CHAPTER 2: LITERATURE STUDY

2.1 SUPPLY ANALYSIS DEFINED

Supply analysis is a term used to refer to a larger set of techniques that evaluate production responses to output-, input prices and other measurable policy and environmental changes. The theory of the firm is the basis from which analysis is conducted [Colman, 1983].

2.2 THE PURPOSE OF SUPPLY ANALYSIS

Analysis of the supply behaviour of firms, aims to improve the understanding of how producers combine inputs in the production process. Agricultural producers are both users of resources and suppliers of agricultural products. The production technology underlying this process can be described through production elasticities, input substitution possibilities, returns to scale and the bias in technology [Thijssen, 1992]. Description of production technology is necessary [Nerlove and Bachman, 1960] and the determinants of agricultural production functions are important as far as they influence the supply of agricultural products and affect the use of inputs (under varying technological conditions) [Thijssen, 1992].

Sadoulet and De Janvry [1995] discuss the application of different techniques, supply analysis amongst others, for quantitative policy analysis as one of the means by which to bridge the gap between the real world and the realm of theory. Many of the complex indirect effects of policies can be captured and better understood through empirical modelling and statistical testing, in order to give clarification on the role of behavioural assumptions and causal relationships that are implicit in the models.

Once a better understanding of supply decision-making exists, analysts can predict how farmers would respond to external factors and policy changes. They can forecast supply- and demand trends, and control and influence supply to induce desired changes in employment patterns, resource allocation, and environmental management [Hassan, 1998].

2.3 APPROACHES TO SUPPLY ANALYSIS

Approaches to supply analysis can be classified into two main groups: normative (programming) and positive (econometric) approaches [Shumway and Chang, 1977; Colman, 1983; Sadoulet and De Janvry, 1995]. Substantial advances occurred in demand analysis, especially in terms of the use and acceptability of complete systems of demand equations [Pope, 1982; Deaton and Muellbauer, 1980]. The continued efforts of analyst to develop comparable complete system approaches for supply[8] response has been less successful because of the higher degree of complexity inherent in supply response [Colman, 1983]. For this reason, various considerably differing approaches (and opinions on their appropriateness) exist for supply response analysis.

2.3.1 NORMATIVE APPROACHES

Normative approaches typically combine historical and artificial (generated) data and impose behavioural assumptions in programming models that attempt to determine optimal choices, i.e. "what ought to be" [Shumway and Chang, 1977; Hassan, 1998].

Linear and Non-linear Programming (LP and NLP) and the method of Data Envelopment Analysis (DEA) enjoy wide support as the most frequently used normative techniques [Colman, 1983; Jaforullah and Whiteman, 1999]. Colman [1983] ascribes the popularity of programming methods to their adherence to the theoretical steps for deriving profit maximising levels of output and inputs.

---

[8] Colman [1983] elaborates on the semantic error of referring to supply response instead of output or production response. In this study, the term "supply response" is used, despite the apparent semantic discrepancy.

Sets of production functions for all the outputs, as well as sets of resource use restrictions are specified. Together with the objective function, (e.g. profit maximisation, with or without risk aversion) these sets are used in programming models that optimise output and input levels to obtain maximum profits or minimum cost, subject to exogenous prices and technologies [Colman, 1983; Shumway and Chang, 1977]. Supply response is derived from parametric programming that solves for different levels of optimal quantities at specified price levels [Sadoulet and De Janvry, 1995]. Iterations are performed for each farm or representative farm in the set over the range of specified prices, to arrive at a set of (partial) price-input-output relations.

Normative supply approaches have a very important advantage over positive methods. At farm level, the normative (programming) approaches are able to take account of all product and input prices, all relevant technological, institutional and physical restrictions [Colman, 1983]. In positive (econometric) studies, incorporation of all these aspects is mostly inhibited because of limited sample sizes. The normative approach suffers from various problems, the most important of which is aggregation bias that can be minimised, yet, not eliminated [Shumway and Chang, 1977; Day, 1963].

A few examples of non-parametric approaches are found in Cowling and Baker [1963], Sheehy and McAlexander [1965], Shumway and Chang [1977], Thompson and Buckwell [1979], Tauer [1998] and Jaforullah and Whiteman [1999], amongst numerous others.

2.3.2 *Positive Approaches*

Positive or econometric approaches are broadly classified into two sub-groups: the primal approach and the dual approach [Colman, 1983; Sadoulet and De Janvry, 1995].

The primal approach involves estimation of the structural production function or frontier from cross-sectional or time-series data. This function is given a largely arbitrary form upon which profit maximising marginal conditions are imposed to derive the supply and demand equations. [Sadoulet and De Janvry, 1995]. A few problems are associated with this method [Colman, 1983]. Simultaneity bias occurs between inputs and outputs unless experimental data is used. This problem arises from the joint and simultaneous determination of levels of inputs and resulting levels of outputs [Lau and Yotopoulos, 1972; Colman, 1983]. In addition, partial adjustments and adaptive expectations are not taken into account, resulting in over-estimation of short run elasticities and the supply elasticities are very sensitive to the chosen functional form [Wipf and Bawden, 1969; Burgess, 1975; Anderson, *et al.*, 1996]. Vide Section 2.4.

The dual or reduced form approach involves estimation of a profit function from either cross-sectional data (that shows inter-farm variation in effective prices) or from long run time series that show variation in fixed factors, or from a combination of the two data types [Sadoulet and De Janvry, 1995]. Supply and factor demand are derived analytically. Alternatively, complete systems of supply response can be estimated from the underlying profit function. In this case, cross-equation restrictions are imposed on parameters so that the system derives rigorously from a profit or cost function [Ray, 1982; Colman, 1983; Higgins, 1986; Thijssen, 1992; Sadoulet and De Janvry, 1995].

The alternative to primal or dual methods is ad hoc specification of supply response (including partial adjustment and expectations formation). This pertains mainly to Nerlovian supply response models that use time series data for one commodity and prices of a few relevant commodities. It has minimal theoretical requirements but consequently also yields unreliable results [Colman, 1983; Sadoulet and De Janvry, 1995]. A final set of methods employ estimation of complete structural models of supply and demand equilibrium (where both prices and quantities are treated as endogenous) for the economy as a whole or for relevant sub-sectors, from which the supply response is derived by simulation (e.g. multimarket or CGE models) [Sadoulet and De Janvry, 1995].

2.3.2.1 THE PRIMAL APPROACH

The first step of the primal approach involves estimation of a production (single output) or transformation (multi-output) function [Debertin, 1986; Chambers, 1988; Färe and Primont, 1995]. In real-world situations (as opposed to experimental conditions), specified models are lesser representations of the true situation, because it is impossible to collect (or include) data on all relevant variables in the production process. This is termed the "hybridity" problem by Heady and Dillon [1961]. Hybridity of estimated functions does not render them useless; but likely represent the path that firms follow during the course of their firms' development process. This is backed by the argument that the theoretical models are subject to many maintained hypotheses and assumptions that may not hold under real world conditions – making the hybrid functions more realistic in the long-run. Therefore, fitted hybrid functions (that sufficiently approximate the true functional forms) are of use as they depict firms' growth paths over time.

The second step in the primal approach (when the purpose is to test for optimality) involves choosing a behavioural model for the economic agents. From this model, production is optimised subject to some objective function (e.g. profit maximisation or cost minimisation) from which optimality conditions are derived [Colman, 1983; Chambers, 1988].

Mansfield [1968] stated that economic growth could largely be ascribed to technological advances (a view typical of the late 1960's). Production functions were said to represent the maximum levels of output producible by combinations of inputs, under given technological conditions. Changes in technology induced shifts in the production function. Therefore, technological change could be measured by the change in output associated with the shifts in the production frontier, given input levels. Mansfield [1968] discussed the problem with this approach to measuring technological change. Observed differences caused by shifts were called residuals, but these residuals captured unexplained effects in the production process, without specifying the cause (source) of variation. To attribute all changes to technological change is erroneous. Much of the responses in agricultural supply are due to policy measures or changes in the business environment that influence output or input prices – relegating these influences to the residual above technical change would be a complete misrepresentation of reality.

As an example of a production function approach to milk supply, Heady and Dillon [1961] devoted a chapter to milk production functions in which they experimented with a variety of functional forms. They maintained that milk production per cow is a function of concentrate intake, forage intake, body size, inherent breed qualities, labour inputs and other unspecified (usually unobserved or immeasurable inputs). The primary objectives of their experimental study was the calculation of, firstly, grain-forage substitution rates, secondly, rates of feed-to-milk conversion under varying rations and levels of production and, thirdly, to evaluate the economic potential of forage-grain substitution.

Heady and Dillon [1961] explained that the input-input and input-output surfaces that depict dairy cows' milk production have special features that should be considered when a functional form is specified. For example, the cows stomach capacity imposes an upper limit on the quantity of feed that can be taken in a given time period. Similarly, physiological requirements determine minimum feed levels required for survival of the animal and these minimum levels in turn depend on the feed ration. Heady and Dillon [1961] found that at high levels of milk production, the slope of the production surface declines for any ration that is converted into milk – thus diminishing returns hold for all feed inputs.

With regard to quasi-fixed (i.e. fixed in the short term) variables in the production function, Heady and Dillon [1961] suggested the use of capital services (rent, depreciation, etc.) as opposed to the value of capital investment, when studies are based on short run data. They argue that only the flow of services (in stead of the stock that creates the flow over the life span of the capital asset) is relevant to the production in an observed period. Yet, using depreciation as part of the service flow includes the bias of a firm towards depreciation – at best the possible errors or bias should be kept in mind. On the same issue, using cash flows that vary in direct proportion to the level of production (e.g. fixed handling rates or commissions) will produce near unity output elasticities and high marginal propensities, but would mean nothing otherwise.

Measurement and quality definitions of labour pose problems in most cases of functional form estimation. Heady and Dillon [1961] point to the discrepancy between actually utilised versus available labour in the specification of a production function. In addition, valuing labour quality always contains some element of bias [Thijssen, 1992]. Heady and Dillon also discussed similar issues pertaining to land and management [1961, pp. 223 – 226].

2.3.2.2 THE DUAL APPROACH

After the popularity of the production function approach in the 1960' s, the 1970' s [Blackorby, *et al.*, 1978; Fuss and McFadden, 1978] saw increased application and development of the duality approach. The simultaneous advances in econometric techniques and electronic data processing capabilities strengthened the increased application of duality theory [Colman, 1983; Chambers, 1988].

Duality between production-, cost- and profit functions implies that there exists a correspondence between the cost- and profit functions and their underlying production function. The correspondence is such that one or both of the former can be used to derive the properties of the latter. This is mainly the case with limited information on the relevant primal variables and possible estimation problems associated with the production function approach [Blackorby, *et al.*, 1978; Fuss and McFadden, 1978; Chambers, 1988; Sadoulet and De Janvry, 1995].

Producer response is determined by two elements: the technological relation between combinations of inputs and the resulting level of output, and producers' behaviour in choosing inputs (given market prices and fixed factor availability). Integration of these two elements leads to (a) definition of the profit maximising or cost minimising functions and (b) to a direct method by which optimal decisions on output supply and factor demand can be determined [Colman, 1983; Sadoulet and De Janvry, 1995].

Rulan Pope [1982] evaluated the practical use of duality theory and argued that it is indeed useful for many practical problems, but it is not always essential. Pope [1982] quotes Diewert [1974] who described the heart of duality theory as the Minkowski theorem: "... every closed convex set can be characterised by its supporting half spaces". Pope [1982] described the essence of the dual approach as the indifference of results to the use of a production function or of Shepard' s Lemma [Beattie and Taylor, 1993] on a cost function in order to arrive at explicit cost minimising demands. Pope [1982] offers some suggestions for the observed prevalence of duality in applied work. Firstly, under the primal approach, input demand and output supply cannot be derived for certain functional forms, but via Hotelling' s theorems [Beattie and Taylor, 1993], the reduced form equations of input demand and output supply are derived directly. Secondly, in some cases (e.g. Leontief technology) the cost function can bridge problems of continuous differentiability imposed on the production or profit function. Thirdly, since inputs are often used in specific proportions or combinations, input demands are probably more collinear than prices, therefore favouring dual approaches to avoid multicollinearity.

The dual approach is based on the specification of a cost or profit function from which the input demand and output supply are derived. The following sub-sections describe the cost and profit function approaches.

2.3.2.2.1 The Cost Function

The cost function represents the firm' s economic behaviour as a cost minimisation problem, subject to a given level of output, under exogenously determined levels of input prices [Chambers, 1988].

Cost elasticity (ε_c) establishes a link between scale elasticity (ε)[9] (from the primal approach) and cost functions. Cost elasticity is the inverse of scale elasticity and it represents the ratio of relative cost increase due to a relative increase in output. It constitutes the ratio of marginal to average cost, and the curve of the former intersects the latter's curve at the point of minimum average cost [Bairam, 1994]. In the case of a traditional U-shaped average cost curve (AC-curve), marginal cost is higher than average cost at points beyond minimum average cost. If the production function is homogenous of degree k (i.e. if $\varepsilon = k$, where λ is constant), a U-shaped AC-curve cannot be derived [Bairam, 1994].

An example of the dual cost function approach is found in Binswanger' s [1974] work. He applied a cost function approach to measure elasticities of factor demand and substitution in the USA agricultural sector, based on cross-sectional data. He justified his choice of a dual approach – specifically the use of a cost function – by discussing six advantages.

Firstly, cost functions are homogenous of degree one in all prices, regardless of the homogeneity properties of the production function. It is therefore unnecessary to impose homogeneity on the underlying production technology. Secondly, Binswanger [1974] argues that the estimation equations treat prices (in stead of quantities) as independent variables. However, entrepreneurs choose input levels (quantities) according to price levels (exogenous at the farm level) and therefore quantities should be endogenous to the models. The third advantage pertains to the increased estimation bias that results from inverting the matrix of production coefficients to calculate elasticities of demand and substitution. A cost function approach does not require any matrix inversions. Fourthly, when dealing with neutral or non-neutral efficiency differences in observational units, or neutral or non-neutral economies of scale, the translog cost function offers a convenient way of handling problems of biased estimates of production parameters [Christensen, Jorgenson and Lau,

---

[9] Under homogeneity assumptions, $\varepsilon = \lambda^k$, where the exponent, k, gives the degree of homogeneity and λ is a constant.

1970]. Fifthly, linearity of all the estimation equations is obtained through the logarithmic transformation, for both the translog cost and production functions. Finally, Binswanger [1974] argues that multicollinearity is lacking in input prices (used in the cost function approach) while it is high between input quantities (used in the production function approach).

Sadoulet and De Janvry [1995] and Chambers [1988] state that the cost function has the same ability as the profit function to fully characterise economic agents' behaviour, whether it be cost minimisation or profit maximisation.

2.3.2.2.2 The Profit Function

The profit function represents a firm's profit as a function of input and output prices (exogenously determined) and quasi-fixed variables. The firm's decision problem is that of choosing levels of output supply and input demand that will maximise the firm's profit. Choosing between a primal or dual approach (i.e. production function versus a cost- or profit function) implies different representations of the underlying technology when the chosen functional form is not self-dual (e.g. as in the case of translog production and cost- or profit functions). Consequently, Thijssen [1992] (an example of a profit function approach to Dutch Milk Supply analysis) employed both the primal and dual approach (using the translog production and profit functions) and he concluded that the calculated elasticities of production and substitution did not differ significantly between the two approaches. Thijssen [1992] refers to the findings of Apelbaum [1978] and Burgess [1975] that stated significantly differing substitution possibilities between inputs, as obtained from translog specifications of the production and cost functions. However, these authors assumed full static equilibrium and they used only aggregated data [Thijssen, 1992].

In Thijssen's study [1992] he used data on decision making units in order to evaluate the influence on dairy product supply resulting from price (input and output) changes, technological change, the quantity of available land, demand for variable and capital inputs, and the supply of farm labour in the

Dutch agricultural sector. The data set incorporated time series and cross-sectional data [Thijssen, 1992] and was analysed as an unbalanced panel.

In a microeconomic approach to input demand and output supply in Irish agriculture, Higgins [1986] applied a profit function approach to data from a cross-section of Irish farms (milk production constituted one of the main activities). The resulting parameters were used to estimate a set of own- and cross-price uncompensated and compensated supply and demand elasticities. His results did not conform to all the expected theoretical requirements, yet he was able to test some of the properties (contrary to other studies were the properties are assumed) and to find acceptable descriptions of inter-relationships between inputs and outputs.

Bouchet, *et al.* [1989] investigated possible short- and long-run sources of growth in French agriculture. They point to the use of both profit and production functions to evaluate the impact of public agriculture expenditure on research on agricultural output (milk features as on of the outputs). Those studies which used the profit function approach, assumed short-run optimisation and exogenously fixed levels of other inputs. Bouchet, *et al.* [1989], however, modifies this method to treat some inputs as quasi-fixed, rather than fixed, thereby creating a way to evaluate long-run profit that comprises variable and quasi-fixed input optimisation.

2.4 FUNCTIONAL FORMS

The choice of a functional form is usually guided by knowledge of the biological nature of production, the economic and environmental (natural or business) factors that influence the production process under study. Generally, a trade-off exists between best-fit considerations and theoretical plausibility of the functional form. Some form of judgement on the part of the analyst is required for most samples, and few of those samples are free of design, measurement or specification errors. Choosing a probability level at which regression coefficients will be accepted as significantly different from zero depends on the phenomena under study. For example, rejecting the coefficient of a quadratic term in a second order polynomial might imply acceptance of a linear relationship. However, this could be implausible under some maintained physical and economic hypotheses [Whitley, 1994].

Quoting Anderson *et al.* [1996], "... the choice of functional form is not a trivial matter". Anderson, *et al.* [1996] point out three functional forms that seem to dominate in empirical production economics literature. Those forms are the translog, the normalised quadratic and the generalised Leontief functions. They concede that economic theory is not sufficient to determine the suitable functional form, although it does aid in identifying relevant variables and homogeneity restrictions. Anderson *et al.* [1996] ascribes the importance of correct specification of a functional form to the impact it has on predicted responses of modelled policy interventions.

According to Bairam [1994], most production functions in empirical research are assumed linearly homogenous, without implementing formal statistical tests on the true structure. Yet, the assumption of linear homogeneity (implying constant returns to scale) is not appropriate for some aspects of production theory. Bairam [1994], Fuss [1978] and Christensen [1973] emphasised the importance of flexible functional structures that require a minimum of maintained hypotheses.

One of the most famous production structures is the Cobb-Douglas function [Cobb and Douglas, 1928], which forms the basis of many subsequent functional forms. The other is the Leontief production function [Leontief, 1947]. Diewert [1971] expanded the Leontief function into a generalised form, while Arrow, Chenery, Minhas and Solow [1961] proposed adaptations [Zellner, Kmenta and Dreze, 1966] of the Cobb-Douglas production function into the Constant Elasticity of Substitution (CES) function [Kmenta, 1967], to incorporate constant elasticity of substitution between inputs, as opposed to the Cobb-Douglas' s restrictive unitary elasticity. Christensen *et al.* [1971, 1973] introduced the transcendental logarithmic ("translog" for short) production function – a non-homogenous function with varying scale elasticity. Under specific parameter restrictions, this function can also represent linear homogeneity (constant returns to scale, CRS).

Heady and Dillon [1961] mention the expansion of the Cobb-Douglas (in it's log-linear transformation) into the translog as one appropriate functional form for milk production modelling. However, when corner solutions are possible [Weaver and Lass, 1989] or when dealing with missing values, functional forms in logarithmic format (such as the Cobb-Douglas and translog) run into computational problems. Heady and Dillon [1961] suggest transformation of these variables (in which corner solutions prevail) by adding a small constant to all observations on the variable, or by replacing zeros with a small positive constant. They warn that care should be taken with this procedure: with high prevalence of zero-values, the consequent adjustments might be unsatisfactory and could induce non-negligible errors.

The translog form has been used more frequently than the generalised Leontief or the normalised quadratic [Anderson *et al.*, 1996]. All three of these flexible forms are linear in their parameters – facilitating estimation with standard statistical procedures. All three are separability inflexible, implying that they are not second-order Taylor series when separability is maintained in a partition of the variables. However, compared to expansion in powers, the translog (as an expansion in logarithms) has a larger region of convergence when large dispersion occurs in the data [Anderson *et al.*, 1996]. Conversely, the translog form is less suitable for technologies that seem to exhibit low elasticities of input substitution [Anderson *et al.*, 1996]. Anderson *et al.* [1996] conclude that the preferred functional form is both data and method specific, thus making testing of alternative forms imperative to the selection process. Flexible functional forms allow for objective testing of structural properties before imposing the properties on the technology, as is done with restrictive functional forms (e.g. linear functions) [Andersen, *et al.*, 1996; Hassan, 1998].

2.5 APPROACHES TO SPECIFICATION AND ESTIMATION OF SUPPLY MODELS

Parametric representation of milk supply response usually takes the form of production functions, or systems of supply equations. From these models, elasticities of supply and input substitution, as well as economies or diseconomies of scale and size are calculated. The concept of econometric representation raises the issues of appropriate specification of functional forms (Section 2.4) and problems inherent to the type of data.

Computational difficulties are associated with different types of data [Heady and Dillon, 1961]. When cross-sectional data is used and inputs are not homogenous over observations, aggregation problems occur [Day, 1963]. Standardisation attempts also poses problems, since the standardisation inherently involves judgements on quality differences.

Using cost and profit functions has made it much easier to obtain reliable estimates than in the case of production functions. Additionally, well-behaved cost and profit functions are supported by an underlying neoclassical production function [Quiggin and Bui-Lan, 1984]. However, the main arguments in these functions are prices, which would not likely vary greatly across firms at one point in time, if the maintained hypothesis of perfect competition were upheld.

If price variations can be ascribed to differences in location, estimations can be performed to capture the inter-regional variations. Conversely, if farms are widely dispersed, the assumption of homogenous technology is violated. Other spurious variation may result from valuation errors (opportunity cost of non-traded inputs), recording errors (prices include transaction cost) and quality differences captured by prices [Quiggin and Bui-Lan, 1984].

Cross-sectional data can be useful on a broad level for analysing policy questions at regional or national levels, but it may not be suitable for inferences about individual firm performances [Heady and Dillon, 1961; Pasour, 1981]. Cross-sectional sets should thus allow for ample variation across firms.

In a study aimed at estimating an aggregate production function for the USA agricultural sector, Griliches [1963] made use of cross-sectional data with several important comments on the use thereof. In his study, Griliches faced a lack of appropriate quantity data for many of the hypothesised variables and hence opted to use values (e.g. feed cost instead of the tonnes fed or price of feed). He justifies this choice through the assumption that cross-sectional price variations are not "...too large..." - as in the case of many commodity prices. However, this assumption is unrealistic in the case of labour and land inputs: price differences are substantial and they reflect quality and geographical variation. Griliches [1963] nevertheless maintain that inclusion of values (as opposed to quantities) represents a much smaller bias than ignoring cross-sectional differences in the cases of land and labour.

On estimation methods, Griliches [1963] defends his use of ordinary least squares (OLS) – applied to the logarithms of the variables, with output as dependent variable based on the inappropriateness of a factor share approach. The latter would assume equilibrium and constant returns to scale (CRS) – which essentially implies prior knowledge of the structure to be investigated.

Due to data limitations and the context of the research problem, Griliches [1963] settled for the use of OLS although it was not optimal. He also raised the problem of simultaneity that results from the possibility that no technique could yield sensible estimates of the input-output relationships. The latter argument stems from the deduction that all firms facing the same production technology and prices, should be at the same point. However, nothing would then have to be estimated. Hence, the issue of specifying the reasons for variation in real world data from the theoretical point arises [Griliches, 1963]. The potentially significant differences between regional input-output combinations ensure estimation possibilities. However, this casts doubt on the representativeness of a single production technology specification for all regions.

Weaver and Lass [1989] presented an estimation strategy that takes account of corner solutions in duality models, based on cross sectional observations of dairy production decisions. Their results suggest that substantial estimation bias occur when corner solutions are not recognised. Even under the assumption that all farms face similar technology options, variation in prices and fixed factor-flows cause differences in the choices that farmers make. For some firms it might be possible to exclude the use of certain factors in the short run – these corner solutions are typically observed in cross-sectional samples of farm budgets [Weaver and Lass, 1989].

Weaver and Lass [1989] emphasised the importance of the short run elasticity of output supply and input demand in the design of effective dairy policies. They apply the corner-solution adapted estimation process to a cross section of Pennsylvania dairy farms. From this they derive a complete set of short run elasticities that are consistent with short run profit maximisation, multiple-input multiple-output technology and fixed input flow hypotheses [Weaver and Lass, 1989].

Jaforullah and Whiteman [1999] followed a non-parametric approach to evaluate the scale efficiencies in New Zealand's dairy industry. Data Envelopment Analysis (DEA) was employed for the sample, under the assumption of multi-product (milk fat, milk solids and milk protein) output supply. Jaforullah and Whiteman [1999] observed the level of detail and the applicability of DEA to individual farmers, tied to the advantages of benchmarking to establish so called "best practice" examples. However, they recognise the advantage of parametric stochastic production frontier methodology that allows for statistical hypothesis testing, for which DEA does not allow. In addition, Jaforullah and Whiteman [1999] concede that the results of DEA do not contradict those obtained by Jaforullah and Devlin [1996] through a parametric approach.

Tauer [1998] used the Malmquist index to decompose the productivity of New York dairy farms into technical improvement and efficiency gains. The index is computed in the DEA framework and is based on the distance function – which is free from the assumptions of profit maximisation or cost minimisation.

For Alabama dairy farms, Smith and Taylor [1998] compared results from Estimated Generalised Least Squares (EGLS) and Finite Mixture Estimation (FME) as it applies to the derivation of size economies and production cost frontiers. They argued that FME provides estimation of a stochastic cost frontier with known statistical properties, which was unattainable with available stochastic frontier estimation packages [Smith and Taylor, 1998].

Dawson and Hubbard [1987] studied the effect of management on the size economies in England and Whales' dairy sector. They econometrically estimated long run average cost curves under different levels of management. Their results pointed to U-shaped average cost curves that display higher economies than diseconomies of size [Dawson and Hubbard, 1987]. Additionally, their results suggested that better managers can produce levels of output at lower average cost (i.e. higher optimum levels of output holds for these farms). Dawson and Hubbard [1987] used cross-sectional farm budget data with incomplete price information. They applied the translog functional form due to its flexibility and minimum of maintained hypotheses. Both heteroskedasticity and multicollinearity occurred in the data. Therefore, ordinary least squares-estimation (OLS) was rejected in favour of the two-stage Aitken estimator procedure, which yielded better fits [Dawson and Hubbard, 1987].

A priori, analysts often have little information, if any, on farmers' inclination towards experimenting or innovations, differences in resource endowments and actual primary enterprise goals. Knowledge of this could very well necessitate specification of different regressions, or inclusion of additional explanatory variables.

2.6 INTERPRETING MEASURES OF EFFICIENCY

Pasour [1981] made a few sobering comments on the measurement and interpretation of efficiency and economies of farm size. He argued that farm management data is inappropriate as basis for specification of inefficient producer behaviour and economies of farm size under real world conditions of uncertainty, imperfect knowledge and costly information.

Efficiency is measured by comparing a predefined norm with observed behaviour. The norm, in agriculture, is often specified as the "perfect market". Therefore, the efficiency norm looses its appropriateness in the face of uncertainty and costly information. Pasour [1981] ascribed the immense difficulty with measuring efficiency to the neglect of the entrepreneur's role in economic theory. Since efficiency involves inevitable valuations, those valuations should reflect the cost and returns as perceived by the decision-maker – not the analyst. The traditional assumptions of perfect knowledge and perfect certainty, under constrained maximisation, render inefficiency a contradiction in terms, according to Pasour [1981].

Pasour [1981] further argued that economies of size studies should not conclude that output levels of farms are inefficient or sub-optimal because changes in farm size distribution over time does not imply that a farm's current size is inefficient. Pasour [1981] suggested that farm survey data can neither yield meaningful results concerning efficiency of decision makers, nor of the effect of increased farm size on the unit production cost (economies of scale).

2.7 EXAMPLES OF MILK SUPPLY RESPONSE STUDIES

According to Tauer [1998], previous dairy studies point to great diversity in production methods and efficiency – inherent to the dairy industry. In addition to the studies by Thijssen [1992], Bouchet, *et al.* [1989] and Higgins [1986] where milk featured as an output, the following selection of studies dealt directly with parametric estimation of supply response in the dairy sectors of various countries.

Studies of milk supply response have focused on changes in the national herd and output. Buckwell [1984] argued that the structure of the herds play an important explanatory role in the understanding of secular output increases in the face of decreasing real milk prices. Buckwell dealt with the England and Whales dairy sector for the period 1962 to 1984. In his model, changes in herd numbers, average herd size and yield were explained. Buckwell [1984] tries to demonstrate that alternative specification of milk supply over time overshadow the role of the real milk price. Previous work applied production function approaches, linear programming based studies and mostly national aggregate econometric approaches. All these models decomposed milk production into two components: cow numbers and yield.

Buckwell [1984] refers to Kislev and Peterson [1982], who explained the growth in farm size, given fixed labour inputs, based on developments in relative factor prices. This approach avoids the controversial issues of technical change and economies of scale.

Buckwell [1984] modified and applied Kislev and Peterson's model. The modified main assumptions are that dairy farms are owner- or hired manager operated units, which constitutes the fixed farm input. Secondly, the ratio of mechanical services to cows is constant, irrespective of the herd size, biological service inputs or milk output. Thus, herd size is determined by the output of mechanical services. The production process combines the two sub-processes (mechanical service formation through labour and capital application; and the biological sub-process in which feed and fertiliser is combined with cows) to yield milk. The mechanical sub-process is seen as a cow-augmenting action. Labour cost is opportunity cost of fixed farm labour, and capital cost is the effective user cost of capital [Buckwell, 1984].

Buckwell' s [1984] adaptation of Kislev and Peterson' s [1982] model is mainly the aggregate focus as opposed to the latter's firm level focus. According to Buckwell [1984], returns in dairy and alternative enterprises as well as per capita earnings in dairy and non-agricultural sectors determine herd size. Cow numbers are a function of the cost of capital and the labour in the dairy sector. Cow numbers, milk price and the feed price, determines yield. Together, yield, cow numbers and average herd size explain the total supply response of milk [Buckwell, 1984].

Buckwell [1984] concluded that economic attractions outside dairying cause the main thrust for dairy producers to leave the sector. Changes in herd size result from changes in factor prices. Observed improvements in yield are partly due to technical change associated with larger herds, but yield also responds to feed and output prices. According to Buckwell [1984], the principal determinant of output growth is not the milk price, but the growth in average herd size, which is in turn sensitive to the price of capital relative to labour cost.

Burton [1984] argued that the livestock sector is subject to substantial interdependence, thus requiring simultaneous determination of endogenous variables. Herd size, numbers and rate of culling, replacement heifer prices and milk prices are said to be endogenous and interdependent [Burton, 1984]. In Burton' s model, heifer demand depends on culling levels and herd size expansion, both of which are dependent on the heifer price. This calls for the first simultaneous determination. Secondly, the herd size and the milk price should be determined jointly.

In lieu of these studies and the surveyed theory on supply response analysis, methodology is proposed through which the given data will be analysed.