1. Introduction

1.1 Role of the livestock sector in the national economy

Mongolia is a Central Asian country with extensive livestock production as the dominant sector in agriculture. The population of Mongolia is two million, over half of which is urban, living in the capital Ulaanbaatar and other major towns. About 42% of the population is involved directly in livestock production living a semi-nomadic lifestyle and making seasonal movements according to grazing requirements of their livestock. In 1995, the livestock population of Mongolia was around 28.5 million, and was made up of horses (2.6 m), cattle including yaks (3.3 m), camel (0.4 m), sheep (13.7 m) and goats (8.5 m). Mongolian livestock produce mainly meat, although wool, milk, cashmere, skin and other products are also important.

In 1990 the rural sector, overwhelmingly pastoral, accounted for 40 percent of total exports and 33 percent of total employment. The importance of agriculture and the livestock industry in terms of their share of national income and employment is shown in Table 1.1.

Table 1.1: Share of Agriculture and Livestock Sector in National Economy (%)

	1970	1990	1995
Share of Agriculture in National Income	25.3	20.3	39.4
Share of Agriculture in National Employment	47.3	33.0	44.6
Share of Livestock Sector in total agricultural output	71.0	72.8	82.9

Source: Central Statistical Board of the MPR (1986, 1995 editions)

The year 1990 represents the peak of the socialist period. The increase of agricultural share form 1990 to 1995 was a result of the dramatic decline in the industries other than agriculture which were more affected by the transitional depression.

1.2 Some major characteristics of the extensive livestock industry

The major characteristics of the Mongolian extensive livestock industry in the context of the current study are its absolute dependence on a harsh and highly variable natural environment and the resulting low and basically constant yield per animal over time.

Amongst the major characteristics of extensive livestock production, its high exposure to, and dependence on a severe natural environment is clearly ranked first.

Mongolia is a land-locked country with severe continental climate located at an average 1580 m altitude and surrounded by mountains and hills. From the point of view of physical geography, Mongolia is divided into three main zones: mountain, steppe and the Gobi desert. The main characteristics of these zones are shown in Table 1.2. Because of the range of highlands, vast steppes and desert, rainfall and temperatures vary greatly, not only seasonally but daily.

Table 1.2: The characteristics of the main geographical zones of Mongolia

	Mountains	Steppe	Gobi
Average monthly temperature in winter °C (Dec., Jan. and Feb.)	-20 to -30	-20 to -25	-15 to -20
Average monthly temperature in summer C (June, July and Aug.)	15 to 20	15 to 20	20 to 25
Number of warm days per year	90	90-100	140
Annual rainfall (mm)	250-350	150-200	50-150
Depth of snow cover (cm)	15-20	10-15	2-5
Average pasture yield (100 kg/ha)	5-8	3-4	1-3

Source: Purev (1990)

A more detailed zoning of Mongolia which takes account not only of topography but of all major components of the natural environment was also developed especially for agricultural purposes. According to this zoning Mongolia is divided into 5 regions consisting of 18 subregions as described in section (4.3.3).

Mongolian livestock get over 95 percent of their annual fodder from natural pastures, utilising them all year round. Use of supplementary fodder is very limited, in 1992-1993 it was only around 10 kg of feed unit[1] per one sheep equivalent which is equal to approximately 20 kg of hay. Pasture yields vary with altitude and location decreasing in quantity and improving in quality from north to south. Pasture resources are highly dependent on erratic rainfall and their availability is subject to snowfall during cold seasons. Pasture growth begins in April and usually reaches a maximum in August. Taking the high season standing pasture as 100 percent, winter production is 50-60 percent and spring production is 30-40 percent. As the season progresses, quality of vegetation decreases 2-3 times and its protein content by 3-4 times. Accordingly, Mongolian animals, well suited for pastoral grazing, accumulate fat, grow and produce during summer and autumn, and they survive by using their fat reserves and by reducing their feed requirements during winter and spring. Reflecting this biological cycle, during the winter-spring season animals loose live weight. Average weight losses are: cattle 13.2%, sheep 20.8% and goat 20.8% compared with autumn maximum live weight (Tserendulam, 1976).

The most extreme natural hazard is *dzuud,* (sudden winter snowfall burying pastures) which can appear out of nowhere and overwhelm the herders who are caught in the sudden dump of snow. They are immediately faced with the starvation of their stock, which being already in poor body condition, may die within days. The next danger is drought which usually happens due to the failure of principal rains in June-August. The consequence of drought is that peak summer pastures are poor, followed by insufficient growth in autumn/winter reserved pastures. As a result animals go into the winter with insufficient body condition and become steadily weak because of the poor status of the reserved pastures. Whereas, *dzuud* can appear overnight and cause almost instantaneous major losses, drought is more insidious in that it comes slowly and its effect is spread over longer period of time. In addition to *dzuud* and drought, low temperatures coupled with strong wind, snow storms, sudden frosts in early summer and late autumn, and continuos cold rains in warm seasons can cause serious losses by making animals sick.

---

[1] feed unit is a uniform measure of feed in Mongolia, equivalent to 1 kg of oat. For example, 1 kg of hay is equal to 0.45 kg of feed unit.

The main and traditional strategy in response to the severe and highly variable natural environment is mobility of herding families in search of better pasture, water and favourable weather. Another reasonable strategy, which was one the main focuses of economic policy during the socialist period, is to invest in fixed capital such as winter and spring shelters and wells, and to provide veterinary services, labour and fodder supplements during especially difficult periods.

The second major characteristic of the extensive livestock industry in Mongolia is its overall low productivity imposed by ecological constraints. It is not surprising, therefore, to see basically constant yields of meat, milk and wool per animal over time.

Table 1.3: Per head production of meat, wool and milk in Mongolia

	1960	1980	1991
Average live weight sold to the state (kg)			
-Cattle	248	217	245
-Sheep	36	33	39
-Goat	28	26	33
Wool yield (gr)			
-Sheep	1186	1390	1243
-Camel	4104	5034	4365
-Cow milk (litre)	344	292	323

Source: Central Statistical Board of the MPR (1970, 1981 and 1992 editions)

1.3 Organisation of production

From the late 1950's up until 1990, when Mongolia began the process of economic liberalisation, all members of the rural population were *negdel* (state cooperative) or state farm employees. The *negdel* was primarily an economic unit responsible for marketing livestock products, supplying inputs and consumer goods as well as fodder and transport services to its members. The *negdel* covered the same territory as a single district. The basic livestock production unit within a *negdel* was a *suuri,* which consists of one to four households. Each *suuri* was generally involved in the production of specialised species of *negdel*-owned herds for which a monthly salary was paid. The average size of a *negdel* in 1985 was around 400 000 ha of pastures, with a total of 60 000 livestock, 43 tractors (in 15 horse-power units) and 15 motocars. The number of *negdel*-owned animals in each *suuri* varied between ecological regions, but it was generally fixed within the same ecological region, averaging 500-600 for small stock (sheep and goat) and 140 for cattle. In addition to *negdel*-owned animals, each household was allowed to have small number of private animals -up to a maximum of 75 in desert and semi-desert regions and 50 in other regions. The *negdel* set production targets for each *suuri,* determining the quantity of meat, wool and other products to be supplied according to annual state procurement orders.

1.4 Intensification in the socialist period

As noted earlier, for the last 30 years up to 1990, the centrally planned system of agriculture moved towards the intensification of production by providing shelter structures and veterinary services for livestock, making supplementary fodder and concentrates and irrigating natural pasture. The intensification process was implemented by a series of large campaigns at tremendous cost. The bulk of these investments was brought about in the late 1960's, peaking in the 1970's and early 1980's. The increasing trend in water and shelter-supply compared with the generally constant livestock numbers is shown in Figure 1 where the total number of livestock is given in millions, and the number shelters and wells in thousands.

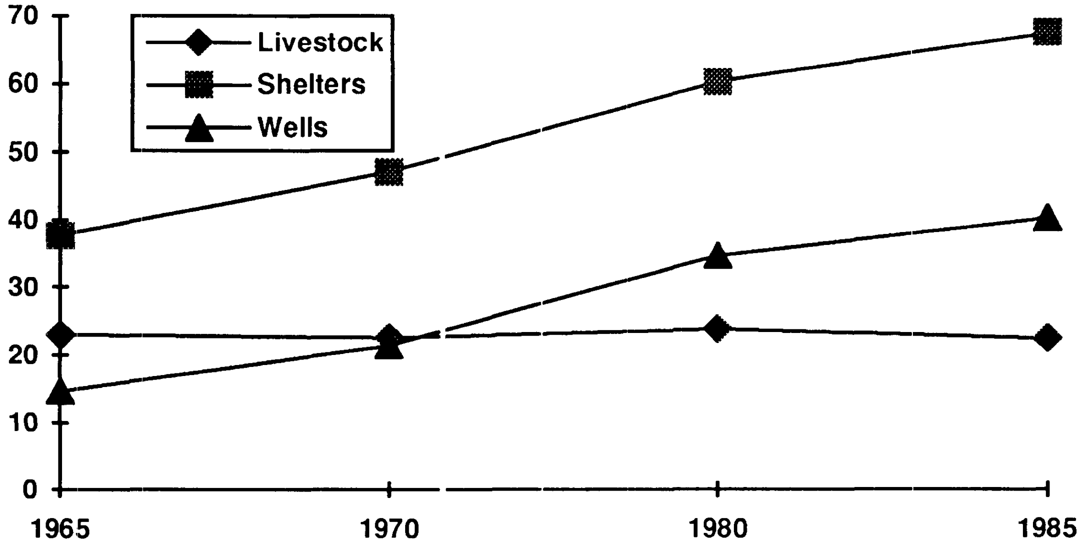

Figure 1.1: The total number of livestock, livestock shelters and wells for pastures in Mongolia.

Source: Central Statistical Board of the MPR (1986)

Likewise, for the same period, the total harvest of natural hay, the main fodder supplement in Mongolia, increased from 522.2 thousand tons in 1970 to 1275.6 thousand tons in 1985. With regard to fodder production in Mongolia two important points to be noted are: (i) fodder is mostly produced in the more productive northern regions, implying that shortage of fodder is more severe in the dry southern regions, (ii) a significant portion of total fodder produced all over the country went to the centrally administered Sate Emergency Fodder Fund (SEFF), whose role was to have stocks of hay on hand to be trucked or flown into affected areas and provide short-term emergency relief. Since the emergency aid from SEFF was delivered after the disaster took place, it was not efficient in many cases.

With regard to wells and shelters, as they were built to fulfil state plans mostly in terms of quantity, there was serious criticism about their location from local people and professionals. The most common case of failure to consider the local conditions was building wells and shelters in pastures that were not used because of remoteness or access difficulties. There were also cases in which wells were dug up in a summer or autumn pasture near a river, where they were absolutely unnecessary.

Finally, referring to the nature of a centrally planned economy, the process of intensification of the extensive livestock industry in Mongolia was imposed on production from outside rather than initiated by producers themselves.

Despite these criticisms, the intensification process undertaken in 1960 to 1990 was claimed to be one of the successes of the communist system in the agricultural area, in the sense that it could decrease the mortality rate and increase the birth rate of pastoral animals.

1.5 The research problem

With the beginning of the transition of the Mongolian economy to a market oriented system in 1990-1991, the extensive livestock production system developed over the past 30 years has collapsed. The government has privatised the ownership of the stock and, in reaction to long years of central planning, has stood back leaving the industry to reorganise itself. In 1995 the share of privately owned livestock constituted 92.9% compared to 31.8% in 1990. As a result of privatisation, the *negdels* have been broken into small household economies with apparently poor abilities to expand production. In 1992 the average livestock per family (5-7 people) were only 110 (around 250 sheep equivalents). This may mean that the productivity of the sector is falling through a decline in labour efficiency and herders are interested only in subsistence production. There is less utilisation of supplementary fodder and veterinary services mainly due to cost considerations and unavailability. There is also an apparent unwillingness of herders to use superior stock for breeding purposes, indicating the danger of a long-term decline in animal productivity and the quality of output.

Thus, the overall productivity of the sector is seen to be at risk, because herders are making individual decisions according to their own economic interests in contrast to the previous system, which could see the national herd as a whole and make management decisions based on national perspectives. Briefly, the sector needs development strategies that will work in the new market economy. In this respect, the question of whether the intensification process undertaken from the 1960's to 1990 led to productivity gains in livestock production is quite interesting and entails serious policy implications. In very general terms, a positive answer to this question would mean that policy makers may be certain about stimulating herders towards the use of technological inputs at a rate to at least match the level in pre-transition period.

Thus, this study is designed to identify and analyse productivity of Mongolian extensive livestock production by specifying a production function.

While the causal factors behind the changes in productivity remain to be determined, account should be taken of the different productivity potential in different ecological regions. These reflect agroclimatic differences and differential resource endowments between regions.

In addition to specific questions addressed in this study, production function analysis has scientific and practical importance. Production functions provide general guidance for farmers' decisions, credit policy formulation, readjustment of agricultural regions, etc. (Heady and Dillon, 1961).

1.6 Objectives of the Study

In relation to the problem outlined, the general objectives of this study are:

1. Specification of livestock production functions most adequate to specific conditions in Mongolia from 1969 to 1990
2. Identification and analysis of productivity in the Mongolian extensive livestock sector over the period 1969 to 1990
3. Suggestion of possible policy initiatives regarding productivity improvement policy for the extensive livestock industry.

1.7 Hypotheses

The following hypotheses are formulated to guide the study.

1. Livestock production was primarily dependent on weather variables

2. Under the socialist regime, the policy of encouraging private ownership of livestock had a positive impact on output of livestock enterprises

3. A positive and significant technical change occurred in the livestock industry during the study period

4. The intensification process undertaken from the 1960's to 1990 increased the natural growth rate of animals as the basic indicator of performance of extensive livestock production, where per head yield of harvested products is generally constant.

5. The pattern of productivity growth was different across agro-ecological regions.

1.8 Organisation of the study

The dissertation is divided into six further chapters. Chapter 2 contains a review of previous work on weather-yield modelling and technical change and, therefore, provides the conceptual framework for the rest of the study.

Chapter 3 discusses econometric issues of specifying and estimating a production function, including the choice of the appropriate functional form, the variables to be included in the production function and techniques for estimating the coefficients of the production function.

Being the first study to estimate a production function and to analyse productivity of the extensive livestock industry in Mongolia, the derivation of the appropriate dependent variable and the independent variables is very important. Because of the industry's absolute dependence on a harsh natural environment, specific attention is paid to the influence of weather on the performance of the industry. These issues and the associated methodology are discussed in Chapter 4.

The empirical results are presented in Chapter 5, where the production function analysis of the extensive livestock industry in Mongolia is carried out in two stages. In the first stage, weather-yield models are estimated and used for deriving a weather-adjusted production measure for cattle and the small stock in Mongolia. In the second stage, the aggregate production function is estimated using weather-adjusted dependent variable derived in stage one. The production function is then used to analyse production structure and output growth in the industry. An attempt is made to separate the total growth of output into two sources - technical change and intensification.

The conclusions, recommendations, limitations and areas of future work are provided in the final chapter.

2. Conceptual Framework and Review of Previous Work

2.1 Introduction

This chapter deals with a revision of previous work on weather-yield modelling and technical change and, therefore, provides the conceptual framework for the rest of the study. Allowing for weather in agricultural production model building is discussed in section 2.2 and technical change and its measurement is discussed in section 2.3. Section 2.4 is concerned with some issues of pastoral nomadism that are beleived to be relevant to the present study.

2.2 Allowing for weather in agricultural production model building

Taking account of weather in agricultural productivity analyses is generally demanded by two factors. First, the validity of the estimates of the production function in agriculture without account of the weather is questionable as the latter significantly contributes to output growth (Shaw, 1964). Second, the incorporation of weather variables in the analysis of yield behaviour has an impact on the trend coefficient as a proxy for technological change. Omitting weather variables may bias the coefficient on the trend variable if weather patterns change. Such an omission is of particular concern because there is evidence that world-wide weather patterns are changing (Offutt *et al.* 1987). For example, if the trend yield in a weather index is derived from time as an independent variable, then yield increases due to trends in benign weather conditions for agriculture will be attributed to technology (Bayer, 1977).

There is a general agreement amongst researches that a time trend should be included in yield-weather models in order to isolate the impacts of weather and technology. However, one should be clear in distinguishing the implications of a time trend variable in two commonly used weather yield models: (i) yield is a function of weather variables plus a time trend or (ii) yield is a function of weather variables, a time trend plus technological inputs. In the first model, the time trend as a measure of the residual, is supposed to capture the combined impact on yield of the two most important processes: technical change and intensification of production. In the second case, the coefficient of the time trend indicates the rate of technical change.

It is widely acknowledged that a useful framework for assessing the climatic suitability of areas for agricultural production is provided by agro-ecological zonation schemes. In particular, the highly variable temporal and spatial precipitation conditions that are characteristic of semi-arid areas make it very difficult to conduct quantitative assessments of agricultural sensitivity to climate over large geographical areas. Zonation schemes provide a means of identifying broadly coherent agroclimatic subregions, within which specific climate impact experiments can be assumed to be of general applicability. Subregions should be sufficiently homogeneous, both in terms of cropping pattern and climate, to provide a basis for further study of crop responses to climatic variations (Carter *et al.* 1988). A series of studies have been undertaken on agro-ecological zonation in Mongolia by Jambaajamts (1975), Shirnen (1978), and Shirnen and Bazargur (1987). Enkh-Amgalan (1990), Myagmarjav (1972) attempted to evaluate the suitability of different ecological systems to the agricultural uses from the economic standpoint using value criteria.

The problems of quantifying the effects of climatic variations on animal health and production are considered to be more formidable than those faced in estimating crop productivity. In part this is because, in addition to the direct effects of climate on animals well-being, the pathway of impacts is dominantly a two-stage process: the climate-vegetation and vegetation-animals stages, each stage characterised by a range of cause and effect relationships. A further problem concerns the type of livestock output affected by climate, for while crop production is usually measured in terms of production per unit area of a single plant component, livestock production can be measured in terms of a range of output parameters, including live weight gain, carcass quality, milk, hides, wool, and reproduction. Moreover, especially in cold regions, the direct impact of unfavourable climatic phenomena such as extreme low temperatures, snow covers, and snow storms significantly increases the number of climatic factors that should be accounted for in the weather-yield models. The direct impact of the weather on animal performance in Mongolia has been studied by Jambaajamts (1975), Sodnoi (1975), Shirnen (1978), Tuvaaansuren, Sangidansranjav and Dagvadorj (1989). For example, Jambaajamts (1975) observes that all species of animals can not use pasture at all if snow depth and density reach 38 cm and 0.35 g/cm in the mountains and 28-30 cm and 0.35 g/cm in the forest-steppes, respectively.

Two general techniques for examining the response of agricultural output to climatic variations can be distinguished: the agroclimatic index and the yield-climate modelling procedure. Both techniques, have been widely employed in the relevant research.

In essence, an agroclimatic index is a derived variable that is defined either by manipulating values of a meteorological variable into a different form or by combining variables with empirically derived coefficients into a composite term. For example, moisture and thermal factors are combined into a single term, such as an aridity index. For the impact analyst such indexes offer two advantages. Firstly, an index constitutes a single term to which crop growth and development is found to be particularly responsive. Secondly, the statistical problem of collinearity amongst meteorological variables is minimised and degrees of freedom are conserved. This kind of weather index is mostly used by meteorologists and agronomists, but rarely by agricultural economists (Oury, 1965). Oury's empirical tests indicated that weather indexes worked about as well as the direct measures of rainfall and temperature on which they were based.

There are numerous reviews of yield-climate modelling in the literature and all attempt some kind of model classification. In general, two broad classes of model can be distinguished: simulation models and empirical-statistical models.

Simulation models generally treat the dynamics of crop growth over the growing season through a set of mathematical expressions tying together the interrelationships of plant, soil and climatic processes. Some of these relationships are understood well enough to be regarded as accepted laws of physics, chemistry and biology and are often referred to as mechanistic functions. Other processes which are either poorly understood or of secondary interest to the analyst are frequently represented by empirical functions. Thus, no simulation model can be described as truly mechanistic since all incorporate at least some empirical (black- box) elements. Since the knowledge about the relationships between animals and climatic processes is not satisfactory, it is clear that simulation models are not applicable to the extensive livestock production in Mongolia. For this simple reason, simulation models are not discussed further.

Empirical statistical models are developed by taking a sample of annual yield data together with a sample of weather data for the same area and time period, and relating them through statistical techniques such as multiple correlation analysis. This procedure is sometimes labelled a 'black box' approach since it does not easily lead to a causal explanation of the relationships between climate and crop yield. This description should not imply, however, that these models are developed blindly or indiscriminately. The most effective empirical statistical models are usually the product of careful and well-informed selection of suitable explanatory variables, based on close understanding of basic crop physiology. Greatest success with this approach has been achieved where one or two environmental variables dominate yield performance (Guise, 1969). The use of simple variables, such as monthly mean or seasonal mean precipitation and temperature, is common in large area studies where the availability of data is restricted and/or computational limitations occur. As a minimum requirement to validate its estimates the outputs from an empirical-statistical model are usually verified against observed yield data.

Generally two approaches in selecting appropriate weather regressors have been used: a priory reasoning and the use of statistical methods such as the stepwise procedure. A priory reasoning is often used when the weather-yield relationships are well understood and the data set is fairly homogeneous. However, a priory reasoning is regarded as a crude approximation to reality, as the net impact of weather on yields is the result of complex interactions and all weather variables have a role in determining yields. The stepwise selection procedure is often used when the data set is massive and includes a large number of crop districts with a variety of soil characteristics and weather variables.

Specification of the appropriate functional relationship is as hard to pin down as identification of the relevant meteorological variables. Because of the complexity of the relationship between yield and weather variables, many authors believe that we are unable to specify satisfactorily this relationship. Nonetheless, other authors support the necessity of specifying weather-yield relationships. Guise (1969) wrote:

> The relationship between crop output and meteorological variables is extremely complex, but no more so than many macro-economic relationships. Despite their complexity, econometricians attempt to estimate macro-economic relationships, using relatively simple functions as approximations to underlying complex

functional relationships of reality, and there seems to be no good reason for rejecting this approach to weather-yield relationships.

The most commonly used approximation to weather-yield relationships has been a straight line. Quadratic and other non-linear forms are rarely used mostly because they drastically increase the number of regressors, thus reducing the available degrees of freedom.

Depending on the data used, empirical statistical models can be classified into two groups. One group of models uses data from experimental plots. An interesting application of this kind of modelling technique is Stallings' weather index (Stallings, 1961 and Shaw, 1964). Stallings uses experimental plot data in which variety and other cultural practices are unchanged during the period studied. A time trend is used to adjust data to account for changes in uncontrolled factors like soil fertility. Then the weather index is derived by dividing the actual yield by the predicted trend yield. This approach has the advantage that changes in production technology can be controlled and therefore do not have to be estimated, as is the case when secondary data are used. The main limitation of this method is that it requires experimental data which are not commonly available. Furthermore, as pointed out by Doll (1967), another objection to this method is that the data sources limit the use of the index to those areas for which an adequate 'sample' of yield series is available. Consequently, most empirical statistical models use non-experimental data. Table 2.1 provides a survey of selected studies on empirical-statistical modelling of weather-yield relationships.

An interesting technique for deriving a weather index is described by Orlan and Lin (1969), Doll (1967), and Desai (1986). The main idea of this technique is that the influence of weather on yield can be calculated by comparing yields predicted for the actual and the average weather. The predicted yields have been calculated by the use of the response coefficients from weather-yield models. However, the method of deriving the weather index slightly differs between the authors. Orlan and Lin (1969) and Doll (1967) calculated the weather index as a ratio of yield predicted for the actual weather to yield predicted for the average weather. Desai (1986) first calculated the weather variability of yield by subtracting yield predicted for the actual weather from yield predicted for the average weather. Then weather adjusted yield was derived by subtracting the estimated weather variability of yield from actual yield. These two methods for deriving the weather index and weather adjusted yield are likely to produce

Table 2.1: A survey of selected studies on empirical statistical modeling of the weather-yield relationships

Author	Desai P. (1986)	Dai Qi *et al.* (1993)	Doll P. (1967)	Duncan, R.C. (1972)	Guise, J. (1969)	Kaylen, M.C. and Koroma, S.S. (1991)
1. Country	1. Soviet Union	1. US	1. US	1. Australia	1. New Zealand	1. US
2. Industry	2. Wheat	2. Corn	2. Corn	2. Sheep	2. Wheat	2. Corn
3. Independent Variable	3. Yield	3. Yield	3. Yield	3. Sheep Number per Man	3. Yield	3. Yield
1. Functional form between yield and weather variables	1. Linear	1. Quadratic and Cubic	1. Linear	1. Linear	1. Linear	1. Not clear
2. Type of time trend	2. Linear	2. Not included	2. Cubic	2. Linear	2. Linear	2. Stochastic Trend
Weather Variables included in the model	Monthly Precipitation and Temperature	Monthly Soil Moisture Index	Weekly Rainfall	Annual Rainfall	Monthly Rainfall, Temperature and Soil Moisture	Monthly Precipitation and Temperature
1. The method of selection of weather variables	1. Priory Reasoning	1. Simulated	1. Priory Reasoning	1. Priory Reasoning	1. Stepwise Method	1. Priory Reasoning
2. Type of Data	2. Non-experimental	2. Experimental	2. Non-experimental	2. Non-experimental	2. Non-experimental	2. Non-experimental

Table 2.1 cont.

Offutt, S.E. Garcia, P. and Pinar, M. (1987)	Orlan, B. and Wuu-Long Lin (1969)	Reid, G.K.R. and Thomas, D.A. (1973)	Sakamato, C.M. (1978)	Thompson, L.M. (1969)	Williams *et al.* (1975)
1. US 2. Corn 3. Yield	1. US 2. Wheat, Corn, Sorghum 3. Yield	1. Australia 2. Sheep 3. Wool and Net Reproductive Rate	1. Australia 2. Wheat 3. Yield	1. US 2. Wheat 3. Yield	1. Canada 2. Wheat 3. Yield
1. Linear and Quadratic 2. Linear	1. Linear 2. Linear	1. Log-Linear 2. Not included	1. Linear 2. Linear	1. Linear 2. Polynomial	1. Quadratic 2. Linear
Monthly Precipitation and Temperature	Drought Severity Index	Seasonal Rainfall and Soil Moisture	Moisture Anomaly Index	Monthly Rainfall, Temperature and Precipitation	Monthly Rainfall and Evapotranspiration
1. Priory Reasoning 2. Non-experimental	1. Priory Reasoning 2. Non-experimental	1. Priory Reasoning 2. Non-experimental	1. Priory Reasoning 2. Non-experimental	1. Priory Reasoning 2. Non-experimental	1. Stepwise Method 2. Non-experimental

similar results. However, the method employed by Desai has the advantage that it gives ready estimates of the weather variability of yield. The weather index derived using this method and Stallings' index are more commonly used by agricultural economists and they are totally different from the weather index discussed earlier. While the latter is a combination of two or three weather variables, the former is claimed to reflect the aggregate influence of weather on agricultural output.

Shaw (1964) pointed out to the importance of the geographical aggregation problem. This problem occurs because the relationships between crop yields and meteorological factors are not monotonic. In other words, crop yields do not always increase as the value of a meteorological factor increases. Assuming that the optimum yields are achieved at four units of the meteorological factor, then three units would have approximately the same effect on yields as five units. The idea is that the results will be better if the weather-yield relationships are analysed at less aggregate levels. But Shaw's next point seems to contradict this idea. He wrote:

> The use of unaggregated meteorological data in attempts to show relationships between weather factors and yields is subject to some difficulties. Practically, the number of observations involved becomes very large when divisions smaller than states are used. Conceptually, the yield estimates for divisions smaller than states are much less accurate than those for larger divisions. As a result, many studies have used state yield averages and meteorological averages as the basic data.

It is not clear why greater number of observations create a problem and yield estimates for smaller divisions are not accurate. Analysts use aggregated data mostly because reliable unaggregated data on smaller divisions are not readily available in many cases.

One interesting direction in the study of the weather-yield relationships has been the analysis of weather variability of agricultural outputs by Desai (1986), Offutt *et al.* (1987) and Mehra (1981).

2.3 Technical Change and its Measurement

It is widely acknowledged that, amongst the factors that explain productivity differences, the most important is technical change. Therefore, most productivity studies are centred on the measurement of technical change. As an explanation for the growing interest in technical change studies Peterson and Hayami (1977) cite two major problems: the significant increase in the supply of agricultural products relative to demand in developed countries leading to depressed farm prices and incomes and the difficulty faced by developing countries in increasing agricultural output.

Before going deeper into a review of literature on agricultural productivity one should be clear in distinguishing between the terms 'productivity growth' and 'technological change' or ' technical change'[1] because they are used interchangeably to label shifts of a production function.

In elementary economic theory, productivity is defined in terms of the rate of output produced per unit of input utilised in the production process. According to Ruttan (1960), technical change in the context of a production function is the creation of a new production function or an upward shift in the production function.

A simple illustration of productivity growth caused by different factors is provided by Capalbo and Antle (1988). This illustration is shown here to distinguish the term 'productivity growth' from the term 'technical change'. Generally, the rate of output depends on three factors: the state of technology, the quantities and types of resources put into the production process, and the efficiency with which those resources are utilised. Figure 2.1 shows single-output neoclassical production functions F(X) and F'(X), which represent the technically efficient combinations of input X and output Q for two different production processes

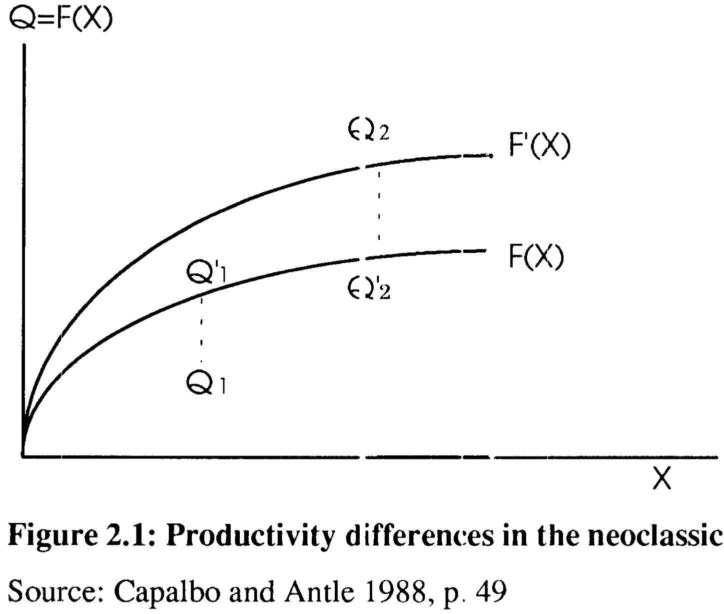

Figure 2.1: Productivity differences in the neoclassical model

Source: Capalbo and Antle 1988, p. 49

---

[1] The terms 'technical change' and 'technological change' are used interchangeably to refer the same thing-input quality improvements. In this study we do not make a difference between them and use the term 'technical change' for convenience.

In Figure 2.1 Q_1 and Q_2 are the outputs observed in periods 1 and 2, and **F** and **F'** are production processes used in period 1 and 2, respectively. Since these two observations lie on different rays from the origin, total factor productivity (TFP), measured as the average product of factor X, is greater in period 2 than in period 1. This measured productivity change can be attributed to three distinct phenomena. First, Q1 is below **F(X)**, indicating technical inefficiency; efficient production would have resulted in output Q'1 . Second, output Q2 was produced with a greater input than was Q1, so there is difference in scale of production, which explains the difference between Q'1 and Q'2. Third, the production function **F** exhibits a higher total productivity than **F**, which explains the gap between Q2 and Q'2. Thus, the observed differences in TFP over time can be explained by differences in productive efficiency, the scale of production, and the state of technology. Therefore, it seems quite safe to state that the term 'productivity growth' refers to the growth of output relative to inputs in general terms. The term 'technical change' refers to a specific form or source of productivity growth. Put in another way, productivity change and technical change are synonymous only if the sources of productivity growth, other than technical change, are assumed to be constant.

It is crucial to understand that in order to have shifts in a production function there must be changes in the quality of inputs. The fact that we observe technical change means that some inputs have changed in quality and these quality changes are not reflected in the total input measure or left as a residual. Thus, the use of the term ‘technical change’ (the residual) is an indication that we do not know where, at least, a part of the output is coming from (Peterson and Hayami, 1977). It is also important to recognise that, as noted by Schultz (1958), technical change is not "manna from heaven", resources must be devoted to improving the quality of inputs. According to Peterson and Hayami (1977), the main sources of technical change in United States' agriculture were *increase in skills of farm people* (quality of human capital); *increase in quality of nonhuman capital* (machinery, equipment and buildings); *increase in quality of other inputs* (fertiliser, chemicals, and more efficient breeds of livestock etc.); *increase in quality of output; and economies of scale.*

Technical change in the production process can be realised in either embodied or disembodied form. Embodied technical change refers to the changes in input quality or the introduction of new processes and new inputs. Disembodied technical change refers to improved methods of utilising existing resources, such that a higher output rate per unit of input is obtained (Capalbo and Antle, 1988). According to Solow (1962), embodied technical change is the more important kind, although this viewpoint has precipitated a substantial amount of controversy and little empirical support (Peterson and Hayami, 1977).

Technical change is often defined as neutral or biased. Technical change is neutral if the marginal rate of substitution between inputs is not affected. Non-neutral technical change is described as either labour saving (capital using) if the marginal product of capital rises relative to the marginal product of labour. According to Hicks (1932), the rise in the price of labour relative to capital tended to cause induced innovation to have a labour-saving bias. Empirical tests of the induced innovation hypothesis were hampered by the difficulty of distinguishing movements along the production function from shifts in the production function. For example, the question of to what extent should the increase in capital intensity be attributed to factor substitution as a result of the change in relative prices, and to what extent to a possible labour-saving bias in technical change remains unanswered (Capalbo and Antle, 1988).

Two approaches to the measurement of technical change have been identified in the literature: the index number and the econometric approaches.

Among the earliest methods of quantifying agricultural productivity were the index number approaches. These productivity measures were ratios of an index of aggregate output to either a single factor, typically labour or an index of all factors. The partial productivity indexes were biased measures of technical progress because they included the effects of factor substitution together with the effects of advances in production techniques (Peterson and Hayami, 1977). The use of partial productivity indexes was replaced by the total factor productivity (TFP) measures in the late 1950s, when output and input indexes were constructed using either linear aggregation with market prices as weights or a geometric aggregation with factor and revenue shares as weights. TFP is defined as a ratio of output to the aggregate of all factor inputs. Solow (1957) was the first to introduce the concept of the aggregate production function into technical change studies. In determining aggregate input and output measures, the method by which raw data are combined into a manageable number of aggregates is important. Recent advances in production economics have made it possible to identify the economic assumptions that are implicit in the choice of indexing procedure (Capalbo and Antle, 1988). For example, the geometric index has been shown to be exact for a Cobb-Douglas production function, and the Tornqvist-Theil index is exact for a homogeneous translog production function.

One disadvantage of the index number approach is that calculations are not based on statistical theory, so statistical methods cannot be used to evaluate their reliability. However, index number calculations can be used when econometric methods are infeasible. For example, very detailed data with many inputs and output categories can be used regardless of the number of observations over time; there are no degrees of freedom problems or statistical reliability problems in working with small samples.

The econometric approach is based on econometric estimation of the production technology and is inferred from shifts in the production function. Technical change that is postulated to be neutral and at a constant rate over time, makes for the simplest possible measurement from the analytical point of view (Brown, 1966). Christensen (1975) and Caves, Christensen and Swanson (1981), however, note that if the assumptions of constant returns to scale and static equilibrium are violated, then estimates of productivity growth include the effect of scale economies and movements toward or away from equilibrium.

Peterson and Hayami (1977) noted that economies of scale, as a more efficient organisation of traditional inputs stemming from an increase in the size of industry, cause an estimation problem because new technology may make it possible to realise scale economies that hitherto could not have been obtained. For example, in poultry production, the development of medicated feeds makes it possible to keep a large number of birds in one location. In this case, the introduction of medicated feeds associated with higher levels of output led to an increase in the marginal productivities of traditional inputs relative to the marginal productivites at lower level of output, i.e. before the introduction of medicated feeds. In such cases the effects of technical change and scale economies are inseparable (Peterson and Hayami, 1977).

In most empirical work, the production function is usually estimated by regression analysis and, so, provides only an average relationship between inputs and output, since the regression line is fitted through the mean of the data set. In contrast, the estimation of a production frontier is claimed to correspond to the formal economic definition of a production function i.e. the *maximum* output that can be produced from a specified set of inputs (Alaudin, Squires and Tisdell, 1993). The idea of measuring technical efficiency using the production frontier was first introduced by Farrell (1957). Technical efficiency refers to the maximum attainable level of output for a given level of production inputs, given the range of alternative technologies available to the producer (Ellis, 1988). A general presentation of Farrell's frontier concept is depicted in Figure 2.2, where the horizontal axis represents the inputs, **X**, associated with producing the output, **Y**.

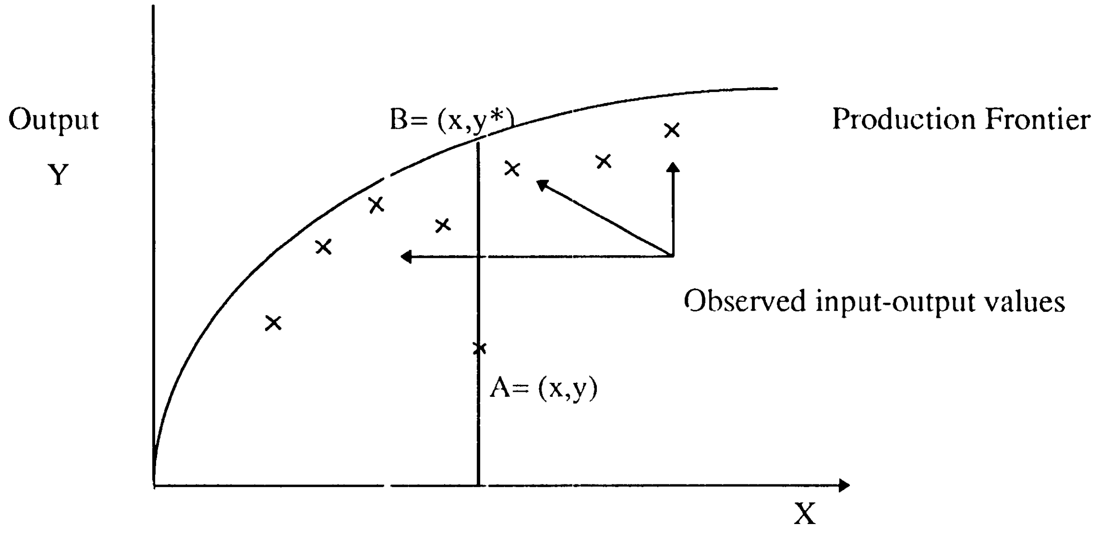

Figure 2.2: Technical efficiency of firms in input-output space.

Source: Battese, 1992

The observed input-output values are below the production frontier, given that firms do not attain the maximum output possible for the inputs involved with the technology available. A measure of the technical efficiency of the firm which produces output **y** with inputs **x** denoted by point A, is given by **y/y***, where **y*** is the 'frontier output' associated with the level of inputs, **x** (point B). The distance between point A and the frontier is a measure of technical inefficiency.

The existence of technical inefficiency of firms engaged in production has been the subject of considerable debate in economics (Battese, 1992). For example, Muller (1974) stated:

> However, little is known about the role of non-physical inputs, especially information or knowledge, which influence the firm's ability to use its available technology fully...This suggests how relative and artificial the concept of frontier itself is ... Once all inputs are taken into account, measured productivity

> differences should disappear except for random disturbances. In this case the frontier and the average function are identical. They only diverge if significant inputs have been left out in the estimation.

However, as Battese (1992) argues, the econometric modeling of frontier production functions provides useful insights into measures by which productive efficiency of different firms may be compared.

Attempts to identify the different sources of productivity growth (factor substitution, scale economies, capacity utilisation, and technical efficiency) are made by Berndt and Khaled (1979), Daly and Rao (1985), Caves, Christensen, and Swanson (1981), Battese and Coelli (1995), Forsand and Hjalmarsson (1979) and others.

A major statistical problem in the estimation of production parameters in the presence of technical change is specification bias caused mainly by approximating technical change inadequately. In applied economic analysis, it has become customary to introduce a smooth time trend as a proxy for technical change. This convention fails when technical change is in fact discrete or cannot be approximated by a statistically manageable function of time (Peterson and Hayami, 1977). Although it has been criticised by some, most studies used a linear trend as a proxy for technical change. Shaw (1964) argued that a straight line time trend variable would misspecify the real structure of technical change, since he believed that the technological trend function followed a steplike pattern, in which shifts occurred with the introduction of new technology. The Shaw's step function hypothesis of technical change was supported by the findings of the study on New Zealand wheat carried out by Guise (1969).

Thus, the econometric approach allows the relaxation of some of the assumptions required for the index number approach, but only at the cost of necessitating other assumptions and methodological difficulties. An econometric production function, such as the translog, can be estimated without making any assumptions about neutrality of technological change, returns to scale, or industry equilibrium. However, estimation of the translog production function with aggregate data requires that the outputs be aggregated into a single index, so input-output separability must be assumed. For sufficient degrees of freedom, and to mitigate multicollinearity problems, it is necessary to aggregate input data into a small number of categories, which can be done only under input separability assumptions (Capalbo and Antle, 1988).

As a result, the choice between the index number and the econometric approaches is based on the research objectives, the data requirements, the availability of data, and the appropriateness of assumptions. Griliches (1963a and 1963b) and Young (1971) acknowledged the superiority of the production function approach arguing that it can explicitly account for the technical characteristics of the production process, like economies of scale and elasticity of substitution. According to Griliches (1963b), many of the problems related to the measurement and sources of technological change can be best analysed and answered within an explicit production function framework. This framework is particularly well adapted to answering questions such as: "What are the variables to be included in an equation explaining output and in what form?" and "What weights should they be combined with to derive a residual measure of technical changes?"

2.4. Pastoral Nomadism

The analysis of specific features of peasants and their economic behavior in connection with modern neoclassical theory of farm production has been carried out by Ellis (1988). He distinguished dependence on crop or livestock activities, family labour, and partial engagement in input and output markets as main features of peasant production. According to Ellis nomads are half-peasants. Widstrand (1975) noted that a pastoral livestock operation is not a capitalistic undertaking aimed at producing a marketable surplus, its aims are rather to provide a good, regular supply of food for the family, to enable it to survive physically and socially. This statement is also in agreement with the viewpoint of Dillon and Hardaker (1984). They wrote that pastoralists regard their livestock as a walking bank, a measure of social status and security, but seldom as an enterprise to be rationally managed to produce profit. In contrast, commercial ranching and dairy farming aim at converting herbage into marketable produce, and that objective is achieved with large herds upon which only a small number of people are dependent (Widstrand, 1975).

One of the main distinguishing characteristics of pastoral economies stems from the relationships between pastoralists and the natural resource base. Animals are owned by individuals but the natural resources necessary for livestock operations, such as grazing and water, are not. As grazing is communal and the ownership of livestock is individual, the perceptions of nomad livestock owners concerning the options open to them leave them little choice but to continue on their present course of trying to increase the size of their herds, even though that course leads to ecological disaster.

Interesting interpretations on the characteristics of nomadic people and possible development paths for them has been found in MacArthur (1980) and Widstrand (1975). According to MacArthur (1980), rainfall is the primary factor determining indigenous ways of using grassland. For example, total nomadism and semi-nomadism correspond to areas where annual rainfall is 50-200 mm and 200-400 mm, respectively. According to this classification, dry southern regions of Mongolia belong to total nomadism and less dry northern regions belong to semi-nomadism. As a development path for semi-nomadism where technical opportunities are more promising, MacArthur (1980) suggested different combinations of ranching. In total nomadism, because of ecological constraints, it is scarcely worth trying to improve the feed situation by cropping fodder, and cross-breeding of animals with more productive types cause more harm than good. Because of these considerations, as well as the fact that 'these particular people have little desire for change', MacArthur states, it is better to leave total nomads undisturbed. While MacArthur may be right in his first reasoning, his second argument is perhaps a bit hasty. It seems quite obvious that many of these people keep total nomadism not because they have no desire for change, rather, they have no other option. In this respect, Widstrand's (1975) statement that 'there is nothing in the culture or value system of livestock peoples that would prevent them from appreciating change -even settlement if they can perceive and appreciate the benefits that would accrue from such a change' - seems more plausible.

Regarding a development path for subsistence agriculture, McLoughlin (1969) pointed out to the importance of emphasizing technical change as a core issue in the movement toward higher levels of farm output and living standards.

3. Econometric issues

3.1 Introduction

Any discussion of technical change in agriculture assumes something about the underlying aggregate production function. In specifying the production function in agriculture one has to make several choices as to what will be estimated and how. These choices include, among others: (i) the choice of the appropriate functional form; (ii) the choice of the variables to be included in the production function; (iii) the choice of a technique for estimating the coefficients of the production function. These choices are discussed in this chapter.

3.2. Choice of functional form

Heady and Dillon (1961) provide the most comprehensive guide to understanding concepts, economic applications, specification and estimation problems of agricultural production functions. More recently, Dillon and Anderson (1990) summarise the new developments in agricultural response analysis. The review of recent developments in production economics in relation to productivity measurement is carried out by Capalbo and Antle (1987).

The problems of functional form in productivity studies are discussed by Capalbo and Antle (1987), Kaneda (1982), Kloot and Anderson (1977) and others. The choice of functional form usually depends on the assumptions and objectives of the study, and the data availability. Most productivity studies use the translog (Berndt and Christensen 1973; Callan 1987; Caves, Christensen and Swanson 1981; McKay, Lawrence, and Vlastuin 1982), the Cobb-Douglas (Hayami 1970; Sidhu 1974; Srivastava, Nagadevara, and Heady 1973) and the constant elasticity of substitution (CES) form (Bairam 1991; Duncan 1972).

The widespread popularity of the Cobb-Douglas form is attributable to the fact that it provides a compromise between (i) adequate fit of data, (ii) computational feasibility and (iii) the need for sufficient degrees of freedom in statistical testing. In addition, it yields estimates of production elasticities directly and permits the phenomenon of decreasing returns to be manifested using a relatively uncomplicated function. Perhaps the most crucial argument against the Cobb-Douglas model is that it imposes the property of unitary elasticity of substitution between input categories. In addition, the Cobb-Douglas form, like the CES function, cannot capture rich implications of substitutability and complementarity relations in multi-input technology (Kaneda, 1982). Accordingly, the use of this form can be generally justified if there is evidence that the elasticity of substitution is unity or close to one.

An important anomaly that keeps reappearing in Cobb-Douglas studies has been observed by Heady (1946), Chowdhury, Nagadevara, and Heady (1975) and Doll (1974). Namely, estimates based on cross-sectional samples, as is the case in most studies, almost typically result in some elasticities for farm labour and land which are negative. For example, the production function analysis conducted by Agrawal and Foreman (1959) and Suryanarayana (1958) resulted in negative elasticities for labour and capital services. These negative coefficients, which are usually difficult to explain, confuse the analysis and leave the researcher more or less empty-handed. One reason for these negative elasticities is the prevalence of multicollinearity in random or unstratified samples, where farmers with larger labour inputs are also those with larger inputs of land and various forms of capital. Another reason is attributed to reporting and measuring biases. For example, the labour variable included in most production function studies is not the labour used (service flows) but rather the labour available for use (Doll, 1974). Although the input of hired labour may be reported accurately, the operator is prone to reporting twelve months labour. Included in these twelve months is the time actually spent at farming and also the slack months in which only a few hours of chore work are done each day (Heady, 1946). It is worth noting that these types of reporting and measuring biases are very common to Mongolian herders. Moreover, the measurement of service flows from capital is further complicated because of the mobility of herds as opposed to the stationarity of most capital.

As a way out of the problem of negative elasticities Chowdhury, Nagadevara, and Heady (1975) used a Bayesian technique with prior restrictions on the sign of land, labour and fertiliser coefficients.

The constant elasticity of substitution (CES) production function, suggested by Arrow *et al.* (1961) is a generalisation of the rather restrictive Cobb-Douglas function by permitting the elasticity of substitution to be equal to any constant (Chung, 1994). The CES function with two inputs, capital (K) and labour (L), can be written as follows.

$$Q = \gamma[\delta L^{-p} + (1-\delta)K^{-p}]^{-1/p} \qquad (3.1)$$

where Q is output, γ is the efficiency parameter ($\gamma > 0$), δ the distribution parameter ($0 < \delta < 1$) and p is the substitution parameter ($-1 < p < \infty$). The elasticity of substitution, ε, derived from (3.1) is

$$\varepsilon = 1/(1+p) \qquad (3.2)$$

From (3.2) it follows that the elasticity of substitution is different from unity as long as the substitution parameter $p \neq 0$. If, however, the parameter p equals zero, the value of the elasticity of substitution ε is exactly one, and the CES function reduces to the Cobb-Douglas form.

Another way of generalising the Cobb-Douglas production function is the transcendental logarithmic (translog) function, which expresses the logarithm of output as a quadratic function of inputs in logarithms. Again with two inputs, a translog function takes the form:

$$\log Q = A(T) + B_K \log K + B_L \log L + B_{KK}(\log K)^2 + B_{LL}(\log L)^2 + B_{LK}(\log K \log L) \qquad (3.3)$$

It is clear that this function reduces to the Cobb-Douglas function if the quadratic terms are disregarded. Thus, the quadratic terms can be regarded as amendments to the Cobb-Douglas assumption of unitary elasticity of substitution. This function allows arbitrary and variable elasticities of substitution among input categories. It provides a second order approximation to an arbitrary production function at any given point (Christensen, Jorgenson and Lau, 1973).

It is also common to estimate production or yield functions as polynomials, usually of first or second degree depending upon the area of the surface to be approximated. The larger the area of the production surface to be approximated, the higher should be the degree of the polynomial.

Some researches choose a function which appeals to the a *priory* hypothesis about the underlying production structure. In the absence of such an priory hypothesis, several algebraic forms are estimated and the one which demonstrates the best performance in terms of goodness of fit, statistical efficiency and economic viability is accepted as the best. However, neither of these types of criteria provides clear and unambiguous guidance, so choice of functional form is inevitably somewhat arbitrary (Upton 1979).

An interesting study on the sensitivity of the estimated coefficient of a time trend as a proxy for technical change to the algebraic form of the function specified and the estimating techniques used is that of Kloot and Anderson (1977). Analysing data from a single sheep grazing property, they found that the estimated rates of technical change varied from - 0.0037 in the Cobb-Douglas function to -0.0051 in the translog specification, and -0.0167 in the CES specification. Calling this variation "model and method variance" they concluded that it is not too obvious what implications this has for students of technical change except, possibly, to engender an increased irreverence for particular point estimates that have appeared and probably will continue to appear in the literature.

3.3 Specification errors

Griliches (1957) wrote that 'sometimes we may know what we want, but even then it may be impossible to obtain it. Either there are no pertinent data or the variables are non-measurable, or our budget and computational facilities are limited. Hence, we exclude variables, accept approximations, aggregate and commit various other sins of omission and commission'.

The omission of relevant variables and inclusion of irrelevant variables are the two most common errors of specification. Considering specification errors of the first type, it is important to note that the model specified can never be estimated (Doll, 1974). The point here is that the correct model should always be specified even though it cannot be estimated. To do otherwise will lead to misinterpretation of the estimated sample parameters. For example, if all inputs are not specified in the theoretical model (as opposed to the statistical model), then the effects of the included inputs may be overestimated. The omission of relevant variables from the model also causes an overestimation of residual variance. This results in larger standard errors, rendering the inferences about the parameters inaccurate. The most common variable that is omitted in production function analysis is 'management'. A broad discussion of this problem is found in Heady and Dillon (1961), who added a management variable to a Cobb-Douglas function that also included land, labour, and capital inputs. Mundlak (1961) combined cross-section and time-series data and utilised dummy variables to measure the efficiency of management.

Including an irrelevant variable into the specification does not introduce biases into the parameter estimates. But the estimates loose precision unless the irrelevant variable, which is included, is orthogonal to the other explanatory variables. Usually, when irrelevant variables are included in the model, estimates of their coefficients should be small, and, on the average over many studies, close to zero (Doll, 1974). In such a case the researcher must be cautious in the interpretation of these coefficients. The researcher might infer that an input category is unproductive when in fact the result is due to misspecification. For example, in cross-section analysis of wheat farms, all input-service flows in wheat production will be perfectly correlated with acres of wheat grown on the farm. Whatever the farmer does on one acre, he does on all. Thus, when wheat acreage is included in the production function, the resulting estimated coefficient will capture the summed effects of all service flows used per acre. So, the service flows of an input are perfectly correlated with quantities of one or more other inputs, then the inclusion of a stock measure of that same input amounts to a misspecification error (Doll, 1974).

3.4 Econometric analysis of panel data

3.4.1 Estimation of Equations with Panel Data

Econometric issues associated with the use of panel data have been broadly investigated in Baltagi (1995), Dielman (1989), Hsiao (1986), and Judge *et al* (1985).

Hsiao (1985) and Baltagi (1995) listed several advantages of panel data over conventional cross-sectional or time-series data sets. These included the following.

1. controlling for individual *heterogeneity*. Panel data suggest that individuals, firms, states or countries are heterogeneous, time-series or cross-section studies not controlling for this heterogeneity run the risk of obtaining biased results.

2. they usually give a researcher a more informative data, more variability, less collinearity among the variables, more degrees of freedom and more efficiency.
3. panel data are more able to capture the dynamics of adjustment.
4. panel data are better able to identify and measure effects that are simply not detectable in pure cross-sectional and time-series data.
5. panel data models allow to construct and test more complicated behavioural models.
6. panel data are usually gathered on micro units, so, biases resulting from aggregation over these units are eliminated.

Despite these advantages, the use of panel data is associated with some special problems. As Dielman (1989) notes, when combining cross-sectional and time-series data one may find the problems of serial correlation and heteroskedasticity occurring simultaneously. In addition, one may find that cross-sectional disturbances for different individuals at the same point in time are correlated. Such contemporaneous correlation is an added problem specific to the analysis of panel data (Dielman, 1989).

Hsiao (1985) and Judge *et al.* (1985) classified the models for panel data into four classes.

1. Slope coefficients are constant, and the intercept varies over individuals:

$$y_{it} = \alpha_i + \sum_{k=1}^{K} \beta_k x_{kit} + u_{it} \qquad (3.4)$$

2. Slope coefficients are constant, and the intercept varies over individuals and time:

$$y_{it} = \alpha_{it} + \sum_{k=1}^{K} \beta_k x_{kit} + u_{it} \qquad (3.5)$$

3. All coefficients vary over individuals:

$$y_{it} = \alpha_i + \sum_{k=1}^{K} \beta_{ki} x_{kit} + u_{it} \qquad (3.6)$$

4. All coefficients vary over individuals and time:

$$y_{it} = \alpha_{it} + \sum_{k=1}^{K} \beta_{kit} x_{kit} + u_{it} \qquad (3.7)$$

where i = 1,2,3, ..., N (cross-sectional units)

t = 1,2,3, ..., T (time-series observations)

α = the intercept term

β = slope parameters

χ_{ki} = the k-th explanatory variable for the i-th cross section in time t

u_{it} = the error term

In each of these four cases, the model can be classified further depending on whether the coefficients are assumed to be random or fixed. Among these the model with constant slopes and variable intercepts (models 3.4 and 3.5) are most widely used when analysing panel data because they provide simple, yet reasonably general alternatives to the assumption that parameters take values common to all agents and all times (Hsiao, 1986). The time and cross-sectionally varying parameter models are the most difficult to handle notationally, computationally, and analytically (Dielman, 1989). The basic assumption of the models (3.4) and (3.5) is that, conditional on observed explanatory variables, the effects of all omitted variables are driven by three types of variables: individual time-invariant, period individual-invariant, and individual time-varying variables. Further, an assumption often made in the case of production function analysis of panel data, is that the different cross-sectional units lie on different levels of efficiency, implying that the intercept terms are different between units, while the response coefficients of the explanatory variables are constant. The differences in efficiency between cross-sectional units might be explained by the differences in ecological environment, management, skills of workers etc. Thus, the model with constant slope coefficients and an intercept that varies over individuals (3.4) can be regarded as the simplest but quite reasonable representation of reality.

Depending upon the assumptions regarding the intercept α_i, which is assumed to capture differences in behaviour over individuals, there are two approaches; (i) the dummy variable model is suggested if α_i is assumed to be constant, (ii) if the intercept is assumed to be random, the error components model is appropriate (Judge *et al.*1985). These models are discussed below.

3.4.2 Dummy variable model

The dummy variable model is obtained by introducing the dummy variable D_{jt} into the model 3.4:

$$y_{it} = \sum_{j=1}^{N} \alpha_j D_{jt} + \sum_{k=1}^{K} \beta_k x_{kit} + u_{it} \qquad (3.8)$$

where D_{jt} takes the values zero if j = i or one otherwise. Under the assumption that $E(u_{it})=0$, $E(u_{it}^2)=\sigma_u^2$ and $E(u_{it}, u_{js})=0$, for $i \neq j$ or $t \neq s$, the ordinary least squares (OLS) give the best linear unbiased estimates (BLUE).

3.4.3 The error components model

Instead of assuming that the intercept α_i in the model 3.4 is a fixed parameter, it is now treated as a random variable equal to a mean plus a random error term:

$$\alpha_i = \alpha_0 + \mu_i$$

In this case model 3.4 can be written as:

$$y_{it} = \alpha_0 + \mu_i + \sum_{k=1}^{K} \beta_k x_{kit} + u_{it} \quad \text{or}$$

$$y_{it} = \alpha_0 + \sum_{k=1}^{K} \beta_k x_{kit} + v_{it} \qquad (3.9)$$

where $v_{it} = \mu_i + u_{it}$ is a composite error term. Because the composite error terms is made of two components, the model 3.9 is commonly referred to as an error components model. In the composite error term, the component u_{it} is the usual error term which varies over both individuals and time and the component μ_i is associated with i-th individual and is a constant over time. The assumption that μ_i is a random variable implies that the N individuals can be regarded as a random sample from a larger population, and μ_i is not correlated with the explanatory variables (Judge *et al*, 1985). It is assumed that the components u_{it} and μ_i have zero means and are independent from each other. It is further assumed that each component exhibits no serial correlation and has constant variance. In this case it can be shown that

$$E(v_{it}) = 0 \text{ and}$$

$$\begin{aligned} E(v_{it} v_{js}) &= \sigma_\mu^2 + \sigma_u^2 && \text{if i=j and t=s} \\ &= \sigma_\mu^2 && \text{if i=j and } t \neq s \\ &= 0 && \text{if } i \neq j \end{aligned}$$

In the case of the error components model, the randomness in the cross-sectional effects requires the use of the generalised least squares estimator (GLSE) for the estimation of the parameters.

3.4.4 Choosing between random or fixed effects

As noted earlier, the choice between the two procedures to estimate model (3.4) depends on whether we treat α_i as fixed or random. If the choice between these two assumptions is clear, then the estimation procedure could be chosen accordingly. For instance, consider an example in which several herders care for their herds. The effects of herders can be assumed random. But if the situation is not that each herder comes and goes, randomly sampled from all herders, but that all are available, and if we want to assess differences between those specific herders, then the fixed-effects model is more appropriate. Similarly, if an experiment involves hundreds of individuals who are considered a random sample from larger population, the random effects model is more appropriate. Thus, the situation to which a model applies and the inferences based on it are the deciding factors in determining whether effects should be treated as random or fixed (Hsiao, 1986). However, the choice between fixed and random effects is often not clear. Mundlak (1978) argues that it is more reasonable to assume that μ_i are correlated with the explanatory variables, x_{it}, and hence, E(μ_i) will not be independent but some function of x_{it}. In this case, the error components model has characteristics similar to those associated with omitted variable misspecification and its estimators are biased. Some arguments advanced against a dummy variable model include (i) it eliminates a major portion of the variation among both the explained and explanatory variables, if the between firm and between time variation is large, (ii) using dummy variables erodes the degrees of freedom, and (iii) it is rarely possible to meaningfully interpret the dummy variables. It has also been pointed out that there is no reason to treat these effects as fixed, since they, like the residuals, are the measures of ignorance (Maddala, 1971). Judge *et al* (1985) suggested a more practical procedure to solve the problem. The following is the brief of this procedure as represented by Griffiths (1996).

1. If T is large, N is small, the assumptions of the error components model hold and E(μ_i x_{it}) = 0, then **b** (represents the estimators from a dummy variable model) and **B** (represents the estimators from an error components model) are both consistent and asymptotically efficient. It follows that we should use the estimator which is computationally easier, namely **b**.

2. If N is large, T is small, the assumptions of the error components model hold and E(μ_i x_{it})=0, then **b** and **B** are both consistent but only **B** is asymptotically efficient. It follows we should choose **B.**
3. If the assumptions of the error components model hold and E(μ_i x_{it})≠0, then **b** is consistent, irrespective of the relative size of N and T. It follows that we should use **b.**

In the present study, correlation between the regional effects and the inputs can be quite logically assumed, as technically efficient regions are likely to use more inputs. In this case, following the procedure described above, the dummy variable model is more appropriate. One thing which needs to be noted with respect to the use of dummy variables in the present study is that the production function is specified in a way that takes account of climate. Accordingly, regional or district dummy variables can be introduced, as mentioned earlier, to take account of the differences in efficiency levels between districts or regions that are conditional on factors other than climate such as vegetation and soil type, management and skills of herders.