor.

A noteworthy feature of the production function emerged from this investigation: that of decreasing returns to scale. The calculated returns to scale are 0.7688327. This implies that a 100 percent increase in both the inputs used in the production sector of our economy will increase production by only 77 percent. This result holds grave consequences for the South African economy and warrants further exploration.

4.4 DYNAMIC SIMULATION: RESPONSE PROPERTIES OF THE MODEL

Next, the dynamic simulation properties of the model are investigated and it is tested for stability and robustness simultaneously. The methodology applied is explained in Appendix 10.

The results of the adjustment process towards either a new long-run equilibrium (in accordance with the elasticities of the respective cointegration relationships) or the baseline equilibrium (in the case of short-run explanatory variables) are shown in the figures below. Vertical axes measure the difference between the outcome of the baseline estimation and the estimation subjected to the exogenous shock, as a percentage of the level of the dependent variable. The speed of adjustment in respective cases is apparent from the graphs. In all instances the adjustment process is completed within the sample range.

Table 4.8 indicates the level of convergence of the dependent variable, gross domestic production. All responses of gross domestic production were consistent with what was expected.

Table 4.8 Difference between the baseline forecast and forecasts with shocked variables, dependent variable: $\ln_bbpfact_90p$

<i>Variable</i>	<i>Coefficient</i>	<i>Expected change (10% of coefficient)</i>	<i>Convergence level (% difference)</i>
\ln_kap_r	0.0788626	0.0078863	0.007544
\ln_n	0.7727844	0.0772784	0.076434
$tecno_index$	0.0448234	0.0058920 ²³	0.005966

The results of the sensitivity tests documented in table 4.8 are portrayed in figures 4.4, 4.5 and 4.6.

Figure 4.4 Dynamic adjustment (percentage change) in real gross domestic product at factor cost ($\ln_bbpfact_90p$) with a 10 percent increase in real capital stock (\ln_kap_r)

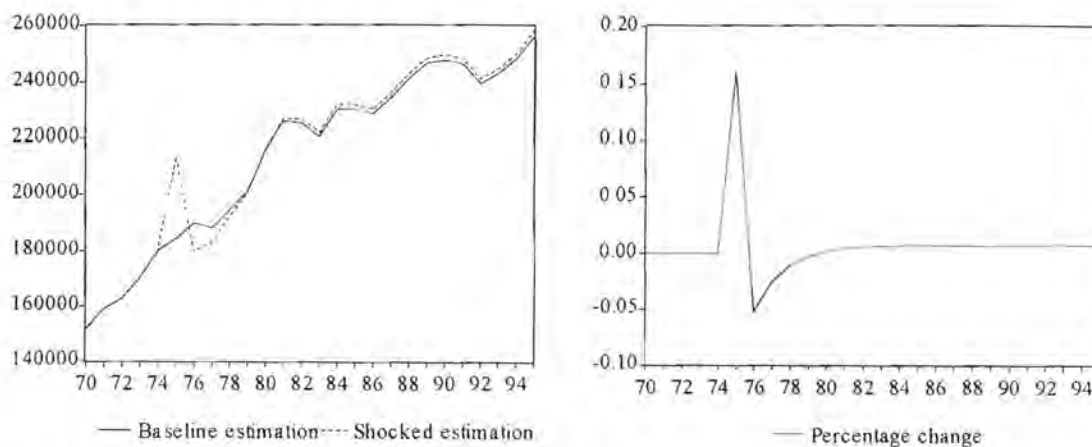


Figure 4.5 Dynamic adjustment (percentage change) in real gross domestic product at factor cost ($\ln_bbpfact_90p$) with a 10 percent increase in the level of employment (\ln_n)

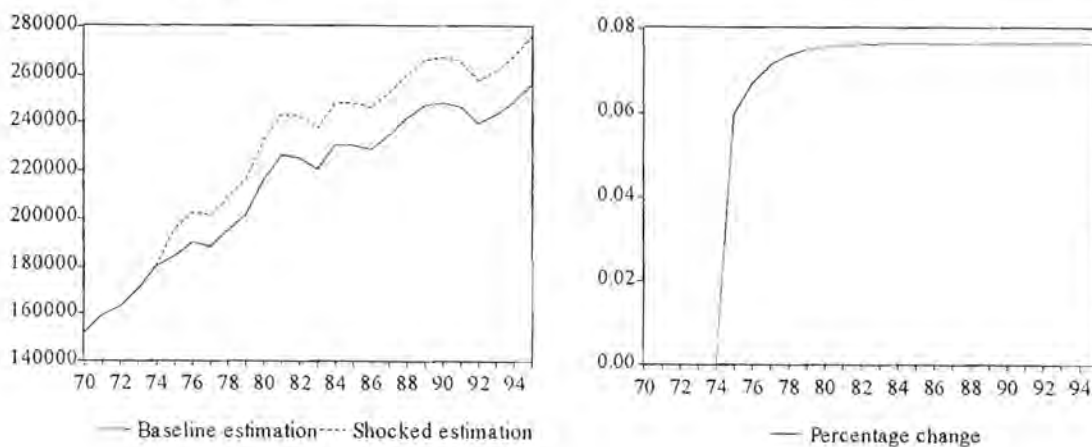
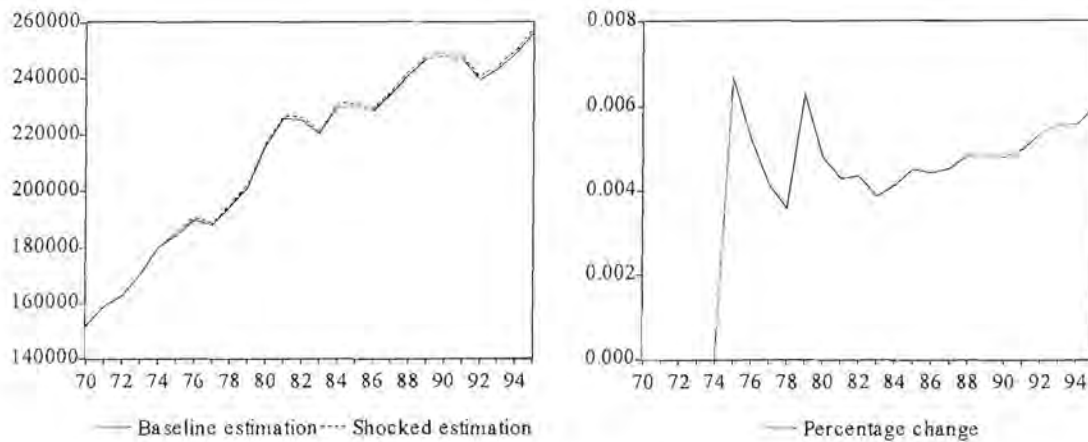


Figure 4.6 Dynamic adjustment (percentage change) in real gross domestic product at factor cost ($\ln_bbpfact_90p$) with a 10 percent increase in technology ($techo_index$)



4.5 TESTING THE VALIDITY OF A COBB-DOUGLAS REPRESENTATION

The Cobb-Douglas functional form has the advantage of being a highly stable and robust production function, which can easily be estimated. It also exhibits duality properties which allows for the easy transformation between the Cobb-Douglas cost and production functions. These properties are not evident in the more flexible functional forms. It remains to be proven, however, that the Cobb-Douglas functional form is indeed a valid representation of the technology used in the South African production sector.

Even though the Cobb-Douglas production function estimated above is satisfactory, the rather restrictive assumptions underlying this function are sufficient motivation to estimate a more flexible form. Given the advantages of the more flexible Translog functional form described earlier, the non-homothetic Translog cost function is estimated and tested for the validity of imposed restrictions to see whether it collapses to the Cobb-Douglas cost function. It is essential that this function collapses to a self-dual functional form, i.e. either a Cobb-Douglas or CES form, for the production function to be derived.

4.5.1 From a Translog to a Cobb-Douglas functional form: theoretical derivation

The goal is to estimate a Translog cost function and impose certain restrictions on the parameters of the estimated equation in order to test whether this function collapses to the Cobb-Douglas cost function estimated initially. The theoretical restrictions for the Translog functional form to collapse to the Cobb-Douglas functional form must be examined before empirical testing commences. (Berndt 1991).

²³ Since the relationship between $\ln_bbpfact_90p$ and $techo_index$ is of semi-log form, the elasticity of $\ln_bbpfact_90p$ with respect to $techo_index$ is not given by the coefficient of $techo_index$, but had to be calculated: $\epsilon = techo_index * \hat{a}$ (Studenmund 1997: 228).

The non-homothetic and non-separable Translog cost function is defined by:

$$\ln C = \ln \alpha_0 + \alpha_q \ln q + \sum_{i=1}^n \alpha_i \ln p_i + \frac{1}{2} \beta_{qq} (\ln q)^2 + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \beta_{ij} \ln p_i \ln p_j + \sum_{i=1}^n \beta_{iq} \ln q \ln p_i$$

with q real output and p_i and p_j input prices ($i \neq j$; $i, j = 1, \dots, n$). (4.20)

The cost function is homogenous of degree one in prices if the following restrictions apply:

$$\sum_{i=1}^n \alpha_i = 1 \text{ and } \sum_{i=1}^n \beta_{iq} = 0, \sum_{i=1}^n \beta_{ij} = \sum_{j=1}^n \beta_{ji} = 0, \text{ and } \beta_{ij} = \beta_{ji}. \quad (4.21)$$

In order for the function to be homothetic, it is necessary and sufficient that $\beta_{iq} = 0$ for all i . Furthermore, a restriction of homogeneity of a constant degree in output occurs if, in addition to these homotheticity restrictions, $\beta_{qq} = 0$. In this case the degree of homogeneity equals $1/\alpha_q$. The function is separable (linearly) if $\beta_{ij} = 0$, for all i and j ($i \neq j$).²⁴ Constant returns to scale of the dual production function occurs when, in addition to the above homotheticity and homogeneity restrictions, $\alpha_q = 1$. Therefore, the Translog function reduces to the constant-returns-to-scale Cobb-Douglas function when all the above restrictions apply simultaneously.

The Translog production function can be defined by:

$$\ln Q = \ln \alpha_0 + \sum_{i=1}^n \alpha_i \ln V_i + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \beta_{ij} \ln V_i \ln V_j \quad (4.22)$$

where $\beta_{ij} = \beta_{ji}$ and $i \neq j$; $i, j = 1, \dots, n$.

For the production function to be homothetic, and homogeneous of a constant degree in output it is necessary and sufficient that $\sum \beta_{ij} = \sum \beta_{ji} = 0, \forall i, j$ (Allen 1997: 28; Chung 1994: 142-143; Nicholson 1995: 342). The function is separable (linearly) and reduces to the Cobb-Douglas or CES production function if $\beta_{ij} = \beta_{ji} = 0, \forall i, j$ (Allen 1997: 37; Chung 1994: 243). Constant returns to scale of the dual production function occurs when, in addition to the above homotheticity and homogeneity restrictions, $\alpha_i = 1$ (Allen 1997: 28; Chung 1994: 142-143; Nicholson 1995: 342).

These theoretical principles can now be applied to an estimated Translog cost function and Wald tests can be performed in order to test the restrictions for validity.

²⁴ If a production function is additive (or separable) and homothetic, the Allen-Uzawa partial elasticities of substitution are equal to the elasticity of substitution (σ) and are therefore equal and constant for all pairs of inputs (Chung 1994: 189, 205). Furthermore, a production function which in addition to being linearly homogeneous and separable, has a Cobb-Douglas or CES structure. (Allen 1997: 19, 35).

4.5.2 Empirical testing of imposed restrictions

The process started with the estimation of an unrestricted, i.e. non-homothetic and non-separable Translog cost function, which is well-behaved in the sense that it is homogeneous of degree one in prices.²⁵

The restrictions of homotheticity²⁶, homogeneity of a constant degree in output²⁷ and linear separability²⁸ were imposed step-by-step and tested for validity by means of Wald tests, in order to establish whether the Translog function breaks down to the Cobb-Douglas functional form.

It was found that, given the empirical testing of the imposed restrictions and the theoretical interpretation of the valid restrictions, the restrictions were valid for a Translog cost function in the South African economy. The results of the Wald tests are reported in table 4.9 below.

Table 4.9 Results of the Wald tests performed on the imposed restrictions on the Translog cost function

<i>Test: restriction (Null Hypothesis)</i>	<i>Probability: F-statistic</i>	<i>Probability: Chi-square statistic</i>	<i>Valid/Not valid</i>
c(7) = 0	0.637064	0.631284	Valid restriction
c(4) = 0	0.971763	0.971367	Valid restriction
c(5) and c(6) = 0	0.149564	0.120575	Valid restriction

The non-homothetic Translog cost function collapses to a cost function which is homothetic, homogeneous of a constant degree in output and linearly separable. The function therefore exhibits a constant elasticity of substitution. That is, the imposed restrictions of homotheticity,

$$^{25} \ln C = \ln \alpha_0 + \alpha_q \ln q + \sum_{i=1}^n \alpha_i \ln p_i + \frac{1}{2} \beta_{qq} (\ln q)^2 + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \beta_{ij} \ln p_i \ln p_j + \sum_{i=1}^n \beta_{iq} \ln q \ln p_i \quad \text{with}$$

$$\sum_{i=1}^n \alpha_i = 1 \quad \text{and} \quad \sum_{i=1}^n \beta_{iq} = 0, \quad \sum_{i=1}^n \beta_{ij} = \sum_{j=1}^n \beta_{ij} = 0, \quad \text{and} \quad \beta_{ij} = \beta_{ji}, \quad \text{for the function to be homogeneous of degree one in prices. The function is empirically estimated by:}$$

$$\begin{aligned} \text{Cost of production} = & c(2) * \text{cost of capital} + (1-c(2)) * \text{cost of labour} + c(3) * \text{gross domestic product} + \\ & 0.5 * c(4) * (\text{gross domestic product})^2 + 0.5 * [c(5) * (\text{cost of capital})^2 + c(6) * (\text{cost of} \\ & \text{labour})^2 + (0-c(5)-c(6)) * [(\text{cost of labour}) * (\text{cost of capital})]] + c(7) * [(\text{gross domestic} \\ & \text{product}) * (\text{cost of capital})] + (0-c(7)) * [(\text{gross domestic product}) * (\text{cost of labour})] + \\ & c(8) * \text{technical index.} \end{aligned}$$

$$^{26} \beta_{iq} = 0 \quad \forall i = 1, \dots, n; \quad \text{i.e. } c(7) = 0.$$

$$^{27} \beta_{qq} = 0, \quad \text{i.e. } c(4) = 0. \quad \text{If a function is homogeneous of a constant degree in output it is possible to derive the degree of homogeneity and the economies of scale, which is equal to } \frac{1}{\alpha_q}, \quad \text{i.e. } \frac{1}{c(3)} \quad (\text{Berndt 1991: 69-70}) \text{ or}$$

$$\frac{1}{1+\alpha_q}; \quad \text{i.e. } \frac{1}{1+c(3)} \quad \text{with the incorporation of technical progress (Berndt 1991: 71-75).}$$

$$^{28} \beta_{ij} = 0; \quad \text{i.e. } c(5) = c(6) = 0.$$

homogeneity and separability on the estimated Translog cost function passes the Wald tests. It is therefore possible to derive the degree of homogeneity and the economies of scale. It can be concluded that the Translog cost function breaks down to either a Cobb-Douglas or CES functional form, exhibiting a constant elasticity of substitution. The next step is to test the validity of a CES function and to determine the value of the elasticity of substitution.

Due to the fact that the Translog functional form is not self-dual, it is plausible to assume that the estimation results of a Translog production function will differ from those of the Translog cost function with regard to production technology. It is therefore necessary to estimate an equivalent Translog production function by imposing the justifiable restrictions as tested for the Translog cost function and determine whether this function collapses to the Cobb-Douglas production function.

4.5.3 Testing for elasticity of substitution

Given the fact that the cost function exhibits the properties of homotheticity, homogeneity of a constant degree in output and linear separability, it is justifiable to assume constant elasticity of substitution in the production sector. The elasticity of substitution is, however, not necessarily equal to unity. Therefore, either a CES or Cobb-Douglas production function could be justified as being representative of the production structure in the South African economy as both these functional forms have a constant elasticity of substitution.

However, it should be established whether Kmenta's Taylor approximation of the Translog functional form into a CES-functional form is valid (Appendix 11) and to use Kmenta's Taylor approximation of the CES function to estimate the elasticity of substitution (Thomas 1993: 331).

For this purpose, it is necessary to estimate the homothetic (although still non-separable) Translog production function²⁹ and to test the validity of the imposed restriction of a constant elasticity of substitution (i.e. the restriction of separability³⁰) (Thomas 1993: 331). The Translog production function passed the Wald test for a constant elasticity of substitution and therefore the validity of Kmenta's Taylor approximation of the CES production function. The Wald test results are reported in table 4.10.

²⁹ The homothetic (although still non-separable) Translog production function is defined by:

$$\ln Q = \ln \alpha_0 + \sum_{i=1}^n \alpha_i \ln V_i + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \beta_{ij} \ln V_i \ln V_j \text{ with } \sum \beta_{ij} = 0; \beta_{ij} = \beta_{ji} \text{ and } i \neq j; i, j = 1, \dots, n. \text{ This}$$

function is empirically estimated by:

$$\text{Gross domestic product} = c(1) * \text{real capital} + c(2) * \text{labour} + c(3) * (\text{real capital})^2 + c(4) * (\text{labour})^2 + (0 - c(3) - c(4)) * [(\text{real capital}) * (\text{labour})] + c(6) * \text{technical index.}$$

³⁰ In order for the homothetic Translog production function to exhibit the property of separability, i.e. a constant elasticity of substitution, it has to obey the restrictions of: $\beta_{ii} = \beta_{jj} = -\frac{1}{2}\beta_{ij}$, i.e. $c(3) = c(4)$ in the case of a homothetic function (Thomas 1993: 331).

Table 4.10 Results of the Wald test performed on the imposed restrictions on the Translog production function

<i>Test: restriction</i>	<i>Probability:</i>	<i>Probability:</i>	<i>Valid/Not valid</i>
<i>(Null Hypothesis)</i>	<i>F-statistic</i>	<i>Chi-square statistic</i>	
$c(3) = c(4)$	0.811379	0.809035	Valid restriction

Kmenta's CES production function could therefore be estimated to test the elasticity of substitution of the production technology.³¹

This estimation of the CES production function resulted in a θ -coefficient (substitution parameter) close to zero ($\theta = 0.007479$). This represents an elasticity of substitution significantly near unity:

$$\sigma = \frac{1}{(1 + \theta)} = 0.99$$

This is an interesting result. The elasticity of substitution can be defined as the percentage change in the capital-labour ratio, relative to the percentage change in the price ratio: $\sigma = \frac{\partial \ln(K/L)}{\partial \ln(p_L/p_K)}$.

The implication of a unitary elasticity of substitution is that the percentage change in the capital-labour ratio is equal to the percentage change in the price ratio.

This is of particular significance in the South African context, in that the price-ratio dictates the capital-labour ratio. The higher the increase in the price of labour relative to the price of capital, the lower the demand for labour relative to the demand for capital. This confirms the phenomenon in the South African economy of an increasing capital-labour ratio as a result of the rising labour-capital price ratio.

It is shown that the CES production function is in fact a Cobb-Douglas production function when the CES production function features unitary elasticity of substitution. It is therefore concluded that the Cobb-Douglas technology is representative of the production technology of the South African economy.

³¹ Kmenta's CES production function was estimated according to the method of Griliches and Ringstad (1971) in order to deal with the possibility of a high degree of multicollinearity between the variables. They rearranged Kmenta's Taylor approximation as:

$$\ln\left(\frac{Q}{V_2}\right) = \ln \gamma + (\nu - 1) \ln V_2 + (\nu \delta) \ln\left(\frac{V_1}{V_2}\right) - \frac{1}{2} \nu \theta \delta (1 - \delta) \left[\ln\left(\frac{V_1}{V_2}\right) \right]^2$$

Estimation of the equation resulted in estimates for γ , ν , δ and θ and therefore information on the properties such as the returns to scale (ν), capital intensity of production (δ) and elasticity of substitution ($\sigma = 1/(1 + \theta)$). The closeness of θ to zero serves as a further test whether the Kmenta approximation to the CES function is valid. The estimation procedure used for the estimation of the CES production function is the Engle and Yoo three-step procedure. Because this estimation is only done in order to test the elasticity of substitution of production technology, the economic and statistical results are not reported. The coefficients are however consistent with economic theory and the residuals from the Engle and Yoo cointegration equation are stationary.

4.6 CONCLUSION

The purpose of this chapter is to estimate an aggregate neoclassical production function for the South African economy and to investigate the long-run properties of the production structure.

Recognising the advantages of such an approach, the analysis was based on the estimation of a cost function for the South African production sector and the subsequent derivation of a production function based on duality principles. However, only homothetic and linearly separable functional forms, such as the more restricted Cobb-Douglas and CES functions, are self-dual.

It had to be proven, therefore, that the production structure of the South African economy features homotheticity and linear separability, i.e. a constant elasticity of substitution.

In order to test the validity of either a Cobb-Douglas or CES functional form as a representation of the technology in the South African economy, a Translog cost function was estimated and tested for the validity of imposed restrictions. The Translog cost function could be collapsed to a homothetic and linearly separable cost function. By making use of Kmenta's Taylor approximation of the CES function, it was further proven that the function not only exhibits a constant elasticity of substitution, but that it is very close to unity. It is therefore concluded that a Cobb-Douglas functional form can be used as a representation (approximation) of the production structure of South Africa.

An evaluation of the estimation results obtained from both the Cobb-Douglas and CES functions, led to a couple of interesting long-run properties of the South African economy. Apart from a unitary elasticity of substitution, which implies that the price-ratio dictates the capital-labour ratio, it was concluded that South Africa produces with decreasing returns to scale.

The following important, although not surprising properties, may be attributed to the South African production and growth structure:

- (i) Production in South Africa is labour intensive with an output-elasticity of 0.77, stressing the importance of all labour-related issues such as wages, level of skill, the role of labour unions and labour legislation.
- (ii) A interesting feature of the production function is the decreasing returns to scale observed. The Engle-Yoo adjusted returns to scale is 0.85. Returns to scale of 0.85 implies that a 100 percent increase in both the inputs used in the production sector of our economy will increase production by only 85 percent. This result holds serious consequences for the South African economy. It must be noted that few empirical studies on the aggregate production structure of the total South African economy have been conducted and information for comparative purposes is not readily available.
- (iii) The fact that the Cobb-Douglas production technology is representative of the South African production structure, is confirmation of an unitary elasticity of substitution. The

4.5.2 Empirical testing of imposed restrictions

The process started with the estimation of an unrestricted, i.e. non-homothetic and non-separable Translog cost function, which is well-behaved in the sense that it is homogeneous of degree one in prices.²⁵

The restrictions of homotheticity²⁶, homogeneity of a constant degree in output²⁷ and linear separability²⁸ were imposed step-by-step and tested for validity by means of Wald tests, in order to establish whether the Translog function breaks down to the Cobb-Douglas functional form.

It was found that, given the empirical testing of the imposed restrictions and the theoretical interpretation of the valid restrictions, the restrictions were valid for a Translog cost function in the South African economy. The results of the Wald tests are reported in table 4.9 below.

Table 4.9 Results of the Wald tests performed on the imposed restrictions on the Translog cost function

<i>Test: restriction (Null Hypothesis)</i>	<i>Probability: F-statistic</i>	<i>Probability: Chi-square statistic</i>	<i>Valid/Not valid</i>
$c(7) = 0$	0.637064	0.631284	Valid restriction
$c(4) = 0$	0.971763	0.971367	Valid restriction
$c(5) \text{ and } c(6) = 0$	0.149564	0.120575	Valid restriction

The non-homothetic Translog cost function collapses to a cost function which is homothetic, homogeneous of a constant degree in output and linearly separable. The function therefore exhibits a constant elasticity of substitution. That is, the imposed restrictions of homotheticity,

$$^{25} \ln C = \ln \alpha_0 + \alpha_q \ln q + \sum_{i=1}^n \alpha_i \ln p_i + \frac{1}{2} \beta_{qq} (\ln q)^2 + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \beta_{ij} \ln p_i \ln p_j + \sum_{i=1}^n \beta_{iq} \ln q \ln p_i \quad \text{with}$$

$$\sum_{i=1}^n \alpha_i = 1 \quad \text{and} \quad \sum_{i=1}^n \beta_{iq} = 0, \quad \sum_{i=1}^n \beta_{ij} = \sum_{j=1}^n \beta_{ji} = 0, \quad \text{and} \quad \beta_{ij} = \beta_{ji}, \quad \text{for the function to be homogeneous of}$$

degree one in prices. The function is empirically estimated by:

$$\begin{aligned} \text{Cost of production} = & c(2) * \text{cost of capital} + (1-c(2)) * \text{cost of labour} + c(3) * \text{gross domestic product} + \\ & 0.5 * c(4) * (\text{gross domestic product})^2 + 0.5 * [c(5) * (\text{cost of capital})^2 + c(6) * (\text{cost of} \\ & \text{labour})^2 + (0-c(5)-c(6)) * [(\text{cost of labour}) * (\text{cost of capital})]] + c(7) * [(\text{gross domestic} \\ & \text{product}) * (\text{cost of capital})] + (0-c(7)) * [(\text{gross domestic product}) * (\text{cost of labour})] + \\ & c(8) * \text{technical index.} \end{aligned}$$

$$^{26} \beta_{iq} = 0 \quad \forall i = 1, \dots, n, \quad \text{i.e. } c(7) = 0.$$

$$^{27} \beta_{qq} = 0, \quad \text{i.e. } c(4) = 0. \quad \text{If a function is homogeneous of a constant degree in output it is possible to derive the degree of homogeneity and the economies of scale, which is equal to } \frac{1}{\alpha_q}; \quad \text{i.e. } \frac{1}{c(3)} \quad (\text{Berndt 1991: 69-70}) \text{ or}$$

$$\frac{1}{1 + \alpha_q}; \quad \text{i.e. } \frac{1}{1 + c(3)} \quad \text{with the incorporation of technical progress (Berndt 1991: 71-75).}$$

$$^{28} \beta_{ij} = 0; \quad \text{i.e. } c(5) = c(6) = 0.$$

homogeneity and separability on the estimated Translog cost function passes the Wald tests. It is therefore possible to derive the degree of homogeneity and the economies of scale. It can be concluded that the Translog cost function breaks down to either a Cobb-Douglas or CES functional form, exhibiting a constant elasticity of substitution. The next step is to test the validity of a CES function and to determine the value of the elasticity of substitution.

Due to the fact that the Translog functional form is not self-dual, it is plausible to assume that the estimation results of a Translog production function will differ from those of the Translog cost function with regard to production technology. It is therefore necessary to estimate an equivalent Translog production function by imposing the justifiable restrictions as tested for the Translog cost function and determine whether this function collapses to the Cobb-Douglas production function.

4.5.3 Testing for elasticity of substitution

Given the fact that the cost function exhibits the properties of homotheticity, homogeneity of a constant degree in output and linear separability, it is justifiable to assume constant elasticity of substitution in the production sector. The elasticity of substitution is, however, not necessarily equal to unity. Therefore, either a CES or Cobb-Douglas production function could be justified as being representative of the production structure in the South African economy as both these functional forms have a constant elasticity of substitution.

However, it should be established whether Kmenta's Taylor approximation of the Translog functional form into a CES-functional form is valid (Appendix 11) and to use Kmenta's Taylor approximation of the CES function to estimate the elasticity of substitution (Thomas 1993: 331).

For this purpose, it is necessary to estimate the homothetic (although still non-separable) Translog production function²⁹ and to test the validity of the imposed restriction of a constant elasticity of substitution (i.e. the restriction of separability³⁰) (Thomas 1993: 331). The Translog production function passed the Wald test for a constant elasticity of substitution and therefore the validity of Kmenta's Taylor approximation of the CES production function. The Wald test results are reported in table 4.10.

²⁹ The homothetic (although still non-separable) Translog production function is defined by:

$$\ln Q = \ln \alpha_0 + \sum_{i=1}^n \alpha_i \ln V_i + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \beta_{ij} \ln V_i \ln V_j \text{ with } \sum \beta_{ij} = 0, \beta_{ij} = \beta_{ji} \text{ and } i \neq j; i, j = 1, \dots, n. \text{ This}$$

function is empirically estimated by:

$$\text{Gross domestic product} = c(1) * \text{real capital} + c(2) * \text{labour} + c(3) * (\text{real capital})^2 + c(4) * (\text{labour})^2 + (0 - c(3) - c(4)) * [(\text{real capital}) * (\text{labour})] + c(6) * \text{technical index.}$$

³⁰ In order for the homothetic Translog production function to exhibit the property of separability, i.e. a constant elasticity of substitution, it has to obey the restrictions of: $\beta_{ii} = \beta_{jj} = -\frac{1}{2}\beta_{ij}$, i.e. $c(3) = c(4)$ in the case of a homothetic function (Thomas 1993: 331).

Table 4.10 Results of the Wald test performed on the imposed restrictions on the Translog production function

<i>Test: restriction</i>	<i>Probability:</i>	<i>Probability:</i>	<i>Valid/Not valid</i>
<i>(Null Hypothesis)</i>	<i>F-statistic</i>	<i>Chi-square statistic</i>	
$c(3) = c(4)$	0.811379	0.809035	Valid restriction

Kmenta's CES production function could therefore be estimated to test the elasticity of substitution of the production technology.³¹

This estimation of the CES production function resulted in a θ -coefficient (substitution parameter) close to zero ($\theta = 0.007479$). This represents an elasticity of substitution significantly near unity:

$$\sigma = \frac{1}{(1 + \theta)} = 0.99.$$

This is an interesting result. The elasticity of substitution can be defined as the percentage change in the capital-labour ratio, relative to the percentage change in the price ratio: $\sigma = \frac{\partial \ln(K/L)}{\partial \ln(p_L/p_K)}$.

The implication of a unitary elasticity of substitution is that the percentage change in the capital-labour ratio is equal to the percentage change in the price ratio.

This is of particular significance in the South African context, in that the price-ratio dictates the capital-labour ratio. The higher the increase in the price of labour relative to the price of capital, the lower the demand for labour relative to the demand for capital. This confirms the phenomenon in the South African economy of an increasing capital-labour ratio as a result of the rising labour-capital price ratio.

It is shown that the CES production function is in fact a Cobb-Douglas production function when the CES production function features unitary elasticity of substitution. It is therefore concluded that the Cobb-Douglas technology is representative of the production technology of the South African economy.

³¹ Kmenta's CES production function was estimated according to the method of Griliches and Ringstad (1971) in order to deal with the possibility of a high degree of multicollinearity between the variables. They rearranged Kmenta's Taylor approximation as:

$$\ln\left(\frac{Q}{V_2}\right) = \ln \gamma + (v-1)\ln V_2 + (v\delta)\ln\left(\frac{V_1}{V_2}\right) - \frac{1}{2}v\theta\delta(1-\delta)\left[\ln\left(\frac{V_1}{V_2}\right)\right]^2.$$

Estimation of the equation resulted in estimates for γ , v , δ and θ and therefore information on the properties such as the returns to scale (v), capital intensity of production (δ) and elasticity of substitution ($\sigma = 1/(1 + \theta)$). The closeness of θ to zero serves as a further test whether the Kmenta approximation to the CES function is valid. The estimation procedure used for the estimation of the CES production function is the Engle and Yoo three-step procedure. Because this estimation is only done in order to test the elasticity of substitution of production technology, the economic and statistical results are not reported. The coefficients are however consistent with economic theory and the residuals from the Engle and Yoo cointegration equation are stationary.

4.6 CONCLUSION

The purpose of this chapter is to estimate an aggregate neoclassical production function for the South African economy and to investigate the long-run properties of the production structure.

Recognising the advantages of such an approach, the analysis was based on the estimation of a cost function for the South African production sector and the subsequent derivation of a production function based on duality principles. However, only homothetic and linearly separable functional forms, such as the more restricted Cobb-Douglas and CES functions, are self-dual.

It had to be proven, therefore, that the production structure of the South African economy features homotheticity and linear separability, i.e. a constant elasticity of substitution.

In order to test the validity of either a Cobb-Douglas or CES functional form as a representation of the technology in the South African economy, a Translog cost function was estimated and tested for the validity of imposed restrictions. The Translog cost function could be collapsed to a homothetic and linearly separable cost function. By making use of Kmenta's Taylor approximation of the CES function, it was further proven that the function not only exhibits a constant elasticity of substitution, but that it is very close to unity. It is therefore concluded that a Cobb-Douglas functional form can be used as a representation (approximation) of the production structure of South Africa.

An evaluation of the estimation results obtained from both the Cobb-Douglas and CES functions, led to a couple of interesting long-run properties of the South African economy. Apart from a unitary elasticity of substitution, which implies that the price-ratio dictates the capital-labour ratio, it was concluded that South Africa produces with decreasing returns to scale.

The following important, although not surprising properties, may be attributed to the South African production and growth structure:

- (i) Production in South Africa is labour intensive with an output-elasticity of 0.77, stressing the importance of all labour-related issues such as wages, level of skill, the role of labour unions and labour legislation.
- (ii) A interesting feature of the production function is the decreasing returns to scale observed. The Engle-Yoo adjusted returns to scale is 0.85. Returns to scale of 0.85 implies that a 100 percent increase in both the inputs used in the production sector of our economy will increase production by only 85 percent. This result holds serious consequences for the South African economy. It must be noted that few empirical studies on the aggregate production structure of the total South African economy have been conducted and information for comparative purposes is not readily available.
- (iii) The fact that the Cobb-Douglas production technology is representative of the South African production structure, is confirmation of an unitary elasticity of substitution. The



implication of a unitary elasticity of substitution is that the percentage change in the capital-labour ratio is equal to the percentage change in the price ratio.

This is of particular significance in the South African context, since the price-ratio dictates the capital-labour ratio. The higher the increase in the price of labour relative to the price of capital, the lower the demand for labour relative to the demand for capital. This confirms the phenomenon in the South African economy of an increasing capital-labour ratio as a result of the rising labour-capital price ratio.

From the above analyses it is clear that South Africa, which has an abundance of relatively expensive unskilled labour, can benefit materially by addressing structural unemployment through education and training. Improved training and education will enhance productivity and ultimately production through both rising employment and technical progress. Increased output growth will, in turn, stimulate the demand for capital (investment), the demand for skilled labour as well as technical progress. The process becomes self-sustaining, as output growth leads to further increases in employment, productivity and ultimately again economic growth.