UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

EMPIRICAL ANALYSIS IN SOUTH AFRICAN AGRICULTURAL ECONOMICS AND THE R-SQUARE DISEASE

by

WALTER HEINRICH MOLDENHAUER

Submitted in partial fulfilment of the requirements for the degree

MAGISTER COMMERCII (AGRICULTURAL ECONOMICS)

in the

FACULTY OF ECONOMIC AND MANAGEMENT SCIENCES

at the

UNIVERSITY OF PRETORIA

DECEMBER 2006

DECLARATION

I hereby certify that, unless specifically stated to the contrary in the text, this dissertation is the result of my own original work.

WALTER HEINRICH MOLDENHAUER

ACKNOWLEDGEMENTS

This research project would not have been possible without the support and encouragement of various individuals. Firstly, I wish to express my sincere appreciation to Prof Johann Kirsten for his supervision and guidance of the work and also for providing the opportunities to complete this study. I would also like thank Prof Deidre McCloskey for her advice especially with respect to the development of the questionnaire and selection of articles necessary for the completion of this thesis.

My appreciation is also extended to Mrs Isabel Klement for her assistance in finding especially some of the first volumes and articles of *Agrekon*. A special word of thanks also goes to Mr Gerrie Mans for scanning photo copied articles of *Agrekon* and converting them to electronic format.

To Yolanda goes my deepest appreciation for her unfailing support and faith in me. To my parents, who have taught me the qualities of hard work self-discipline and perseverance. I thank them for their love and support and dedicate this study to them. Finally, our Creator, who has provided us with the talents we can use, to him all the praise.

Walter Heinrich Moldenhauer
December 2006

ABSTRACT

EMPIRICAL ANALYSIS IN SOUTH AFRICAN AGRICULTURAL ECONOMICS AND THE R-SQUARE DISEASE

by

Walter Heinrich Moldenhauer

Degree: Magister Commercii (Agricultural Economics)
Department: Agricultural Economics, Extension, and Rural Development
Supervisor: Prof. J.F. Kirsten

The South African agricultural sector underwent a significant amount of institutional and structural changes during the past two decades, especially in the aftermath South Africa's first democratic elections in 1994 and the deregulation of the agricultural marketing environment in 1996/97. These changes meant that South African agricultural economics scholars had to adapt to these changes. The increased need towards more quantified output in agricultural economic research has led agricultural economic scholars to "borrow" econometric models from their fellow scholars abroad to apply to South African research problems in order to fulfil the need for more quantified research output.

However, the development of econometrics has over the years given rise to a disenchantment with the way in which econometrics have been applied in economic research. Consequently it is believed that a large body of literature has entered the public domain without being properly reviewed because South African agricultural economic scholars do not have the necessary insight and knowledge of the problems believed to be at the root of the disenchantment with the manner in which econometrics have been applied.

The general objective of this dissertation is to investigate the disenchantment with the manner econometrics has been applied in economic and agricultural economic scholarship in order to

identity the main drivers of this disenchantment, and to use this knowledge gained to evaluate the application of econometrics in South African agricultural economic scholarship as portrayed in *Agrekon,* one of South Africa's agricultural economics peer review journals.

The study is conducted by means of a review of the literature on the history of econometrics, the development of econometric methodologies and the disenchantment with econometrics in economics and agricultural economics. Applied econometrics portrayed in *Agrekon* is evaluated by means of a survey of papers published in this journal.

The main findings of this study revealed that the key drivers of disenchantment can mainly be ascribed to the following:

- The misuse of statistical significance tests in applied studies.
- Problems with data underlying econometric analyses.
- The problems associated with replication.
- Data mining
- The "Black box ideology" in applied econometrics and
- Scholasticism and associated preference falsification.

A survey of papers published in *Agrekon* based on a sample of 65 papers, which were sampled by means of stratified random sampling, revealed that elements behind the disenchantment with econometrics are present in South African agricultural economic scholarship. It was also found that the data underlying econometric analyses are a major point of concern in South African agricultural economics and it seem as if South African agricultural economics scholars have adopted a lackadaisical attitude towards data. The study concludes with recommendations for future studies into to the application of econometrics in South African agricultural economics.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND

Economics and its sub-disciplines has become increasingly quantified, applied in some form of econometric model or other quantitative method. Indeed, economics, as practiced today, is hardly any different from its sub-discipline, econometrics. It has become a blend of economic theory, mathematical economics and statistics. Today literally thousands of articles published in economic and agricultural economics journals contain applied econometrics.

Econometrics is that subdivision of economics which explicitly unites deduction, induction and statistical inference; its methodology concerns the procedures adopted in the testing and where, applicable, the quantification of economic theories. Econometric models made their first appearance during the 1930s, at a time when the world was experiencing one of the most-well-known disruptive economic problems of the 20th century. It was a time when western economies experienced massive unemployment and greatly reduced incomes. It was clear that a theory, model or structure was needed, to offer possibilities that could reduce the economic hardship faced by so many people during the Great Depression (Zalm, 1998). Tinbergen (1939) developed the first macro econometric model of the Dutch economy in 1936. He later also prepared similar models for the economies of the United States and the United Kingdom (Zalm, 1998). Tinbergen's efforts were soon followed by Stone and Stone (1939) and later by Klein (1950) when they too, started to develop macro econometric models.

Econometrics has come a long way since these early attempts. Not only has there been substantial development in terms of theory, methodology and technique, but also because of the continued contribution of technology. The explosion in technical growth since 2000 has had a tremendous impact on econometrics and statistics. When econometrics was still in its infancy, researchers and analysts were restricted to electro-mechanical desk calculators and large, slow and expensive mainframe computers. However, since the 1970s an increase in the expansion and availability of microcomputers started to emerge. These advances in technology meant that researchers were able to increase the scope and scale of their econometric models as well as the techniques used in developing them. New improved "tools" were coupled with the increasing availability of improved and more adequate economic statistics. The evolution in computer software programs during the 1990s allowed researchers to pursue more ambitious and more comprehensive research projects. Arguably one of the most valuable contributions of these software programs was the ability to provide faster results to the researcher as econometric or simulation models were being developed. Friedman & Schwartz (1991, p. 48) provide a good example of just how much faster the evolution in technology has made econometrics. They note:

> "Today's statisticians will be interested to know that, not counting data insertion, it took 40 hours to calculate a regression that I can now calculate on my desktop computer in less than 30 seconds – my favourite story to illustrate what has happened to our computer power."

In their comparative assessment of modelling approaches in agricultural trade modelling, Van Tongeren *et al.* (1991) conclude that the most important innovations during 1996 to 2006 were neither theoretical, nor technological. They argue that the most significant changes were of an institutional nature, supported by recent computer and communication technologies.

Ferris (1998) argues that the great improvements in computer hardware and software may also have shortened the distance between the decision maker and the modeller. Improved technology has enabled decision makers to become modellers or decision makers' support staff to tap computer models for answers to questions requiring quick response. In fact, econometric models and statistical techniques have today become part of the day-to-day tasks of many decision makers, especially in agricultural economics. This has had an impact on the training needs of agricultural economics students who are being prepared by universities before they venture into the agricultural sector. In fact, an internet search on syllabi of the undergraduate courses in agricultural economics revealed that all the major universities offering agricultural economics in South Africa (Universities of Stellenbosch, Free State, Natal and Pretoria) include some introductory econometrics course as part of their undergraduate program. This situation is also not any different for post graduate programs. Here, students have the opportunity to take advanced econometric and statistics courses.

Unfortunately the development of econometrics to what it is today has not been without any problems. During the 1970s, the large scale econometric models which were developed as tools to cure the economic ills of the world began to fail. The inadequacy of these models' ability to deal with large external shocks such as the oil crisis shook the trust of policy makers. This has led some commentators to take a nihilistic stance towards econometrics. In fact, some argue that as the number of articles containing applied econometrics increased, so to did the number of articles expressing authors' cynicism with the way econometrics have been practised. In a squib Kennedy (1992, p. 82) illustrates:

> "Economists' search for "truth" has over the years given rise to the view that economists are people searching in a dark room for a non-existent black cat; econometricians are regularly accused of finding one."

If researchers consider the expanding analytical capacity with new computer hardware and software programs, the common exchange of computer programs and databases through institutional innovations, and the importance of food in world economic development, it has now, more than ever, become crucial that agricultural economists exhibit the ability to transform agricultural statistics into accurate, reliable and useable information for policy and other decision makers in South Africa.

1.2 PROBLEM STATEMENT

The South African agricultural economics discipline has evolved over the years in response to the important economic problems facing the agricultural sector and to developments in economic theory, quantitative methods, and the computational capacity to deal with these problems (Kirsten, 2002). Kirsten (2002) reflects on the evolution of agricultural economics scholarship in South Africa. He argues that in living up to the challenges facing farmers, agribusinesses and rural communities, the agricultural economics profession remained relevant and focussed on the needs of the country and the industry, but has moved the focus away from the 'frontier-pushing' research and theoretical work of agricultural economists in the United States and in Europe. In this sense agricultural economics in South Africa has often borrowed from these scholars and applied the models and methodologies to local problems (p. 256).

However, borrowed models and methodologies have highlighted what many have perceived as limited quantitative skills and application in South African agricultural economics (Kirsten, 2002, p. 256). This is a major problem since agricultural economic scholars abroad have expressed concern regarding the application of econometrics in agricultural economics; precisely those scholars from whom South African agricultural economic scholars have "borrowed" models. Tomek (1993, p. 6) for example, stressed that "the strength of agricultural economics rests on its capacity to combine theory, quantitative methods, and data to do useful analysis of problems faced by society." He expresses his concern that there is a growing awareness that agricultural economists are not, in fact doing this very well. In fact, the 75th Anniversary Issue of the *American Journal of Agricultural Economics* (1993) was devoted entirely to these problems. Tomek (1993, p. 14) concludes his paper by urging the agricultural economics fraternity to set higher standards of excellence in empirical research.

> "In sum, agricultural economics (and applied economics in general) must set higher standards of excellence in empirical research. Higher quality output requires both more and better inputs in terms of model specification, data, and the researcher's intellectual input. It is my view that adding conformation as an initial component of the research agenda will improve quality. Admittedly, it will lengthen the research process and probably result in fewer published reports, but the profession profoundly needs to establish higher standards for published empirical results."

Indeed, it should be a major problem and concern to the agricultural economics discipline in South Africa that only a few scholars are well trained to provide sufficient oversight of the sometimes ruthless application of econometric techniques in order to get some publishable result. Consequently it is possible that a large body of literature has entered the public domain over the last twenty years without being properly reviewed because South African agricultural economists may not have the necessary insight and knowledge of the reasons for the disenchantment with the way in which econometrics have been applied. This is the major problem area this study intends to address.

1.3 OBJECTIVES

The intended outcome of this study is to analyse studies in South African agriculture in order to determine how these studies have been conducted, if the researcher(s) made any reference to the econometric methodologies and specification used in individual analyses and the process of deriving the final results. Before 1962 scholarly work by agricultural economists in South Africa were published in a variety of journals and it is, therefore, very difficult to review econometric studies in the early years. The establishment of the Agricultural Economics Association of South Africa and Agrekon in 1962 provided a home for the main body of South African agricultural economic literature (Kirsten, 2002). Thus, this study reviews articles on econometric studies which were published in *Agrekon*.

1.4 METHODOLOGY

This study consists of a literature review of applied econometrics. The aim of the literature review is to gain the necessary insight into the problems which have persisted since the development of econometrics and also the reasons for this persistence. The work of McCloskey, more specifically her work on the use of statistical significance testing will also be explored in some detail. Ziliak and McCloskey (1996 and 2004a) have developed a 18 question questionnaire which they applied in their review of the use of statistical significance testing in articles published in the *American Economic Review* during the 1980s and 1990s. The knowledge gained from the review of applied econometrics as well as McCloskey and Ziliak's (1996 and 2004a) studies on the use of statistical significance testing will then be applied in a review of articles published in *Agrekon* in order to establish if there is in fact, reason for concern over the manner in which econometrics have been applied in South African agricultural economics. This is achieved by developing a questionnaire based on Ziliak and McCloskey's original questionnaire and applying this questionnaire in assessing the use of regression analysis in articles published in Agrekon during the period 1962 to 2005.

Papers containing regression analysis published in *Agrekon* between 1962 and 2005 were grouped into five strata. Stratified simple random sampling was applied to draw a sample of 65 papers from the universe of papers containing regression analysis. Each paper was then evaluated by means of the questions contained in the questionnaire.

1.5 OUTLINE OF THE STUDY

The study is divided into six chapters. Chapter two provides a framework of the literature by reviewing the history of econometrics and the different approaches to econometric modelling. Chapter three and four attempt to identify the key drivers of the disenchantment with applied econometrics. Chapter five provides a crossover from economics in general to agricultural economics and then specifically agricultural economics in South Africa. Chapter six consists of an analysis of papers published in *Agrekon* containing applied econometrics. The study is concluded in Chapter seven. Recommendations are made and suggestions for future studies are identified

CHAPTER 2

APPLIED ECONOMETRICS IN EMPIRICAL STUDIES: SCIENCE OR ILLUSION?

2.1 INTRODUCTION

It is necessary to understand what spurred the evolution of econometrics, what drove the development of econometric models and what methods have been developed for the construction of econometric models. Hence, this Chapter provides a starting block for the investigation by doing just that: providing background information of developments and events in the history of econometrics.

Research in the methodology of economics is a branch of enquiry which has a long history and which has attracted much interest over the years[1]. Interest in the methodology and history of economics mounted especially during the 1970s through to the 1980s. Much of the interest may have been prompted by a spate of publications that are critical of the subject's standing and which display dissatisfaction with the perceived accomplishments of economics[2]. However, applied econometrics has also been an area of much interest and examples of applied econometric work intended to provide both a test of economic theory as well as a test of quantitative calculus are widely published in academic journals. The various texts on applied econometrics, such as Rao and Miller (1971), Wallis (1969 and 1973), Mayes (1981) and Thomas (1985) can attest to the interest in applied econometrics.

A great deal of the voluminous literature on the methodology of economics is devoted to discussing the role of empirical analysis, which has largely been conducted in terms of the use of data as a test of theory or as a source of hypothesis formation. Empirical analysis has also been used as a test to determine whether data should be used as an attempt to verify or falsify economic theory (Darnell and Evans, 1990, p3 – p22).

---

[1] See for example Blaug (1980), especially Part II: "The history of economic methodology"
[2] Examples abound, see, for example, Blaug (1980), Hausman (1981 and 1984) , Leontief (1971) and Darnell and Evans (1990).

The primary purpose of this chapter is, therefore, to discuss the methodological foundations of econometrics. This is to be achieved through a critical appraisal of the different schools of econometric methodology. The purpose of this appraisal is to provide the foundation on which some of the econometric studies in agricultural economics in South Africa can be assessed and reviewed in the following chapters of this study. The chapter starts with a short overview of the rise in the popularity of econometrics in the optimistic climate of 'positive economics'. The remainder of the chapter examines the various econometric practises which currently prevail: the 'traditional textbook' approach, the no-estimation approach and those practices championed by Hendry, Leamer, and Sims.

2.2 THE EVOLUTION OF ECONOMETRICS AS A SCIENCE

Reviews of the history and development of empirical economics, or more specifically the literature on the history of econometrics, are found in De Marchi and Gilbert (1989), Darnell and Evans (1990), Morgan (1990), Colander and Brenner (1992) and Hendry and Morgan (1995). De Marchi and Gilbert (1989) discuss some of the early work in econometric methodology and have an extended discussion of the British approach to econometrics relating to time series data. Morgan's (1990) book provides an excellent historical perspective on the theory and practice of econometrics, with an in-depth discussion on the early contributions of Haavelmo to econometrics. In the same spirit, Hendry and Morgan (1995) have collected seminal writings in econometrics to illustrate the evolution of econometrics over time. Further, Colander and Brenner (1992) present a critical, at times agnostic, view of economic teaching and practice.

Empirical analysis in economics has a long and rich history whose origin can be traced as far back as the sixteenth century when the "political arithmeticians" led by Sir William Penny analysed problems such as taxation and international trade with quantitative information (Darnell and Evans, 1990). However, the body of literature on econometric history highlights the 1930s as the starting point for the evolution of econometrics as the term is currently understood. Not only were the 1930s characterised by great hardship brought about by the Great Depression, but they also marked the founding of the Department of Applied Economics at Cambridge in the UK, while the Cowles Commission was founded in the USA during the 1940s. The 1930s and 1940s was a period in which there was great optimism that economics could be given empirical content (Darnell and Evans, 1990). Econometrics was seen as the answer to the problem of identifying most of the important factors in modelling economic reality. The methods of classical statistical inference were thought to be useful both to test, but also to quantify theoretical parameters of economic models. In his celebrated work: "*The Probability Approach in Econometrics*", Haavelmo (1944) argued that theoretical propositions could be formulated in the context of a well-defined statistical model.

However, optimism at the time was restricted only to the few who had "converted" to the "new age" of economic quantification. At the time, the application of econometric techniques to economic problems was still relatively slow. To some extent this was due to a lack of adequate computing facilities capable of handling complex calculations of econometric techniques, but the value of such work to economics as a whole was also questioned. Yule (1926) points out that early application of correlation analysis to economic time series had raised the problem of spurious regression. Sceptics, therefore, found it relatively easy to point out these problems and to dismiss much of the work. An illustrative example of this scepticism is found in Havelmo (1958, p. 383) when he summarised the dichotomous early perceptions of economists in his Presidential address to the Econometric Society in 1957, by stating:

> "Some people hailed the regression technique as a miracle tool for surprise discovery of economic laws. Others sensed the danger of a mechanical approach and created the bogy of "spurious correlation"".

Scepticism soon made way for increased optimism when, during the 1950s and 1960s, econometric techniques started to receive widespread interest and acclaim throughout the economics profession. It was a time when most scholars in economics felt the need to raise economics to a level in which it would be regarded as a science; in which quantification was seen as a way of claiming that status. Economic literature made continued reference to the natural sciences, using scientific approaches as reference for evaluating economic methodology. Perhaps the strongest motivation came from the desire to resolve theoretical controversy at the time. The Keynesian Revolution not only gave the economics profession new interest in a range of macro-behavioural equations, but it had also given rise to the

Keynes versus Classics debate, which did not seem capable of resolution at the theoretical level.

Various philosophies of science define the role and use of hypotheses, the most important being logical positivism. This holds that the approximate nature of reality can be learned from observing behaviour in a world of unknown and unknowable true causal relationships. Scholars at the time were greatly influenced by the works of Popper (1968). According to Popper's demarcation criteria, in which scientific hypotheses are falsifiable, unfalsifiable propositions belong to the domain of metaphysics, not science, and imply that in order to be scientific, we must test hypotheses.

The notion of testing hypotheses was advocated in Friedman's 1953 essay "The Methodology of Positive Economics" and Lipsey's textbook. These two economists were responsible for informing and 'converting' most economics students to a so-called scientific view of their subject, and, thereby, firmly establishing the label "Positive Economics" on both sides of the Atlantic (Darnell and Evans, 1990).

So it seemed then, that during the mid 1960s, objectives for the next generation of economists were set. Economics should be regarded as a quantified science and economic knowledge should be expressed in a form making it amenable to testing (Darnell and Evans, 1990). An economist's work was, therefore, perceived to be that of pursuing rigorous analyses to test hypotheses. Examples of the 1960s' optimism in applied econometric work are plentiful. Good examples can be found in early published work of members of the M^2T group (Methodology, Measurement and Testing). De Marchi (1988) lists some 26 publications which he regards as M^2T-related work: some of these studies are applied econometric studies. As an example, consider Lipsey's well-known reconsideration of Phillips' work on the relation between the unemployment rate, the rate of change of unemployment and the rate of change of money wage rates (Lipsey, 1960). Other examples include Klein and Kosobud (1961) in which they sought evidence on the 'fundamental ratios' of economics.

Yet not all scholars were and are convinced that observation has had the same profound effect on economic theory as had observation on theory in other sciences (e. g. Summers 1991). Mayer (1980), furthermore, suggested that, to a large extent, the poor ability to test hypotheses distinguishes economics from hard empirical science. Indeed the theory and observation link in economics is complicated by human behavioural patterns, such as individual choice, social interaction and the non-experimental nature of observations (Townsend, 1997). Darnell and Evans (1990, p. 96) identified three main difficulties with the positive approach:

- Many aspects of economic theory do not imply either strong or qualitative predictions.

- Economics is made up of interlinked propositions: thus the main hypothesis is insulated from testing by the range of ancillary hypotheses necessitated in making it testable.

- Refutation is difficult because hypotheses are probabilistic and errors (rejecting a true hypothesis or not rejecting a false hypothesis) are always possible. Formally, refutation requires the rejection of a theory if the researcher is confronted with contrary evidence; however it is difficult to know what proportion of such incidences are required before the theory is rejected.

These problems popularised the view that econometrics could not result in the rejection of many hypotheses and, therefore, the emphasis should turn to estimation of the parameter values of economic theory and comparability through predictive performance (Archibald, 1966). Gradually, many economists and econometricians moved towards this more relativist position. The desire to resolve theoretical controversy and the Keynes versus Classics dispute through quantified analysis was lost in the difficulties of carrying out the exercises and was soon replaced by an emphasis on verification and forecasting. Provided of course, that the ‘evidence was consistent with the hypothesis’.

Another main development in the use of econometrics during the 1960s was the creation of ambitiously large-scale macroeconometric models. The increasing availability of macroeconomic data created optimism among econometricians that they could build simultaneous equation models of an economy and use them to both test hypotheses and provide conditional forecasts. The forecasts would be useful for policymakers and they would quite often be willing to pay for them.

By the early 1970s, however, the heyday of large economy-wide models was beginning to fade. The glamour about those models subsided owing to developments in the world economy, especially the oil price shocks of 1973 and 1979. Lucas's (1976) book entitled: "*Econometric Policy Evaluation: A Critique*", which later became known as the *Lucas Critique*, emphasised the way that the forecasting performance of the relatively complex structural equation models was not particularly good, especially for time horizons greater than six quarters. Consequently, the emphasis on structural estimations weakened.

At the same time simplified forecasting models based on either '*reduced form*' or *time series* characteristics of economic variables gained in standing. An example of this is found in Anderson and Carlson (1970) who provide an example of how the *St. Louis* model expressed the growth rate of nominal income as a distributed lag function of the growth rates of money and government expenditure. On the other hand, time series analysts (using the Box-Jenkins methodology), promised relatively cheap, straightforward models for the generation of forecasts. Economists also learned more about autocorrelative structure and stationarity in time series (Darnell and Evans, 1993). Time-series models often outperformed the large econometric models and have served as the foundation for the now popular 'a-theoretical macroeconomics' associated with Christopher Sims (Sims, 1980a).

A general awareness that econometrics was not able to resolve the economic dispute also started to develop. By the 1980s however, a renewed optimism about the role of econometrics was discernable and many authors began to express their dissatisfaction of the way econometrics was practised. Since the 1980s economic journals have carried articles where the authors express their dissatisfaction with how econometrics is being conducted. Examples of how authors have expressed their concern, dissatisfaction and in some cases utter disgust; can be found in Leontief (1971), Mayer (1980), McCloskey (1983), Leamer (1983), McCloskey(1985a and b, 1992, 1994, 1995a and b, 1997, 1998, 1999, 2002), and Ziliak and McCloskey (1996, 2004a and b).

While many have spent a great deal of their research on evaluating and criticising econometric practice, other have devoted themselves to developing proposed solutions to the problems in econometrics. To some extent, this has kindled a strong confidence among econometricians that they have something to offer. In particular, there are some four strands of current econometric practice which display this confidence. These strands are the topic of the next section of this chapter.

2.3 ECONOMETRIC METHODOLOGY IN ECONOMICS

When econometric practices were first adopted by the economics fraternity, little, if any, attention was paid to the methodological foundations which underpin econometrics. Commentators on the state of economics and econometrics during the 1960s and 1970s often suffered from a failure to distinguish method (i.e. technique) from methodology. The 1960s and 1970s was characterised by a marked growth in econometric techniques; however, a few contributions were made to the issue of methodological status of econometrics as part of the methodology of economics (Darnell and Evans, 1990). As described earlier, there was a somewhat superficial appeal to Popperian falsificationism and hope was placed that it would be possible to place economics on an empirical footing similar to that of the natural sciences. But the increasing use of mathematics in economics and references to the work of Popper at the time were, and are, insufficient to guarantee that economics is a hard science: what was lacking was a genuine methodological discussion of the way in which economics generates and tests hypotheses. References to Popperian falsificationism concerned more the role of econometrics rather than the methodology of economics. Economists and econometricians thought that the adoption of the regression model from the experimental sciences would bring more rigour to economics and provide it with a reliable method of testing hypotheses and decisively rejecting false hypotheses. However, this proved not to be the case and as Mayer (1980) observed: ‘On all too many questions we are buried in an inchoate mass of seemingly contradictory evidence.’

This section focuses on the different approaches to econometrics which developed from the growth in econometrics during the 1960s and 1970s, fuelled by developments which have been discussed earlier in this chapter. The topic of econometric methodology is vast and controversial and it is not possible to discuss all the related approaches and topics within the scope of this chapter.

In *New Directions in Economic Practise*: *General to Specific Modelling, Cointegration and Vector Autoregression*, Charemza and Deadman (1992) critiqued the traditional approach to econometrics and gave a detailed exposition of new approaches to econometric methodology. Darnell and Evans (1990) wrote about the "*limits of econometrics*" and provided a balanced discussion of the various methodological approaches to econometrics. Pagan (1987) provides a comparison of the three econometric methodologies associated respectively with Hendry, Sims and Leamer. His study provides a statement of the main steps to be followed in using each of the methodologies and comments upon the strengths and weaknesses of each approach.

Hylleberg and Paldam (1991) provide a somehow different analysis of the different schools of thought. Instead of focussing only on the three methodologies, they identify six research strategy schools of thought, namely the analytical economic history/comparative storytelling approach, the calibration/simulation school, the Traditionalists, the Hendry School, the Leamer School and the Sims School. For a more recent opnion (recent in context of the literature) Townsend (1997) presents a review of the debate between economic theory and observation and describes their use in alternative econometric methods. His focus is again on the different methodologies in econometrics and it is the only study to date published in a South African Agricultural Economics Journal (*Agrekon*) which focuses on the different methodologies in econometrics.

2.3.1 No-estimation approaches

The two no-estimation approaches are the analytical economic/comparative storytelling approach and the calibration / simulation school.

The analytical economic/comparative storytelling approach

The analytical economic history/comparative storytelling school has a long tradition in economics and, as Hylleberg and Paldam (1991) suggest, re-emerged as a research strategy during the early 1990s. They suggest that in many cases this strategy appears to be the only (or best) strategy possible. Economics quite often deals with subjects that are in an early stage of intellectual clarification or the research topic is so large and nebulous that it is hard to apply precise models. They also argue that in some cases the relevant experience may be scattered and can be of a fragmentary character; analysing the data in a *comparative setting* can often prove to be fruitful. Nonetheless, the method often requires considerable space as the story of a set of events has to be told and compared, although the result is often very readable. Sargent (1981) provides a powerful example. Other examples of how sometimes a whole book is necessary for such an approach includes Friedman and Schwartz's 1963 book; *A Monetary History of the United States 1867 – 1960.*

Summers (1991) suggests that: "the only empirical research that has influenced thinking about substantive questions has been based on methodological principles directly opposed to those that have been characterised by attempts to gauge the strength of associations rather than to estimate structural parameters, verbal characteristics of how causal relationships might operate rather than explicit mathematical models, and the skilful use of carefully chosen natural experiments rather than sophisticated statistical techniques to achieve identification". He takes the view of emphasising qualitative rather than quantitative conclusions from empirical work. It appears that the main thrust of Summers' critique is not against econometric technique, but rather against making too rigorous demands on theoretical models in empirical work.

The calibration/simulation school

The method of calibration – also known as the 'synthetic approach' – starts from basic general equilibrium beliefs such as minimising representative agents; then key empirical facts are found in the correlation structure between variables, and the models are thus calibrated to contain these correlations. Then simulation can be said to provide interesting evidence about a stylised world having exactly the desired properties (Townsend, 1997). It is argued that this approach is preferable to the systems-of-equations approach, i.e., the macroeconomic modelling approach (Hylleberg and Paldam, 1991). Anderson (1991) suggests that another advantage of these models is that calibration is a useful way of illustrating properties of theoretical models having a complexity, which precludes an analytic solution.

Calibration is widely used in international trade models and computable general equilibrium models (CGE models). Most applied trade modellers resort to calibration to generate a set of parameters that is consistent with both the benchmark data and the model's theory. This approach takes the initial estimates of elasticities from outside sources and adjusts certain other parameters in the given functional forms to the initial equilibrium dataset. Calibration, therefore, exploits theoretical restrictions, equilibrium assumptions and assumption of functional forms to arrive at a proxy estimate (Van Tongeren *et al.*, 2001).

Van Tongeren *et al.* (2001) provide a comparative assessment of alternative modelling approaches, considering a total of 16 partial equilibrium and general equilibrium models. The assessment includes theoretical modelling foundations, datasets employed and institutional aspects, such as model maintenance and dissemination of results. A typology of models is provided by structuring the assessment along a clear set of evaluation criteria. Their analysis revealed that of the 16 models under review, the parameters of only two of the models were estimated. The parameters of the remaining 14 fourteen models were all calculated by means of some form of calibration procedure. Meyer and Kirsten (2005) provide an example of a partial equilibrium model for the South African wheat sector. Analysis of their study suggests that they too, have resorted to calibration methods to arrive at their estimates.

2.3.2 Estimation approaches

The traditional approach

The traditional approach of econometric modelling stems from the work of Tinbergen (1939) who argued that economic theory should be quantified through mathematics, thus converting theory into relationships capable of statistical tests (Townsend, 1997). The dichotomy between theoretical and empirical activities was central to this approach; the theorist provided the model and the econometrician estimated and tested it. The strategy was effective because the theory involved could be expressed in the form of a linear or simple non-linear regression (Townsend, 1997).

Darnell and Evans (1990) argued that the specific regression strategy used by economists of the 1950s and 1960s led to the (now) familiar ordering of the typical econometric theory textbook: the reader is first confronted with a 'properly specified' single equation model, which has fixed regressors and a zero mean, non-autocorrelated and homoscedastic error term. They argue that under such conditions the method of ordinary least squares (OLS) yields the 'best linear unbiased estimators' (BLUE); any deviations from the assumptions made regarding the error term invalidate the property of 'BLUEness'. Thus, if the error term does not have a zero mean, then the unbiased property of the estimators is violated; if the error term exhibits either autocorrelation or heteroscedasticity then the efficiency of the estimators is reduced; and if the regressors are stochastic there may be implications for both the unbiasedness and the efficiency of the estimators. The standard approach was then to proceed to evaluate estimation techniques which improve upon OLS in such conditions and, having introduced the 'ideal' regression model, such text proceeded to deal with the 'problem' cases. Darnell and Evans (1990) argue that the 'problem' was often very narrowly defined: if the original assumptions were violated then the optimality of the estimators derived by OLS was denied.

This traditional textbook treatment then indicated that, as a consequence, estimation techniques other than OLS were called for when the distributional assumptions of the probabilistic model were shown to be unwarranted. Instrumental variables, full information maximum likelihood, two stage least squares, three stage least squares, non-spherical disturbances, generalised least squares estimators like Cochrane-Orcutt and seemingly unrelated regression analysis are all examples of techniques designed to improve upon OLS (Pesaran, 1987).

Townsend (1997) is of the opinion that, as a consequence, the traditional approach was driven by the assumed distributional properties of the error term, but failed to suggest that any correspondence existed between those assumptions and the nature of the underlying economic theory. Townsend (1997) continued and suggested that this method also failed to discuss why the original specification of the equation might have been incorrect but concentrated instead on developing estimation techniques to be used in preference to OLS. By using econometrics as a device for estimation, rather than a device for testing, the truth of hypothesised models was maintained throughout the process of estimation. Only those equations which 'failed to refute' the hypothesised model were deemed 'successful' and the process of re-specification and re-estimation was insufficiently detailed.

Hylleberg and Paldam (1991) suggested that the reason why this strategy was discarded by most economists during the 1970s, 1980s and still today, can be summarized as follows:

(i) It does not describe what actually takes place when empirical research is carried out. They argue that strict adherence to this strategy would lead to rejection of all theories and the practice has thus developed into a trial and error process. This is evident from the development of the alternative techniques discussed above, designed to improve upon OLS. This is also illustrated on the right-hand side of Figure 2.1, adopted from Hylleberg and Paldam (1991). Unfortunately, the analysis is seldom described in such a way that the reader could recognise a trial and error process. It is rather described as following the main line on the left-hand side of Figure 2.1.

(ii) The strategy obstructs free and valuable exchange and integration of information from different sources such as theory, data and the measurement system.

(iii) Economic theories and the corresponding theory models are extremely simple approximations that cannot be expected to explain more than a fraction of what is actually going on. At the same time economic data series are usually quite short or 'noisy' and seldom measure exactly what it is supposed to be in the latent theoretical variables. As illustrated in Figure 2.1, the development of alternative techniques often ends in a trial and error process. The trial and error process has often been criticised for data mining. Data mining refers to running a large number of regression equations which differ according to specification and explanatory variables used (Townsend, 1997). The equation that best supports the theory under consideration is the one deemed worthy of reporting (Lovell, 1983).

(iv) Finally, decision-makers complained that the models were ineffective for practical purposes of forecasting and policy analysis. Models were seen to be 'statistically inadequate', 'theoretically inconsistent' and 'practically irrelevant' (Townsend, 1997)

Despite its shortcomings and the critique against it, Darnell and Evans (1990) advocated the traditional approach as the preferred econometric strategy. However, they recognised that the typical presentation of the traditional approach has a number of critical weaknesses and argued that these weaknesses have been significant in the development of the alternative strategies, which are to be discussed below. In *The Limits of Econometrics,* Darnell and Evans recast the traditional approach into a methodologically acceptable strategy with the modifications summarized below. They argued that with a few modifications it is possible to restate the ‘traditional econometric modelling’ within a falsificationist methodology. The modifications include *inter alia* the recognition of the role of auxiliary hypotheses in the testing of main hypothesis and the recognition of the status of the ‘error term’ in a regression equation. They point out that any main hypothesis may be phrased within a large variety of regression equations, depending on the particular treatment of the auxiliary hypotheses (Darnell and Evans, 1990, p. 148). As far as the error term is concerned, they adopted an approach where any regression equation is treated as a decomposition of the determinants of the dependent variable into that equation due to a set of regressors and an error term which is wholly non-systematic.

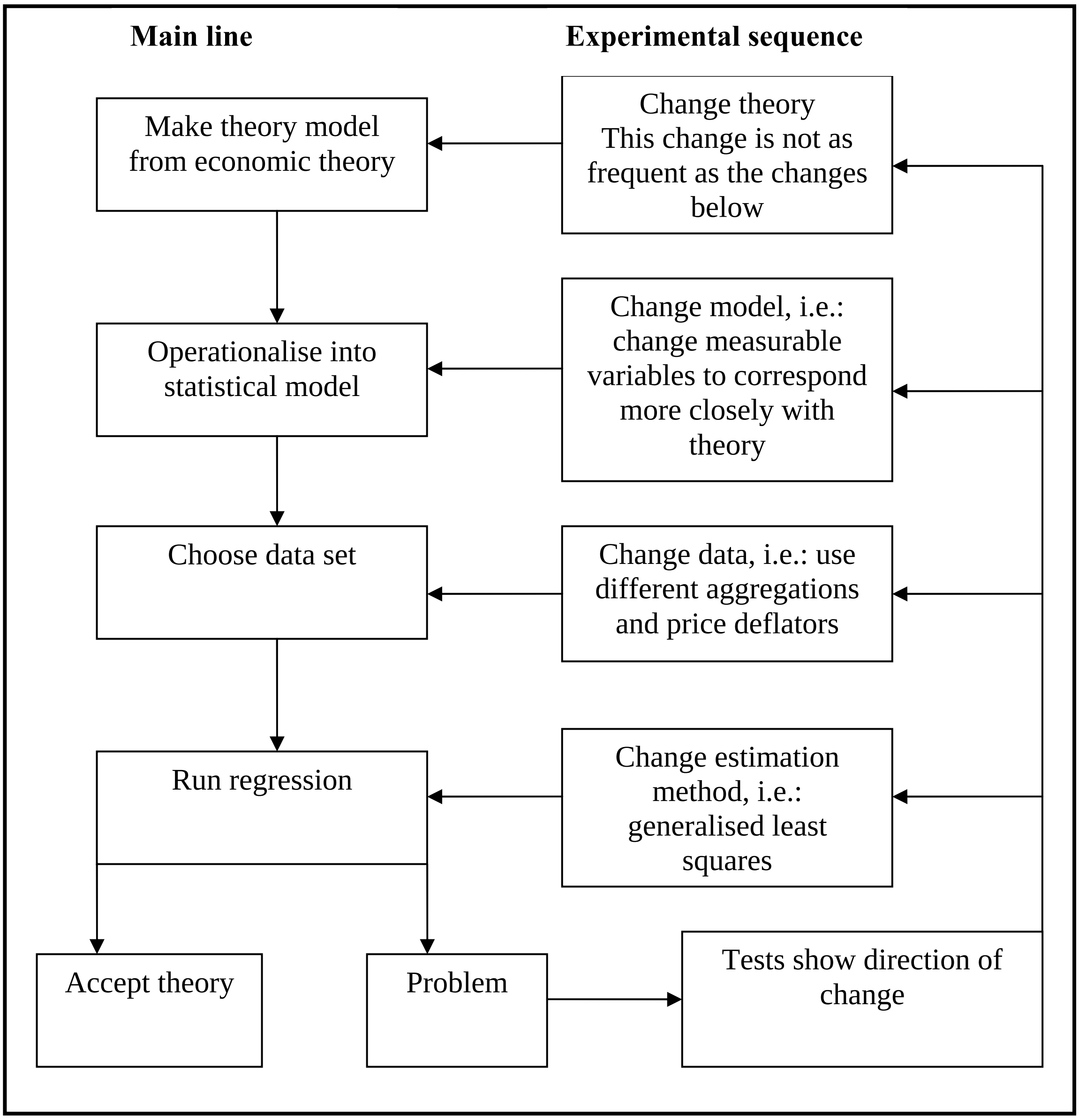

Figure 2.1: The traditional approach in econometrics

Source: Adapted from Townsend, 1997 and Hylleberg and Paldam, 1991

Thus, within this framework, it is necessary to confirm the correctness of the specification of a regression equation as a prerequisite to the testing of the main hypothesis of interest. Therefore, in addition to the familiar issue of practical falsification, it is essential that the specification of the model in which the main hypothesis is tested has been confirmed as correct. Practical falsification therefore requires not only methodological norms which set the criteria of rejection, but also methodological norms which set the criteria of conformation (Darnell and Evans, 1990, p. 148). This ‘adjusted’ approach requires both an iterative procedure and the use of methodological norms within the familiar framework of statistical testing.

The Hendry approach to econometric modelling

Perhaps the closest of all the methods to the traditional approach of investigation is the ‘Hendry methodology’ (Pagan, 1987). Owing much to Sargan’s (1964) seminal paper, it also reflects an oral tradition which was developed largely at the London School of Economics during the 1960s and 1970s.

In a nutshell, Hendry’s methodology comprises four steps namely:

- *Formulate* a general model that is consistent with what economic theory postulates are the variables entering any equilibrium relationship and which restricts the dynamics of the process as little as possible, i.e. establish the long-run equilibrium relationship.

- *Re-parameterize* the model to obtain explanatory variables that are near orthogonal and which are ‘interpretable’ in terms of the final equilibrium. Near orthogonal variables are variables that have zero collinearity.

- *Simplify* the model to the smallest version that is compatible with the data, i.e. congruent.

- *Evaluate* the resulting model by extensive analysis of residuals and predictive performance, aiming to find the weakness of the model designed in the previous step.

Theory and data frequently interplay in this methodology and unless there are good reasons for believing otherwise, it is normally assumed that theory suggests which variables should enter a relationship, and the data is left to determine whether this relationship is static or dynamic (Pagan, 1987), i.e. in the sense that once disturbed from equilibrium, it takes time to be re-established. Hendry's approach mostly describes a re-parameterisation of the distributed lag model. As a result, the dynamics of the equation is usually re-written as an 'error correction mechanism' (ECM) (Townsend, 1997). Steps one and two, therefore, demand a clear statement of what the variables in the equilibrium relation should be, as well as a choice of parameterisation (Pagan, 1987). Hendry's second step is described in most detail by 'the master' in Hendry (1986). Yet, Pagan (1987) argues that the perusal of the source leaves the impression that this step is more of an art than a science, and consequently difficult to codify. Pagan further argues that Hendry tends to blur steps one and two in his applied work, with the re-formatted equation sometimes seeming to derive from an inspection of the parameter estimates in the original equation; which, in those cases, seem to be both simplified and re-arranged at the same time.

Although the first two steps in the methodology seem unexceptionable, difficulties arise when progressing to the third and fourth steps. The difficulties relate to question of how does the researcher move from a very general to a more specific or simplified model? Or, to state the question differently: How does the researcher decide the size of the lag? According to Hendry and Richard (1983), a simplified model should satisfy the following six criteria:

- *Be data admissible*: That is, predictions made from the model must be logically possible.

- *Be consistent with theory*: It must make good economic sense.

- *Have weakly exogenous regressors*: That is, the regressors must be uncorrelated with the error term.

- *Exhibit parameter constancy*: The values of the parameters should be stable, otherwise forecasting will be difficult.

- *Exhibit data coherency*: That is, the residuals estimated from the model must be purely random (white noise). If not, there is some specification error in the model.

- *Be encompassing:* The model should encompass or include all rival models in the sense that it is capable of explaining their results. Thus, other models cannot be an improvement over the chosen model.

Hendry and Ericsson (1991) put a great deal of emphasis on whether one 'model' encompasses another. They assert that "a necessary condition for encompassing is variance dominance, where one equation variance dominates another if the former has a smaller error variance". A footnote attached to his sentence states: "Formally, variance dominance refers to the underlying (and unknown) error variances. Without loss of clarity, we often will say a model variance-dominates another if the estimated residual variance of the former is smaller than the latter."

Obviously, in choosing such a model, the researcher will have to try several specifications (i.e., choose different values of the lag) before the researcher can finally "settle" on the "final" model (i.e., search for the elusive "Holy Grail"). That is why Hendry's methodology is also known as the *TTT methodology*, that is, as Hendry states: "test, test and test". It seems, therefore, that modelling according to this approach is a matter of design, while 'tests' are used as design criteria (Hendry and Richard, 1983).

Whether such a stringent test procedure is possible in practice is open to debate and criticism. McAleer *et al*. (1985) argue that an exact description of the decisions taken in moving from the general to the specific is imperative in any application of the methodology. Pagan (1987) goes on and states that this rarely involves a single decision, although it would be possible to act as if it did by just comparing the finally chosen simplified model and the original one, thus ignoring the path followed to the simplified version. This is precisely what Hendry seems to do in his various applications of his methodology. He normally only provides the value of a test statistic comparing the two models at each end of the scale, with very little reference of how he went from the one end to the other. Pagan's argument against this stance is that it is hard to have much confidence in a model if little is explained about its origin as well as if not much is documented on the path followed in the simplification process. Townsend (1997) reiterates this view and states that there is little explicit connection between observation and theoretical statements. He is of the opinion that the economic exercise fails to discriminate between clearly stated competing hypotheses. Reviewing Hendry's applied works reveals an attitude that the manner in which a final model is derived is largely irrelevant. It is either useful or not useful and that characteristic is independent of whether it comes purely from whimsy, some precise theory, or a very structured search (Hendry and Mizon, 1985). Though true in a sense, it is cold comfort to someone implementing the methodology or attempting a replication study. Subsequent to reading other critical appraisals of Hendry's methodology, such as Darnell and Evans (1990), the researcher is left with the question whether this methodology is nothing else than a 'glorified' data-mining exercise.

Friedman and Schwartz (1991) provide another shrewd discussion on one of Hendry's applications of his methodology. The article was written in response to Hendry and Ericsson's (1991) article. Hendry and Ericsson (1991) re-estimated the regressions produced by Friedman and Schwartz (1991) according to the general to specific approach. Friedman and Schwartz (1991, p. 47) responded to Hendry and Ericsson's study by remarking the following:

> "After years of experiments, Hendry and Ericsson's econometric techniques produced a series of models that confirm some of our principles, contradict none, and are less successful than our equations in terms of their own criterion of variance dominance."

Friedman and Schwartz (1991, p. 48) conclude by stating that they do not regard their statistical tests as demonstrating the validity of their statistical estimates.

> "The real proof of their pudding is whether it produces a satisfactory explanation of data not used in baking it – data for subsequent or earlier years, for other countries or for other variables."

In order to alleviate some of the problems experienced by Hendry, Leamer has suggested a process of sensitivity analysis.

Leamer's attempt at taking the con out of econometrics.

The second contender for alternative econometric methodologies made no secret of his dissatisfaction as to how econometrics was being practised at the time. In probably one of his most well read articles: "Let's take the con out of econometrics", Leamer (1983) provides an entertaining as well as perceptive analysis of the ills of econometrics. He notes that "the econometric art as practised at the computer terminal involves fitting many, perhaps thousands of statistical models. One or several that the researcher finds pleasing are selected for reporting purposes. This search for the model is well intended, but there can be no doubt that such a specification search invalidates the traditional theories of inference. The concepts of unbiasedness, consistency, efficiency, maximum likelihood estimation, in fact, all the concepts of the traditional model utterly lose their meaning" (Leamer, 1983, p. 36).

Yet, although Leamer's work proved to be very entertaining, providing a succinct description of his methodology is a great deal more difficult than doing so for Hendry's methodology. Pagan (1987, p. 9) notes "basically, the problem lies in a lack of applications of the ideas; consequently it is hard to infer the general principles of the approach from any classic studies of how it is to work in practise." As was the case with the Hendry variant, Pagan reduces Leamer's methodology to four distinct steps:

- *Formulate* a general family of models.

- *Decide what inferences are of interest*, express these in terms of parameters, and form 'tentative' prior distributions that summarise the information not contained in the given data set.

- *Consider the sensitivity of inferences* to a particular choice of prior distributions, namely those that are diffused for a specified sub-set of the parameters and arbitrary for the remainder. Pagan states that this constitutes the extreme bounds analysis (EBA) of Leamer (1983) and Leamer and Leonard (1983). The process is sometimes terminated at step three, but when it appears that the inferences are sensitive the prior specification of the third step is only a 'warm-up' for the next step.

- *Try to obtain a narrower range for the inferences.* In some places this seems to involve fixing a prior mean and interval for prior covariance matrices. If the restrictions in this latter step needed to get a narrow range are too 'implausible', one concludes that any inference based on this data is fragile.

As expressed in steps one to four, Leamer is of a Bayesian stance. Yet, in practise, the limited appeal of Bayesian methods to econometricians seems to have been based on the difficulties coming from a need to formulate high-dimensional priors in any realistic model.

To explain Leamer's extreme bound analysis, suppose in a regression model there are some regressors that the investigator regards as *free* (i.e. key) regressors and some as *doubtful* (i.e., of secondary importance); the terms free and doubtful are Leamer's. Regressions are consequently run on the key variables including or excluding all combinations of the doubtful variables. In this exercise the coefficients of the key variables will change from regression to regression. Therefore, for the coefficient of each key variable there will be several estimates; the lowest and highest values of the estimate will constitute a bound or a range. The bound or range can be regarded as a confidence interval for the coefficient in question, a confidence interval reflecting model specification uncertainty. This is, of course different from the conventional confidence interval, which represents sampling uncertainty within a given model specification (Darnell and Evans, 1990, p. 109).

Like the other approaches, Leamer's approach also has it limitations. Hendry and Mizon (1978) view this extreme bounds analysis as an odd method of analysing sensitivity. Furthermore, they consider the method to be one which is non-informative on the coefficients of the doubtful variables. If the doubtful variables are, in fact, of any importance, then models excluding them would be manifestly invalid. They cannot accept that extreme bounds analysis is a route towards increasing the credibility of econometric evidence, either constructively, as done by a researcher to help him select 'non-fragile' models, or destructively, since invalid models could well seem 'non-fragile and excellent ones fragile' (Townsend, 1997).

McAleer *et al.* (1985) highlight two important limitations to this approach as a method for evaluating fragility. Firstly, the bounds depend on which variables are treated as doubtful and, secondly, if a "doubtful" variable is actually crucial in accounting for the behaviour of the dependent variable, then fragile results will be formed. They conclude by arguing that extreme bounds are generated by the imposition of highly arbitrary, and generally unknown, restrictions between the parameters of a model. They show that extreme bounds analysis demands a general, adequate model from which variables are critical to a relationship. They also observe that these conventions are highly questionable and go on to illustrate, both theoretically as well as empirically, that deviations from these conventions almost completely negate the utility of extreme bound analysis.

They conclude by confessing (McAleer *et al.*, 1985, p. 306):

> "We are only too aware that what have been described are the necessary rather than sufficient conditions for taking the con out of econometrics. As any users of corporate accounts will be aware, there are many ways around standards. But that is not to deny their value. It serves only to highlight the need".

In Hylleberg and Paldam (1991), Yitzhaki criticises Leamer's proposals because they add to an already abundant "toolbox" of econometrics. Yitzhaki claims that econometrics faces a credibility crisis and that a skilful econometrician can prove or disprove any theory by using the "proper" technique.

Sims' methodology: Atheoretical time series merchants

The last contender for alternative econometric methodologies developed his theories and methods at a time characterised by a climate of great dissatisfaction with the large-scale macroeconomic models. These models were not living up to expectations and predictive performance was alleged to be poor and identification of these simultaneous equation models was questionable. Simultaneous macroeconomic models were the focus point of the Cowles Commission at that time. Sims rejected the then familiar 'Cowles style' of identification and expressed unease about *a priori* restrictions on lag length for the identification of rational expectation models (Sims, 1980a).

After an analysis of Sims's work, it is evident that much of his work was founded and inspired from an article by Liu (1960). Liu's argument touches a chord with anyone involved in the construction of computable general equilibrium models. If decisions on consumption, labour supply, and portfolio allocations are all determined by individuals maximising lifetime utility subject to a budget constraint, each relationship would be determined by the same set of variables (Pagan, 1987). As a consequence, theoretical considerations would predict no difference in the menu of variables entering different equations, although the quantitative importance of individual variables is most likely to vary with the type of decision (Pagan, 1987, p. 15). In fact, this action is little different to what is done in any attempt to model reality by capturing the major influences at work. This was precisely what Fisher (1961) replied to Liu, which Pagan (1987) considers as pertinent during the late eighties as when it was written.

Sims seems to put a lot of emphasis on the issue of identification and the reader is left to question if it is, in fact, as serious as Sims suggests. Examples in applied work where identification is the likely suspect when accounting for poor results, are not that easy to find. Rather the issues of specification and data quality seem to be of far greater importance.

Whereas the Colwes Commission methodology would have derived 'structure free' conclusions through the reduced form, Sims chooses to work with a vector autoregressive representation (VAR) for the endogenous and exogenous variables.

Sims defines his VAR in the form:

$$Z_t = \sum_{j=1}^{p} A_{jZ\,t-j} + e_t. \tag{2.1}$$

Sims also manipulates (1) for use in later stages in the methodology. (1) is inverted to give the moving average form:

$$Z_t = \sum_{j=0}^{\infty} A_j e_{t-j} \tag{2.2}$$

$A_0 = \mathrm{cov}(e_t)$. Since $\bar{A}_0$ is a positive definite matrix there exists a non-singular lower triangular matrix *P* such that: $PA_0P' = I$, allowing the definition $\eta_t = Pe_t$, where η_t has zero mean and covariance matrix *I*. (2) may then be re-written in terms of η_t as:

$$Z_t = \sum_{j=0}^{\infty} A_j P^{-1} P e_{t-j} = \sum_{j=0}^{\infty} D_{j\eta_t - j} \tag{2.3}$$

Where η_t are the orthogonalised innovations.

As with the two previous methodologies, Pagan (1987) summarised Sims's VAR method into four steps:

- Transform data to such a form that a VAR can be fitted to it.
- Choose as large a value of p and $\dim(z_t)$ as is compatible with the size of the data set available and fit the resulting VAR.
- Try to simplify the VAR by reducing p or by imposing some arbitrary 'smoothness' restrictions upon the coefficients.
- Use the orthogonalised innovations representations to address the question of interest.

This model contains no exogenous variables, assuming that no economic theory can be used to set elements of the structural form matrices to zero. The work of Sims has been labelled 'atheoretical', choosing to model economic behaviour on the basis of no particular economic theory. Each variable is measured either in levels or in first differences, is treated symmetrically and explained by lagged values of itself and other variables in the system. In this approach, the only role of theory is to specify the variables to be included.

In step two both p and the number of variables in z_t need to be specified. Most of Sims' methodology applications have put p between four and ten. The selection of variables to appear in z_t is also important. An example can be found in Sims (1980a & b) where his conclusions about the role of money in Sims (1980a) were severely modified in Sims (1980b) by expanding z_t to include an interest rate. Essentially, the second step is the analogue of step one in Hendry's and Leamer's methodologies, and the need to begin with a model that is general enough haunts all three methodologies (Pagan, 1987). Yet in Sims' case, the difficulties might be greater; as he wants to model reduced form rather than a single structural equation.

As stated in the summary steps, step three requires simplification of the VAR by reducing *p* or by imposing some arbitrary 'smoothness' restrictions on the coefficients. The third step is necessary precisely because of the fact that both *p* and $\dim(z_t)$ need to be large, and so the number of unknown parameters, $p\text{x}\dim(z_t)$, can easily become too large to be estimated from the available data. This has important implications for the degrees of freedom in the model. If *p* is large, then a degrees of freedom problem may occur and this may also limit the number of variables included in the model. If *p* is too small, then the relevant information in the lagged values may be lost thereby leading to biased results. Inefficiency associated with the over-parameterisation must be balanced against the biases associated with a parsimonious parameterisation (Townsend, 1997). A number of methods of lag length determination have been used, namely the modified likelihood ratio, the Akaike Information criteria (AIC) and the Schwarz criteria (SC). Townsend (1997) elaborates on the detail of the AIC and SC approaches. Essentially, the lag which maximises the value of the AIC and SC is chosen. Gewke and Messe (1981) prefer the Schwarz criteria as it yields a consistent estimator with the probability of under- or overestimating the true lag length approaching zero as the sample size approaches infinity. The final step has been the subject of a number of critiques. The VAR approach has broadly been used for two uses, forecasting and policy analysis. In this sense, policy analysis is interpreted narrowly to mean the addition of a known innovation shock to the model (Townsend, 1997, p. 342).

As Cooley and LeRoy (1985) point out, to ascribe any meaning to impulse responses for theses innovations, it is necessary that the latter be treated as exogenous variables, which requires the imposition of prior restrictions upon causal structure of the system in exactly the same fashion as was done by the Cowles Commission. The strong claims the methodology makes to being free of prior information therefore seem to be largely illusory. However, less pessimistic reviews such as the work by Johansen (1995), suggests that VAR models do not represent the truth about economic phenomena but should be considered useful in describing the statistical variation of the data, such that insight can be gained on the interrelationships between economic variables.

Sims' aim is a much broader analysis of issues than what Hendry or Leamer try to do, but encounters difficulties with this wideness. Whereas Hendry's and Leamer's methodologies seem to provide useful information in a concise form, Sims' methodology is the worst when it comes to reporting. Most of Sims' reporting only consists of pages of graphs and impulse response functions.

The question now remains if any of these approaches have been used in practice, especially in the field of agricultural economics? Examples of calibration methods have already been stated. To note, are the large scale models such as the FAPRI model, FAO world model, the SWAPSIM model and the AGLINK model, all well documented in Van Tongeren (2001).

As the name suggests, the traditional model is probably the most widely used of all approaches. Examples of its application also abound in agricultural economics. Here, the reader is reminded of some of the seminal papers such as the paper by Nerlove (1956). Suits' (1955) paper on an econometric model for the watermelon market also provides a good example of how agricultural economists were 'up to speed' with developments and approaches in econometrics during the 1950s and 1960s.

Although few would deny that in the hands of the "masters" themselves, the methodologies of Hendry, Leamer and Sims appear to perform impressively, it's not quite certain that this is the case when these approaches are applied by their "disciples". One reason for this (or least as far as Hendry's and Leamer methodologies are concerned) is the fact that there are not many examples of the approaches being applied in fields of economics, even more so in agricultural economics. However, the underlying approaches of both Hendry's and Leamer's methodology are well represented. Examples of Error Correction Mechanisms (ECM) being applied in agricultural economics abound. The same can also be said for Leamer's Bayesian methods. Yet, both have not been the sole developers of these approaches and could, therefore, not receive full credit for them. Sims on the other hand seems to be the most well represented of the three methodologies. Examples of how Sims's VAR approach was applied in agricultural economics can be found in Orden and Fackler (1989), Mount (1989), Todd (1989), Kaylen (1988), Devadoss and Meyers (1987), Featherstone and Baker (1987), Bessler & Kling (1986), Chambers (1984) and Bessler (1984). The paper by Bessler and Kling (1988) also provides an example of where the VAR approach has been combined with a Bayesian perspective, almost Leamer's and Sims' methodology combined.

2.4 CONCLUSION

This chapter briefly reviewed the history and development of econometrics over the last five decades. Econometrics unites deduction, induction and statistical inference; its methodology concerns the procedures adopted in the testing and, where applicable, the quantification of economic theories. Its development is largely a post World War II phenomenon, exploiting the increased availability of economic data, the phenomenal development of high speed computer technology as well as the appropriate software.

Econometrics' development was also highly influenced by an increased need from economic scholars to bring economics to a level where it would be regarded as a hard science. Indeed, econometric methodology is, perhaps, best understood via the sophisticated falsification of Sir Karl Popper. Popper (1968, 1972) offered a demarcation criterion between science and non-science, designating science as that body of synthetic propositions regarding the real world which, at least in principle, are capable of refutation through the use of empirical observations.

The 1950s and 1960s were characterised by econometric methodology which appeared to be based upon a superficial appeal to Popperian falsificationism. Economists held the hope that econometrics would facilitate the establishment of an empirical base similar in content to that of the hard sciences. But, econometric development during the 1950s and 1960s was also greatly influenced by the increased interest in positivism. The proponents of positivism, Friedman and Lipsey, greatly advocated hypothesis testing in their works. Many saw in econometrics what they perceived as the provision of a rigorous and reliable method of testing hypotheses, a clear-cut route by which 'poor' theories would be weeded out to be replaced by better theories. Yet, the majority of econometric investigations in the 1960s and 1970s were directed more towards the estimation of economic models than towards testing of hypotheses. Therefore, in practice, econometrics became a vehicle for verification, but applied the rhetoric of falsification.

However, during the seventies, the large scale econometric models which were developed as tools to cure the economic ills of the world began to fail. The inadequacy of these models' ability to deal with large external shocks such as the oil crisis shook the trust of policy makers. Scholars began to question the application of econometrics and many criticised the way in which econometrics was being applied empirically.

As a consequence, various econometric practices were developed in an attempt to restore confidence in how econometrics was being practised. The remainder of the chapter examined the econometric approaches which currently prevail: no estimation approaches and estimation approaches. The section on estimation approaches focussed on the traditional approach of econometric application as well as those approaches championed by Hendry, Leamer and Sims. It was pointed out that each of these approaches still has methodological weaknesses. Granted that no approach has managed to present a completely flawless method for the application of econometrics, what can be said about the whole debate on different approaches? Firstly, there needs to be a substantial clarification of procedures used in model selection and verification as well as auxiliary concepts such as exogeneity (Pagan, 1987). Second, there needs to be a clear understanding of the limits of econometric modelling (Darnell and Evans, 1990). An astutely critical attitude towards econometrics, which is held by critics such as McCloskey (1985a) and Mayer (1980), might interpret the differing methodologies as a tacit admission of a complete failure of econometrics, rather than as constructive attempts to improve it. It is important to rid econometrics of the 'black box' mentality that always besets it. Instead of taking some complex econometric approach out for a walk, scholars should rather focus their attention on the economic problem at hand.

What does the future hold for econometric scholarship? Most of this chapter has dealt with events and trends in economics during the 1950s to 1990s and, therefore, did not touch on any developments in the new millennium. Robert Solow (1997) concluded his summary of economics near the end of the twentieth century with a phrase of Oscar Wilde's description of a fox hunt – 'the unspeakable in pursuit of the inedible" – noting that perhaps economics was and example of "overeducated in pursuit of the unknowable". Colander (2000) sketches a picture of economics in the next century and argues that despite the ongoing controversies in the field of economics, even more so in econometrics, today, "New Millennium economists" are far more comfortable with what they do after the changes in structure and contents of economics over the last half century. Colander (2000, p. 131) concludes:

> "Rather than bounding after the knowable, and trying to deduce analytically models that hold for all times, economics has reduced its search to what it believes is

> knowable. New Millennium economists search for patterns in data, try to find temporary models that fit the patterns, and study the changing nature of those patterns as institutions change. In some ways, the economics profession has come to a full circle back to the more descriptive and institutional approach which was common a full century ago, in the middle of the 20th century. The underlying mathematical structure of models and computational techniques that economists use in 2050 is, of course, much more complicated, but most economists are being trained to use these tools, not to derive them. This frees the training of graduate students to focus on what textbooks of the 1940s focussed on – melding together insights, numerical examples, classification, and simulations to arrive at sensible discussions of policy – and allows me to describe economics in 2050 as the "appropriately educated in search of the knowable.""

With the overview of the history and development of econometrics now complete, the question of what are the problems of econometrics that have sparked controversy and criticism, still remains. This will be the focus of the next chapter, which will explore problems and criticism associated with the way which econometrics has been applied in economics and agricultural economics.

CHAPTER 3

THE SECRET SINS OF ECONOMETRICS

3.1 INTRODUCTION

So far, this study has reviewed the history and development of econometrics since the early 1920s. The review highlighted the increased application of mathematics and statistics in economics and also pointed to the rise in criticisms of such applications. Attempts at developing alternative methods and approaches to counter these problems have also been covered. This chapter shifts the focus from the history and development of econometrics to the criticism associated with the way in which econometrics has been and still is being applied. "Vices" of empirical economic scholarship will therefore be the main focus of the remainder of the study.

Today, literally thousands of applied econometrics papers appear in almost every economic journal, econometrics journal as well as agricultural economics journals. However, with the decades of churning out applied econometric studies and papers, also came an increasing sense of dissatisfaction and critique regarding the way in which econometrics were and are being practised. Many authors have expressed their dissatisfaction and concern on the topic. Mayer (1980) queries "*Economics as a hard science: Realistic goal or wishful thinking?*" Hendry (1980) questions "*Econometrics: Alchemy or Science?*" Sims (1980a) suggests blending "*Macroeconomics and reality*". Black (1982) writes of "*Trouble with Econometric Models*". Leamer (1984) suggests "*Let's Take the Con out of Econometrics*". More recently, McCloskey (2002) writes, with zeal close to that of a TV evangelist, of "*The Secret Sins of Economics*". Many journals, such as the *Journal of Socio Economics*, the *Journal of Economic Methodology*, the *Journal of Econometrics* and the *American Economic Review*, to name but a few, have devoted entire issues to the problems associated with the empirical application of econometrics. The dissatisfaction has also spread to the more specialised economic disciplines such as agricultural economics. Agricultural economists too, have expressed their concern over the application of applied econometrics in agricultural economics. The editors of the *American Journal of Agricultural Economics* also devoted an issue to this problem in 1993. Hoch (1984) suggests "*Retooling the mainstream*", McCloskey (1990) writes of "*Styles of persuasion in agricultural economics*". McGuirk and Driscoll (1995) contemplate the "*Hot air in* R^2 ". South African agricultural economists have, over the years, also expressed their concern over the issue with Nieuwoudt (1973) addressing the "*Data problems in agricultural economic research*". Other South African studies also include Groenewald (1990), Nieuwoudt (1992) and Kirsten (2002).

Reviewing the literature on the dissatisfaction with applied economic studies quickly reveals that, over the years, there have been a couple of scholars such as Leamer, Leontief, McCloskey and Mayer who have devoted some of their scholarship to analysing applied economic studies. Quite often these scholars have spurred huge debates. Another, yet on a somewhat lighter note, is the fact that the majority of studies on this topic have proven to be very entertaining reading. Studies by Leamer, Mayer and McCloskey immediately come to mind as they seem to be pioneers in this regard. However, it appears that this observation is far from new. Recordings of this phenomenon, for example, are found in Leijonhufvud's (1973) "*Life among the Econ.*"

The purpose of this chapter then, is to explore the reasons behind the disenchantment with applied econometrics. A review of the literature reveals that the reasons behind the disenchantment could be ascribed mainly to:

- The use or misuse of statistical significance tests in applied studies.
- The problems associated with the quality of the data used in applied studies.
- The problems associated with replication.
- Data mining
- The "Black box ideology" in applied econometrics and
- Scholasticism and associated preference falsification.

The last vice is perhaps, at first, not that apparent, but is well documented in studies such as Davis (2004), McCloskey (1999) and Coupé (2003) Mayer's (1987) study is perhaps the best example as he briefly addresses all the problems. The focus of this study is on the first vice, namely, statistical significance tests in econometrics. The remaining vices will form part of the discussion in this chapter and the remainder of this study. This chapter, then, starts off by looking at some of the literature on the use of statistical significance in economics. The aim is to identify the main problems with the use of statistical significance, discuss them in detail, and then proceed towards identifying why these problems have persisted for so long.

3.2 THE USE AND ABUSE OF STATISTICAL SIGNIFICANCE TESTS IN ECONOMETRICS

McCloskey (1985, p. 182) contends that "no proposition about economic behaviour has yet been overturned by econometrics". Many authors are doubtful of the value added by econometric testing (see for example, Summers, 1991). Moreover, many econometricians have become increasingly worried about the credibility gap between econometric theory and applied economics. Leamer notes: "Like elaborately plumbed birds who have long since lost the ability to procreate but not the desire, we preen and strut and display our *t*–values" (p. 37). Mayer too, expressed his concern, "…. econometric work includes far too many *examples of game playing*, or what Firsch (1970) in his criticism of certain types of mathematical economics has called "*playometrics*" (Mayer, 1980, p. 169, emphasis added). He later goes on by citing: "the automatic pardon for the crime of using upside-down significance tests" (p. 171).

Indeed, *t*-values, *F*-ratios and null hypotheses have played an inherent role in the image of economics as portrayed above. Testing hypotheses are among the basic pastimes of econometricians. It is a compulsory topic in any course in introductory statistics and econometrics. Many authors agree that economists have followed the Neyman-Pearson procedure, published in 1933, which calls for specification of a null and an alternative hypothesis with associated distributions. Then, using a predetermined critical value for the test statistic, the null hypothesis is accepted or rejected based on the sample data. However, econometric practice seems closer to the approach of Sir R. A. Fisher, although he is rarely mentioned (apart from references to the *F*-test) (Keuzenkamp & Magnus, 1995). The Fisher approach is called statistical inference; the Neyman-Pearson approach is called statistical decision making (Tweeten, 1983). In fact, one of the hostile disputes in science is that of the Fisher versus Neyman-Pearson controversy. The differences between the two approaches are discussed in detail in Keuzenkamp and Magnus (1995, p. 12 – 16). In short, the Fisherian theory of significance testing contains the following characteristics:

- Reliance on tail areas (P – values),
- Intended for small samples,
- Instruments for inductive scientific inference, i.e. statistical inference.

The following points characterise the Neyman-Pearson methodology:

- Emphasis on size and power,
- Applications to contexts of repeated sampling,
- Instruments for inductive behaviour and decision making.

The shortcomings of both approaches are discussed in Keuzenkamp & Magnus (1995) and Tweeten (1983). But, the Fisherian or Neyman-Pearson approach aside, it is especially the fact that in so many instances, statistical significance has become the cornerstone for accepting or rejecting hypotheses, which is alarming.

It is precisely this use of tests of statistical significance as the key criterion to establish the analytical importance of empirical results that has been the centre of the whole debate behind statistical significance in economics. Hitherto, it has been a mainstay across disciplines for many decades and in spite of severe criticism of this practice, there has been barely any improvement over time. Typically, little evidence apart from statistical significance is provided to establish the importance of empirical findings; and if additional evidence is provided, such as coefficient size or intervals, such information is not much discussed. This procedure unites empirical practitioners from economics, ecology, sociology, psychology, and medicine (Altman, 2004).

Indeed, the practice of statistical significance has produced a powerful consensus in use. Thus, the focus of scholarly debate is rarely about sampling issues, the size of coefficients, result replicability, or missing variables (possible by-products of misconstructed theories), which speak to the analytical importance of the researcher's empirical results. The focus is rather on whether or not the results are statistically significant. Needless to say, given the nature of statistical significance, few are convinced by such a discourse. This shows how this misconception encompasses most of the vices listed previously, for example, replicability and data problems. Analytically significant and insignificant results can both be statistically significant. Yet, statistically insignificant results might nevertheless be suggestive of analytical significance.

The prevalence of the misuse of statistical significance is well documented across fields such as economics, ecology, sociology, psychology and medicine. Anderson, Burnham and Thompson (2000) document the problem in the *Journal of Wildlife Management*, a premier journal in ecological studies. Morrison and Henkel (1970) review the sociology and psychology literature. Fidler *et al*. (2004) review the psychology and medical literature. Other contributions on the use of statistical significance in psychology also include Thompson (2004). Sadly, the only major reforms with regards to the use of statistical significance tests have been instituted in psychology and medicine.

In psychology, decades of criticisms eventually resulted in the American Psychological Association (APA) revising its publication manual in 1994, encouraging authors to go beyond the use of statistical significance tests (Altman, 2004, p. 653). According to Fidler (2002), the American Psychological Association (APA) is responsible for 27 journals and, moreover, the APA manual is the primary publication guide for no less than 1 000 other journals in psychology, the behavioural sciences, nursing and personnel administration. In 1996, the Board of Scientific Affairs (BSA) of the American Psychological Association (APA) convened a committee named the Task Force on Statistical Inference who's charge was to "elucidate some of the controversial issues surrounding applications of statistics including significance testing and its alternatives; alternative underlying models and data transformation; and newer methods made possible by powerful computers" (Wilkinson *et al*., 1999, p. 598). Fundamentals of the Task Force were published in Wilkinson *et al.* (1999) and further reforms recommended, encouraging journal editors to be ever more receptive and encourage other measures of significance, such as size and confidence intervals. The professional psychology community has probably gone the furthest in confronting and dealing with the misuse of statistical significance tests. However, in spite of its various efforts at reform and moral suasion not much has changed in the world of publishing. Thompson (1999; see also Fidler *et al*., 2004) concludes that the often reported misuse of statistical significance tests in psychology persists and has remained unabated over time.

As detailed by Fidler *et al*. (2004), there have been various attempts made in medical research to encourage movement away from significance tests as the main determinant of analytical significance since the late 1970s, even prior to the efforts made in psychology, starting with the efforts made by the *New England Journal of Medicine* in 1977. The *British Medical Journal* implemented a policy encouraging the reporting of confidence intervals in 1986. In 1988, the International Committee of Medical Journal Editors revised their manuscript submission requirements for the biomedical journals to discourage the traditional focus of statistical significance tests and redirect energies to alternative measures. Nevertheless, the determination of the effectiveness of the new standard for excellence was at the Editor's discretion and the gist of the recommendations was in the context of moral suasion (Altman, 2004, p. 654). Nonetheless, moral suasion has not been enough to displace the use of statistical significance tests as the dominant determinant of analytical significance, though some changes with regard to reporting of results did occur. Fidler *et al*. (2004, p. 124) conclude that even when non-statistical test variables are reported, such as confidence intervals, statistical significant tests remained the bedrock of the analytical narrative. For example, confidence intervals did, in fact, become more noticeable on paper, but more as a form of window dressing required or encouraged by journal editors or editorial policies.

3.2.1 Significance testing in economics

Significance testing has also been widely applied in economics and as is the case with the other social sciences, economic scholars have also expressed concern over the way in which significance tests have been applied. For twenty years, since the publication of the first edition of *The Rhetoric of Economics* (1985a), McCloskey has campaigned tirelessly to convince the economics profession that it is deeply confused about statistical significance. Yet, she has not been the first to look at significance tests in economics. Zellner (1979) contains a small survey of twenty two quantitative articles in five issues of different leading economic journals in 1978. He finds that significance testing is very popular, that 1% and 5% significance levels dominate, and that power considerations are rarely discussed, despite the dominance of Neyman-Pearson methodology in the training of economists. In another survey, of Canterbery and Bukhardt (1983), of 542 empirical papers published in the *American Economic Review, Journal of Political Economy*, *Economic Journal*, and *Quarterly Journal of Economics* from 1873 – 1978, only three papers attempted to refute the hypothesis under investigation. Keuzenkamp and Magnus (1995, p. 20) note: “In most cases rejection of economic hypotheses is easy; whereas verification is hard (anyone with experience in economic modelling knows how difficult it can be to obtain models that are ‘satisfactory’).”

Keuzenkamp & Magnus (1995) explored the different aims of testing namely: theory testing, validity testing, simplification testing, and decision making. They also reviewed the approaches of Fisher and Neyman-Pearson and discussed each approach’s shortcomings. For the purpose of investigating the significance tests, they surveyed the papers in the *Journal of Econometrics*. In total, 668 papers were counted, of which eventually 137 papers were selected and reviewed. They found that 72% made use of significance tests, the 5% level of significance was the most used and that sample sizes were not always reported nor was the size of a test considered.

As noted previously, it is perhaps McCloskey who has investigated the use of significance in economics in the greatest detail. McCloskey (1985) reviewed 10 full length papers, using statistical significance tests, published in the 1981, 1982 and 1983 volumes of the *American Economic Review*. She concludes that most authors confuse statistical and substantive significance noting: “If we do not wish to leave science to chance we must rethink the use of statistical significance in economics” (McCloskey 1985, p. 204).

In 1996, the study was extended in Ziliak & McCloskey’s “*The Standard Error of Regression*”. They reviewed 182 full length papers that used regression analysis, published during the 1980s in the *American Economic Review*. Each article’s use of statistical significance was evaluated by means of a 19 question questionnaire. Their review was, however, not only limited to the universe of article publications. They also evaluated the use of statistical significance tests in typical statistical textbooks, which is where the students who become publishers derive so many of their norms for excellence in applied research. They find that in economics it is only the exceptional statistical textbook which discusses in any detail and places much emphasis on the distinction between statistical significance and indicators of analytical significance. They quote Granger (1994) who reviewed four leading books in 1994 namely: those of Goldberger 1991; Davidson and MacKinon 1993; Greene 1993; and Griffiths, Hill and Judge 1993. Granger (1994, p. 118) notes that:

> "When the link is made [in Goldberger, between the economics and the technical statistics] some important insights arise, as for example the section discussing "statistical and economic significance," a topic not mentioned in other books."

McCloskey and Ziliak (1996, p. 111) conclude:

> "In a squib published in the *American Economic Review* in 1985 one of us claimed that "Roughly three-quarters of the contributors to the *American Economic Review* misuse the test of statistical significance" (McCloskey 1985, p. 201). The full survey confirms the claim, and in some matters strengthens it. We would not assert that every economist misunderstands statistical significance, only that most do, and these some of the best economic scientists."

Yet, it did not seem that economic scholars were going to part with their use of significance tests. McCloskey acknowledged the fact, when, in her column in the *Eastern Economic Journal*, she compared herself to a figure from Greek Mythology: Cassandra. "I have to admit guys, I feel like Cassandra" (McCloskey 1999, p. 357). Cassandra was the most beautiful of the daughters of Priam, king of Troy. The god Apollo became attracted to her and made her a prophetess. In exchange he wanted sexual favours, which she refused to give. This resulted in him cursing her in a most peculiar way. Her curse was that although she would be correct in her prophecies, nobody would believe her.

This became evident when in 2004 Ziliak & McCloskey applied the same 19-item questionnaire of their 1996 paper to the 137 full-length papers, using regression analysis, published in the *American Economic Review* during the 1990s. In fact, they felt it was getting worse, sighting that of the 137 relevant papers, 82% mistook statistically significant coefficients for economically significant coefficients (as against 70% in the 1980s). Their 2004 article was the lead article of a special issue in the *Journal of Socio-Economics* on the widespread use of statistical significance. The special issue also contained detailed comments on McCloskey and Ziliak's study. Elliott and Granger (2004), Horowitz (2004), Leamer (2004), Lunt (2004), Wooldridge (2004) and Zellner (2004) all provide detailed comments on the lead article. According to Ziliak and McCloskey (2004b), though some did not agree entirely with their findings, in principle, every one of the commentators agreed with their two main points:

- That economic significance usually has nothing to do with statistical significance, and

- That a supermajority of economists does not explore economic significance in their research.

However, not every economic scholar has been entirely happy with McCloskey's demolition charge on the use of statistical significance in economics over the past 20 years. More recently, Hoover and Siegler (2005) wrote of "*Sound and Fury: McCloskey and Significance testing in economics.*" They argue that McCloskey's analysis of the state of significance testing in economics is apocalyptic. They analysed her charges and rejected them, arguing that "statistical significance is not economic significance is a jejune and uncontroversial claim, and there is no convincing evidence that economists systematically mistake the two." (Hoover and Siegler, 2005, p.1). They also claim that other elements of McCloskey's analysis of statistical significance are ill-founded, and her criticisms of the practices of economists are based on inaccurate readings and tendentious interpretations of their work. They reiterate that if properly used, significance tests are a valuable tool for assessing signal strength, for assisting in model specification, and for determining causal structure.

Their study reminds the reader of one important use of statistical significance, namely that statistical significance tests are key in establishing whether the assumptions underlying the model used in the analysis are violated. Econometric models are not solely developed on the basis of economic assumptions; the statistical assumptions underlying the model are also crucial. Recall that in Chapter two it was discussed how most applied econometric studies follow the standard approach. Furthermore, it was also shown that this approach is based on a set of assumptions developed to ensure that the estimators obtained from the model are unbiased and have minimum variance, i.e. they are best linear unbiased estimators (BLUE). Significance tests are therefore paramount in establishing if, in fact, the estimators satisfy the criteria of being BLUE or BUE (in the case of non-linear models).

McCloskey largely ignores this critical application of statistical significance tests in regression and econometrics. Instead, she adopts a *ceteris paribus* approach to these issues as illustrated below (Ziliak and McCloskey, 1996, p.98):

> "An estimated coefficient β is of course a random variate, and the accuracy of its estimated mean depends on the properties of the error term, the specification of the model, and so forth. But to fix ideas suppose that all the usual econometric problems have been solved."

As Hoover and Siegler (2005) illustrate, McCloskey's dark sketch of the use of statistical significance tests in economics is not entirely that dark. Statistical significance does have its place in applied econometric analysis, especially in establishing whether the critical statistical conditions underlying the econometric model have been satisfied. Rather, the real concern at hand is whether economic scholars have been applying significance tests correctly. In his comments on McCloskey and Ziliak's "Size Matters", O'Brien (2004) reviewed economic history to see whether economic historians fare any better in using statistical significance than the authors in the *American Economic Review* articles examined by McCloskey and Ziliak. O'Brien examined every full-length paper published in the *Journal of Economic History* and *Explorations in Economic History* in the years 1992, 1996, 2001 and 2002. In total, 185 papers were reviewed of which 118 papers used regression analysis. O' Brien, (2004, p.569) concludes his findings:

> "Finally, many of these 185 papers would fail one or more of McCloskey and Ziliak's tests for the proper use of statistical significance. Yet, the conclusions of only a few—eight by my count—are cast in doubt by these errors. Authors, editors, and reviewers should pay more attention to when statistical significance matters and when it doesn't, and whether quantitative analysis can bear the weight of the conclusions being drawn from it. But the relative lack of oomph from errors in using statistical significance may

> help explain why McCloskey and Ziliak found so little progress in the eight years between their two papers. Unless, of course, economic historians are doing their economics better than the authors in the *AER*."

Agricultural economists too have examined the use of regression analysis and statistical significance tests in papers published in leading Agricultural Economics Journals. Gardner (1983) surveyed the agricultural economics profession's use of statistical data in its scientific work in the *Journal of Farm Economics* and the *American Journal of Agricultural Economics*. Volumes of 1950, 1960, 1970 and 1981 were included in the sample. Though not the focus of his paper, he found that econometrics and even more so algebraic expression of models and hypotheses have evolved from an oddity to the standard mode of research reporting. Though his study did not focus on the use of statistical significance, it did show that as in economics, econometrics was becoming far more popular. Furthermore, his study hints at another vice of economics, namely data problems, which will be discussed in a later chapter of this study. Hoch (1984) also surveyed the volumes of the *Journal of Farm Economics* and the *American Journal of Agricultural Economics* published in 1950, 1966 and 1983. His focus was on the primary tools of analysis used in articles published in the two journals, noting a substantial increase in the use of econometric techniques. His article was strongly influenced by the works of Leamer, Tweeten and McCloskey. On the use of regression Hoch (1983, p. 795, emphasis added) notes:

> "In contrast to the textbook admonition to specify your model beforehand and then to accept the results, however they turn out, the regression man uses trial and error to select from among many variables the "best" subset, best on such criteria as highest *R^2, "reasonable" coefficient signs, and the "best" significance levels*. The process has the unpleasant consequence that standard errors, *t*-ratios, and other test statistics are no longer strictly applicable. Yet even the critics of "specification search" see reason for it. How we "know" what we know is the crux of the matter, for we must necessarily bring much outside information to bear in carrying out our work and in reaching our conclusions."

Again the author discussed some of the more general concerns with econometrics in agricultural economics and did not really zoom in on the misuse of statistical significance.

This task was left to none other than McCloskey. In "*Styles of persuasion in agricultural economics*", McCloskey (1990) surveyed articles published in the 1989 volume of the *American Journal of Agricultural Economic*s and the 1929 volume of the *Journal of Farm Economics*. She notes that almost all empirical work published in the 1989 volume made use of regression analysis. "Regression analysis seems to have a tighter hold on the empirical imagination of agricultural economics than it has in other applied fields, probably because of the agronomical origins of the statistics. R.A. Fisher, who named most of them, worked at an agricultural experiment station." (McCloskey, 1990, p. 1126) Her opinion of the way in which agricultural economists use statistical significance tests in the *American Journal of Agricultural Economics* is the same as that of economists using it in the *American Economic Review.* She claims that every one of the twenty one articles that applied regression analysis in the 1989 issue of the *American Journal of Agricultural Economics* grossly misused it. "They take statistical significance to be the same thing as scientific significance." (p. 1127) As in her other articles (McCloskey, 1985, 1992, 1996, 2004) she states that all econometric work will have to be redone.

It seems that significance tests do have a function in econometrics, though, as suggested by the literature, this function is not clearly understood and in many cases abused. Another point that is also visible from the literature is the relation between the misuse of statistical significance tests and the other vices of economics, those being data mining, replication, data quality, "black box" ideology and scholasticism. It is especially the relation between data quality, data mining and replication which is of great interest to this study. As seen from the literature, these vices are quite often overshadowed by the misuse of statistical significance.

3.3 SIGNIFICANCE TESTS

At this point it would be fitting to revisit the assumptions underlying the standard regression approach; in fact, most regression approaches in some way or the other rely on these assumptions. Though McCloskey and Ziliak, in their studies, treat these assumptions in the form of "other things equal"[1], it does play a key role as far as statistical significance tests'

[1] It is perhaps ironic that McCloskey's column in the *Eastern Economic Journal* also shares this name: "Other things equal".

function in econometrics as well as the relation between significance and the other vices are concerned.

In order to draw inferences about the true estimators of a regression model, it is necessary to not only specify the functional form correctly, but also make certain assumptions about the manner in which the dependent variable is generated. This requirement is easily explained if the reader looks at the population regression function (PRF), say: $Y_i = \beta_1 + \beta_2 X_i + u_i$. This shows that Y_i depends on both X_i and u_i. Therefore, unless the researcher is specific about how X_i and u_i are created or generated, there is no way the reader can make any statistical inference about Y_i and also β_i. The assumptions made about the X_i variable(s) and the error term are extremely critical to the valid interpretation of the regression estimates.

The Gaussian, standard, classical linear regression model (CLRM), the cornerstone of most econometric theory, makes ten important assumptions. (Gujarati, 1998, p.59 and 2003, p. 65)

Assumption 1: The regression model is linear in the parameters. That is, in matrix notation of the general k-variable linear regression model:

$$y = X\beta + u \tag{3.1}$$

Assumption 2: X values are fixed in repeated sampling. Values of the X variable are considered fixed in repeated samples; that is, they are nonstochastic. This entails that the regression analysis is conditional on the given values of the regressors X_i.

Assumption 3: Zero mean value of the disturbance term u_i. Coupled with Assumption two's fixed values of X_i, the mean, or expected value, of the random disturbance term u_i is zero and therefore:

$$E(u_i \mid X_i) = 0 \tag{3.2}$$

Effectively this assumption says that the factors not included in the model, and therefore subsumed in u_i, do not systematically affect the mean value of Y. That is, the positive u_i

values cancel out the negative u_i values so that their average or mean affect on Y is zero. A more technical reason why this assumption is so important is discussed in Malinvaud (1966). The assumption then implies that if $E(u_i \mid X_i) = 0$, then $E(u_i \mid X_i) = X\beta$.

Assumption 4: Homoscedasticity or equal variance of u_i. Given the value of X, the variance of u_i is the same for all observations, i.e. the conditional variances of u_i are identical. Symbolically:

$$\begin{aligned}\operatorname{var}(u_i | X_i) &= E[u_i - E(u_i) | X_i]^2 \\ &= E(u_i^2 | X_i) \\ &= \sigma^2\end{aligned} \tag{3.3}$$

Technically Equation 3.3 represents the assumption of homoscedasticity, or equal spread, or equal variance. This means that the Y populations corresponding to the various X values have the same variance. Assumption four also implies that the conditional variances of Y_i are also homoscedastic. That is,

$$\operatorname{var}(Y_i | X_i) = \sigma^2 \tag{3.4}$$

Assumption 5: No autocorrelation between the disturbances. Given any two X values, X_i and $X_j (i \neq j)$, the correlation between any two u_i and $u_j (i \neq j)$ is zero. Symbolically,

$$\begin{aligned}\operatorname{cov}(u_i, u_j | X_i, X_j) &= E[u_i - E(u_i) | X_i][u_j - E(u_j) | X_j] \\ &= E(u_i | X_i)(u_j | X_j) \\ &= 0\end{aligned} \tag{3.5}$$

This assumption postulates that the errors u_i and u_j are uncorrelated. It is what is technically referred to as the assumption of no serial correlation or no autocorrelation. Given X_i, the deviations of any two values of Y values from their mean value do not exhibit any patterns.

Assumption 6: Zero covariance between u_i and X_i, or $E(u_i X_i) = 0$ More formally,

$$\begin{aligned}
\operatorname{cov}(u_i, X_i) &= E[u_i - E(u_i)][X_i - E(X_i)] \\
&= E[u_i(X_i - E(X_i))], \text{ since } E(u_i) = 0 \\
&= E(u_i X_i) - E(X_i)E(u_i), \text{ since } E(X_i) \text{ is nonstochastic} \\
&= E(u_i X_i), \text{ since } E(u_i) = 0 \\
&= 0, \text{ by assumption}
\end{aligned} \tag{3.6}$$

This assumption states that the error term u_i and the explanatory variable X_i are uncorrelated. What this assumption essentially means is that X_i and u_i have separate influences on the dependent variable Y_i. If X_i and u_i are correlated, it is not possible to assess their individual effects on Y_i. Assumption six is automatically fulfilled if X_i is non-random or nonstochastic and if Assumption three holds, since then $\operatorname{cov}(u_i, X_i) = [X_i - E(X_i)]E[u_i - E(u_i)] = 0$. Gujarati (1998, p. 65) notes that this assumption is not that critical as it holds true even if *X's* are stochastic, provided they are independent or at least uncorrelated with the disturbances u_i.

Assumption 7: The number of observations *n* must be greater than the number of parameters to be estimated. Not as innocuous as it seems, Assumption seven is of critical importance, especially for statistical inference.

Assumption 8: Variability in X values. The X values in a given sample must not all be the same, i.e. **var**(X) must be a finite positive number.

Assumption 9: The regression model is correctly specified or alternatively, there is no specification bias or error in the model used in empirical analysis. Although all the assumptions are equally important, Assumption nine is critical. In a sense it is the crux of the whole problem or disenchantment with the way econometrics have been applied in empirical research.

Fair enough, the assumption warns that the researcher should be sure of the theory used in specifying the model. Yet, even if the researcher was to apply Hendry's general-to-specific approach and he or she is thorough in specifying the model according to the theory, some judgement is still needed in choosing the final number of variables entering in the model. To some extent, there is some trial and error involved in choosing the "right" model for empirical analysis. But, researchers should be wary of the fine line between the trial and error and what is known as data mining. Data mining is trying every possible model with the hope that at least one will fit the data well. Again, it stresses why it is essential that there be some economic reasoning underlying the chosen model and that any modifications in the model should have some economic justification. In short, there should be a fine line between the art of model building and the science behind the model. To some extent this assumption almost summarises McCloskey's whole effort, i.e. models should be correctly specified based on sound economic as well as statistical reasoning behind it, even though she might be emphasising the statistical reasoning.

Assumption 10: There is no perfect multicollinearity. That is, there are no perfect linear relationships among the explanatory variables.

The last assumption is essential for multiple regression analysis, where the regression contains several regressors, or explanatory variables.

Under these assumptions it can be shown that the estimators of the vector of parameters $\boldsymbol{\beta}$ and σ^2, $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$ satisfy several desirable statistical properties, such as unbiasedness and minimum variance, i.e. they are best linear unbiased estimators (BLUE). To note another point, since these are estimators, their values will change from sample to sample. Therefore, these estimators are *random variables*.

But estimation is only half the battle. Hypothesis testing is the other half. As explained in any introductory or intermediate econometrics course, the objective of regression is not only to estimate the sample regression function, but also to use it to draw inferences about the population regression function. Thus, the aim is also to establish how close $\hat{\boldsymbol{\beta}}$ is to the true $\boldsymbol{\beta}$ or how close $\hat{\sigma}^2$ is to the true σ^2.

Therefore, since $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$ are random variables, their probability distributions should be investigated, for it will not be possible to relate them to their true values without knowledge of their probability distributions. Since hypothesis testing is also one of the objectives, the researcher has to specify the probability distribution of the disturbances u_i. The answer to why such a specification for u_i is necessary, is not difficult. The disturbances are random by assumption (white noise), since the estimators vector $\hat{\boldsymbol{\beta}}$ are linear functions of the dependent variable Y and Y in itself is a linear function of u_i, and therefore $\hat{\boldsymbol{\beta}}$ are ultimately linear functions of u_i. Thus, the sampling distributions of the ordinary least squares (OLS) estimators will depend upon assumptions made about the probability distribution of u_i. Since the probability distribution of these estimators is necessary to draw inferences about their population values, the nature of the probability distribution of u_i assumes an extremely important role in hypothesis testing. Notwithstanding the Gauss-Markov theorem, the method of OLS is of little help for the purpose of drawing inferences about the population from the sample. It is therefore important to assume that the *u*'s follow some probability distribution. It is usually assumed that the *u*'s follow the normal distribution. Gujarati (1998, p. 103) justifies this appropriateness by stating the following reasons:

- Nestled in the principal of parsimony, u_i represents the combined influence of a large number of independent variables that are not explicitly introduced in the regression model. By the celebrated central limit theorem of statistics it can be shown that if there are a large number of independent and identically distributed random variables, then, with a few exceptions, the distribution of their sum tends to a normal distribution as the number of such variables increases indefinitely. It is the central limit theorem which provides a theorem that provides theoretical justification for the assumption of normality of u_i.

- A variant of the central limit theorem states that even if the number of variables is not very large or if these variables are not strictly independent, their sum may still be normally distributed.

- With the normality assumption, the probability distributions of the OLS estimators can be easily derived because one property of the normal distribution is that any linear function of normally distributed variables is in itself normally distributed.

- Finally, the normal distribution is a relatively simple distribution involving only two parameters; it is well-known, and its theoretical properties have been extensively studied in mathematical statistics.

Given the assumption of normality, the OLS estimators have minimum variance in the entire class of unbiased estimators, whether linear or not. This result is shown by Rao (1965, p. 258) and unlike the Gauss-Markov theorem it is not restricted to the class of linear estimators only. It can, therefore, be stated that the least-squares estimators are best unbiased estimators (BUE).

Are these assumptions realistic? Furthermore, are they of any use? The “reality of these assumptions” is an age-old question in the philosophy of science. It boils down again to economics’ search for recognition as a hard science. What matters are the predictions based on these assumptions (Gujarati, 1998, p.69). Some argue that it does not matter whether the assumptions are realistic. Of note among the “irrelevance-of-assumptions theses”, is Milton Friedman. According to Friedman (1953, p. 14), unreality of assumptions is a positive advantage: “to be important … a hypothesis must be descriptively false in its assumptions.” Again, the reader is reminded of Popper’s falsification criterion as discussed in Chapter two.

These assumptions are therefore extremely important. Furthermore, it seems that it is critical to test if these assumptions hold when constructing an econometric model. Testing these assumptions is done through statistical significance tests, thereby confirming that statistical significance does in fact a have a cause to be in econometrics and, moreover, to assume that these assumptions hold when evaluating applied econometric studies, might not be that fruitful.

Although Ziliak and McCloskey regard these assumptions in an “other things equal” kind of approach, their 19 point questionnaire might be worth exploring.

3.4 THE STANDARD ERROR OF REGRESSION AND SIZE MATTERS: ZILIAK AND MCCLOSKEY'S QUESTIONNAIRE

Ziliak and McCloskey asked 19 questions about the use of statistical significance in their survey, which was answered "yes" (sound statistical practice) or "no" (unsound practice) or "not applicable." The survey questions were:

1. "Does the paper use a small number of observations, such that statistically significant differences are not found at the conventional levels merely by choosing a large number of observations?" (Ziliak and McCloskey, 1996, p. 101)

Their 2004 paper title says it all: "Size matters". McCloskey and Ziliak show much concern for size, both sample size as well as coefficient size. Indeed, the most important component affecting the statistical power is the sample size. In fact, there is a little room to change a test size (significance level). Important too, is the fact that it is also difficult to control effect size. According to Cohen (1988) an effect size is a deviation of the hypothesized value in the alternative hypothesis from the baseline in the null hypothesis.

A larger sample size generally leads to a parameter estimate with smaller variances, a larger standardized effect size, and eventually, a greater ability to detect a significant difference. But, as Ziliak and McCloskey warn, in instances of large sample sizes, authors need to pay attention to the trade-off between power and the size of a test, and the economic significance of the power against alternatives. If too many observations are used even a trivial effect will be mistakenly detected as a significant one (High, 2000).

However, it is doubtful whether in economics, or agricultural economics for that matter, the researcher will have such a large number of observations such that this effect will actually prevail. The odds are greater that the researcher will have too few observations at his or her disposal. In such a case, it may be difficult to detect a meaningful effect even if it does, in fact, exist. It is, therefore, important to have a reasonable number of observations in order to do any meaningful analysis. Or as Assumption seven of the classical regression model states, have at least as many observations as the number of regressors in the model. In such small sample cases the researcher might be better off in applying the approach of calibration as discussed in the section on alternative approaches in Chapter two.

2. "Are the units and descriptive statistics for all regression variables included?" (p. 102).

Empirical work in economics is measurement. Ziliak and McCloskey (1996) emphasise that it is, therefore, elementary to include units of the variables, and then also to supply the means. This question is important, though in practice often neglected. In fact, the data underlying the analysis has been deemed one of the vices because of its neglect. A number of observers argue that economic (including agricultural economic) scholars apply their "tools" with little concern and respect for the underlying data which the "tools" employ (See for example Leontief 1971 and 1982 and Mayer 1980), Agricultural economic observers such as Bonnen (1983, p. 188) stress the importance of being careful about one's numbers and cite many examples of bad practice in the field. For example, economists no longer even look at secondary data; they get them on magnetic tape and let the computer do the looking (Bonnen, 1975, p. 755.). As far as agricultural economics is concerned, Bonnen (1976, p. 761) concludes:

> "Agricultural economists have a tradition of inquiry that prevents innocence of the empirical. Even we, however, are increasingly failing in individual and institutional research to do the hard, unglamorous slogging in data collection that often is most productive of new knowledge."

The dangers in not being concerned about the data underlying the analysis cannot be overestimated. Whether agricultural economic scholars have paid attention to the data they use, will be investigated in a later chapter of this study.

But, this question is not only relevant to the problems with the data underlying the analysis. Another important issue which has not been receiving the needed attention is the issue of replication. Reporting on the units and descriptive statistics will ease replication of the study. At least, a researcher will able at see what data was used if he or she aims to replicate a study. Many authors argue that replication of econometric studies is often difficult to undertake. Mayer (1980, p. 170), for instance, notes on data problems and replication:

> "A related point is the apparently frequent non-replicability of results, i.e., the inability to determine what data were used in a published paper, and to use these same data to reproduce the results."

Tomek (1993) discussed the meaning and benefits of conformation and replication in agricultural economics. His study identified the difficulties of conformation and replication. Conformation and replication of published results requires duplication of the data set, models, and differences in computer codes as well as the effect on colleagues. He identified the actual data used in the analysis as the principal reason for the difficulty in replicating previous work. Tomek (1993, p. 9) argues, "citations to data sources frequently are vague. Thus, the original data cannot be reconstructed." The editors of the *Journal of Money, Banking and Credit* found that even when they had requested the data from authors (which was almost never done at this time, because of the primitive state of computing power) they could not replicate the results of the studies (Dewald *et al.*, 1986). Tomek too, has devoted some of his research to replication of econometric studies in agriculture.

Mayer (1987) finds this rather surprising, since it is one of the most basic rules of scientific practice that one's methods must be reproducible, so that the results can be verified. He surveyed the criteria used by scientists in evaluating scientific publications and found that 62 per cent of natural scientists considered "replicability of research techniques" to be "essential" (Mayer, 1987, p. 170).

3. "Are coefficients reported in elasticity form, or in some interpretable form relevant at hand and consistent with economic theory, so that readers can discern the economic impact of regressors?" (Ziliak and McCloskey, 1996, p. 102)

Ziliak and McCloskey (1996) cite Wallis and Roberts (1956), who long ago complained that "sometimes authors are so intrigued by tests of significance that they fail even to state the actual amount of the effect, much less to appraise its practical importance" (1956, p. 409) Ziliak and McCloskey (1996 and 2004a) found that only 67 per cent of the papers surveyed in the 1980s and 87 per cent of the papers surveyed during the 1990s did reported coefficients in elasticities, or some other form useful for economic interpretation.

4. "Are the proper null hypotheses specified?" (p. 102).

The most common approach is to test against a null of zero. For example, a null hypothesis is commonly tested in empirical work as $H_0 : \beta_2 = 0$, that is, the slope coefficient is zero. This "zero" null hypothesis is a kind of straw man, the objective being to find out whether Y is related at all to X. Fisher viewed the null hypothesis as the hypothesis to be nullified. However, the term null too frequently has been interpreted to mean zero. Tweeten (1983) warns that constant use of the particular use of the null hypothesis that the true parameter is zero leads to biased working estimates in day-to-day applications. Instead, he suggests that such a bias could be reduced by using the average results from previous experiments as the null hypothesis. With this procedure, the expected value of the null hypothesis is the true parameter. This null hypothesis will not be rejected even if the sample size is very small. Thus, the working hypothesis for making day-to-day management and policy decisions will tend to be unbiased.

Alternatively, the Bayesian approach is to use prior information to formulate hypotheses, but "it is common practice first to estimate parameters from sample information and then to give reasons why these results are correct" (Leamer, 1975, p. 88), though, a problem with this approach is that researchers tend to remember only past results consistent with their results. Another is the problem of selective reporting of research (Tweeten, 1983, p.550). Gujarati (2003, p. 139) argues that editors of reputable journals do not find it exciting to publish an empirical piece that does not reject the null hypothesis. Ziliak and McCloskey (1996 and 2004a) are proponents of the Neyman-Pearson approach since it specifies the null hypothesis as something the researcher believes to be true.

5. "*Are coefficients carefully interpreted*?"

As with question three, this question emphasizes the importance of interpreting one's results correctly.

6. "Does the paper eschew reporting all t – or F – statistics or standard errors, regardless of whether a significance test is appropriate?"

Many observers note that only coefficients of which the *t*-statistics are significant are reported in journal papers. Ziliak and McCloskey (1996, p. 102) suspect that referees enforce the proliferation of "meaningless *t*- and *F*-statistics out of a belief that statistical and substantive significance are the same".

7. "Is statistical significance at the first use, commonly the scientific crescendo of the paper, the only criterion of "importance"?"

Their explanation of "crescendo" is that place in the paper where the author comes to what she evidently considers the crucial test.

8. "Does the paper mention the power of the tests?" (p. 103)

McCloskey and Ziliak place a lot of emphasis on statistical power. It is rather peculiar why they did not ask this question directly after or in conjunction with the question on sample size (question 1) since sample size is one of the components of power analysis.

What is the *power of a test*? The power of a statistical test is the probability that it will correctly lead to the rejection of a false null hypothesis (Greene 2000). The statistical power is the ability of a test to detect an effect, if the effect actually exists (High 2000). Cohen (1988, p. 4) says it is the probability that it will result in the conclusion that the phenomenon exists. Put differently, the power of a test is its ability to reject a false null hypothesis and thus not committing a Type II error. The relation is illustrated in Table 3.1.

Table 3.1: Type I and II errors and the power of a test.

	State of nature	
Decision	H_0 **is true**	H_0 **is false**
Reject	Type I error Denoted by the size of a test (α : significance level)	No error Denoted by $1-\beta$: The power of a test
Do not reject	No error Denoted by $1-\alpha$: The Confidence level	Type II error Denoted by β

Source: Own summary

Their questions on statistical power indirectly address the issue of loss functions. The only situations in which McCloskey appears to accept the usefulness of statistical significance is when it is cast in a strict, decision-theoretic Neyman-Pearson framework, the marker of which is the existence of an explicit loss function: "You can't run science without a loss function" (McCloskey 1998, p. 118; cf. 1985b; 1992, p. 359; 1999, p. 361; 2002, p. 58; Ziliak and McCloskey 2004a, p. 543). As their survey pointed out, not many authors in the 1980s (only 4.4 per cent) nor the 1990s (only 8.0 per cent) considered the power of a test.

9. "If the paper mentions power, does it do anything about it?"

As seen from the manner in which question 8 was answered, power analysis is not frequently discussed or considered in journal articles containing regression analysis. Indeed, the power of a test does have its difficulties. For one, it requires an explicit alternative hypothesis. Furthermore, it involves somewhat complex calculations and it is not regularly computed by commonly used statistical packages. Len and Krebs (1997) list and review statistical power analysis software used in zoology. Gujarati (1998, p. 132 and 2003, p. 137) suggests instead reporting the exact probability (*p* value) of committing a Type I error, or alternatively, the lowest significance level at which a null hypothesis can be rejected. It is then preferable to leave it to the reader to decide whether to reject the null hypothesis at the given *p* value.

10. "Does the paper eschew "asterisk econometrics," that is, ranking the coefficients according to the absolute size of the t-statistics?"

It seems, according to McCloskey and Ziliak's (2004) study, that scholars have improved on their question citing that 75 per cent of papers during the 1980s and only 33 per cent during the 1990s were "guilty" of this practice.

11. "Does the paper eschew "sign econometrics", that is, remarking on the sign but not the size of the coefficients?"

McCloskey and Ziliak (1996, p. 103) notes:

> "Sign is not *economically* significant unless the magnitude is large enough to matter. Statistical significance does not tell whether the size is large enough to matter. It is not true, as custom seems to be arguing, that sign is a statistic independent of magnitude."

12. "Does the paper discuss the size of the coefficients?"

Once regression results are presented, the question is whether the paper makes a point that some of the coefficients and their variables are economically influential, while others are not. This question is closely linked to questions 3 and 5. It again reiterates the notion of specifying the proper target of research – the estimation of relevant parameters

13. "Does the paper discuss the scientific conversation within which a coefficient would be judged "large" or "small"?" (p. 104)

Again, question 13 emphasises the same point as made in questions 12, 5 and 3.

14. "Does the paper avoid choosing variables for inclusion solely on the basis of statistical significance?"

This question is closely linked to assumption 9 of the classical linear regression model, i.e. the model should be correctly specified. It also points to the fine line between correct specification and data mining.

15. "After the crescendo, does the paper avoid using statistical significance as the criterion of importance?"

Again the proper specification of the target of the research paper comes to mind. But, perhaps too, in order for the paper to be persuasive the author might have included a significance test or two in order to accede to the insistence of journal referees.

16. "Is statistical significance decisive, the conversation stopper, conveying the sense of an ending?"

As is the case with question 15, questions 16 and 18 imply the abuse of statistical significance. In a sense it also raises the issue as to why this abuse has persisted for so long. The issue of scholasticism and preference falsification will be explored at a later stage in the study.

17. "Does the paper ever use a simulation (as against a use of the regression as an input into further argument) to determine whether the coefficients are reasonable?"

McCloskey and Ziliak suggest that simulations using regression coefficients can be informative, but statistical significance should not be used as screening device for input. Yet, in most cases, successful simulations rest on the probability distribution of the parameter to be simulated.

18. "In the "conclusions" and "implications" sections, is statistical significance kept separate from economic, policy and scientific significance?"

19. "Does the paper avoid using the word "significance" in ambiguous ways, meaning "statistically significant" in one sentence and large enough to matter for policy or science" in another?"

Questions 18 and 19 echo the call that statistical significance and substantive significance should not be merged. If the assumptions of the classical linear regression model have been adhered to properly, this should not be a problem.

3.5 CONCLUSION

This chapter continued to explore the literature on econometrics in economics and its sub-disciplines. It became apparent in Chapter two that, over the years, there have been a growing number of scholars who were dissatisfied with the way which econometrics have been applied. This chapter explored the literature on this dissatisfaction. It became apparent through the review of the literature that the disenchantment could be mainly ascribed to the use of statistical significance tests in regression, the data underlying the analysis, data mining, replication, and scholasticism and preference falsification. Statistical significance has been the main focus of this chapter. Literature surveying the use of statistical significance in economics was reviewed. But, problems with the use of statistical significance tests were not only limited to economics. It was shown through a review of the literature that this problem also persisted in other social sciences, specifically in psychology. It was also shown that in economics, statistical significance tests were mostly based on the Neyman-Pearson approach, though in practise remnants of Fisher's approach were also found in the literature.

McCloskey has been vigorous in trying to persuade fellow economic scholars that the use of statistical significance in economics is all wrong. Nevertheless, this chapter also showed that statistical significance tests have their place in economics or more specifically in regression analysis in economics. It was shown that in order to adhere to the assumptions underlying the econometric approach, statistical significance tests are necessary.

Moreover, the review also included agricultural economics literature. Literature reviewing the use of econometrics in agricultural economics suggests that in agricultural economics too, significance tests have been abused, justifying an analysis to see whether the same can also be said of its use in South African agricultural economics.

The literature review pointed to problems other than the use of statistical significance tests. Closely related is the use of tests of goodness of fit in econometric analyses, a topic of discussion in the next chapter.

What remains then is to look at the use of goodness of fit tests and then to evaluate econometric analyses in South African agricultural economics. Lastly, there should at least be suggestions on how this situation could be improved. In short then, this is to be the topic of the remainder of this study.

CHAPTER 4

THE STRANGE CASE OF DR JEKYLL AND MR HYDE IN ECONOMICS AND DATA MINING

"ECONOMETRICS ANONYMOUS.
One of the major trends of the past decade has been the proliferation of redundant and useless econometric models and analysis. This new professional body has been formed to enable an economist, when he feels the urge to run multiple regressions far into the night, to telephone a fellow member of E.A. who will come over and sit up with him until the desire to regress passes." – Leonard Silk –

4.1 INTRODUCTION

In 1886 Scottish author Robert Louis Stevenson published a novel: *The Strange Case of Dr Jekyll and Mr Hyde and Other Tales of Terror*. It is about a London lawyer who investigates strange occurrences between his old friend, Dr Henry Jekyll, and the misanthropic man Edward Hyde. The work is known for the vivid portrayal of the duality in man's nature, the two aspects within man – good and evil – and the psychopathology of a split personality. In mainstream culture the very phrase "Jekyll and Hyde" has come to signify wild or polar behaviour.

As noted in previous chapters of this study, criticisms of statistical significance tests are almost as old as the methods themselves (for example, Boring, 1919; Berkson, 1938). These criticisms have been voiced in disciplines as diverse as psychology, education, wildlife science, and economics, and the frequency with which such criticisms are published is increasing (Anderson *et al*., 2000). However, it was also noted that statistical significance tests do have a place in econometrics, if applied correctly. Fair enough, if applied correctly, statistical significance tests do perform an important role in econometrics. But, as the literature reviewed in Chapter three revealed, it does not seem as if statistical significance tests have been applied correctly.

It seems, therefore, that there exists a case of Jekyll and Hyde in econometrics, the good and the evil. Researchers most certainly start off with good intentions (Dr Jekyll) to apply an econometric model to an economic problem under study, but somehow during the process, the process turns into the evil Mr Hyde, abusing the statistical significance and thereby creating an "illusion"[1]. In fact, the problem is far greater since the economic research community as a whole has become riven with an evil vice, namely the obsession with significant *t*- and *F*-values and high R-squared statistics. And as speculated in Chapter three, this vice often encapsulates the other vices of economics as researchers attempt to obtain those illusive high R-squared and significant *t*- and *F*-values. Indeed, as so many authors have noted (for example, Leamer, 1983) scholars have been using statistical significance as a method of "persuasion", citing *t*-values, *F*-values and R^2, so it seems, to ensure acceptance of their articles in journals as well as to gain respect "among the elders" and peers. Constructing an [almost] intimidating Almost Ideal Demand System (AIDS) which is statistically significant in every facet would almost certainly ensure that one's article will be accepted by the journal editor; who in his (her) captive state of baffled amazement won't even notice the lack of attention paid to the data underlying the "brilliant" model or the fact that this model is the only one which did not end up on the mine dump of computer print out after a long and strenuous data mining exercise.

As countless examples in the literature suggest, there are indeed elements of good and evil in economic research, though it seems that evil has got the upper hand. But evil did not become so well entrenched in economics through some scientific potion as was the case in Stevenson's novel. No, "Mr Hyde's" well-rooted presence is the result of a serious disease, which has caused economic researchers to become misguided and obsessed with statistical significance. It is a disease which is to be known as the R-square disease.

This chapter, then, continues the review of the vices of economics and since the disease has been christened the R-square disease, this statistic is discussed in some detail. The R^2 statistic too has become a subject of abuse in applied economics including agricultural economics (see for example, Hoch, 1984). The R^2 statistic is probably the most well known and most widely used goodness of fit measure in regression analysis. Today, R^2 statistics are computed automatically by even the most elementary statistical software program. Indeed, the use of R^2

[1] Recall the paper by Summers (1991), "The Scientific Illusion of Empirical Macroeconomics".

is so well entrenched in applied economic scholarship that it is not only invariably quoted, but quite often used as "evidence" that the reported model is a good one. Yet, as with statistical significance tests, over-reliance on this measure does not ensure a model which is economically significant nor correctly specified.

The R^2 statistic is discussed and it is shown that the over-reliance on this measure of model selection introduces what is known as a pre-test bias, which might destroy some of the properties of the model estimators. This form of application of the R^2 statistic, therefore, represents the fourth vice of economics namely data mining, the topic of discussion in the last section of this chapter.

4.2 THE MECHANICS OF R-SQUARE

R^2, or more formally the coefficient of determination, is defined as the percentage of the dependent variable variance explained by the regression function, usually a linear combination of *k* independent variables. Consulting any elementary econometric textbook reveals that R^2 is defined as:

$$R^2 = \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2} \tag{4.1}$$

In words, 4.1 is defined as the explained sum of squares (ESS) divided by the total sum of squares (TSS). The total variations of the actual Y values about their sample mean (TSS) consist of the estimated values about their mean (ESS) and the unexplained or residual variation (RSS). R^2 is alternatively defined as:

$$R^2 = 1 - \frac{\sum u_i^{\,2}}{\sum(Y_i - \bar{Y})^2} = 1 - \frac{RSS}{TSS} \tag{4.2}$$

Gujarati (1998, p. 77) notes two properties of the R^2:

- It is a nonnegative quantity

- The value of R^2 lies between 0 and 1. An R^2 of one denotes a perfect fit, that is, $\hat{Y}_i = Y_i$ for each *i*. On the other end of the scale, a R^2 of zero means that there is no relationship between the dependent variable and the regressor whatsoever. In this case, the best prediction of any *Y* value is simply its mean value. The regression line will therefore be horizontal to the *X* axis.

Another important property of R^2 is that it is a non-decreasing function of the number of explanatory variables present in the model. As the number of explanatory variables increases, R^2 almost invariably increases and never decreases. This can easily be illustrated by means of Equation 4.2. TSS is independent of the number of X variables in the model since it is simply $\sum(Y_i - \bar{Y})^2$. On the other hand, RSS, $\sum \hat{u}_i^2$, depends on the number of explanatory variables present in the model. Intuitively, it is obvious that as the number of explanatory variables increases, $\sum \hat{u}_i^2$ is likely to decrease; hence R^2 will increase. With this in mind, comparing two regression models with the same dependent variable but a differing number of explanatory variables, researchers should be cautious of choosing the model with the highest R^2.

Thus, when comparing two R^2 terms, the researcher has to consider the number of explanatory variables in the model. An alternative coefficient of determination, the adjusted R^2, can be applied which accounts for the number of explanatory variables present in the model. It is calculated as follows:

$$\bar{R}^2 = 1 - \frac{\sum \hat{u}_i^2 /(n-k)}{\sum y_i^2 /(n-1)} \tag{4.3}$$

Where k = the number of parameters in the model including the intercept. It is adjusted because it adjusts for the number of k variables entering into the model. This implies that as the number of explanatory variables increases, $\bar{R}^2$ increases less than the unadjusted R^2; furthermore, it can also be shown that $\bar{R}^2$ can be negative.

Which R^2 should researchers use in practice? Theil (1978, p. 135) notes:

> "… it is good practice to use $\bar{R}^2$ rather than R^2 because R^2 tends to give an overly optimistic picture of the fit of the regression, particularly when the number of explanatory variables is not very small compared with the number of observations."

Yet, not everyone shares Theil's view. Goldberger (1991) – Ziliak and McCloskey's (1996, p. 99) Zeus of econometric textbooks – argues that a "modified" R^2 will do just as well. He also emphasises going beyond just reporting R^2, and reporting *n* and *k* also so that the reader can decide how to adjust R^2 by allowing for *n* and *k*. Again the reader is pointed to the relation to the assumptions of the classical linear regression model. Reporting *n* would turn the focus on the number of observations on which the model is constructed (Assumption 7), which in turn would at least let the reader think about the data underlying the model. It also relates to Gujarati's advice on reporting the exact level of significance (*p*-values), thereby allowing for greater credibility and transparency, not to mention the fact that it would ease replicability of one's results.

As discussed in Chapter three, Assumption 2 states that the *X* values or explanatory variables are fixed in repeated sampling and the regression analysis is therefore conditional upon the given values of the regressors X_i. This is typically the case in economic research. However, the statistics literature makes it clear that when Y_i and X_i are random, R^2 is more appropriately interpreted as an estimate of $\Re^2$, the population coefficient of determination. This notion is discussed in Pierce (1979), Ranney and Thigpen (1981), Helland (1987), and McGuirk and Driscoll (1995). Y_i is therefore assumed to have a conditional distribution, given $X_t\{f(Y_t \mid X_t)\}$, with a conditional mean $E(Y_t \mid X_t = x_t) = \beta_0 + \beta_t' x_t$ and a conditional variance $E(Y_t \mid X_t = x_t) = \sigma^2$ (McGuirk and Driscoll, 1995, p. 320). The population coefficient, $\Re^2$ is then defined as:

$$\Re^2 = 1 - \frac{\sigma^2}{\sigma_{11}} \tag{4.4}$$

where σ_{11} is the unconditional variance of Y_t. If $X_t = x_t$ provides no useful information in terms of "explaining" Y_i, the conditional variance, σ^2, will be identical to the unconditional variance of Y_t, σ_{11}. The more information garnered by the conditioning, the smaller σ^2 will be relative to σ_{11}; consequently, the higher will be $\Re^2$ (McGuirk and Driscoll, 1995).

4.3 THE PROBLEMS WITH R^2

Now that the R^2 statistic has been defined and its mechanics explained, it is time to focus on the problems associated with it. Why should researchers be wary of the R^2 statistic? McGuirk and Driscoll (1995) argued that over-reliance on R^2 as a model selection tool is inappropriate since correctly specified values can have a "low" R^2 and incorrectly specified models often have "high" R^2 values. They illustrated the problems associated with R^2 by means of five *monte carlo* experiments. Their experiments illustrated that firstly, the R^2 associated with a correctly specified model is not necessarily large; and conversely, a large R^2 does not guarantee a correct model. Secondly they also showed that the usual estimators of $\Re^2$ (the population R^2) will be inconsistent whenever the mean of the dependent variable is not stationary. McGuirk and Driscoll (1995, p. 327) conclude:

> "We have shown that explained variation in the dependent variable may be consistently estimated by R^2. As long as the size of R^2 is not used to gauge the adequacy of specification, this measure may be used without apology."

Thus, a high R^2 does not necessarily point to a correctly specified model. Moreover, the R^2 may be inconsistent whenever the data analysed exhibit trends and other forms of non-stationarity. It is these observations that warrant a discussion of the R^2. As with the problems associated with statistical significance, this boils down to the underlying assumptions of the model. The importance of correctly specifying one's model links to the first problem with the R^2 as argued by McGuirk and Driscoll (1995). Paying attention to the data used in the construction of a regression model is emphasised by the second problem. Economic data are prone to be non-stationary since trends such as seasonality and business cycles do occur over time. Not much stays the same over time, nor does economic data. So then, what about these problems or dangers?

4.4 THE GAME OF MAXIMISING R^2

Mayer (1980) refers to "game playing" in econometrics. A common symptom of the R^2 disease is that it awakens Mr. Hyde in researchers who tricks them into playing the game of maximising $\bar{R}^2$, choosing the model yielding the highest $\bar{R}^2$. As shown by McGuirk and Driscoll (1995) the objective to obtain a high $\bar{R}^2$ is not a good one. Yet, more important, in empirical analysis it is not unusual to obtain very high $\bar{R}^2$ but find that some of the regression coefficients either are statistically insignificant or have signs that are contrary to *a priori* expectations. However, on the other hand, it might also be a useful instrument for detecting violations of the assumptions underlying the model. Indeed, one of the signs pointing to the presence of multicollinearity (Assumption 10) is that one or more coefficients is statistically insignificant, while the $\bar{R}^2$ is still very high.

But the practice of choosing a model on the basis of the highest $\bar{R}^2$ introduces a pre-test bias, which might destroy some of the properties of the ordinary least squares estimators of the classical linear regression model. This pre-test bias is discussed in detail in Judge *et al.* (1993). More relevant to this study, though, is the fact that selecting for the highest $\bar{R}^2$ has to do with another vice in economics namely, data mining.

4.5 DATA MINING

Data mining is almost never defined, only infrequently defended, much criticised but widely practiced (Mayer, 2000). Moreover, data mining is not confined to economics; it is also found in sociology, psychology and in the testing of new drugs (Morrison & Henkel, 1970 and Altman 1996). Lovell (1983) investigated, empirically, the consequences of data mining. He too held the opinion that the data miner's research strategy is usually not defined in the textbooks. However, he noted that data mining is clearly revealed by considering some typical quotations culled from leading professional journals. Lovell (1983, p. 1) cited some examples:

> "Because of space limitations, only the best of a variety of alternative models can be presented here"
>
> "The precise variables included in the regression were determined on the basis of extensive experimentation (on the same body of data). . . ."
>
> "The method of step-wise regression provides an economical way of choosing from a large set of variables . . . those which are most statistically significant..."
>
> "Since there are no firmly validated theories of the process . . . we consciously avoided *a priori* specification of the functions we wished to fit . . ."
>
> "We let the data specify the model. . . ."

However, many authors are not as candid as these cited by Lovell (1983). Evidence of "experimentation" may only become apparent once the researcher has been asked why one model was employed rather an equally plausible alternative. Applied researchers are usually quite modest in describing how industrious a search was undertaken in generating reported results; however, the criterion by which variables have been selected is usually left unspecified. In fact, work by "serious" data miners often exhibits all the makings of another Highlander movie, since "in the end there can be only one", for many instances only the final, what Leamer (1984) refers to as "a rose", is reported.

Lovell's (1983) paper examined the likely consequences of using standard regression procedures when the investigator's choice of explanatory variables is not inhibited by well-defined *a priori* considerations. His simulations considered three alternative selection criteria namely: the stepwise regression procedure, a minimum – maximum $\bar{R}^2$ selection criterion and a maximum – minimum $|t|$ criterion. The simulations revealed that of the three alternative selection criteria considered, the maximum – minimum $|t|$ is likely to uncover explanatory variables yielding the most impressive regression results, a substantially higher yield of "significant" regression coefficient with only a modest sacrifice in goodness of fit (precisely the aftermath of the disease). Unfortunately, the maximum – minimum $|t|$ criterion was a disaster in terms of correctly identifying the correct variables. Furthermore, the maximum – minimum $|t|$ is also particularly prone to Type I errors, rejecting the null hypotheses when true, 81% of the time at a claimed significance level of 0.05. (McCloskey would be jumping up and down by now shouting power, power!) Again, the relationship between the vices is apparent, especially in the presence of the disease. On the other hand, the remaining criteria proved to be more successful. The stepwise procedure was the most successful, as 70% of the selected variables that appeared in the significant actually participating in the generation of the dependent variable. The minimum – maximum $\bar{R}^2$ was almost as successful, correctly selecting 52% the variables participating in the generation of the dependent variable. However, this might be somewhat contradicting to the findings of McGuirk and Driscoll (1995).

More recently, Hoover and Perez (2000) also believed that the question of data mining needs to be tackled empirically. They repeated Lovell's simulation experiment but added a general-to-specific procedure to the three criteria used by Lovell (1983). As with Hoover and Siegler's (2005) study to refute McCloskey's crusade on significance tests, Hoover and Perez (2000) embrace data mining too as an essential activity. They found that, apart from instances where signal-to-noise ratios are very low, the general-to-specific approach works very well. They conclude that fears of those who worry about data mining are misplaced. Their study was part of a special issue of the *Journal of Economic Methodology* on data mining and included many other papers on the issue.

Mayer (2000) argues that data mining is, in a sense, a rhetorical problem and suggests that there is in principle no objection to the searching of data for the specification that fits best. Rather the problem is that when data is mined in such way that it becomes difficult to know how to interpret standard statistical tests which are based on the assumption that only a single specification has been tested. Since in most cases, only one specification is reported, readers will be given a false idea of how it was obtained and will, therefore, be given a misleading view of its significance. Mayer (2000) contends that this is the reason why econometric results are less robust than reported diagnostic statistics would suggest. In his view, the scope for data mining arises because a variety of possible specifications for any economic theory always exists. Even if investigators were committed to testing theories according to the strict canons of Neyman-Pearson testing, they would inevitably have to mine the data to a certain extent. Indeed, a great many specifications may exist for a particular problem, but would it not then be easier to reject one possible specification in favour of another? Shouldn't the researcher's objectives be based on scientific practice? After all, it was economics' desire for becoming a hard science that led to the explosion of applied econometrics in the first place.

Mayer's (2000) solution involves a distinction that follows from the way he set up the problem – between "objective" and "biased" data mining. He argued that there is no need to change econometric methods, but merely to change the way in which results are reported. In short, he takes the view that data mining does affect the significance that should be attached to statistical results but he believes that economists know how to take account of the problems it poses. The issue is one of enabling readers to reach the conclusion they would reach if they were fully informed about how results were generated, once again reiterating the necessity to ensure that results are replicable.

In another study published in the special data mining issue Pagan and Veall (2000) make three substantive points about data mining. Firstly, they observe that data mining is a problem that arises solely because of small samples. The reader is again reminded of the, by now, *cliché*, 'look at the assumptions underlying the regression model' (Assumption 7). Secondly, they argue that if samples of data were sufficiently large, it would always be possible to test models against virgin data, with the result that data mining would not be a problem. They also observe that it is important to distinguish between two stages in economic research: the estimation of statistical models and using those models to tell stories about the economic world. Data mining is generally accepted at the first stage, though it is in the second stage that worries over the use of data mining arise. This links to their third point, which involves shifting the emphasis away from the individual to the "industry" level. They argue that emphasis on the "industry" has two components. One is the imposition of suitable standards by journal editors, referees and others. The other is competition. They are of the opinion that where econometric results matter (for example, where they are important to policy) they will be replicated by researchers. They suggest that recent advances in information technology could lower entry barriers and enhance competition to reduce the possibility for what Mayer (2000) terms "biased" data mining to have any serious affect. However, the dissatisfaction with data mining has been expressed for more than 30 years, and in that period technology has been advancing at an alarming pace; and, at least by Ziliak and McCloskey's (2004a and b) standards, things haven't got any better.

Greene (2000) is a bit more explicit than Mayer, Hoover and Perez, and Pagan and Veall. He argues that, far from being a problem, data mining is the only way researchers can learn from the data. He continues by arguing that no sane economist should claim that they do not belong to a data mining discipline. Rather, the problem is that the use of statistical tests as design criteria for models has been confused with their use in testing theories, to the extent that the vitally important aspect of testing – confronting theories with new data – has become widely forgotten (the symptoms of the R^2 disease). Instead, in Greene's (2000) opinion, researchers should mine the data in order to design their models, to learn lessons from the past, and then test them by using them to predict what will be found in the new data. Greene (2000) too, focuses on the "industry" rather than on the individual researcher. Yet, he is somehow pessimistic still, that, where empirical results really matter, they will rapidly be replicated and tested. Instead, he argues that the "industry" needs to be restructured so as to place much greater emphasis on encompassing previous empirical results. In his view, encompassing of previous empirical results should become a universal activity as should encompassing of theoretical results. This again points to replication and its importance. Greene (2000) has one objection to data mining, namely that it distracts attention from the main problem, which is testing theories against new data. He argues that when faced with a choice between research at the "intensive" and "extensive" margins, economists have allocated excessive resources to the former, and not enough to the latter.

Perhaps the paper most sceptical about the notion of the R^2 disease is Spanos (2000). He believes that data are performing a double duty, of leading the investigator to a claim and then providing evidence in favour of that claim. He explains the problem of data mining by using the analogy of shooting at a blank wall and then drawing a bull's eye round the bullet hole. He argues that such practice teaches nothing about the skill of the person shooting at the wall, because the probability of the shot being in the bull's eye is one. It illustrates the illusion which the disease creates. Of course, if the researcher would just adhere to the underlying assumptions of the regression model, he or she would be able to get round this problem. As Assumption 9 states, the model should be correctly specified, thus requiring researchers to draw the target before the shooting at it, i.e. formulating the model before looking at the data.

A unifying feature of these papers is that they generally agree that data mining is undesirable and they all respond with a search for procedures that will mitigate, or even remove the undesirable effects which are believed to be associated with data mining, i.e. a call for reform of econometric practice. However, a more radical solution is to start from the observation that, as data mining is so universally undertaken, there might be something wrong with the methodology on which injunctions against it are based. To an extent, Backhouse and Morgan (2000) view the problems associated with data mining in such a potentially radical setting. They argue that the roots of the conventional attitude towards data mining seem to lie in what they call “the somewhat dated philosophy of science” that constituted the “received view” of the subject in the 1940s through to the 1960s, the time when modern econometric practices were being established (Backhouse and Morgan, 2000, p. 176). As discussed in Chapter two, it rests on a sharp distinction between theoretical statements and the observation statements against which theories are to be tested[2]. Theoretical laws could not be derived by generalising from the data but had to be formulated as hypotheses derived from a theory, and then tested against observed facts. In this setting, the logic of both discovery and justification seems to require a certain independence of theory from observation. From here, it is but a short step to injunctions against data mining (Backhouse and Morgan, 2000).

However, a more recent philosophy of science attends much more to the details of how science works in practice and, as a result, offers more measured accounts of these issues. This view goes some way towards defending the practice of data mining. Backhouse and Morgan’s (2000) paper draws on a few of the most relevant themes of this view.

A theme that has arisen in the philosophy of science literature is the importance of distinguishing between phenomena and data. Backhouse and Morgan cite Bogen and Woodward (1988) who produced detailed arguments and evidence to support the claim that science aims to establish theories that explain phenomena rather than data. Bogen and Woodward suggest that, by distinguishing carefully between phenomena and data, science is about establishing: i.) facts about phenomena; and ii.) explanations of those phenomena. Bogen and Woodward (1988) emphasise how non-observational most data are and how difficult it is to establish facts about phenomena from them.

But how does this differentiation between data and phenomena help the researcher with the question of data mining? Backhouse and Morgan (2000) note that some data-manipulation

[2] The reader could also consult Suppe (1977), and Carnap (1995) for a discussion of these developments.

procedures that are commonly associated with data mining would appear to fall well within the realm and definition of procedures designed to establish facts about phenomena. They also argue that these activities are all part of the daily work of official statistical offices in their preparation of economic data released to the public. Researchers are thus merely extending this data preparation process in order to establish facts other than those provided by the statistical offices about the phenomena that they are interested in explaining. The rest of their paper explores some of the other themes which they have highlighted. In short, all these themes suggest that econometrics should both be understood as a process of interacting with phenomena of interest in order to understand and explain those phenomena rather than primarily as "theory testing" vehicles. They also reported that detailed accounts of science practice suggest that theory testing usually proceeds via a whole research project of many experiments, involving the establishment of facts about phenomena and attempts to explain the behaviour of phenomena, rather than resting on any one single experiment. Furthermore, they argue that replication of experimental results, whereby facts become established, are no mechanical process but rely on the judgements (informed by theory) of those engaging in experimental work and involves simultaneous negotiation over facts and theories (Backhouse and Morgan, 2000, p.179).

A striking feature comes to light upon examining the paper cited above; there seems to be a strong relationship between the vices of data, data mining and replication. Fixing data mining would, so it appears, also involve fixing the other vices.

4.6 GOODNESS OF FIT BEYOND THE R-SQUARE STATISTIC

Given the problems associated with R^2 and its relation to data mining, the exclusive use of R^2 as a measure of goodness of fit is clearly not recommended. Besides R^2 and adjusted R^2, other criteria are often used to judge the adequacy of a regression model.

These include Akiake's Information Criteria (AIC), the Schwarz criterion, Ameniya's PC measure, Hocking's S_p measure, Mallow's C_p measure, the Hannan-Quinn criterion, and the Shibata criterion. A discussion of these measures is found in Maddala (1992). Nowadays prominent statistical software packages such as SHAZAM, SAS and TSP publish one or more of these statistics. The Excel based package SIMETAR also provides the AIC statistic.

These are not the only alternatives to R^2. An ad hoc remedy, more frequently recommended than applied, is to reserve a portion of the data for post-sample prediction. In fact, some authors believe that this approach is the only method to truly test for good fit. Friedman and Schwarz (1991, p. 48) notes: "… the real proof of their pudding is whether it produces a satisfactory explanation of data not used in baking it – data for subsequent or earlier years, for other countries, or for other variables." There are several measures for evaluating out-of-sample goodness of fit or forecast accuracy. These include Theil's inequality coefficients, the Mean Absolute Percentage Error (MAPE) and McLaughlin's Batting Average. Makridakis *et al.* (1998) discuss these statistics and their application in forecasting.

4.7 CONCLUSION

This chapter continued to explore the literature on the disenchantment with the way in which econometrics has been applied in economics and its sub-disciplines. The chapter focussed more specifically on the most common measure used to analyse goodness of fit, the R^2 and the related $\bar{R}^2$ as well as data mining. The 'incorrect' application of this statistic together with statistical significance testing has been christened the R-square disease of economics. It is a disease because the literature has shown that incorrect application of statistical significance tests and goodness of fit tests has been misleading, thereby creating a strange case of Dr Jekyll and Mr Hyde in economics. Countless accounts of examples in the literature exist where these analytics have been applied without regarding the underlying assumptions of the regression approach used by the researcher. Instead, or so the literature suggests, these analytics are used as tools of persuasion because the economic research environment has become more interested in significant statistics and high R^2 statistics. These analytics have become criteria for acceptance among one's peers and journal editors; and it this role that is the disease, since it misleads economic research away from its true objectives.

The relationship between the disease and data mining has also been discussed. A review of the literature on data mining revealed that the general attitude of researchers is that data mining is undesirable but that the private practice of econometrics is clearly riven with it. The common response among the papers reviewed in this chapter is to search for procedures that will mitigate, or even remove the undesirable effects that are believed to be associated with data mining. Some of the reviewed papers also held the view that if the "industry" were to be reformed or restructured, the undesirable effects of data mining could be kept under better control. A paper by Backhouse and Morgan (2000) also explored the relation between data mining and the philosophy of economic science; suggesting that recent developments in the philosophy of science might have a direct bearing on econometric practice.

Whatever the argument for and against data mining, the rationale behind the use of these analytics should be the real source of concern, the effects of the disease so to speak. The problems associated with statistical significance tests and data mining would be mitigated if they were to be used within the correct context, i.e. supporting the underlying assumption of the regression model. The disease of significant *t*- and *F*-values and a high R^2 should not be the objective of an econometric research study.

Again, the linkages between the different vices were highlighted. Eradicating the disease by conducting studies within the correct context and motives would also aid in eradicating the vices of econometrics. The other issue that would ensure eradication of the disease is that of restructuring the "industry"; a topic to be explored in a later chapter.

What remains, then is to explore the agricultural economics literature in South Africa in order to see if elements of the disease can be found. Chapter five investigates the agricultural economics literature in South Africa. Apart from looking at the industry and its role in the persistence of the disease, the data underlying one's analysis also requires a more detailed investigation.

CHAPTER 5

ANALYSING FOLKLORE AND FACT IN AGRICULTURAL ECONOMIC SCHOLARSHIP IN SOUTH AFRICA

5.1 INTRODUCTION

As noted in the previous chapter, what remains is to investigate if elements of the R-square disease exist in South African agricultural economics. This study has so far only looked at the disenchantment with applied econometrics in economics in its broader sense. From this chapter onwards, the focus will turn to agricultural economics, specifically agricultural economics in South Africa. This chapter, therefore, unifies the problem which had been identified in Chapter one with the discussions in Chapters two to four.

Conclusions drawn from the previous chapters suggest that the disease has spread throughout economics and its sub-disciplines. Indeed, opinions of authors cited in Chapter three also suggest that elements of the disease are prevalent in agricultural economics (see for example Hoch 1984 and McCloskey 1990). Can the same be said for agricultural economics in South Africa or would such a notion be pure folklore?

The first objective of this chapter is to explore the literature applying econometrics in South African agriculture in order to find facts substantiating the opinions of authors who suggest that elements of the disease do exist in South African agricultural economics. To date no empirical studies have been undertaken exploring the disenchantment with applied econometrics in South African agricultural economics. This chapter explores articles published in *Agrekon* between 1962 and 2005 and sets the stage for the formal analysis of published articles in Chapter six.

The second objective of this chapter is concerned with the data underlying econometric analysis, raised in many of the articles cited in the previous chapters. This chapter explores agricultural statistics in South Africa, specifically looking at the discrepancies in agricultural statistics reported by different official sources

5.2 DISENCHANTMENT IN A SOUTH AFRICAN SETTING

South African agricultural economists, too have expressed their concern over the manner in which econometrics has been applied in agricultural economics. Of note are the concerns expressed by Groenewald (1990), Nieuwoudt (1992) and Kirsten (2002). Groenewald (1990) evaluated the effectiveness and efficiency of agricultural economists as experts. He wrote about a predilection of some agricultural economists toward tool-orientated work; arguing that South African research has largely chosen research methodologies or tools and sought problems to solve thereby. He added that faddism has repeatedly plagued agricultural economists in their choice of analytical tools. Kirsten (2002) supported this argument and writes that, to some extent, agricultural economics in South Africa always had the ‘luxury’ of being able to pursue problem solving and applied research. This preoccupation with problem solving, he argued, has taken South African researchers’ time away from the “frontier-pushing” research and theoretical work of their European and American colleagues. In this sense agricultural economics in South Africa has often borrowed from these scholars, and applied their models and methodologies to local problems. Kirsten (2002, p. 256 – 257) noted:

> “The influence from scholars from abroad has also highlighted what many of us perceive as limited application of quantitative skills in our discipline in South Africa. The examples set by leading journals such as the *American Journal of Agricultural Economics*, *Journal of Agricultural Economics*, *Agricultural Economics*, *World Development* and *Economic Development and Cultural Change* have put agricultural economists in South Africa in pursuit of more quantitative and perceivably more ‘rigorous’ output. It is also a function of the fact that after the process of democratisation normal relationships with leading universities and scholars abroad have become possible and has put many agricultural economists in touch with the latest theories and quantitative methods. As a result, there has been a continued jockeying for position in terms of the institution or researcher with the ‘best’ analytical tools or models.”

However, it is the third author who perhaps has grasped the idea of the disease best. Nieuwoudt (1992, p. 45) too, expressed concern for what he called “preoccupation with technology”. In his opinion researchers tend to concentrate on specific techniques which become and remain fashionable for some time. Nieuwoudt (1992) went further than just identifying in South African agricultural economics the common problems which have been discussed in such detail thus far. He argues that the preoccupation with technology has led to users often not fully understanding implicit assumptions of techniques. Unlike McCloskey, Nieuwoudt (1992) understands the importance of not treating these assumptions as merely "things equal" when reviewing the disenchantment with econometrics.

Indeed these authors have, from different angles, identified one of the vices which have not yet been discussed namely, the preoccupation with techniques of "black box econometrics". It is black box econometrics in a sense because this form of the disease leads the researcher to pursue an econometric technique no matter what, no matter if such a technique is suitable for the problem at hand nor if it suits the data. Many authors have expressed their dismay over this symptom of the disease. Mayer (1980, p. 177) whom this study has cited so many times before, views much of the published economic research as taking a new technique out for a walk rather than solving a problem. Groenewald (1990, p. 246) contends that the tail has often wagged the dog and that this form of inefficiency stems from a mental immaturity and is often a symptom of a desire to gain peer adoration irrespective of whether the analysis aids in understanding any problem whatsoever. This poses the question: Why has this phenomenon persisted for so long? Butz (1989) commented that the preoccupation with technology has caused journals to become more important to writers than to readers, while members cannot and do not read them. Nieuwoudt (1992, p. 45) also commented on this notion by noting that promotion in academic departments depends on the length of publication in a candidate's CV "It is only natural to concentrate effort where gain can be maximised. This preoccupation with technology leads to the assertion that the technique was looking for a problem to be solved."

Again, it points to the disease polluting the researcher in applying technique and moreover statistical significance testing to persuade rather than to be scientific. This problem is precisely what the last vice is all about, namely the vice of scholasticism and preference falsification. The idea of scholasticism and preference falsification in economics is a vast and involved topic reaching beyond the scope of this study. Yet, relevant to this study is that to its followers, this view has produced many answers as to why the problems associated with applied econometrics have persisted for so long. There are quite a number of authors who have discussed this view in great detail. Altman (2004) writes about path dependency by arguing that the inappropriate socially sub-optimal and inefficient practice of econometrics is path dependent and represents a market failure often resulting in misleading research findings and misguided public policy. The idea of scholasticism has also led some authors to view the economic profession as a group of "clubs" or "villages". Klein (2004, p. 145) views economic academia as "a self-organizing, self-validating club". He argues that people with degrees from top departments hold most of the academic positions and publish most of the top-journal articles mainly because the top departments attract the best students and do the best job at training them. Klein (2004, p. 140) notes:

> "Graduate education is a formative period for a young economist. He or she learns directly from professors and learns to emulate them, in order to get a PhD degree. He or she depends on them for resources and for entrée in the journals and job market. He or she continues to depend on them throughout his or her career. The dependence resides in a rich nexus of relationships, exemplified by the letter of recommendation. Academics is an ongoing and circular system of validations, and even full professors need validation to obtain publication contracts, prestige awards, large grants, and jobs at universities further up the pyramid. Thus, the ties and imprint of graduate training is lasting. One's PhD degree is a marker of the sub-group that raised you. It is a way of drawing the social connections between members of the encompassing society of Economics: "Oh you got your degree at Columbia, so you must know . . ."

The view of economics being a tribe has been best portrayed by Leijonhufvud's 1973 paper entitled: "*Life among the Econ*". An extract illustrates (Lejonhufvud, 1973, p. 327, emphasis added):

> "The Econ tribe occupies a vast territory in the far North. Their land appears bleak and dismal to the outsider, and travelling through it makes for rough sledding; but the Econ, through a long period of adaptation, have learned to wrest a living of sorts from it. They are not without some genuine and sometimes even fierce attachment to their ancestral grounds, and their young are brought up to feel contempt for the softer living in the warmer lands of their neighbours, such as the *Polscis* and the *Sociogs*. Despite a common genetical heritage, relations with these tribes are strained – the distrust and contempt that the average Econ feels for these neighbours being heartily reciprocated by the latter – and social intercourse with them is inhibited by numerous taboos. The extreme clannishness, not to say xenophobia, of the Econ makes life among them difficult and perhaps even somewhat dangerous for the outsider. This probably accounts

> for the fact that the Econ have so far not been systematically studied. Information about their social structure and ways of life is fragmentary and not well validated. More research on this interesting tribe is badly needed."

Apart from entertaining reading, Lejonhufvud's (1973, p. 330) satirical portrayal of economic scholarship illustrates the key as to why problems with econometrics have persisted for so long.

> "The young Econ or "grad," is not admitted to adulthood until he has made a "modl" exhibiting a degree of workmanship acceptable to the elders of the "dept" in which he serves his apprenticeship. Adulthood is conferred in an intricate ceremony the particulars of which vary from village to village. In the more important villages, furthermore, (the practice in some outlying villages is unclear) the young adult must continue to demonstrate his ability at manufacturing these artefacts. If he fails to do so, he is turned out of the "dept" to perish in the wilderness.
>
> This practice may seem heartless, but the Econ regard it as a manhood right sanctioned by tradition and defend it as vital to the strength and welfare of the dept. If life is hard on the young, the Econ show their compassion in the way that they take care of the elderly. Once elected an elder, the member may need to do nothing and will still be well taken care of."

Ambition and, to some extent, survival urge scholars to conform, and this has been the main reason for the persistence of the obsession with rigorous output. In the end, the situation is just as Mayer (1980, p. 173) illustrates: "… my own attitude towards econometrics is like that of a person who upon being told that the craps game he was about to participate in is crooked; replied, "Sure, I know that, but it is the only game in town."'

The vices of preoccupation with techniques and scholasticism are not the only sources of disenchantment perceived by Groenewald (1990), Nieuwoudt (1992) and Kirsten (2002) to exist in South African agricultural economic scholarship. As discussed by so many critics in economics as well as critics of agricultural economic scholarship abroad, they too express concern over the manner in which data has been treated in South African agricultural economic scholarship. Groenewald (1990, p. 246) argues that the efficiency of agricultural economists in the US and most certainly in South Africa has been 'significantly' eroded by a cavalier approach to data. He went further noting that "the mental or academic snobbery related to elegant, refined statistical or mathematical models has also led many agricultural economists astray, and has yielded a false aura of excellence around refined manipulation of third-rate data. Too many have forgotten of the "Garbage In – Garbage out" adage." Kirsten (2002) notes that availability of good quality, relevant and timely data has always been a problem, even more so in the aftermath of deregulation. Groenewald (1990) also notes the other vice, data mining, and cites Tweeten (1983) who noted that this reverses the scientific method of using statistical analysis to determine the hypotheses.

5.3 ASSESSING FORTY FIVE YEARS OF AGRICULTURAL ECONOMIC SCHOLARSHIP IN SOUTH AFRICA

South Africa's agricultural economics journal, *Agrekon*, of which the first issue was published in 1962, is now in its 45th year. Much has happened during this period and much has changed. Political, economic and agricultural conditions have changed drastically in South Africa as well as around the globe.

By the nature of things, the journal also had to undergo considerable changes. It is especially the changes in content with which this study is concerned. But first, it is perhaps fitting to briefly sketch the journal's history and its development.

Wissing and Groenewald (1987) reviewed the first twenty five years of *Agrekon*. They note that agricultural economic research and agricultural economic services in South Africa can largely be traced to the foundation of the Division of Economics and Markets of the Department of Agriculture in 1925. This Division, later renamed the Division of Agricultural Economic Research, and subsequently divided into the Directorates of Agricultural Production Economics and Agricultural Economic Trends, was for a long time the most important source of economic research and information concerning South African agriculture. It also served as a practice-school where many young graduates gained their first research experience (Wissing and Groenewald, 1987, p. 1). Today, much has changed; the Directorates have again undergone a name change and are now known as the Directorates Production and Research Economics and Agricultural Statistics respectively. Their role has also changed;

from being an important source of economic research and information to mostly an important source of information on agricultural economic phenomena.

The name *Agrekon* and the late Mr. S.J.J. De Swart, a previous chief of the Division of Economics and Markets of the Department of Agriculture and later Secretary of Agricultural Economics and Marketing, are synonymous. During the nineteen fifties it was particularly De Swart who contrived the establishment of an official journal for agricultural economic matters. The result was the publication of the first volume of *Agrekon* in 1962.

The mission of the journal originally comprised the dissemination of information, while the reporting of independent research probably took second place in terms of priority. However this turned around to the point where today *Agrekon* is completely a research journal.

Agrekon was (from 1962 to 1989) published by the Division of Economics and Markets of the Department of Agriculture, later known as the Division of Agricultural Economic Research. However, this period saw an inexplicable decline in interest in the journal in the late nineteen seventies and early eighties. The number of articles presented during this period declined and it began to appear as if, in a sense, both the agricultural economics profession and the journal had entered a period of stagnation. During this period the number of issues per year was also reduced from four to three in 1981 and finally to two in 1982. Fortunately, a gradual recovery started in 1984. Wissing and Groenewald (1987) noted that the recovery was both in quantity as well as the quality of articles submitted. The process of recovery received an impetus when, in 1985, the Department of Education decided to recognise *Agrekon* as a scientific journal for university subsidisation purposes. *Agrekon* has had a number of editors since its foundation:

- Vol. 1 No. 1 to Vol. 4 No. 3: Messrs H.J. Van Rensburg and O.E. Burger.
- Vol. 4 No. 4 to Vol. 9 No. 1: Dr A.J. Beyleveld
- Vol. 9 No. 2 to Vol. 25 No. 3: Mr G.J. Wissing
- Vol. 26 No. 1 to Vol. 28 No.3: Miss A.M. Visagie

- Vol. 29 No 1 to Vol. 31 No. 4: Prof J. Van Zyl and Prof N. Vink
- Vol. 32 No 1 to Vol. 36 No. 4: Prof. J. A. Groenewald
- Vol. 37 No. 1 to Vol. 40 No. 4: Prof. T.I. Féynes
- Vol. 41 No. 1 to present: Prof. N Vink

As noted earlier, there has been a movement from a partly information dissemination and partly research journal to completely a research journal. Apart from recognition as a scientific journal, it can also be argued that the change in the responsibility for publishing the journal could also have had an influence in this movement. The Agricultural Economics Association of South Africa (AEASA) took over the responsibility for publishing the journal in 1990. The Department of Agriculture was however responsible for the financing of the journal for the first three years.

The institutional changes within the journal are of great importance, but it is the changes in content with which this study is predominantly concerned, searching for traces of the disease in agricultural economics. These changes are perhaps more clearly explained by means of a time line. As seen from Figure 5.1, the number of articles published in *Agrekon* has varied substantially over the period 1962 to 2005.

There are specifically four conspicuous years namely: 1982, 1990, 1999 and 2002. As discussed earlier, 1982 was the year in which the least number of articles were published. The number of articles published in the period subsequent to 1982 started to recover until it reached a peak in 1990; when the responsibility for publishing the journal was taken over by the Agricultural Economics Association of South Africa. After South Africa became a democracy in 1994, the number of articles published reached another peak in 1999. This was the result of the release of a special issue which subsequently led to a greater than usual amount of articles published that year. After 1999, the number of articles again started to decline until reaching a turning point in 2002. In 2002, it was decided to exclude papers presented at the South African Agricultural Economics Association's conferences.

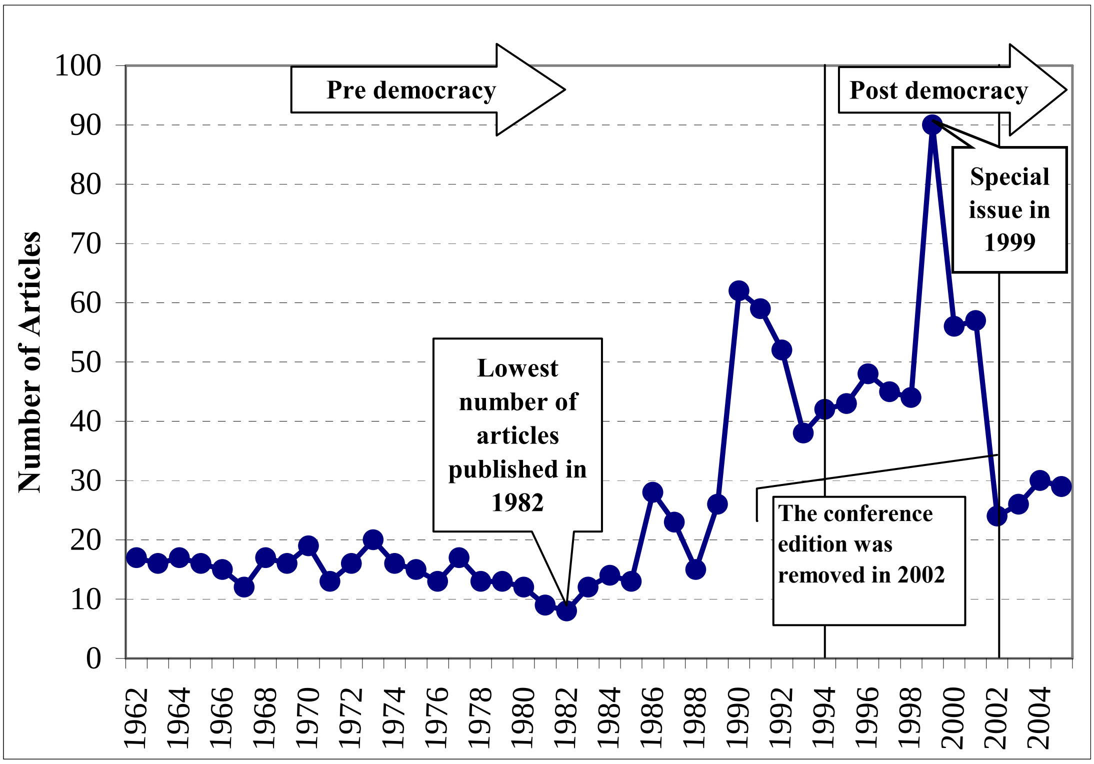

Figure 5.1: Number of articles published in Agrekon from 1962 to 2005

Source: Agrekon Volumes 1 to 44

These years were tested for breakpoints by means of the Chow breakpoint test (Gujarati, 1998, p.265 – 263). The results of the Chow breakpoint test are shown in Table 5.1. The PROC AUTOREG procedure in SAS version 9.1.3 was used to test for structural breaks.

Table 5.1: Results of the Chow test for a structural break

Break point	F Value	P value
1982	2.84	0.070
1990	22.2	<0.0001
1999	11.58	0.000
2002	3.43	0.042

Source: Own calculations using SAS version 9.1.3

The Chow test is used to test the null hypothesis of structural stability against the alternative hypothesis of no structural stability. In this case, H_0: The period between 1962 and 2002 was structurally stable. The alternative hypothesis (H_1): The period between 1962 and 2002 was not structurally stable.

Table 5.1 indicates that the *p*-values for all four periods are all at least less than 0.10 and thus statistically significant at the 90% level of confidence. H_0 is therefore rejected, and the number of articles published in the five periods is therefore statistically different.

However, of interest is the content. What happened to the content during this period? The period under review (between 1962 and 2005) saw a total number of 1 186 articles published in *Agrekon*. As with Ziliak and McCloskey's studies (1996 and 2004a), an attempt was made to identify the number of articles containing some form of regression analysis. Though subjective, it does provide a view of how econometrics has been applied in South African agricultural economics. The number of articles containing some form of regression or mathematical programming models published in volumes 1 to 44 totalled 377 or 32% of the total number of articles published. Examples of techniques such as Discriminant Analysis, Principal Component Analysis, Chi-square analysis, Analysis of Variance (ANOVA), and Multi Analysis of Variance (MANOVA) were also among these articles.

5.3.1 Scholasticism and path dependency

The analysis of these 377 articles starts off by first looking at the last vice of economics: scholasticism. As noted earlier, while it is not possible (within the scope of this study) to assess whether the form of scholasticism has been detrimental to South African agricultural scholarship, the results do show that scholasticism did and still exists in *Agrekon*. The first article employing simple regression, in which coefficients were subjected to statistical testing, only appeared in the third volume of publication, Groenewald's (1964) article. The ensuing years saw similar techniques being used, but only sporadically. The first article applying multiple regression analysis appeared two years later. Nieuwoudt and Döckel's (1966) article also contained the first empirical production function to be published in *Agrekon*. The first econometric demand study followed three years later in Vosloo and Groenewald (1969). The proportion of articles containing some form of regression analysis or mathematical programming is illustrated in Figure 5.2.

These results demonstrate a couple of interesting points on the issue of scholasticism, especially in terms of what Klein (2005) has noted. As cited earlier, Klein (2004, p. 140) notes:

> “One’s PhD degree is a marker of the sub-group that raised you. It is a way of drawing the social connections between members of the encompassing society of Economics: “Oh you got your degree at Columbia, so you must know . . .”

A closely related argument is found in Altman (2004, p. 657), who writes about cultural embeddedness, referring to the social context in which decision making is embedded. Thus, if an individual finds her or himself in a network wherein behaving in one fashion is most highly regarded (utility augmenting) or behaving alternatively is negatively viewed (utility reducing) the individual will tend to choose the former course of action.

The results are perhaps a good example of scholasticism, path dependency and cultural embeddedness. It indicates that Groenewald and Nieuwoudt have to a degree “introduced” regression analysis to *Agrekon*. The two scholars could be regarded as Leijonhufvud’s “*elders*” in “Econoland.”. Yet in this case, “Econoland” is much further away from Leijonhufvud’s Econland. Gazing upon the literature published in *Agrekon*, it appears that they have determined the path for their protégés. A common approach of investigating scholasticism is to study the papers in terms of the authors who published these papers in a journal. Examples of authors applying this approach are found in Klein (2005), Coupé (2000 a and b, 2001 and 2003 a and b) and Lazear (2000). Table 5.2 contains the number of articles published, respectively, by Groenewald and Nieuwoudt in *Agrekon* from volumes 1 to 44.

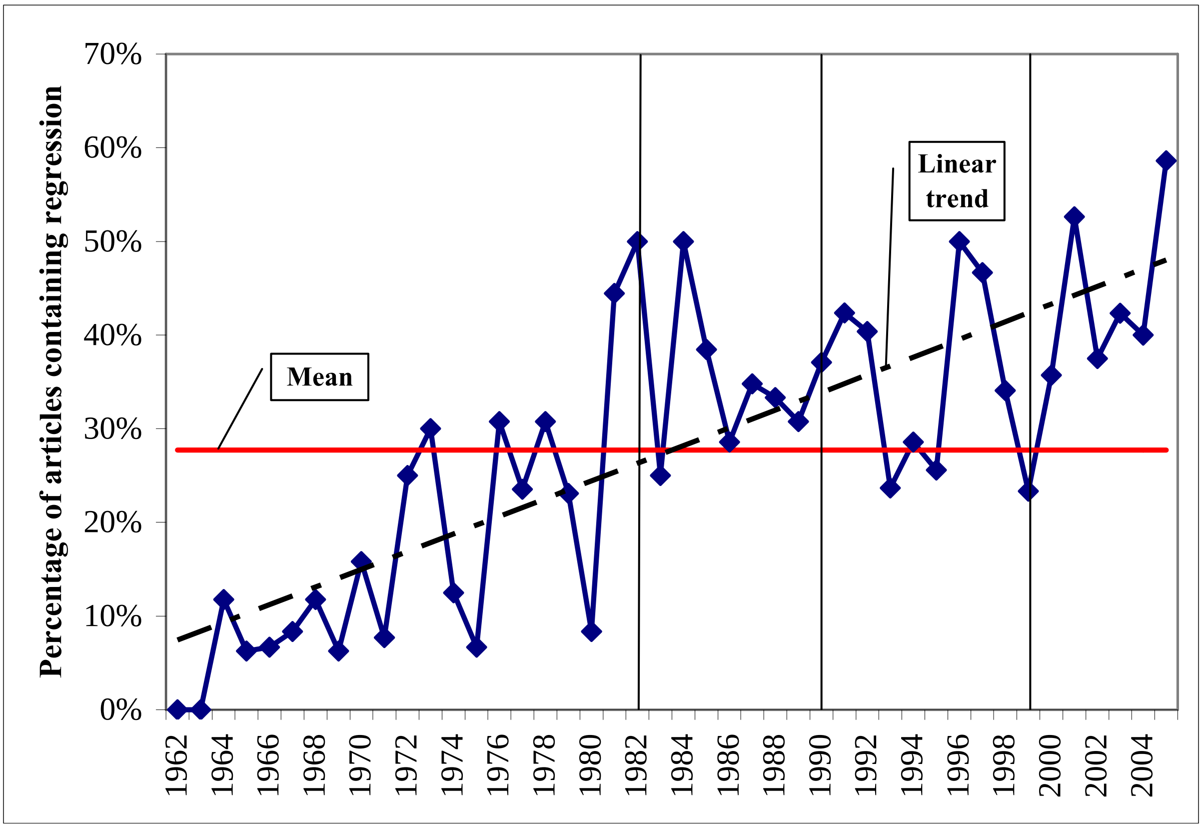

Figure 5.2: Percentage of articles containing regression analysis or linear/non-linear programming

Source: Agrekon Volumes 1 to 44

Table 5.2: Number of articles published by Groenewald and Nieuwoudt, 1962 – 2005

	Groenewald	Nieuwoudt	Total
Number of individually authored articles	23	24	47
Number of co-authored articles	88	51	139
Total	**111**	**75**	**186**

Source: Calculated from articles published in Agrekon Volumes 1 to 44.

As seen from Table 5.2, between the two authors they have been responsible for the publication of 186 articles. Table 5.3 contains the number of articles containing regression analysis or mathematical programming and shows that regression and mathematical programming have been applied in half of their articles published in A*grekon*.

Table 5.3: Number of articles containing regression or mathematical programming published by Groenewald and Nieuwoudt, 1962 – 2005

	Groenewald	Nieuwoudt	Total
Number of individually authored articles	3	11	14
Number of co-authored articles	39	40	79
Total	**42**	**51**	**93**

Source: Calculated from articles published in Agrekon Volumes 1 to 44.

5.4 DATA PROBLEMS IN AGRICULTURAL ECONOMIC RESEARCH IN SOUTH AFRICA

The focus turns yet again to the data underlying the analysis. Problems with data have been stressed throughout this study. References have been cited where authors have expressed their concern with the way in which data underlying econometric studies have been handled. The situation is also not any different in agricultural economics. Indeed, Gardner (1983, p. 887, emphasis added) stressed the point, noting in his primary conclusion: "I believe we are at the stage where *our lack of attention to the ammunition we use in our sophisticated analytical artillery* is an important impediment to the progress of agricultural economics as a science; and the silliness of our penchant to drop any bullet-shaped item down the barrel and see what happens when we pull the trigger is becoming dangerous."

Indeed, agricultural statistics has been under a lot of pressure in the United States, especially during the early 1980s when reduced budgetary allocations to federal statistics agencies saw a decline in the available resources to produce statistics on important agricultural phenomena (Gardner, 1983). The changes in United States federal statistics led to a fair amount of criticism. Bonnen (1983) asked the question: "Federal Statistical Coordination Today: A Disaster or Disgrace?" Others viewed the problem from a different angle. Just (1983) focussed on the public versus private good aspect of information, while Bullock (1976) measured "The Social Costs Caused by Errors in Agricultural Production Forecasts". On the same note Gunnelson, Dobson and Pamperin (1972) analysed the accuracy of USDA crop forecasts, while Sumner and Mueller (1989) focussed on whether harvest forecasts are news.

Has the situation been any different in the South African context? As was the case with regression in *Agrekon*, Nieuwoudt (1973) was one of the first authors to take note of this problem. Addressing fellow scholars at the Agricultural Economics Association of South Africa's conference in 1973, Nieuwoudt addressed the data problems in agricultural economics research. His paper discussed the problems associated with the collection and interpretation of data, the use of agricultural censuses, structural changes resulting in quality changes as well as the measuring and interpretation of economic data. Groenewald (1990) argues that the efficiency of agricultural economics in South Africa has been substantially eroded by a cavalier approach to data.

The South African agricultural sector has undergone significant changes since 1973, especially in terms of the availability of agricultural data. The 1980s and especially 1990s saw extensive deregulation of the agricultural marketing sector, with the abolition of the marketing boards. One of the most important functions of the past regulated agricultural market was an excellent data collection system. In fact, the South African controlled environment was useful for information gathering (Kirsten and Van Zyl, 1996). This situation changed dramatically since the advent of deregulation. In some instances, data collection was taken over by producer organisations, for example, organisations such as the Milk Producer's Organisation. The garbage in, garbage out (GIGO) principle comes to mind here since this is the consequence of unreliable data.

The garbage in, garbage out principle is precisely why it is so important to analyse data before feeding it into a model. Gujarati (1998, p. 467 – 472) discussed the consequences of errors of measurement in data. In fact, errors of measurement have different effects on dependent variables and explanatory variables. Gujarati notes, "although the errors of measurement in the dependent variable still give unbiased estimates of the parameters and their variance, the estimated variances are now larger than in the case where there are no such errors of measurement." The OLS estimators are therefore no longer best estimators since they do not have minimum variance. The case of errors of measurement in the explanatory variables is somewhat more severe, since the OLS estimators are not only biased but also inconsistent, that is, they remain biased even if the sample size n increases indefinitely. Gujarati notes further that even if errors of measurement are detected or suspected, the remedies are often not easy. Although, theoretically, the use of instrumental variables is attractive, it is not always practical. Again, it illustrates the crucial importance of paying much needed attention to data before attempting any econometric analysis. Moreover, it is very important in practice that the researcher be careful when referencing his or her data to elaborate on how the data was collected and what definitions were used.

Has there been any reason to suspect that South African agricultural data are, in fact, prone to errors or is it mere folklore? Groenewald (1989) expressed his concern over the reliability of agricultural statistics. His study compared statistics on the primary production of fresh produce in South Africa from two official sources, namely the agricultural census reports as compiled by the Central Statistics Service (now Statistics South Africa (STATSSA)) and the data from the Directorate: Agricultural Economic Trends (now Directorate: Agricultural Statistics). Groenewald (1989, p. 91) concluded:

> "The inevitable conclusion of the above-mentioned is that there is reason for serious concern over the quality of data in respect of fresh produce. It would also benefit researchers in other fields of agriculture first to investigate the quality of data available to them. This problem is also not new, it existed already two decades ago (Brand, 1969)"

A decade later, the situation does not seem to be any different. Analysing the data on production of fresh produce as published in the *Abstract of Agricultural Statistics* highlights an earlier comment made on reporting footnotes supplied under tables containing statistics on agricultural phenomena. Revisiting statistics on fresh produce published in the *Abstract of Agricultural Statistics* and the *Census of Commercial Agriculture* proved to be a daunting task precisely because of footnotes and definitions or, in this case, the lack there of. The latest available agricultural census information, the *Census of Commercial Agriculture 2002*, only contains statistics from the commercial agricultural sector. However, statistics on fresh produce as reflected in Tables 34 to 57 of the *Abstract of Agricultural Statistic* contains total production. The footnote for the "Total production" column only states the length of the production year. Stating whether total production includes commercial as well as non-commercial production is omitted. This renders the notion of "*comparing apples with apples*" very important and any comparison of the data from the two sources dangerous.

Data on other primary products also seem to be plagued by problems. In particular, statistics on livestock numbers in South Africa. Livestock statistics are also inundated by problems, especially discrepancies in livestock numbers from the different sources within the Department of Agriculture and Statistics South Africa.

5.4.1 Comparing livestock estimates from three official sources

Livestock estimates by the Directorate: Agricultural Statistics

The Directorate: Agricultural Statistics currently estimates livestock numbers for cattle, sheep, pigs and goats. The estimates are conducted on a quarterly basis for the quarters ending in February, May, August and November. The currently applied methodology was initially developed to estimate the trend in livestock numbers and not the level of livestock numbers.

The base for this trend approach is Statistics South Africa's 1983 Agricultural census. The livestock numbers obtained from the census are, therefore, used as basis. The August quarter is used as base for each year and the change in livestock numbers from one year to the next is represented by the change in livestock numbers from August to the following August. These changes represent the trend, which forms the crux of the methodology. The question remains: If the 1983 Census is the base for this approach, how is the annual change in livestock numbers from one year to the next calculated?

This forms the second part of the methodology. A sample of approximately 4 400 commercial livestock producers is used for calculating the change. Mail surveys are conducted each quarter for February, May, August and November. Respondents are asked to complete a questionnaire, stating the number of cattle, sheep, pigs and goats they own at the end of each quarter. Again, August is used as the base quarter. The change in the sample's livestock numbers is then calculated by means of establishing the percentage change in livestock numbers from the previous August to the current quarter. For example: For the February 2006 survey; the researcher would calculate the percentage change in livestock numbers from August 2005 to February 2006, etc. The percentage change from August 2005 to August 2006 would, consequently, reflect the annual change in livestock numbers from 2005 to 2006. This change in the sample's livestock numbers is then applied to the base estimate in order to obtain an estimate of the total number of livestock in the country. The same procedure is also applied in deriving estimates for each quarter. Consequently, for February 2006 the reseacher would multiply the August 2005 estimate by the percentage change in the sample's numbers from August 2005 to February 2006. The same method would also be followed to obtain estimates for May 2006 and November 2006. The only difference is that the November 2006 estimate would be obtained by multiplying the August 2006 estimate by the sample's % change in livestock from August 2006 to November 2006.

However, this approach depends on two crucial factors, namely: the assumption that the 1983 census was complete and accurate and that the sample used to calculate the percentage change is representative of the total livestock population in South Africa. Another problem associated with this method is the issue of benchmarking because the only time the entire livestock population was surveyed, was in 1983. Any errors made in estimates for years after 1983 would therefore have escalated over the years.

Rebasing the results from the 1988 Agricultural Census was considered, but it was soon apparent that livestock numbers obtained from the 1988 census were unreliable. Comparisons revealed that the 1988 Census figures were much lower than the Directorate's estimated livestock figures for 1988. That could have been because of the fact that:

- The 1983 census figures were too high,
- The 1988 census figures were too low,
- The estimated trends of the directorate were too steep,
- Or a combination of the possibilities mentioned.

An analysis of the 1988 Census data by the directorate revealed many inaccuracies. Furthermore, during discussions between the directorate and the Meat Board, it seemed as if the reproduction rate of cattle, for example, would be too high should the 1988 Census be accepted. The Meat Board stated that the trends indicated by the directorate's estimates, in general corresponded with cycles in the livestock industry and that they were willing to use these figures until the matter has been resolved (Directorate Agricultural Statistics). Since deregulation, similar analyses have not been undertaken, nor have there been discussions with the industry over the issue of the accuracy of livestock numbers.

Another important element to consider when analysing the directorate's livestock estimates, is the inclusion (or exclusion) of livestock numbers from the developing agricultural sector and the former homelands. In the past, South Africa was not allowed to undertake surveys in the former homelands and the directorate therefore excluded those areas from its statistics. However, information on livestock numbers was received from agricultural organisations in those areas up until 1992. These figures were included in the RSA figures and were adjusted according to the estimated changes in the total livestock numbers of commercial farmers in the province in which the relevant former homeland was situated. In 2003, the directorate also looked at livestock numbers as estimated by the Directorate: Animal Health, in order to establish a base for livestock numbers from the developing agricultural sector and the former homelands. Yet, adjusting livestock numbers from the former homelands according to estimated changes in livestock numbers from the commercial sector might not be very accurate. Changes in livestock numbers from these two sectors are not necessarily affected by the same factors and, therefore, might not change in the same order of magnitude or direction. Applying the same magnitude of change to both sectors would, consequently, not produce estimates of a high level of accuracy. However, because no benchmark figures are available, measuring the accuracy of these estimates would be almost impossible.

Livestock numbers are estimated on a national as well as a provincial level and include information on the number of cattle, sheep, pigs and goats. At present, the directorate does not have any information available on the number of horses, ostriches or poultry.

Livestock numbers by the Directorate: Animal Health

The Directorate: Animal Health is the Department of Agriculture's second source of statistics on livestock numbers in South Africa. Animal numbers are compiled by the provincial state veterinary offices in South Africa. These animal numbers include information on the number of cattle, sheep, pigs, goats, horses, donkeys and mules, ostriches, poultry, dogs and cats. Animal numbers are collected on an annual basis. The Directorate: Animal Health compiles an annual report containing the total livestock numbers per province. Livestock numbers are also published per magisterial district and further broken down into commercial livestock numbers as well as livestock numbers from the communal sector (in essence the former homelands).

Unfortunately, the methodology applied in compiling these estimates is not well documented and it seems as if every provincial office applies its own methodology. This creates a serious problem concerning consistency of methods among the nine provinces. Furthermore, a personal analysis of the historical data also revealed that there seemed to be inconsistency in terms of the methodology applied over time. For example, animal numbers for Gauteng province included animal numbers from communal areas in 2004, but these numbers were excluded in 2003 and included in 2002. Another example is where, in some provinces, horses and donkeys are grouped together in one year and split in the following year. These problems render comparisons over time almost impossible.

Livestock numbers from the Census of Agriculture, 2002

The 2002 Census of Commercial Agriculture also included livestock numbers. However, it differs from censuses conducted in the past because the population of the 2002 Census of Commercial Agriculture differed from populations targeted by past censuses. The 2002 census covered the entire country, including the former TVBC states (Transkei, Venda, Bophuthatswana and Ciskei) and self-governing territories. Moreover the census was also based on a new business register containing all businesses registered for Value Added Tax (VAT) with the South African Revenue Service (SARS).

All enterprises are legally compelled to register for VAT when their turnover for a period of 12 months equals or exceeds R300 000. However, those with a turnover of less than R300 000 may register for VAT voluntarily. From the commercial farming units registered for VAT, a total of 45 818 were identified as live and active at the time of the census and formed the population for the census.

Comparison of livestock numbers

Livestock numbers as estimated by the three official sources are contained in the tables below. Table 5.4 contains estimated livestock numbers by the Directorate: Agricultural Statistics and the Directorate: Animal Health. Total animal numbers in South Africa for 2002 are grouped according to cattle, sheep, pigs and goats. Estimated animal numbers from the developing sector are also included in the totals. The last column of Table 5.4 represents the differences between the two sources. The Agricultural Statistics value is subtracted each time by the Animal Health value. As seen from Table 5.4, apart from estimated pig numbers, estimates by the Directorate: Agricultural Statistics seem to much higher compare to those by the Directorate: Animal Health.

Table 5.4: Total livestock numbers in South Africa, 2002

Livestock type	Agricultural Statistics	Animal Health	Difference
	Number		
Cattle	13 634 981	10 884 446	2 750 535
Sheep	25 727 114	22 363 296	3 363 818
Pigs	1 663 051	2 337 017	-673 966
Goats	6 451 889	5 777 507	674 382

Source: Directorate: Agricultural Statistics and Directorate: Animal Health

Table 5.5 provides a comparison of estimates from the Directorate: Agricultural Statistics, Directorate: Animal Health and the 2002 Census of Commercial Agriculture. These figures represent the total number of commercial livestock in South Africa for the 2002 calendar year. Commercial livestock numbers were derived from the estimated total livestock numbers by the directorates Agricultural Statistics and Animal Health in order to enable comparison with the estimates from the 2002 Census.

Table 5.5: Commercial livestock statistics for 2002 from three sources

Livestock type	Agricultural Statistics	Animal Health	2002 Census
	Number		
Cattle	8 083 363	7 044 223	4 367 187
Sheep	22 576 362	21 717 553	10 896 306
Pigs	1 470 966	2 119 757	722 718
Goats	2 168 536	2 622 906	988 474

Source: Directorate: Agricultural Statistics, Directorate: Animal Health and Statistics South Africa (2002)

The differences between the estimates from Agricultural Statistics and Animal Health, the differences between Agricultural Statistics and 2002 Census and the differences between

Animal Health and 2002 Census are shown in Table 5.6. Again, the estimates from the Directorate: Agricultural Statistics seem to be the highest for all livestock types except pigs. It must, however, be borne in mind that the estimates from the 2002 Census only represent animal numbers from farms that have been registered for Value Added Tax (VAT). It could, therefore, be expected that these estimates would be lower than those by the directorates Agricultural Statistics and Animal Health.

Table 5.6: Discrepancies in commercial livestock statistics from three sources

Livestock type	Difference between:		
	Agricultural Statistics and 2002 Census	Agricultural Statistics and Animal Health	Animal Health and 2002 Census
Cattle	3 716 176	1 039 140	2 677 036
Sheep	11 680 056	858 809	10 821 247
Pigs	748 248	-648 791	1 397 039
Goats	1 180 062	-454 370	1 634 432

Source: own calculations

Although rudimentary, the results from the comparison confirm that there are indeed fairly large discrepancies in the data. There is no doubt that these discrepancies could affect future projections by farmers as well as Government's ability to effectively determine and analyse policy. It is also of the utmost importance that researchers are aware of these discrepancies. Furthermore, the comparison also reiterates Groenewald's (1989) conclusion that there is reason for serious concern over the quality of data, not only in respect of fresh produce but, as this basic analysis points out, also in respect of livestock and livestock products. The situation is rather alarming, considering that this problem was identified more than three decades ago. The reader cannot help but to think that research, in order to improve official data series, has not received a very high priority in the public sector over the past 40 years. This is in spite of the important role that reliable statistics play in the successful growth of the agricultural sector; a fact recognized by Frick and Groenewald (1999a and b, 2001).

A final note on recognising the importance of data underlying econometric analysis is the unavailability of statistics on important phenomena. Nieuwoudt (1973, p. 22) noted that more information is needed about the income of farmers from outside sources, and the institutions providing part-time employment. Thirty three years later it seems as if, at least to some extent for non-farm income, things have improved slightly. Kirsten and Moldenhauer (2006, p. 75) comment on the treatment of non-farm income by concluding their paper:

> "From casual observation, it has become evident that currently a large number of commercial farmers, especially after the period of decentralization, occupy non-farm jobs or alternatively generate income out of activities on the farm other than the sale of crops and livestock suggesting that more attention should be paid to the nature and composition of the "other income" component of future surveys. In addition new entrants into farming and beneficiaries of the land reform programme frequently continue their current non-farm career such as teachers, taxi operators, and carpenters while they establish their newly acquired farms. In these cases, non-farm income is either used to sustain the livelihood of the household or to assist with the establishment cost of the farm – therefore another reason why an integrated concept of household income should be used in future. The agricultural census of 2002 took a first but modest step in trying to capture total farm household income but the capturing and recording of this information has to be improved in future."

5.5 CONCLUSION

This chapter was the medium through which the discussion in Chapters two to four was brought into context of agricultural economics and more specifically, agricultural economics in South Africa. The chapter started off by trying to establish if there is any reason to believe that the problems persisting in economics also persist in agricultural economics. Investigating the literature on South African agricultural economics revealed that this is in fact the case. Specifically, three authors have expressed their disenchantment with the manner in which econometrics has been applied in South African agricultural economics. Reviewing their articles highlighted that all the elements of the R-square disease do seem to be prevalent in South African agricultural economics, at least as it has been portrayed in *Agrekon.*

The focus then turned to the medium for publication of agricultural economic scholarship in South Africa. *Agrekon*, South Africa's agricultural economics journal, has had a rich history and has endured the many changes in the South African agricultural environment since the journal's launch in 1962. Applying a similar approach to that of Ziliak and McCloskey (1996 and 2004a) an analysis of the number of published articles during the period 1962 to 2005 revealed that about 32% contained some form of regression analysis or mathematical programming technique. These articles contained many examples of leading applied econometric techniques applied to South African as well as other parts of Africa's agricultural problems.

A discussion of the last vice, scholasticism and path dependency, shed some light on the reason for this persistence. An analysis of articles containing regression or mathematical programming revealed that scholasticism and path dependency are very much alive in South African agricultural economics.

The data underlying econometric analysis was also discussed in the context of South African agriculture. There seem to be quite a number of problems with data in South African agricultural economic research. Scholars such as Nieuwoudt (1973) have been concerned with this problem for many years. In fact, agricultural statistics in South Africa have been plagued by huge discrepancies in sources of statistics, data errors and reduced availability as the result of structural changes in the South African agricultural sector. The chapter has therefore, highlighted the need to pay attention to these problems as well as establishing how scholars have dealt with them

CHAPTER 6

THE TRUTH IS OUT THERE: FINDING THE VICES IN AGRICULTURAL ECONOMIC SCHOLARSHIP

6.1 THE INSTRUMENT: SEARCHING FOR THE PERVASIVENESS OF THE VICES IN AGRICULTURAL ECONOMIC SCHOLARSHIP

It appears from reviewing the literature cited in Chapter five that the vices of economics also exist in South African agricultural economics. Yet, much of what has been drawn from the three papers by Groenewald (1990), Nieuwoudt (1992) and Kirsten (2002) could be made off as mere speculation. Armstrong (1978) writes about folklore and fact in econometric forecasts. According to Armstrong (1978, p.550) folklore persists because people who hold viewpoints on an issue tend to perceive the world so as to reinforce what they already believe; they look for "confirming" evidence and avoid "disconfirming" evidence.

This chapter has the task to establish whether folklore persists in agricultural economics in South Africa. If, indeed, the facts are found to substantiate the authors' viewpoints, then their views are not folklore. Instead, the notion that everything is in order in applied econometrics in South African agricultural economics would then rather be regarded as folklore. How would researchers go about investigating these views?

South Africa's only agricultural economics peer reviewed journal is reviewed to search for the existence of the vices of economics and in order to substantiate Groenewald, Nieuwoudt and Kirsten's disenchantment. A sample of 65 full-length papers that used regression analysis is taken from volumes 1 to 44 of *Agrekon.* The sampling process, the population and sample size is discussed in the next section; thereafter the survey questionnaire and results of the survey are presented. This does not necessarily provide a total review of agricultural economic scholarship in South Africa, since articles are also published in the *South African Journal of Economics* and *Development Southern Africa* and other international journals.

6.2 METHODOLOGY

Although critics of applied significance testing in economics have intensely reviewed the application thereof in prominent economic journals, very little has been said as to how samples for their analyses were derived.

Zellner (1979) contains a small survey of 22 quantitative articles in five issues of different leading economic journals in 1978. Yet, the survey did not elaborate much about the process of selecting the 22 articles. Canterbery and Burkhardt (1983) reviewed 542 empirical papers that appeared in the *American Economic Review, Journal of Political Economy, Economic Journal*, and the *Quarterly Journal of Economics* from 1873 to 1978. Keuzenkamp and Magnus (1995) surveyed the papers in the *Journal of Econometrics*, Volumes 1 to 46 (1973 – 1990). Their survey was more of a census since they reviewed all the 668 papers published during 1973 to 1990. Ziliak and McCloskey (1996, p. 101) took a selection of full-length papers published in the American Economic Review. They comment on their use of the word "selection" saying that by not using "sample" they will not use tests of statistical significance.

However, any attempt at examining the use of statistical significance testing in papers published in *Agrekon* should be based on a scientific approach. A survey of papers not based on a scientific sampling process would not survive the scrutiny of both the journal editor and critics.

Figure 5.2 (p. 87) of this study illustrates that there has been an increasing trend in the percentage of articles published in *Agrekon* containing some form of regression analysis or mathematical programming. Any sample selected from this population of published articles would therefore need to reflect this trend.

6.2.1 Sampling methodologies

Steyn *et al.* (1998) distinguishes between two main classes of sampling methods namely probability sampling methods and non-probability sampling methods.

Probability sampling

In probability sampling every element of the population has known positive probability of being drawn for the sample. Sampling methods of this class enable the user to obtain the sample distribution of the estimator under consideration; so that an estimate of the sample variance can be obtained from the sample and valid conclusions be reached concerning characteristics of the population. The precision of an estimate can therefore be estimated from the sample itself (Stoker, 1984). Well-known probability sampling techniques include simple random sampling, systematic sampling, stratified sampling, cluster sampling and complex sampling.

Simple random sampling is the most basic form of probability sampling and provides the theoretical basis for the more complicated sampling techniques. In simple random sampling each element in the population has the same probability of being selected at each draw. The sample can either be drawn with replacement in which the same unit may be included more than once in the sample, or without replacement in which all sampling units are distinct. The practical requirements for simple random sampling are that every element or unit in the population must be clearly identifiable and a sampling frame must be available. The advantages of simple random sampling include its simplicity and the fact that it is the only assumption-free sampling method. Disadvantages include that fact that it requires a complete, up to date, sampling frame. Furthermore, simple random sampling does not guarantee a representative sample (Neethling, 2001).

Cassell and Rousey (2003) argue that before the increased application of computers, systematic or sequential sampling was an extremely useful tool. By ordering the sampling frame sequentially and selecting a sampling point, an entire sample could be selected without generating any more random numbers. As an example, consider a population of N elements numbered from one to N. let n be the required sample size and suppose $k = N/n$, an integer or nearest integer. Draw the starting point (first element) randomly from all the population elements and take the first drawn element together with n-1 elements obtained by taking every k^{th} element proceeding round the circumference of the circle. Advantages of systematic or sequential sampling include its simplicity and even distribution of the elements of the systematic sample across the population. The elements are therefore intuitively more representative of the population than in the case of simple random sampling. The disadvantages include the fact that systematic sampling requires the availability of a complete, up to date, sampling frame. The occurrence of periodicity among the population elements on the population list may cause misleading or biased results, Cassell and Rousey (2003, p. 4) also note that a single systematic sample has no unbiased variance estimator.

Stratified random sampling is used when the population can be divided into non-overlapping subpopulations, called strata, each of which is more homogenous than the entire population. A sample of elements is drawn from each stratum. Every population element belongs to only one stratum. Typical stratification variables are regional variables (for example, provinces, and geographical area), demographic variables (for example, gender and age group) and socio-economic variables (for example, income groups). The advantage of stratified random sampling is that it can be used to ensure a representative sample. Another useful advantage of stratified random sampling is that within each stratum, different sampling methods may be used depending on the sampling problems experienced in the various strata.

In cluster sampling the population elements are grouped together in non-overlapping groups, called clusters. The size of the clusters need not be the same. Every population element belongs to only one cluster. According to Cassell and Rousey (2003, p.5) there are two main reasons for using cluster sampling. The first is logistical: sometimes it is not possible to build a complete, reliable sampling frame for the entire population. In such cases, clusters of the population can be constructed and sampling frames can be built for those clusters. The second reason is budgetary: it is often far more cost-effective to sample all students within a school or all discharges from a hospital than to randomly sample a few here and there from thousands of institutions.

Complex or multistage sampling consists of a combination of the four probability sampling methods discussed above. A multistage sample might be a sample in which a first-stage sample is first drawn, as if cluster sampling is performed. Then, in the second stage, subsets of the clusters are created. The elements composing the frame from which the first sample is drawn are called the primary sampling units. The second stage sample is built by taking a subset of elements from each of the primary sampling units.

Non-probability sampling methods

In the case of non-probability sampling methods, the probability that a particular sample element will be drawn is unknown and can't be determined. Consequently, the estimation of the variance of an estimator of a population characteristic serves no purpose since it can't be interpreted statistically. Steyn *et al.* (1998, p.38 – 40) discuss the following well-known non-probability sampling methods:

- **Convenience sampling:** The sample is confined to a part of the population that is easily accessible. The advent of the internet has made it possible for anyone to collect survey data just by putting up a webpage, an example of a sample of convenience.

- **Haphazard sampling:** The sample is drawn haphazardly. For example, the haphazard selection of students sitting or walking on a university campus or persons passing by on a street corner.

- **Judgemental sampling:** The elements are chosen subjectively and deliberately by an "expert" to obtain a "representative" sample. For example, the researcher selects a "representative" number of farms, using his or her own judgement. A serious deficiency of this method is that other "experts" have different ideas on which population to select.

- **Purposive sampling:** Certain elements are chosen purposely to form part of the sample.

- **Quota sampling:** This well-known sampling procedure can be regarded as a combination of convenience sampling and judgemental sampling. Subclasses are formed by control variables such as geographical area, gender, and age. The sizes of theses subgroups are determined from census data or other available information. The quotas of the number of sample elements in each subclass are then determined proportionally to the population subclass sizes. These sample quotas are then divided among interviewers who attempt to find persons complying with the stated requirements, irrespective of the way in which these respondents are selected.

Lastly, only sample results obtained by using probability sampling can be generalised to the population in a statistically valid manner.

It was shown in Chapter five (p. 84) that there are four conspicuous years in the history of the number of papers published in *Agrekon* namely: 1982, 1990, 1999 and 2002. The number of articles was consequently divided into five strata namely, 1962 – 1982, 1983 – 1990, 1991 – 1999, 2000 – 2002 and 2003 – 2005. The number of articles published in *Agrekon* during 1962 to 2005 is illustrated in Figure 6.1. The trend lines for each period are clearly different from one another.

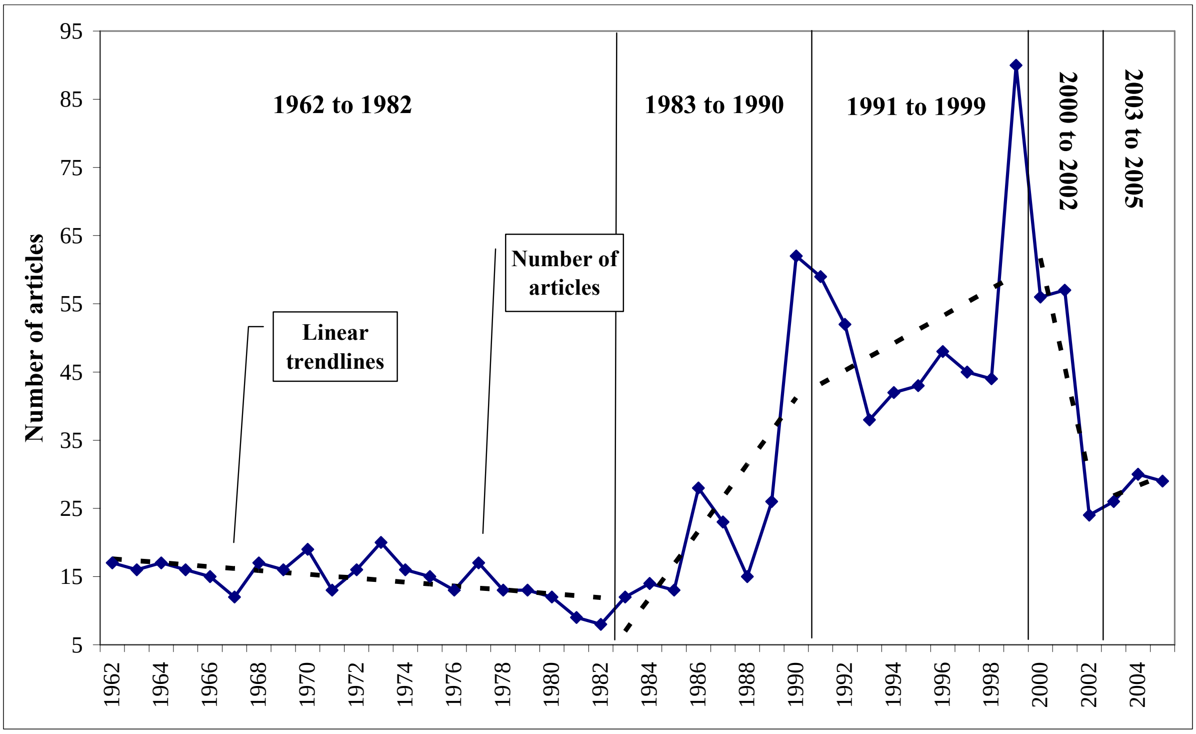

Figure 6.1: Number of articles published in Agrekon, 1962 - 2005

Source: Agrekon Volumes 1 to 44

Stratified random sampling was consequently applied to derive a sample of papers containing regression analysis. Since Ziliak and McCloskey's (1996 and 2004) questionnaire was applied to *Agrekon* papers, it was decided to exclude papers applying techniques such as Discriminant Analysis, Principal Component Analysis, Chi-square analysis, Analysis of Variance (ANOVA), and Multi Analysis of Variance (MANOVA). The reason for the exclusion being that these procedures are not that commonly applied in economics. More mathematically orientated techniques such as linear programming were also excluded from the analysis. The subsequent population of articles was therefore reduced to only articles containing regression analysis, a total of 276. The number of articles containing regression analysis per stratum is shown in Table 6.1.

Table 6.1: Number of articles containing regression analysis per stratum

Stratum	Period	Number of articles containing regression	Percentage of total
One	1962 – 1982	39	14.13
Two	1983 – 1990	49	17.75
Three	1991 – 1999	104	37.68
Four	2000 – 2002	44	15.94
Five	2003 – 2005	40	14.49
Total	**1962 – 2005**	**276**	**100.00**

Source: Own calculations

The first step was to calculate the required sample size for the stratified random sample of articles containing regression analysis. Since this study is interested in articles containing regression analysis, the aim is to estimate a sample of articles that have a certain characteristic, the characteristic being the use of regression analysis in the paper. A variable y_i was therefore defined as

$y_i = 1$ if the paper has this characteristic

$= 0$ otherwise

According to Neethling (2001, p.18), the population proportion can then be defined as

$$P = \frac{1}{N}\sum_{i=1}^{N} y_i = \overline{Y} \qquad (6.1)$$

which is estimated by the sample proportion

$$p = \frac{1}{n}\sum_{i=1}^{n} y_i = \overline{y} \qquad (6.2)$$

Further,

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \overline{Y})^2 = \frac{1}{N-1}(\sum_{i=1}^{N} y_i^2 - 2P\sum_{i=1}^{N} y_i + NP^2) = \frac{N}{N-1}P(1-P) \qquad (6.3)$$

Lohr (1999, p.45) shows that the variance of the mean $\overline{y}$ from a simple random sample is

$$V(\overline{y}) = E(\overline{y} - \overline{Y})^2 = \frac{(N-n)}{N}\frac{S^2}{n} = (1-f)\frac{S^2}{n} \qquad (6.4)$$

where

$f = \dfrac{n}{N}$ and $S^2 = \dfrac{\sum_{i=1}^{N}(y_i - \overline{Y})^2}{N-1}$, the population variance.

Proof: Lohr (1999, p.45)

From equation (6.4)

$$V(p) = \left(\frac{N-n}{N-1}\right)\frac{P(1-P)}{n} \qquad (6.5)$$

So that

$$v(p) = \left(1 - \frac{n}{N}\right)\frac{p(1-p)}{n-1} \qquad (6.6)$$

with

$$s^2 = \frac{n}{n-1}p(1-p) \qquad (6.7)$$

A sample size was thus estimated for a proportion, the proportion being the number of articles containing regression analysis.

In the case of estimating sample size for a proportion Neethling (2001, p.20) defines the formula for sample size as

$$n_0 = \frac{z^2_{\alpha/2}}{d_2} p(1-p) \qquad (6.8)$$

where

n_0 is the sample size

$z^2_{\alpha/2}$ is the critical value

d is the precision of the estimate

By applying equation (6.8) the sample the size was estimated as

$$n_0 = \frac{1.96^2}{0.10^2} 0.23(1-0.23) = 65$$

The sample size was multiplied by the fraction for each stratum in order to obtain a sample size for each stratum. The sample size per stratum is shown in Table 6.2

Table 6.2: Estimated sample size per stratum

Stratum	Period	Sample size of articles containing regression	Percentage of total
One	1962 – 1982	10	14.13
Two	1983 – 1990	12	17.75
Three	1991 – 1999	23	37.68
Four	2000 – 2002	13	15.94
Five	2003 – 2005	7	14.49
Total	**1962 – 2005**	**65**	**100.00**

Source: Own calculations

Simple random sampling was used to draw the samples out of each stratum. The simple random sample procedure was performed using PROC SRUVEYSELECT in SAS version 9.1.3. The list of sampled papers can be found in the appendix.

6.2.2 The questionnaire

Ziliak and McCloskey "admit" in their 2004 paper that a smaller number of questions could also serve the purpose. They note (p. 544)

> "Requiring referees to complete a 19-item questionnaire would probably go against the libertarian grain of the field; a short form would do: "Does the paper focus on the size of the effect it is trying to measure, or does it instead recur to irrelevant tests of the coefficient's statistical significance?"

Since this study drew heavily on the ideas from Ziliak and McCloskey's (1996, 2004) paper it was decided to apply all of their original questions in this survey. Their 19 questions, as discussed in Chapter three (p. 52 – p. 59) of this study, are summarised in Table 6.3. Since their questionnaire does not implicitly cover the use of the R^2 statistic, additional questions on the use of the R^2 statistic were added in this survey's questionnaire. These additional questions appear in Table 6.4

Table 6.3: Questions from the Ziliak and McCloskey survey questionnaire used in their 1996 and 2004 studies to be applied to the sample of articles

Survey question	
Does the paper ...	
1.	Use a small number of observations, such that statistically significant differences are not found merely by choosing a very large sample?
2.	Report descriptive statistics for regression variables?
3.	Report coefficients in elasticities, or in some other useful form that addresses the question of "how large is large"?
4.	Test the null hypotheses that the authors said were the ones of interest?
5.	Carefully interpret the theoretical meaning of the coefficients? For example, does it pay attention to the details of the units of measurement, and to the limitations of the data?

Survey question	
Does the paper ...	
6.	Eschew reporting all standard errors, *t*-, *p*-, and *F*-statistics, when such information is irrelevant?
7.	At its first use, consider statistical significance to be one among other criteria of importance?
8.	Consider the power of the test?
9.	Examine the power function?
10.	Eschew “asterisk econometrics,” the ranking of coefficients according to the absolute value of the test
11.	Eschew “sign econometrics,” remarking on the sign but not the size of the coefficient?
12.	Discuss the size of the coefficients?
13.	Discuss the scientific conversation within which a coefficient would be judged large or small?
14.	Avoid choosing variables for inclusion solely on the basis of statistical significance?
15.	Use other criteria of importance besides statistical significance after the crescendo?
16.	Consider more than statistical significance decisive in an empirical argument?
17.	Do a simulation to determine whether the coefficients are reasonable?
18.	In the conclusions, distinguish between statistical and economic significance?
19.	Avoid using the word “significance” in ambiguous ways?

Source: Ziliak and McCloskey (2004, p. 529, questions sorted in ascending order)

Table 6.4: Two additional questions used in this study’s survey of applied econometrics in *Agrekon*

Survey question	
20.	Are candid examples pointing to data mining, as discussed in this study (Chapter 4, p. 68), found in the text?
21.	Does the paper go beyond the sole reporting of R-square as a measure of goodness of fit, i.e. are other goodness of fit measures also considered?

6.3 RESULTS OF THE SURVEY OF *AGREKON*

The results of the survey are discussed in the light of Ziliak and McCloskey’s results of their 1996 and 2004 studies. How does the results of this study compare to that of Ziliak and McCloskey’s? In a footnote, Mayer (1980, p.166) cites Leontief (1971) and speculates that the "problem" might be less severe in agricultural economics. Is this true? The results of the Agrekon survey are portrayed in Table 6.5.

The survey results indicate that the problem might be less severe in agricultural economics, or at least as far as agricultural economics (as published in *Agrekon*) in South Africa are concerned. South African agricultural economists did much better than economics colleagues who published in the *American Economic Review*.

Table 6.5: Searching for the standard error of regression analysis in *Agrekon*

Survey question	Papers containing regression and published in *Agrekon*						Ziliak and McCloskey's 1996 and 2004 surveys	
Does the paper ...	**1962 - 1982**	**1983 - 1990**	**1991 - 1999**	**2000 - 2002**	**2003 - 2005**	**1964 - 2005**	**1990s**	**1980s**
1. Use a small number of observations, such that statistically significant differences are not found merely by choosing a very large sample?	0.00	0.00	4.17	10.00	0.00	3.08	67.9	85.7
2. Report descriptive statistics for regression variables?	50.00	16.67	29.17	40.00	22.22	30.77	66.4	32.4
3. Report coefficients in elasticities, or in some other useful form that addresses the question of "how large is large"?	100.00	100.00	91.67	100.00	100.00	96.92	86.9	66.5
4. Test the null hypotheses that the authors said were the ones of interest?	100.00	100.00	100.00	100.00	100.00	100.00	83.9	97.3
5. Carefully interpret the theoretical meaning of the coefficients? For example, does it pay attention to the details of the units of measurement, and to the limitations of the data?	100.00	100.00	100.00	100.00	100.00	100.00	81	44.5
6. Eschew reporting all standard errors, *t*-, *p*-, and *F*-statistics, when such information is irrelevant?	80.00	16.67	37.50	50.00	44.44	43.08	12.4	8.3
7. At its first use, consider statistical significance to be one among other criteria of importance?	20.00	0.00	0.00	0.00	11.11	4.62	36.5	47.3
8. Consider the power of the test?	0.00	0.00	4.17	0.00	0.00	1.54	8	4.4
9. Examine the power function?	0.00	0.00	0.00	0.00	0.00	0.00	45.5	16.7
10. Eschew "asterisk econometrics," the ranking of coefficients according to the absolute value of the test	10.00	16.67	8.33	0.00	0.00	7.69	32.8	74.7
11. Eschew "sign econometrics," remarking on the sign but not the size of the coefficient?	10.00	25.00	33.33	0.00	0.00	18.46	19	46.7
12. Discuss the size of the coefficients?	50.00	66.67	45.83	70.00	77.78	58.46	78.1	80.2
13. Discuss the scientific conversation within which a coefficient would be judged large or small?	10.00	50.00	29.17	30.00	33.33	30.77	54	28
14. Avoid choosing variables for inclusion solely on the basis of statistical	90.00	83.33	91.67	80.00	88.89	87.69	25.5	68.1

Survey question	Papers containing regression and published in *Agrekon*						Ziliak and McCloskey's 1996 and 2004 surveys	
Does the paper ...	**1962 - 1982**	**1983 - 1990**	**1991 - 1999**	**2000 - 2002**	**2003 - 2005**	**1964 - 2005**	**1990s**	**1980s**
significance?								
15. Use other criteria of importance besides statistical significance after the crescendo?	50.00	83.33	79.17	80.00	77.78	75.38	28.5	40.7
16. Consider more than statistical significance decisive in an empirical argument?	10.00	0.00	4.17	0.00	11.11	4.62	18.2	29.7
17. Do a simulation to determine whether the coefficients are reasonable?	0.00	0.00	20.83	10.00	44.44	15.38	35	13.2
18. In the conclusions, distinguish between statistical and economic significance?	40.00	91.67	91.67	100.00	77.78	83.08	52.6	30.1
19. Avoid using the word "significance" in ambiguous ways?	100.00	83.33	91.67	100.00	77.78	90.77	37.2	41.2
20. Are candid examples pointing to data mining, as discussed in this study (Chapter 4, p. 68), found in the text?	0.00	33.33	12.50	0.00	0.00	10.77	NA	NA
21. Does the paper go beyond the sole reporting of R-square as a measure of goodness of fit, i.e. are other goodness of fit measures also considered?	0.00	25.00	0.00	10.00	22.22	9.23	NA	NA

Source: Sample of 65 full length papers using regression analysis in *Agrekon*.

In general, the results show that economic theory was the key driver in the selection of variables and choice of regression technique. Most of the papers surveyed applied statistical significance to test the validity of the assumptions underlying the regression model and to test strength of the explanatory power of variables included in the models. Another noteworthy observation is that most of the surveyed papers seem to have followed Hendry's general to specific approach which this study addressed in Chapter two. Papers would generally start off with a general model which would include all the variables that, according to the economic theory underlying the model could have an effect on the dependent variable. Variables would then be eliminated on the basis of testing their economic and statistical significance in order to arrive at a specific model.

Only three per cent of the papers surveyed used a small number of observations, such that statistically significant differences would not be found at the conventional levels merely by choosing a large number of observations. In fact, only two papers replied "yes" to this question, one paper in stratum three (1991 – 1999) and another in stratum four (2000 – 2002). However, one paper did acknowledge the fact that the use of the small number observations might be problematic. Liebenberg *et al*. (1991, p.95) notes: "More observations may have caused more coefficients to be significant at the five per cent level of significance."

Agricultural economists generally seemed not be as specific about the data used on their analysis. In total, only 31 per cent of the papers surveyed did report descriptive statistics for the regression variables used in their regressions. This is lower than the 66 and 32 per cent found by Ziliak and McCloskey (1996 and 2004) in their surveys of papers published during the 1990s and 1980s. Papers published in *Agrekon* during 1962 to 1982 (stratum one) had the highest occurrence of reporting descriptive statistics for their regression variables. Papers published during 1962 to 1982 also have another important feature, especially the earlier papers. Data were mostly obtained through surveys. The early papers also applied regression models mostly on farm level in order to estimate optimum levels of input use, for example, the optimum level of fertilizer. In general, authors who made use of surveys to collect their data would generally report descriptive statistics for all variables used. For example, Groenewald and Graham (1968. p.31), Nieuwoudt and Behrmann (1976, p.15), Van Schalkwyk *et al.* (1993, p.253), and Mahabile *et al.* (2005, p.108) all make use of surveys and they all report descriptive statistics on the regression variables. Thus, there seems to be a correlation between obtaining data for regression through the use of surveys and reporting descriptive statistics for regression variables.

Almost all of the papers surveyed reported the coefficients in elasticities or in some other useful form to enable the reader to answer the question of "how large is large"? No evidence could be found that the null hypotheses tested were not those hypotheses that the authors said were the ones of interest.

The theoretical meaning of the coefficients was generally well interpreted. All of the papers did consider the details of the units of measurement of the data. Although better than the results obtained by Ziliak and McCloskey (1996 and 2004a), the reporting of all the standard errors, *t*-, *p*-, and *F*-statistics was not that well reported. Only 43 per cent of papers surveyed did report all the standard errors, *t*-, and *F*-statistics. Stratum 3 (1983 – 1990) recorded the lowest score, only 17 per cent. However, the *p*-value was not that well reported. Only a few of the sample of articles did report *p*-values, for example, Randela *et al*. (2000, p.650). Gujarati (1998, p.132 and 2003, p.137) suggested reporting *p*-values. The reader would then be able to decide whether to reject the null hypothesis at the give *p*-value. Most articles generally reported more than one statistical significance level. Statistics were mostly reported at the 99 and 95 per cent levels of significance.

Statistical significance at first use was commonly not the only criterion of importance. Only five per cent of the surveyed papers used statistical significance as the only criterion of importance. Generally, economic significance was the scientific "crescendo" of the paper.

The power of the test was not generally addressed. Only one paper indirectly referred to the power of the test. In their study of the demand for eggs in South Africa Cleasby and Ortmann (1991, p.35) note: "As the correlation between PE and PB is high, the confidence intervals for the relevant population parameters are expected to be overstated, increasing the probability of accepting the false hypothesis (i.e. type II error)." Although indirect, it was the only reference made to the power of a test. The power function was also not examined by any of the papers surveyed.

Asterisk econometrics was not that prevalent among the surveyed papers, with only eight per cent of the papers committing such a "crime". Stratum two (1983 – 1990) contained most of the papers committing asterisk econometrics. However, sign econometrics was more prevalent (18 per cent) than asterisk econometrics. Almost 33 per cent of the papers in stratum three remarked on the sign but not the size of the coefficients. This figure is more or less in line with the 48 per cent reported by Ziliak and McCloskey (1996) for their survey of papers published during the 1980s. Size did matter for most authors, 58 per cent did include a discussion on the size of the coefficients. It also seems that this trend has been on the increase because 78 per cent of the papers in stratum five (2003 – 2005) included a discussion of the size of the coefficients. What would be judged large or small was generally addressed by comparing coefficients to those obtained from similar studies and studies from abroad. For example, Sartorius von Bach and Van Zyl (1994, p.149) included a discussion as well as a table of elasticities for the demand for carbohydrates in South Africa. In their table, they compared elasticities from other studies to elasticities derived from their analysis.

Variables included solely on the basis of statistical significance were not that common. About 88 per cent of the sample avoided choosing variables for inclusion solely on the basis of statistical significance. 75 per cent of the sample made use of other criteria besides statistical significance. Economic theory was generally the main criteria of importance. Statistical significance was therefore not the conversation stopper. Only five per cent of the sample applied statistical significance so that it was decisive in the empirical argument.

Simulation in order to determine whether coefficients are reasonable was less common, only 15 per cent of the sample applied simulation. Simulation would be useful in validating the economic significance of the model, as well as determining whether the model is sensitive to changes in the data. Meyer and Kirsten (2005, p. 234) applied simulation to generate "what if" scenarios for the wheat sector in South Africa.

Authors, in their conclusions, generally distinguished between statistical and economic significance. Furthermore, the use of statistical significance in ambiguous ways was also not that prevalent. 91 per cent of the sample avoided the use of statistical significance in ambiguous ways. In their conclusion, Niebuhr and Van Zyl (1992, p.133) note that their regressions yielded significant results. Somehow it is unclear if the regressions yielded economically or statistically significant results.

Apart from the questions included in Ziliak and McCloskey's (1996 and 2004a) studies, two additional questions on the R^2 were also included. The sample generally reported R^2 statistics above 70 per cent, Agbola (2001, p.555) even reported an R^2 statistic of a hundred per cent or one. Can these results be attributed to data mining or excellent skills among contributors to *Agrekon*? The reader should decide. However, there were examples of papers reporting lower scores such as Randela *et al.* (2000, p.650) and Conradie (2005, p.149). Candid examples pointing to data mining as discussed in Chapter four (p.68) of this study were not that prevalent among the sample of papers. Cleasby and Ortmann (1991, p.35) note: "A number of models were estimated in an attempt to find the best possible model." Another candid example was also found in Niebuhr and Van Zyl (1992, p.131). These examples also point to the "Highlander approach" since although many models were estimated; only one model was reported. Only nine per cent of the sample did consider other goodness of fit measures apart from the R^2 and *F*-statistic. Berry *et al.* (1990, p.88) reported the predicted residual sum of squares (PRESS) and the root mean square (RMS). Meyer and Kirsten (2005, p.233) included the root mean square error (RMSE) and Theil's inequality coefficient as additional measures of goodness of fit.

6.4 CONCLUSION

The results of the survey show that Leontief's (1971) speculation that the "problems" with applied econometrics might be less severe in agriculture, might not be that far fetched. South African agricultural economic scholars seem to be more thorough than their economist colleagues in the United States and United Kingdom in applying statistical significance testing. Papers included in this survey seem to have grasped the idea of looking not only at statistical significance, but also at the economic theory behind the model and that statistical significance *per se* does not imply economic significance. In fact, surveyed papers performed generally well in questions focussing on the basic elements such as considering more than one statistical significance level, reporting coefficients in elasticity form or some other interpretable form relevant to the problem at hand.

However, when applied econometrics is explored in greater detail, elements of the disease start to surface. The survey revealed that the following main errors or elements of the disease are present in South African agricultural economic scholarship:

- **Statistical significance:** Although South African agricultural economic scholars performed better than their fellow economics scholars surveyed by Ziliak and McCloskey, there is some room for improvement. The lack of attention paid to statistical power is a point of concern. However, this situation might improve as the processing power of computers improves. Reporting *p*-values will be a huge step in the right direction of improving this situation.

- **The apparent obsession with R^2:** This problem also seems to be as common in South African agricultural economic scholarship as it is in economics. It is, however, an institutional problem. South African agricultural economic scholars should revert from solely relying on high R^2 values as an indication of goodness of fit and they should stop treating high R^2 values as a status symbol.

- **Data mining:** Results seem to suggest that data mining may not have been so candidly applied, yet its presence is still well rooted. Arguably impossible to eradicate, care should still be taken when applying the procedure. The researcher should not be overpowered by the dark side of the disease when applying this technique.

- **Data underlying the analysis:** Data is by far the biggest problem in applied econometrics in South African agricultural economics. Results suggest that scholars adopted a lackadaisical attitude towards the data underlying their analyses. Scholars should pay much more attention to the numbers they feed into their models.

Lastly, it should be noted that this survey was based on articles of one journal. South African agricultural economists would also have been able to publish papers in other South African journals such as the *South African Journal of Economics* and *Development Southern Africa*. This might constitute biased results. Further studies would thus be required to also study the use of econometrics in these journals. To conclude, although this survey revealed that agricultural economists seem to have performed better in applying statistical significance in regression analysis, there is some room for improvement.

CHAPTER 7

SUMMARY AND CONCLUSION

The general objective of this study was to analyse the disenchantment with the way econometrics have been applied empirically and to establish whether these sources of disenchantment are present in South African agricultural economic scholarship.

The first part of this dissertation provided a historical foundation of the developments and events in the history of econometrics. Econometrics is that subdivision of economics which explicitly unites deduction, induction and statistical inference; its methodology concerns the procedures adopted in the testing and where, applicable, the quantification of economic theories. Econometrics' development is largely a post-Second World War phenomenon, exploiting the increased availability of economic data, high speed computers and appropriate software programs. Its development was structured by an increasing awareness amongst economists of the work in the philosophy of science, especially that of Sir Karl Popper. The methodology is perhaps best understood via the sophisticated falsificationism of Popper. Popper (1968, 1972) offered, among other things, a demarcation criterion between science and non-science, designating science as that body of synthetic propositions regarding the real world which, at least in principle, are capable of refutation through the use of empirical observations.

During the 1960s especially, the methodology of economics appeared to be based on the superficial appeal of Popperian falsificationism and some economists held the hope that econometrics would facilitate the establishment of an empirical base similar in content to that of the hard sciences. In econometrics many saw what they perceived as the provision of a rigorous and reliable method of testing hypotheses, a clear-cut route by which "poor" theories would be weeded out to be replaced by better theories. This hope was firmly based upon a falsificationist methodology in which econometrics was to provide the evidence on which refutation would take place.

Econometric investigations in the 1960s and 1970s were directed more towards the estimation of economic models than towards the testing of hypotheses Thus, in practice, econometrics became a vehicle for verification, but used the rhetoric of falsification.

However, during the seventies, the large scale econometric models which were developed as tools to cure the economic ills of the world began to fail. The inadequacy of these models' ability to deal with large external shocks such as the oil crisis shook the trust of policy makers in these models. This has led some commentators to take a nihilistic stance towards econometrics.

Scholars began to develop various econometric practices in an attempt to restore confidence in how econometrics was being practised. Chapter two focussed on the different approaches currently prevailing in econometrics: no estimation approaches and estimation approaches. The review of estimation approaches included the traditional approach of econometric application as well as those approaches championed by Hendry, Leamer and Sims. The review pointed out that each of these approaches had methodological weaknesses and could, therefore, not present a completely flawless method for the application of econometrics. It was subsequently argued that there needs to be a substantial clarification of procedures used in model selection and verification. Furthermore, there needs to be a clear understanding of the limits of econometric modelling.

The nihilistic stance developed by certain commentators evolved into a school astutely critical of the manner in which econometrics were being applied in empirical studies. Chapter three explored the secret sins of econometrics and explored the literature on the dissatisfaction with econometrics. Here is was found that the disenchantment can mainly be ascribed to the use of statistical significance tests in regression, the data underlying the analysis, data mining, replication, scholasticism and preference falsification. Literature on the use of statistical significance testing in economics highlighted that the problems experienced with the use of statistical significance testing in economics were not unique. The problem also persisted in other social sciences specifically in psychology. McCloskey has been vigorous in trying to persuade fellow economic scholars that the way in which statistical significance is used in economics is wrong.

Literature reviewing the use of econometrics in agricultural economics suggests that in agricultural economics too, significance tests have been abused, justifying an analysis to see whether the same can also be said of its use in South African agricultural economics.

Chapter four focussed specifically on the most common measure used to analyse goodness of fit, the R^2 and the related $\bar{R}^2$ as well as data mining. The ‘incorrect’ application of this statistic together with statistical significance testing has been christened the R-square disease of economics. It is a disease because the literature has shown that incorrect application of statistical significance tests and goodness of fit tests have been misleading, thereby creating a strange case of Dr Jekyll and Mr Hyde in economics. Countless accounts of examples in the literature exist where these analytics have been applied without regarding the underlying assumptions of the regression approach used by the researcher. Instead, or so the literature suggests, these analytics are used as tools of persuasion because the economic research environment has become more interested in statistically significant coefficients and high R^2 statistics. These analytics have become criteria for acceptance among one’s peers and journal editors; and it is this role which is the disease, since it misleads economic research from its true objectives.

Chapter four also covered the relation between the disease and data mining. A review of the literature on data mining revealed that the general attitude of researchers is that data mining is undesirable but that the private practice of econometrics is clearly riven with it. Some of the reviewed papers also held the view that if the “industry” were to be reformed or restructured, the undesirable effects of data mining could be kept under better control. A paper by Blackhouse and Morgan (2000) also explored the relation between data mining and the philosophy of economic science; suggesting that recent developments in the philosophy of science might have a direct bearing on econometric practice.

The review of statistical significance and data mining highlighted that the rationale behind the use of these analytics should be the real source of concern. It was argued that the problems associated with statistical significance tests and data mining would be mitigated if they were to be used within the correct context, i.e. supporting the underlying assumption of the regression model. The disease of significant *t*- and *F*-values and a high R^2 should not be the objective of an econometric research study.

Again, the linkages between the different vices were highlighted. Eradicating the disease by conducting studies within the correct context and motives would also aid in eradicating the vices of econometrics.

The second part of this study brought the issues discussed in the first part into context with agricultural economics and more specifically agricultural economics in South Africa. Chapter five started off by trying to establish if there is any reason to believe that the problems persisting in economics also persist in agricultural economics. Investigating the literature on South African agricultural economics revealed that this is in fact the case. Specifically three authors have expressed their disenchantment with the econometrics that has been applied in South African agricultural economics. Reviewing their articles highlighted that all the elements of the R-square disease do seem to be prevalent in South African agricultural economics. The focus then turned to the medium for publication of agricultural economic scholarship in South Africa. *Agrekon,* South Africa's agricultural economics journal has had a rich history and has endured the many changes in the South African agricultural environment since the journal's launch in 1962. Applying a similar approach to that of Ziliak and McCloskey an analysis of the number of published articles during the period 1962 to 2005 revealed that about 32 per cent contained some form of regression analysis or mathematical programming technique. These articles contained many examples of leading applied econometric techniques applied to South African as well as other parts of Africa's agricultural problems.

Chapter five also covered the vices of scholasticism and path dependency which shed some light as to the reason for this persistence. An analysis of articles containing regression or mathematical programming showed that scholasticism and path dependency are very much alive in South African agricultural economics.

The data underlying econometric analysis was also discussed in the context of South African agriculture. There seem to be quite a number of problems with data in South African agricultural economic research. Scholars such as Nieuwoudt (1973) have been concerned with this problem. In fact, livestock statistics in South Africa have been plagued by huge discrepancies in data from the different official sources of statistics, data errors and reduced availability as the result of structural changes in the South African agricultural sector.

The last chapter investigated the presence of the disease in South African agricultural economic scholarship by surveying papers published in *Agrekon* containing regression analysis. Following the same approach as Ziliak and McCloskey (1996 and 2004), it was found that traces of the disease do exist in South African agricultural economic scholarship. Although most scholars generally seem to grasp the importance of justifying their analysis in terms of economic as well as statistical significance, it is especially data mining and a perceived obsession with "R^2" values which is rather concerning. Although not as serious as expected, the application of statistical significance testing should also receive some attention since it was found that most scholars applied statistical significance testing without considering the power of a test.

7.1 RECOMMENDATION FOR FUTURE RESEARCH

However, the most alarming fact established through this survey is that scholars seem to have adopted a lackadaisical attitude towards the data underlying their rigorous output and thereby showing that they, in fact, do not quite grasp the notion of "*Garbage in – Garbage out*"

The study concludes by looking at the recommendations for future research. What should be the next step now that the key drivers of the problems with the application of econometrics have been identified?

Firstly, a more detailed study of applied econometrics in South African agricultural economics is required. This study has shown that there are problems: these problems should be investigated further. Agricultural economic scholars should take note of the problems identified in this study and learn from the mistakes made in the past, no matter how insignificant they might seem. More detailed studies should look at each of the vices in greater detail. It is of the utmost importance that agricultural economic scholars take note of these problems and ensure that these problems do not persist. Agricultural economic scholars have, in the past, devoted much of their research in analysing the inefficiencies of the previously regulated agricultural system. In order to ensure that these inefficiencies do not persist, the methods in analysing them should be rid of any errors which might cast a shadow of doubt on the results derived from them.

The next step should be to further investigate scholasticism and path dependency in South African agricultural economics. An understanding of the reason for the persistence would assist in establishing a path of reform in applied econometric scholarship in South Africa. Studies investigating path dependency and scholasticism usually look at the demand side, more specifically the "market" for academic articles which allows for the persistence of inefficient and sub-optimal practices. Journal editors and organisations which sponsor academic research can play a critical role in changing the incentive structure in the empirical sciences towards a high quality scientific product.

The findings of this study could therefore assist in such reform. The vices of economics identified in this study could be applied to the development of a segmentation model. Segmentation modelling has been applied successfully in the marketing sciences in order to classify consumers into different market segments. A segmentation model could be developed by which an author can be segmented into segments according to the econometric vice present in his/her article. This would allow peer reviewers to point authors to the specific problems in their articles before they are published in *Agrekon*.

REFERENCES

ABBOTT, M. &. AHMED, A (1999). The South African wool supply response. *Agrekon*, 38(1): 90-104

AGBOLA, F.W. (2001). A dynamic adjustment model for South African agriculture: 1965-97. *Agrekon*, 40(4): 549-562

ALEMU, Z.G., OOSTHUIZEN, K. VAN SCHALKWYK, H.D. (2003). Grain-supply response in Ethiopia: an error-correction approach. *Agrekon*, 42(4): 389-404

ALTMAN, M. (2000). A behavioural model of path dependency: the economics of profitable inefficiency and market failure. *Journal of Socio Economics*, 29(2): 127 – 145.

ALTMAN, M. (2004). Statistical significance, path dependency, and the culture of journal publication. *Journal of Socio Economics*, 33(5):651 – 633.

ANDERSON, D.R., BURNHAM, K.P. & THOMPSON, W.L. (2000). Null hypothesis testing: problems, prevalence, and an alternative. *Journal of Wildlife Management*, 64(4): 912 – 923.

ANDERSON, E. (1991). Danish Experience with Macroeconometric Models. *Scandinavian Journal of Economics*, 93(2): 315 – 322.

ANDERSON, L.C. & CARLSON, K. (1970). A monetarist model for economic stabilization. Federal Reserve Bank of St. Louis. *Review*, 52: 7 – 25.

ANIM, F.D.K. (1999). Organic vegetable farming in rural areas of the Northern Province. *Agrekon*, 38(4): 645-658

ARCHIBALD, G.C. (1966). Refutation or comparison? *British Journal for Philosophy of Science*, 17(3): 279 – 296.

ARMSTRONG, J.C. (1978). Forecasting with Econometric Methods: Folklore versus Fact. *Journal of Business*, 51(4): 549 – 564.

BACKHOUSE, R.E. & MORGAN, M.S. (2000). Introduction: is data mining a methodological problem? *Journal of Economic Methodology*, 7(2): 171 – 181.

BAKER, J.D. (Jr.). & PAARLBERG, D. (1952). How Accurate Is Outlook? *Journal of Farm Economics*, 34(4): 509 – 519.

BERRY, C.G., DICKS, H.M. & ORTMANN, G.F. (1990). Derivation of appropriate functions for the economic analysis of maize yield responses to fertiliser and rainfall variations at Dundee. *Agrekon*, 29(2): 85 – 92.

BEYERS, L. & HASSAN, R.M. (2001). The structure of South African milk production technology: A parametric approach to supply analysis. *Agrekon*, 40(4): 609 – 618.

BESSLER, D.A. & KLING, J.L. (1986). Forecasting Vector Autoregressions with Bayesian Priors. *American Journal of Agricultural Economics*, 68(1): 144 – 151.

BESSLER, D.A. (1984). Relative Prices and Money: A Vector Autoregression on Brazilian Data. *American Journal of Agricultural Economics*, 66(1): 25 – 30.

BLACK, F. (1982). The Trouble with Econometric Models. *Financial Analysts Journal*, 38(2): 29 – 37.

BLAUG, M. (1980). *The Methodology of Econometrics. Or How Economists Explain*, Cambridge: Cambridge University Press.

BOGEN, J. & WOODWARD, J. (1988). Saving the phenomena. *Philosophical Review*, 97: 303 – 352.

BONNEN, J.T. (1975). Improving Information on Agriculture and Rural Life. *American Journal of Agricultural Economics*, 57(5): 753 – 763.

BONNEN, J.T. (1983). Federal Statistical Coordination Today: A Disaster or a Disgrace? *American Statistician*, 37(3): 179 – 192.

BONNEN, J.T. (1983). The Dilemma of Agricultural Economists: Discussion. *American Journal of Agricultural Economics*, 65(5): 889 – 890.

BOWMAKER, P.A. & NIEUWOUDT, W.L. (1990). Demand equations for selected South African agricultural commodities. *Agrekon*, 29(1): 46 – 50.

BRAND, S.S. (1969). *The contribution of agriculture to the economic development of South Africa since 1910.* DSc(Agric). dissertation, University of Pretoria, Pretoria.

BREITENBACH, M.C. & FÉNYES, T.I. (2000). Maize and wheat production trends in South Africa in a deregulated environment. *Agrekon*, 39(3): 292 – 312.

BULLOCK, J.B. (1976). Social Costs Caused by Errors in Agricultural Production Forecasts. *American Journal of Agricultural Economics*, 58(1): 76 – 80.

BUTZ, E.L. (1989). Research That Has Value in Policy Making: A Professional Challenge. *American Journal of Agricultural Economics*, 71(4): 1195 – 1199.

CAHN, R.M (1982). The Trouble with "The Trouble with Econometric Models". *Financial Analysts Journal*, 38(3): 78.

CAMERON, S. (1993). Why is the R Squared Adjusted Reported? *Journal of Quantitative Economics*, 9(1): 183 – 186.

CANTERBERY, E.R. & BURKHARDT, R.J. (1983). What do we mean by asking whether economics is a science? In: Alfred S. Eichner, eds., *Why economics is not yet a science*. London: MacMillan Press.

CASSELL, D.L. & ROUSEY, A. (2003). *Complex Sampling Designs Meet the Flaming Turkey of Glory*. SUGI 28 Conference Proceedings, SAS Institute.

CHADWICK, J.B. & NIEUWOUDT, W.L. (1985). The demand for bananas and the economic effect of supply restriction. *Agrekon*, 24(2): 14 – 20.

CHAMBERS, R.G. (1984). Agricultural and Financial Market Interdependence in the Short Run. *American Journal of Agricultural Economics*, 66(1): 12 – 24.

CHANTYLEW, D. & BELETE, A. (1997). A statistical analysis of demand for beef, mutton/goat, pork and chicken in Kenya 1961-1991. *Agrekon*, 36(1): 1 – 9.

CHAREMZA, W.W. & DEADMAN, D.F. (1992). *New Directions in Econometric Practice: General to Specific Modelling, Cointegration and Vector Autoregression*, Hants, England: Edward Elgar Publishing Limited.

CLEASBY, R.C.G. & ORTMANN, G.F. (1991). Demand analysis of eggs in South Africa. *Agrekon*, 30(1): 34 – 36.

CLEASBY, R.C.G., DARROCH, M.A.G. & ORTMANN, G.F. (1993). Factors affecting the demand for and supply of South African yellow maize exports. *Agrekon*, 32(1): 1 – 5.

COHEN, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Second Edition, Hillsdale, New Jersey: L. Erlbaum Associates.

COLANDER, D. & BRENNER, R. (Eds.). (1992). *Educating Economists*. Michigan: University of Michigan Press, Ann Arbor.

COLANDER, D. (2000). New Millennium Economics: How Did it Get This Way, and What Way is It? *Journal of Economic Perspectives*, 14(1): 121 – 132.

CONRADIE, B. (2005). Wages and wage elasticities for wine and table grapes in South Africa *Agrekon*, 44(1): 138 – 156.

COOLEY, T.F. & LEROY, S.F. (1985). Atheoretical macroeconometrics: A critique. *Journal of Monetary Economics*, 16(3): 283 – 308.

COUPÉ, T. (2000a). *Revealed Performances. World Wide Rankings of Economics Departments and Economists*. Mimeo ECARES, Université Libre de Bruxelles.

COUPÉ, T. (2000b). *Statistics 101 for (Wannabe) Economists*. Working paper, available online at: http://student.ulb.ac.be/~tcoupe/

COUPÉ, T. (2003a), *The Prize is Right – An Analysis of Best Paper Prizes*. Mimeo EERC, NAUKMA, Kyiv.

COUPÉ, T. (2003b). *What Do We Know about Ourselves? On the Economics of Economics*. Working paper, available online at: http://student.ulb.ac.be/~tcoupe/

DARNELL, A.C. & EVANS, J.L. (1990). *The Limits of Econometrics*. Hants: England, Edward Elgar Publishing Limited.

DARNELL, A.C. (1997). Imprecise Tests and Imprecise Hypothesis. *Scottish Journal of Political Economy*, 44(3): 247 – 268.

DAVID, P.A. (1985). Clio and the Economics of QWERTY. *The American Economic Review*, 75(2): 332 – 337.

DAVIDSON, R. & MACKINNON, J. (1993). *Estimation and inference in econometrics*. Oxford: Oxford University Press.

DAVIS, W.L. (2004). Preference Falsification in the Economics Profession. *Econ Journal Watch*, 1(2): 359 – 368.

DE MARCHI, N. (1988). Popper and LSE economists, in De Marchi, N. (ed.) *The Popperian Legacy in Economics*, Cambridge: Cambridge University Press, pp. 139 – 166.

DE MARCHI, N., GILBERT, C. (Eds.). (1989). *History and Methodology of Econometrics*, New York: Oxford University Press.

DEATON A & MUELBAUER J (1980). An Almost Ideal Demand System. *The American Economic Review*, 70(3):312–326.

DEATON A & MUELBAUER J (1999). *Economics and consumer behaviour*. Cambridge: Cambridge University Press.

DENNISON, D.B. & LYNE, M.C. (1997). Analysis and prediction of water treatment costs at the DV Harris Plant in the Umgeni Catchment Area. *Agrekon*, 36(1): 27 – 45.

DEPARTMENT OF AGRICULTURE (2002). *Various reports with appendices.* Pretoria: Directorate: Agricultural Statistics.

DEPARTMENT OF AGRICULTURE (2003). *Livestock Figures*. Pretoria: Directorate: Animal Health.

DEVADOSS, S. & MEYERS, W.H. (1987). Relative Prices and Money: Further Results for the United States. *American Journal of Agricultural Economics*, 69(4): 838 – 842.

DEWALD, W.G., THURSBRY, J.G. & ANDERSON. R.G. (1986). Replication in Empirical Economics: *The Journal of Money, Credit and Banking* Project. *The American Economic Review*, 76(4): 587 – 603.

DIRECTORATE: AGRICULTURAL STATISTICS. (2002). *Abstract of Agricultural Statistics.* Pretoria: Department of Agriculture.

DIRECTORATE: AGRICULTURAL STATISTICS. (2003). *Abstract of Agricultural Statistics.* Pretoria: Department of Agriculture.

DIRECTORATE: AGRICULTURAL STATISTICS. (2004). *Abstract of Agricultural Statistics.* Pretoria, Department of Agriculture.

DIRECTORATE: AGRICULTURAL STATISTICS. (2005). *Abstract of Agricultural Statistics.* Pretoria: Department of Agriculture.

DIRECTORATE: AGRICULTURAL STATISTICS. (2006). *Abstract of Agricultural Statistics*. Pretoria: Department of Agriculture.

DöCKEL, J.A. & GROENEWALD, J.A. (1970). The demand for food in South Africa. *Agrekon*, 9(4): 15 – 20.

DU TOIT, J.P.F. & DöCKEL, J.A. (1978) An analysis of the demand for bananas in South Africa. *Agrekon*, 17(1): 5 – 10.

DUSHMANITCH, V.Y. & DARROCH, M.A.G. (1992). The impacts of monetary policy on the maize and beef sectors of South Africa II: Model estimation and simulation results. *Agrekon*, 31(1): 3 – 11.

ELLIOTT, G. & GRANGER, C.W.J. (2004). Evaluating significance: comments on “size matters”. *Journal of Socio-Economics*, 33(5): 547 – 550.

ESSA, J.A. & NIEUWOUDT, W.L. (2001). Determinants of hybrid maize seed and fertiliser adoption by emerging farmers in communal areas of KwaZulu-Natal. *Agrekon*, 40(4): 537-548.

FEATHERSTONE, A.M. & BAKER, T.G. (1987). An Examination of Farm Sector Real Asset Dynamics: 1910 – 85. *American Journal of Agricultural Economics*, 69(3): 532 – 546.

FERRIS, J.N. (1998). *Agricultural Price and Commodity Market Analysis*. New York: McGraw-Hill.

FIDLER, F. (2002). The fith edition of the *APA Publication Manual*: why its statistics recommendations are so controversial. *Educational and Psychological Measurement*, 62(5): 749 – 770.

FIDLER, F., CUMMING, G., BURGMAN, M. & THOMASON, N. (2004). Statistical reform in medicine, psychology and ecology. *Journal of Socio-Economics*, 33(5): 615 – 630.

FIRSCH, R. (1970). Econometrics in the World Today. In Mayer, T. (1980). Economics as a hard science: Realistic goal or wishful thinking? *Economic Inquiry*, 18(2): 165 – 178.

FISHER, F.M. (1961). On the cost of approximate specification in simultaneous equation estimation. *Econometrica*, 29(2): 139 – 170.

FRIEDMAN, M. & SCHWARTZ, A.J. (1963). *A Monetary History of the United States, 1867 – 1960*. Princeton: Princeton University Press (for National Bureau of Economic Research.

FRIEDMAN, M. & SCHWARTZ, A.J. (1991). Alternative Approaches to Analysing Economic Data. *American Economic Review*, 81(1): 39 – 49.

FRIEDMAN, M. (1953). *The Methodology of Positive Economics, in Essays in Positive Economics*. Chicago: University of Chicago Press.

FRICK, A. & GROENEWALD, J.A. (1999). The need for agricultural information and data: Researchers and policy makers. *Agrekon*, 40(1): 104 – 117.

FRICK, A. & GROENEWALD, J.A. (1999a). The need for agricultural information in the new South Africa. *Agrekon*, 38(2): 241 – 254.

FRICK, A. & GROENEWALD, J.A. (1999b). Agricultural statistics needs of agribusiness in South Africa. *Agrekon*, 38(4): 526 – 534.

GARDNER, B. (1983). Fact and Fiction in the Public Data Budget Crunch. *American Journal of Agricultural Economics*, 65(5): 882 – 888.

GARNAP, R. (1995). *An Introduction to the Philosophy of Science*. Edited by Martin Gardner, New York: Dover Publications.

GAY, S.H. & W.L. NIEUWOUDT. (1999). Results of a trade simulation model for the South African fresh orange industry. *Agrekon*, 38(4): 707 – 715.

GEWEKE, J. & MESSE, R. (1981). Estimating regression models of finite but unknown order. *International Economic Review*, 22(1): 55 – 70.

GOEDECKE, E.J., DUSHMANITCH, V.Y. & ORTMANN, G.F. (1992). Exchange rate determination using a linear regression model: A monetary approach. *Agrekon*, 31(1): 33 – 35.

GOLDBERGER, A.S. (1991). *A course in econometrics*. Cambridge: Harvard University Press.

GRANGER, C.W.J. & NEWBOLD, P. (1976). The use of R^2 to determine the appropriate transformation of regression variables. *Journal of Econometrics*, 4(3): 205 – 210.

GRANGER, C.W.J. (1994). A Review of Some Recent Textbooks of Econometrics. *Journal of Economic Literature*, 32(1):115 – 122.

GREENE, C.A. (2000). I am not, nor have I ever been a member of a data-mining discipline. *Journal of Economic Methodology*, 7(2): 217 – 230.

GREENE, W.H. (2000). *Econometric Analysis*. Fourth Edition, Prentice Hall.

GRIFFITHS, W.E., HILL, R.C. & JUDGE, G.G. (1993). *Learning and practicing econometrics*. Wiley: New York.

GROENEWALD, J.A. (1964). Changes in primary resources in South African agriculture. *Agrekon*, 3(3): 22 – 29.

GROENEWALD, J.A. (1990). Effectiveness and efficiency of experts: An evaluation of agricultural economists. *Agrekon*, 29(4): 243 – 247.

GROENEWALD, J.A. (1989). Discrepancies in agricultural data from two official sources. *Agrekon*, 28(1): 90 – 91.

GROENEWALD, J.A. & GRAHAM, J.P (1968). The use of multiple regression to determine grazing densities from survey data. *Agrekon*, 7(2): 31 – 32.

GUJARATI, D. N. (1998). *Basic Econometrics*, Third Edition, Singapore: McGraw-Hill Inc.

GUNNELSON, G., DOBSON, W.D. & PAMPERIN, S. (1972). Analysis of the Accuracy of USDA Forecasts. *American Journal of Agricultural Economics*, 54(3): 639 – 645.

HAAVELMO, T. (1944). The Probability Approach in Econometrics. Supplement to *Econometrica*, 12: 1 – 118.

HAAVELMO, T. (1958). The role of the econometrician in the advancement of economic theory, *Econometrica*, 26(3): 351 – 357.

HANCOCK, P.J., NIEUWOUDT, W.L. & LYNE, M.C. (1984). Demand analysis of meat in South Africa. *Agrekon*, 23(2): 26 – 29.

HARVEY, A.C. (1984). A Unified View of Statistical Forecasting Procedures. *Journal of Forecasting*, 3(3): 245 – 275.

HAUSMAN, D. (1984). (ed.). *The Philosophy of Economics: An Anthology*. Cambridge: Columbia University Press.

HAYWARD-BUTT, P.R.N. & ORTMANN, G.F. (1994). Demand analysis of oranges in South Africa. *Agrekon*, 33(3): 141 – 144.

HELLAND, I.S. (1987). On the Interpretation and Use of R^2 in Regression Analysis. *Biometrics*, 43(1): 61 – 69.

HENDRY, D.F. (1980). Econometrics – Alchemy or Science? *Economica*, 47(188): 387 – 406.

HENDRY, D.F. (1986). Econometric modelling with cointegrated variables: An overview. *Oxford Bulletin of Economics and Statistics*, 48(3): 201 – 212.

HENDRY, D.F. & ERICSSON, N.R. (1991). An Econometric Analysis of U.K. Money Demand in Monetary Trends in the United States and the United Kingdom by Milton Friedman and Anna J. Schwartz. *American Economic Review*, 81(1): 8 – 38.

HENDRY, D.F. & KROLZIG, H. (1999). Improving on 'Data mining reconsidered' by K. D. Hoover and S. J. Perez. *Econometrics Journal*, 2(2): 202 – 219.

HENDRY, D.F. & KROLZIG, H. (2004). Automatic model selection: a new instrument for social science. *Electoral Studies*, 23: 525 – 544.

HENDRY, D.F. & MIZON, G.E. (1978). Serial Correlation as a Convenient Simplification, Not a Nuisance: A Comment on a Study of the Demand for Money by the Bank of England. *Economic Journal*, 88(351): 549 – 563.

HENDRY, D.F. & MIZON, G.E. (1985). Procrustean econometrics. In Pagan, A. R. (1987). Three econometric methodologies: A critical appraisal. *Journal of Economic Surveys*, 1(1): 3 – 24.

HENDRY, D.F. & MORGAN, M.S. (1995). *The History of Econometric Ideas*. New York: Cambridge University Press.

HENDRY, D.F. & RICHARD, J.F. (1982). On the formulation of empirical models in dynamic econometrics. *Journal of Econometrics*, 20(1): 3 – 33.

HENDRY, D.F. & RICHARD, J.F. (1983). The econometric analysis of economic time series. *International Statistical Review*, 51(1): 3 – 33.

HIGH, R. (2000). http://cc.uoregon.edu/cnews/summer2000/statpower.html

HOCH, I. (1984). Retooling the Mainstream. *American Journal of Agricultural Economics*, 66(5): 793 – 794.

HOOVER, K.D. & PEREZ, S.J. (2000). Data Mining Reconsidered: Encompassing and the General to Specific Approach to Specification Search. *Econometrics Journal*, 2(2): 167 – 191.

HOOVER, K.D. & SIEGLER, M.V. (2005). *Sound and Fury: McCloskey and Significance Testing in Economics*. Available online at: http://econpapers.repec.org/paper/wpawuwpem/0511018.html

HORROWITZ, J.L. (2004.) Comments on “size matters”. *Journal of Socio-Economics*, 33(5): 551 – 554.

HYLLEBERG, S. & PALDAM, M. (1991). New Approaches to Empirical Macroeconomics. *Scandinavian Journal of Economics*, 93(2): 121 – 128.

JANSEN, A.A., SWANEPOEL, G.H. & GROENEWALD, J.A. (1972). The measurement of managerial inputs in agriculture IV : Applications with business results. *Agrekon*, 11(2): 5 – 14.

JOOSTE, A & VAN SCHALKWYK, H.D. (1999). The impact of different trade scenarios on selected agricultural commodities in South Africa *Agrekon* , 38(Special edition): 414 – 424.

JOFFE, H. (2003). Analysts turn spotlight on new Sanlam CE’s promise of tight focus. *Business Day*, 31 March 2003, p. 12

JOHANSEN, S. (1995). *Likelihood-based inference in cointegrated vector-autoregressive models*. New York: Oxford University Press.

JOOSTE, A., ALIBER, M. & VAN SCHALKWYK, H.D. (1998). Indirect effects of different agricultural trade scenarios: A South African case study. *Agrekon*, 37(4): 572 – 581.

JOUBERT, J.S.G. & VILJOEN, P. (1974). Production costs of crops in the North-Western Free State and the factors which influence these costs. *Agrekon*, 13(3): 10 – 19.

JUST, R.E. (1983). The Impact of Less Data on the Agricultural Economy and Society. *American Journal of Agricultural Economics*, 65(5): 872 – 881.

KALABA, M. & HENNEBERRY, S.R. (2001). The effects of a free trade agreement on South African Agriculture: Competitiveness of fruits in the EU market. *Agrekon*, 40(4): 794 – 809.

KAYLEN, M.S. (1988). Vector Autoregression Forecasting Models: Recent Developments Applied to the U. S. Hog Market. *American Journal of Agricultural Economics*, 70(3): 701 – 712.

KENNEDY, P.A. (1992). A Guide to Econometrics. Third edition, Cambridge: The MIT Press.

KEUZENKAMP, H.A. & MAGNUS, J.R. (1995). On tests and significance in econometrics. *Journal of Econometrics*, 67(1): 5 – 24.

KILLE, G.S. & LYNE, M.C. (1993). Investment on freehold and trust farms: Theory with some evidence from Kwa Zulu. *Agrekon*, 32(3): 101 – 109.

KIRK, R.E. (1995). *Experimental Design: Procedures for the Behavioral Science*. 3rd Edition, Pacific Grove, California: Brooks/Cole Publishing.

KIRSTEN, G.C.J. & GROENEWALD, J.A. (1970). Changes in karakul pelt prices between 1952 and 1969. *Agrekon*, 9(4): 9 – 11.

KIRSTEN, J. F. & VAN ZYL, J. (1996). The contemporary agricultural policy environment: undoing the legacy of the past. In: J. F Kirsten, J. Van Zyl and H.P. Binswanger, eds., *Agricultural land reform in South Africa: policies, markets and mechanisms*. Cape Town: Oxford University Press.

KIRSTEN, J.F. (2002). Forty years of agricultural economics scholarship and practice in South Africa: a time to challenge the consensus and refocus our intellectual work: presidential address. *Agrekon*, 41(4): 251 – 284.

KIRSTEN, J.F. & MOLDENHAUER, W.H. (2006). Measurement and Analysis of Rural Household Income in a Dualistic Economy: The Case of South Africa. *Agrekon*, 45(1): 60 – 67.

KLEIN, D.B. (2004). The Ph.D. Circle in Academic Economics. *Econ Journal Watch*, 2(1): 133 – 148.

KLEIN. L.R. (1950). *Economic Fluctuations in the United States, 1921 – 41*. John Wiley & Sons, New York.

KLEIN, L.R. & KOSOBUD, R.F. (1961). Some econometrics of growth: Great ratios in economics. *Quarterly Journal of Economics*, 75(2): 173 – 198.

KUHN, M.E., DARROCH, M.A.G. & ORTMANN, G.F. (2000). Assessing the efficacy of a South African microlender's loan screening mechanism. *Agrekon*, 39(4): 699 – 707.

LANGLEY, D.S. & DU TOIT, J.P.F. (1978). A statistical analysis of the supply of maize in the Transvaal. *Agrekon*, 17(1): 11 – 17.

LATT, E.A. & NIEUWOUDT, W.L. (1985). A supply and demand analysis of regular Black labour in Natal. *Agrekon*, 24(2): 1 – 8.

LAZEAR, E.P. (2000). Economic Imperialism. *Quarterly Journal of Economics*, 115(1): 99 – 146.

LEAMER, E.E (1983). Let's take the con out of econometrics *American Economic Review*, 73(1): 31-43.

LEAMER, E.E. & LEONARD, H. (1983). Reporting on the fragility of regression estimates. *Review of Economics and Statistics*, 65(2): 306 – 317.

LEAMER, E.E. (2004). Are the roads red? Comments on "Size Matters". *Journal of Socio-Economics*, 33(5): 555 – 557.

LEAVER, R. (2004). Measuring the supply response function of tobacco in Zimbabwe. *Agrekon*, 43(1): 113 – 131.

LEIJONHUFVUD, A. (1973). Life among the econ. *Western Economic Journal*, 11(3): 327 – 337.

LENTZ, W. & VAN ZYL, J. (1990). Dynamic optimal control of plant production: An application to determine optimal greenhouse temperature strategies. *Agrekon*, 29(3): 171 – 177.

LEONTIEF, W. (1971). Theoretical assumptions and non-observed facts. *American Economic Review*, 61(1): 1 – 7.

LEONTIEF, W. (1971). Theoretical assumptions and non-observed facts, *American Economic Review*, 61(1): 1 – 7.

LIEBENBERG, G.F., VIVIER, F.L. & GROENEWALD, J.A. (1991). Exchange rates and South African wool prices. *Agrekon*, 30(2): 95 – 96.

LINEHAN, K. (ed.)(2003).*Strange case of Dr. Jekyll and Mr. Hyde : an authoritative text, backgrounds and contexts, performance adaptations, criticism*. New York: Norton.

LIPSEY, R.G. (1960). The relation between unemployment and the rate of change of money wage rates in the United Kingdom, 1862 – 1957: A further analysis. *Economica*, 27(105): 1 – 31.

LIPSEY, R.G. (1975). An introduction to positive economics. Fourth edition, London: Weidenfeld and Nicolson.

LOHR, S. L. (1999). *Sampling: Design and Analysis*. California: Duxbury Press.

LOVELL, M.C. (1983). Data mining. *Review of Economics and Statistics*, 65(1): 1 – 12.

LUCAS, R.E. (1976). Econometric Policy Evaluation: A Critique. In Brunner, K. & Meltzer, A. H. (Eds.). *The Phillips Curve and the Labor Market*. Amsterdam, North Holland, Carnegie Rochester Conference Series, 1: 19 – 46.

LUI, T.L. (1960). Underidentification, Structural Estimation, and Forecasting. *Econometrica*, 28(4): 855 – 865.

LUNT, P. (2004). The significance of the significance test controversy: comments on 'Size Matters'. *Journal of Socio-Economics*, 33(5): 559 – 564.

LYNE, M.C. & GRAHAM, D.H. (2001). The impact of land redistribution on tenure security and agricultural performance in KwaZulu-Natal. *Agrekon*, 40(4): 655 – 667.

LYNE, M.C., THOMSON, D.N. & ORTMANN, G.F. (1996). Institutional change to promote land rental markets in the developing regions of Southern Africa. *Agrekon*, 35(1): 12 – 19.

MACHLUP, F. (1978). *Methodology of Economics and Other Social Sciences*. New York: Academic Press San Diego.

MADALA, G.S. (1992). *Introduction to Econometrics*. Second edition, New York: Macmillan.

MAHABILE, M.; LYNE, M., PANIN, A. (2005). An empirical analysis of factors affecting the productivity of livestock in southern Botswana. *Agrekon*, 44(1): 99 – 117.

MALINVAUD, E. (1966). Statistical Methods of Econometrics. Chicago: Rand McNally.

MAKRIDAKIS, S., WHEELWRIGHT, S.C. & MCGEE, V.E. (1998). *Forecasting: Methods and Application.* Second edition, John Wiley and Sons.

MAYER, T (1980) Economics as a Hard Science: Realistic Goal or Wishful Thinking? Economic Inquiry, 18(2): 165 – 178.

MAYER, T. (2000). Data mining: a reconsideration. *Journal of Economic Methodology*, 7(2): 183 – 194.

MAYES, D. G. (1981). *Applications of Econometrics*. London: Prentice-Hall.

MCALEER, M. PAGAN, A. R. & VOLKER, P. A. (1985). What will take the con out of econometrics? *American Economic Review*, 75(3): 293 – 307.

MCCARL, B.A. & ONAL, H. (1989). Linear Approximation Using MOTAD and Separable Programming: Should It Be Done? *American Journal of Agricultural Economics*, 71(1): 158 – 166.

MCCLOSKEY, D.N. & J. ZECHER, J.R. (1984). The Success of Purchasing Power Parity. In: Bordo, M. D. & Schwartz, A. J., eds. *A Retrospective on the Classical Gold Standard, 1821-1931*. Chicago: University of Chicago Press and the National Bureau of Economic Research.

MCCLOSKEY, D.N. (1983). The Rhetoric of Economics. *Journal of Economic Literature*, 21(2): 481 – 517.

MCCLOSKEY, D.N. (1985a). *The Rhetoric of Economics*. First edition. Madison: University of Wisconsin Press.

MCCLOSKEY, D.N. (1985b). The Loss Function Has Been Mislaid: The Rhetoric of Significance Tests. *American Economic Review*, 75(2): 201 – 205.

MCCLOSKEY, D.N. (1990). *Agon* and *Ag Ec*: Styles of Persuasion in Agricultural Economics. *American Journal of Agricultural Economics*, 72(5): 1124 – 1130.

MCCLOSKEY, D.N. (1992). Other Things Equal: The Bankruptcy of Statistical Significance. *Eastern Economic Journal*, 18(3):359 – 361.

MCCLOSKEY, D.N. (1994). *Knowledge and Persuasion in Economics*. Cambridge: Cambridge University Press.

MCCLOSKEY, D.N. (1995a). The Insignificance of Statistical Significance. *Scientific American*, 272(4):32 – 33.

MCCLOSKEY, D.N. (1995b). Computation Outstrips Analysis. *Scientific American*, 272(7): 26.

MCCLOSKEY, D.N. (1997). Other Things Equal: Aunt Deirdre's Letter to a Graduate Student. *Eastern Economic Journal*, 23(2): 241 – 244.

MCCLOSKEY, D.N. (1998). *The Rhetoric of Economics.* Second edition, Madison: University of Wisconsin Press.

MCCLOSKEY, D.N. (1999). Other Things Equal: Cassandra's Open Letter to Her Economist Colleagues. *Eastern Economic Journal*, 25(3): 357 – 363.

MCCLOSKEY, D.N. (2002). *The Secret Sins of Economics*. Chicago: Prickly Paradigm Press. Available online at http://www.prickly-paradigm.com/paradigm4.pdf

MCCLOSKEY, D.N. (2005). The Trouble with Mathematics and Statistics in Economics. In Hoover, K.D. & Siegler. M.V. *Sound and Fury: McCloskey and Significance Testing in Economics.* EconPapers Working Paper, http://129.3.20.41/eps/em/papers/0511/0511018.pdf

MCGUIGAN, S.M. & NIEUWOUDT, W.L. (2001). Projecting international demand for and supply of protein feed. *Agrekon*, 40(3): 361 – 392.

MCGUIRK, A.M. & DRISCOLL, P. (1995). The Hot Air in R^2 and Consistent Measures of Explained Variation. *American Journal of Agricultural Economics*, 77(2): 319 – 328.

MCGUIRK, A.M., DRISCOLL, P. & ALWANG, J. (1993). Misspecification Testing: A Comprehensive Approach. *American Journal of Agricultural Economics*, 75(4): 1044 – 1055.

MEYER, F.H. & KIRSTEN, J.F. (2005). Modelling the wheat sector in South Africa. *Agrekon*, 44(2): 225 – 237.

MIZON, G.E. (1977). Model selection procedures. In M. J. Artis and A. R. Nobay (eds.). *Studies in Modern Economic Analysis*. Oxford: Basil Blackwell.

MKHABELA, T.S. (2003). Economic feasibility of using composted feedlot manure on dryland maize *Agrekon*, 42(1): 60 – 70.

MORGAN, M.S. (1990). *The History of Econometric Ideas*, New York: Cambridge University Press.

MOHAMMED, M.A., ORTMANN, G.F. (2005). Factors influencing adoption of livestock insurance by commercial dairy farmers in three Zobatat of Eritrea. *Agrekon*, 44(2): 172 – 186.

MORRISON, D.E. & HENKEL, R.E. (1970). *The Significance Test Controversy: A Reader*. Chicago: Aldine.

MOUNT, T.D. (1989). Policy Analysis with Time-Series Econometric Models: Discussion. *American Journal of Agricultural Economics*, 71(2): 507 – 508.

MURPHY, K.R. (1998). *Statistical Power Analysis: A Simple and General Model for Traditional and Modern Hypothesis Tests*. Mahwah: L. Erlbaum Associates.

MUSHUNJE, A. & BELETE, A. (2001). Efficiency of small scale communal farmers of Zimbabwe. *Agrekon*, 40(3): 344 – 360.

NEETHLING, A. (2001). *Sampling in practice*. Pretoria: University of Pretoria.

NERLOVE, M. (1956). Estimates of the elasticity of supply of selected agricultural commodities. *Journal of Farm Economics*, 38(2): 496 – 512.

NIEBUHR, H.G. & VAN ZYL, J. (1990). 'n Ekonomiese ontleding van die aanbod van koring in Suid-Afrika : 'n Streeksbenadering. *Agrekon*, 29(4): 324 – 334.

NIEBUHR, H.G. & VAN ZYL, J. (1992). Die vraag na brood in Suid-Afrika, 1984-90: 'n Streeksbenadering. *Agrekon*, 31(3): 128 – 135.

NIEUWOUDT, W.L. (1973). Agricultural policy research: Some thoughts concerning methodology and technology. *Agrekon*, 12(1): 20 – 26.

NIEUWOUDT, W.L. (1973). The maize/meat price gap. *Agrekon*, 12(4): 37 – 40.

NIEUWOUDT, W.L. (1983). An economic analysis of demand and supply functions for wheat (bread) in South Africa, 1948-1981. *Agrekon*, 22(2): 19 – 23.

NIEUWOUDT, W.L. (1992). Agricultural policy research: Some thoughts concerning methodology and technology. *Agrekon*, 31(1): 43 – 46.

NIEUWOUDT, W.L. & BEHRMANN, H.I. (1976). The effect of the weather on the economic optimum level of fertiliser use. *Agrekon*, 15(4): 14 – 16.

NIEUWOUDT, W.L. & DöCKEL, J.A. (1966). The determination of the economic optimum level of fertiliser application of silage production in the Eastern Cape. *Agrekon*, 5(4): 4 – 18.

NIEUWOUDT, W.L., BULLOCK, J.B. & MATHIA, G.A. (1976). An Economic Evaluation of Alternative Peanut Policies. *American Journal of Agricultural Economics*, 58(3): 485 – 495.

NIEUWOUDT, W.L. & HOWELL, J. (2000). Farming without drought relief: Time to revisit and income equalisation deposit scheme? *Agrekon*, 39(3): 332 – 356.

NIEUWOUDT, T.W. & NIEUWOUDT, W.L. (2004). Privatising agricultural R&D, an example from the South African sugar industry. *Agrekon*, 43(2): 228 – 243.

NIEUWOUDT, T.W. & NIEUWOUDT, W.L. (2004). The rate of return on R&D in the South African Sugar Industry, 1925 – 2001. *Agrekon*, 43(3): 265 – 275.

OELLERMANN, R.G. & DARROCH, M.A.G. (1994). Estimating wetland preservation values : A Wakkerstroom case study. *Agrekon*, 33(4): 202 – 205.

ORDEN, D. & FACKLER, P.L. (1989). Identifying Monetary Impacts on Agricultural Prices in VAR Models. *American Journal of Agricultural Economics*, 71(2): 495 – 502.

ORTMANN, G.F. (1982). Demand analysis of vegetables and subtropical fruit in South Africa. *Agrekon*, 21(2): 15 – 21.

PAGAN, A.R. (1987). Three econometric methodologies: A critical appraisal. *Journal of Economic Surveys*, 1(1): 3 – 24.

PAGAN, A.R. & VEALL, M.R. (2000). Data mining and the econometrics industry: comments on the papers of Mayer and of Hoover and Perez. Journal of Economic Methodology, 7(2): 211 – 216.

PANIN, A. (1999). The economic impacts of education on smallholder crop production systems in Africa: Empirical evidence from Botswana. *Agrekon*, 38(2): 183 – 193.

PESARAN, H.M. (1987). Econometrics. In Eatwell, J., Milgate, M. and Newman, P. (eds.). *The New Palgrave: A Dictionary of Economics* (4 vols.) New York: Stocktron Press; London Macmillan Press, 9 – 22.

PHORORO, H. (1996). The supply of wool in Lesotho. *Agrekon*, 35(1): 20 – 28.

PIERCE, D.A. (1979). R^2 Measures for Time Series. *Journal of the American Statistical Association*, 74(368): 901 – 910.

PINDYCK, S. & RUBINFELD, D.L. (1998). *Econometric Models and Econometric Forecasts*, Fourth Edition. McGraw Hill.

POONYTH, D. & VAN ZYL, J. (2000). Representing the production structure of South African agriculture. *Agrekon*, 39(4): 620 – 632.

POONYTH, D. & VAN ZYL, J. (2000). The impact of real exchange rate changes on South African agricultural exports: An error correction model approach. *Agrekon*, 39(4): 673 – 685.

POONYTH, D. & VAN ZYL, J., MEYER, F. (2000). Forecasting the market outlook for the South African maize and sorghum sector using econometric modelling. *Agrekon*, 39(4): 607 – 619.

POPPER, K.R. (1968). *The logic of scientific discovery*. Second edition, New York: Haper & Row.

POPPER, K.R. (1972). *Objective Knowledge. An Evolutionary Approach.* London: Oxford University Press.

RANDELA, R., LIEBENBERG, G.F., KIRSTEN, J.F. & TOWNSEND, R.F. (2000). Demand for livestock tick control service in the Venda region, Northern Province. *Agrekon*, 39(4): 644 – 655.

RANNEY, G.B. & THIGPEN, C.C. (1981). The Sample Coefficient of Determination in Simple Linear Regression. *The American Statistician*, 35(3): 152 – 153.

RAO, P. & MILLER, R.L. (1971). *Applied Econometrics*. Belmont: Wadsworth.

ROODT, I.J. & LUBBE, W.F. (1990). Bemarkings-doeltreffendheid in die akwakultuurbedryf. *Agrekon*, 29(4): 341 – 346.

SAMUELSON, P.A., KOOPMANS, T.C. & STONE, J.R.N. (1954). Report of the

SARGAN, J.D. (1964) Wages and prices in the United Kingdom: a study in econometric methodology. In Hart, P. E., Mills, G. and Whitaker, J. K. (eds) *Econometric Analysis for National Economic Planning*. London: Butterworth.

SARGENT, T.J. (1981). Interpreting Economic Time Series. *Journal of Political Economy*, 89(2): 213 – 248.

SARTORIUS VON BACH, H.J., TOWNSEND, R.F. & VAN ZYL, J. (1998).Technical inefficiency of commercial maize producers in South Africa : A stochastic frontier production function approach. *Agrekon*, 37(2): 162 – 170.

SARTORIUS VON BACH, H.J. & VAN ZYL, J. (1994). Human carbohydrate demand in South Africa. *Agrekon*, 33(3): 145 – 150.

SAS INSTITUTE. (2004). *Getting Started with the SAS Power and Sample Size Application*. Cary, North Carolina: SAS Institute.

SAS INSTITUTE. (2004). *SAS/STAT User's Guide, Version 9.1*. Cary: North Carolina, SAS Institute.

SCHIMMELPFENNIG, D., THIRTLE, C. & VAN ZYL, J. (1996). Crop level supply response in South African agriculture: An error correction approach. *Agrekon*, 35(3): 114 – 122.

SIMS, C.A. (1980a). Macroeconomics and reality. *Econometrica*, 48(1): 1 – 47.

SIMS, C.A. (1980b). Comparison of Interwar and Postwar Business Cycles: Monetarism Reconsidered. *American Economic Review*, 70(2): 250 – 257.

SOLOW, R.M. (1997). How did economics get that way and what way did it get? *Daedalus*, Winter, 126(1): 39 – 58

SPANOS, A. (2000). Revisiting data mining: 'hunting' with or without a license. *Journal of Economic Methodology*, 7(2): 231 – 264.

STATISTICS SOUTH AFRICA (2005). *Census of commercial agriculture 2002: Financial and production statistics*. Report No. 11-02-01 (2002), Pretoria: Statistics South Africa.

STONE, R. & STONE, W.M. (1939). The marginal propensity to consume and the multiplier: a statistical investigation. *Review of Economic Studies*, 6: 1 – 24.

STEENKAMP, P.J.D., SARTORIUS VON BACH, H.J. VIVIER, L. & MILLARD, S. (1994). Marketing margin analysis of South African potatoes. *Agrekon*, 34(3): 109 – 119.

STEYN, A.G.W., SMIT, C.F., DU TOIT, S.C.H. & STRASHEIM, C. (1998). *Modern statistics in practice.* Pretoria: J. L. Van Schaik.

SUITS, D.B. (1955). An Econometric Model of the Watermelon Market. *Journal of Farm Economics*, 37(1): 237 – 251.

SUMMERS, L.H. (1991). The Scientific Illusion in Empirical Macroeconomics. *Scandinavian Journal of Economics*, 93(2): 129 – 148.

SUMNER, A. & MUELLER, A.E. (1989). Are Harvest Forecasts News? USDA Announcements and Futures Market Reactions. *American Journal of Agricultural Economics*, 71(1): 1 – 8

SUPPE, F. (1977). The search for philosophic understanding of scientific theories. In F. Suppe (ed.). *The Structure of Scientific Theories*. Second edition. Illinois: University of Illinois Press.

TALJAARD, P.R., ALEMU, Z.G. & VAN SCHALKWYK, H.D. (2004). The demand for meat in South Africa: An almost ideal estimation. *Agrekon*, 43(4): 430 – 443.

THEIL, H. (1978). *Introduction to econometrics*. New Jersey: Prentice-Hall.

THOMAS, L. & KREBS, C. J. (1997). A Review of Statistical Power Analysis Software. *Bulletin of the Ecological Society of America*, 78(2): 126 – 139.

THOMAS, R.L. (1985). *Introductory Econometrics*. London: Longman.

THOMPSON, B. (1999). Why ‘encouraging’ effect size reporting is not working: the etiology of researcher resistance to changing practices. *The Journal of Psychology*, 133(2): 133 – 141.

THOMPSON, B. (2004). The “significance” crisis in psychology and education. *Journal of Socio-Economics*, 33(5): 607 – 613.

THUROW, L.L. (1983) *Dangerous Currents: the State of Economics.* Oxford University Press.

TINBERGEN, J. (1939). *Statistical Testing of Business Cycle Theories, Vol. I and II.* Geneva: League of Nations.

TODD, R.M. (1989). Policy Analysis with Time-Series Econometric Models: Discussion. *American Journal of Agricultural Economics*, 71(2): 509 – 510.

TOMEK, W.G. (1993). Conformation and Replication in Empirical Econometrics: A Step Towards Improved Scholarship. *American Journal of Agricultural Economics*, 75: 6 – 14.

TOWNSEND, R. (1996). Macro-economic impacts on commercial farm income in South Africa. *Agrekon*, 35(4): 236 – 239.

TOWNSEND, R.F. (1997) Econometric Methodology I: The Role of Theory and Observation in Agricultural Economics Research. *Agrekon*, 36(3): 325 – 347.

TOWNSEND, R.F., VAN ZYL, J. & THIRTLE, C. (1997). Assessing the benefits of research expenditures on maize production in South Africa. *Agrekon*, 36(4): 585 – 597.

TOWNSEND, R.F., VAN ZYL, J. & THIRTLE, C. (1997). Machinery and labour biases of technical change in South African agriculture : A cost function approach. *Agrekon*, 36(4): 492 – 500.

TWEETEN, L (1983). Hypothesis Testing in Economic Science. *American Journal of Agricultural Economics*, 65(3): 548 – 552.

VAN DER RIET, D.F. & DARROCH, M.A.G. (1993). Impact of land prices on export demand for South African deciduous fruit: 1972-1992. *Agrekon*, 32(4): 247 – 251.

VAN SCHALKWYK, H.D. & GROENEWALD, J.A. (1992). Die indiensneming van plaaswerkers in die Republiek van Suid-Afrika. *Agrekon*, 31(4): 199 – 204.

VAN SCHALKWYK, H.D. & SWANEPOEL, L. (1997). How free is information in the agricultural market? *Farmer's Weekly*, 9 May 1997: 30 – 32.

VAN SCHALKWYK, H.D., VAN ZYL, J. & SARTORIUS VON BACH, H.J. (1993). Management and returns to farm size: Results of a case study using parametric and non-parametric methodology to measure scale efficiencies. *Agrekon*, 32(4): 252 – 256.

VAN TONGEREN, F., VON MEIJL, H. & SURRY, Y. (2001). Global models applied to agricultural and trade policies: a review and assessment. *Agricultural Economics*, 26(2): 149 – 172.

VAN ZYL, J. (1986). A statistical analysis of the demand for maize in South Africa. *Agrekon*, 25(3): 45 – 51.

VAN ZYL, J. & GROENEWALD, J.A. (1984). Farm size and flood damage proneness : Joint effects on cash flow. *Agrekon*, 23(1): 27 – 31.

VAN ZYL, J. & STAPELBERG, J.S. (1990). Die optimum oordragvoorraad vir mielies in Suid-Afrika. *Agrekon*, 29(3): 185 – 193.

VOSLOO, J.J. & GROENEWALD, J.A. (1969). The demand for apples in South Africa : A statistical analysis. *Agrekon*, 8(4): 21 – 25.

WALLIS, K.F. (1969). Some recent developments in applied econometrics. *Journal of Economic Literature*, 7(1): 71 – 96.

WALLIS, K.F. (1973). *Topics in Applied Econometrics*. London: Gary-Mills.

WILKINSON, L. & Task Force on Statistical Inference APA Board of Scientific Affairs. (1999). Statistical methods in psychology journals: guidelines and explanations. *American Psychologist*, 54: 594 – 604.

WISSING, G.J. & GROENEWALD, J.A. (1987). Agrekon: The First Quarter Century. *Agrekon*, 26(2): 1 – 5.

WOOLDRIDGE, J.M. (2004). Statistical significance is okay, too: comment on “Size Matters”. *Journal of Socio-Economics*, 33(5): 577 – 579.

YULE, G.U. (1926). Why do we Sometimes get Nonsense-Correlations between Time-Series? – A Study in Sampling and the Nature of Time-Series, *Journal of the Royal Statistical Society*, 89(1): 1 – 63.

ZELLNER, A, 1979, Posterior odds ratios for regression hypotheses: General considerations and some results; Reprinted in: *Basic issues in econometrics* Chicago: University of Chicago Press.

ZALM, G. (1998). The relevance of economic modelling for policy decisions. *Economic Modelling*, 15: 309 – 316.

ZELLNER, A. (2004). To test or not to test and if so, how? Comments on “size matters”. *Journal of Socio-Economics*, 33(5): 581 – 586.

ZILIAK, S.T. & MCCLOSKEY, D.N. (1996). The Standard Error of Regression. *Journal of Economic Literature*, 34(1): 97 – 114.

ZILIAK, S.T. & MCCLOSKEY, D.N. (2004a). Size matters: The standard error of regressions in the American Economic Review. *Journal of Socio Economics*, 33(5): 527 – 546.

ZILIAK, S.T. & MCCLOSKEY, D.N. (2004b). Significance redux. *Journal of Socio Economics*, 33(5): 665 – 675.

APPENDIX

The 65 papers sampled for the survey of papers published in *Agrekon*

Author(s)	Title	*Agrekon*		Year
		Vol.	**No.**	
GROENEWALD, J.A.	Changes in primary resources in the South African agriculture.	3	3	1964
GROENEWALD, J.A. & GRAHAM, J.P	The use of multiple regression to determine grazing densities from survey data.	7	2	1968
VOSLOO, J.J. & GROENEWALD, J.A.	The demand for apples in South Africa: A statistical analysis.	8	4	1969
KIRSTEN, G.C.J. & GROENEWALD, J.A.	Changes in karakul pelt prices between 1952 and 1969.	9	4	1970
JANSEN, A.A., SWANEPOEL, G.H. & GROENEWALD, J.A.	The measurement of managerial inputs in agriculture IV: Applications with business results.	11	2	1972
JOUBERT, J.S.G. & VILJOEN, P.	Production costs of crops in the North-Western Free State and the factors which influence these costs.	13	3	1974
NIEUWOUDT, W.L. & BEHRMANN, H.I.	The effect of the weather on the economic optimum level of fertiliser use.	15	4	1976
DU TOIT, J.P.F. & DöCKEL, J.A.	An analysis of the demand for bananas in South Africa.	17	1	1978
LANGLEY, D.S. & DU TOIT, J.P.F.	A statistical analysis of the supply of maize in the Transvaal.	17	1	1978
ORTMANN, G.F.	Demand analysis of vegetables and subtropical fruit in South Africa.	21	2	1982
NIEUWOUDT, W.L.	An economic analysis of demand and supply functions for wheat (bread) in South Africa, 1948-1981.	22	2	1983
VAN ZYL, J. & GROENEWALD, J.A.	Farm size and flood damage proneness: Joint effects on cash flow.	23	1	1984
HANCOCK, P.J., NIEUWOUDT, W.L. & LYNE, M.C.	Demand analysis of meat in South Africa.	23	2	1984
JANSE VAN RENSBURG, B.D.T.	Land prices in South Africa: 1960-1979.	23	2	1984
LATT, E.A. & NIEUWOUDT, W.L.	A supply and demand analysis of regular Black labour in Natal.	24	2	1985
CHADWICK, J.B. & NIEUWOUDT, W.L.	The demand for bananas and the economic effect of supply restriction.	24	2	1985
McKENZIE, C.C. & NIEUWOUDT, W.L.	Estimation of demand and supply functions for fresh and industrial milk in South Africa.	24	2	1985
VAN ZYL, J.	A statistical analysis of the demand	25	3	1986

Author(s)	Title	*Agrekon* Vol.	*Agrekon* No.	Year
	for maize in South Africa.			
BOWMAKER, P.A. & NIEUWOUDT, W.L.	Demand equations for selected South African agricultural commodities.	29	1	1990
BERRY, C.G., DICKS, H.M. & ORTMANN, G.F.	Derivation of appropriate functions for the economic analysis of maize yield responses to fertiliser and rainfall variations at Dundee.	29	2	1990
NIEBUHR, H.G. & VAN ZYL, J.	'n Ekonomiese ontleding van die aanbod van koring in Suid-Afrika : 'n Streeksbenadering.	29	4	1990
ROODT, I.J. & LUBBE, W.F.	Bemarkings-doeltreffendheid in die akwakultuurbedryf.	29	4	1990
CLEASBY, R.C.G. & ORTMANN, G.F.	Demand analysis of eggs in South Africa.	30	1	1991
LIEBENBERG, G.F., VIVIER, F.L. & GROENEWALD, J.A.	Exchange rates and South African wool prices.	30	2	1991
NIEBUHR, H.G. & VAN ZYL, J.	Die vraag na brood in Suid-Afrika, 1984-90 : 'n Streeksbenadering.	31	3	1992
VAN SCHALKWYK, H.D. & GROENEWALD, J.A.	Die indiensneming van plaaswerkers in die Republiek van Suid-Afrika.	31	4	1992
CLEASBY, R.C.G., DARROCH, M.A.G. & ORTMANN, G.F.	Factors affecting the demand for and supply of South African yellow maize exports.	32	1	1993
VAN DER RIET, D.F. & DARROCH, M.A.G.	Impact of land prices on export demand for South African deciduous fruit : 1972-1992.	32	4	1993
VAN SCHALKWYK, H.D., VAN ZYL, J. & SARTORIUS VON BACH, H.J.	Management and returns to farm size: Results of a case study using parametric and non-parametric methodology to measure scale efficiencies.	32	4	1993
HAYWARD-BUTT, P.R.N. & ORTMANN, G.F.	Demand analysis of oranges in South Africa.	33	3	1994
SARTORIUS VON BACH, H.J. & VAN ZYL, J.	Human carbohydrate demand in South Africa.	33	3	1994
OELLERMANN, R.G. & DARROCH, M.A.G.	Estimating wetland preservation values: A Wakkerstroom case study.	33	4	1994
STEENKAMP, P.J.D., SARTORIUS VON BACH, H.J. VIVIER, L. & MILLARD, S.	Marketing margin analysis of South African potatoes.	34	3	1994
PHORORO, H.	The supply of wool in Lesotho.	35	1	1996
SCHIMMELPFENNIG,	Crop level supply response in South	35	3	1996

Author(s)	Title	Agrekon Vol.	Agrekon No.	Year
D., THIRTLE, C. & VAN ZYL, J.	African agriculture : An error correction approach.			
CHANTYLEW, D. & BELETE, A.	A statistical analysis of demand for beef, mutton/goat, pork and chicken in Kenya 1961-1991.	36	1	1997
DENNISON, D.B. & LYNE, M.C.	Analysis and prediction of water treatment costs at the DV Harris Plant in the Umgeni Catchment Area.	36	1	1997
TOWNSEND, R.F., VAN ZYL, J. & THIRTLE, C.	Machinery and labour biases of technical change in South African agriculture: A cost function approach.	36	4	1997
TOWNSEND, R.F., VAN ZYL, J. & THIRTLE, C.	Assessing the benefits of research expenditures on maize production in South Africa.	36	4	1997
SARTORIUS VON BACH, H.J., TOWNSEND, R.F. & VAN ZYL, J.	Technical inefficiency of commercial maize producers in South Africa: A stochastic frontier production function approach.	37	2	1998
ABBOTT, M. &. AHMED, A	The South African wool supply response.	38	1	1999
PANIN, A.	The economic impacts of education on smallholder crop production systems in Africa: Empirical evidence from Botswana.	38	2	1999
ANIM, F.D.K.	Organic vegetable farming in rural areas of the Northern Province.	38	4	1999
GAY, S.H. & W.L. NIEUWOUDT.	Results of a trade simulation model for the South African fresh orange industry.	38	4	1999
JOOSTE, A & VAN SCHALKWYK, H.D.	The impact of different trade scenarios on selected agricultural commodities in South Africa	38	SI	1999
BREITENBACH, M.C. & FÉNYES, T.I.	Maize and wheat production trends in South Africa in a deregulated environment.	39	3	2000
NIEUWOUDT, W.L. & HOWELL, J.	Farming without drought relief: Time to revisit and income equalisation deposit scheme?	39	3	2000
MOHAMMED, M.A., ORTMANN, G.F.	Factors influencing adoption of livestock insurance by commercial dairy farmers in three Zobatat of Eritrea	44	2	2005
RANDELA, R., LIEBENBERG, G.F., KIRSTEN, J.F. & TOWNSEND, R.F.	Demand for livestock tick control service in the Venda region, Northern Province.	39	4	2000
POONYTH, D. & VAN	The impact of real exchange rate	39	4	2000

Author(s)	Title	*Agrekon* Vol.	*Agrekon* No.	Year
ZYL, J.	changes on South African agricultural exports: An error correction model approach.			
MUSHUNJE, A. & BELETE, A.	Efficiency of small scale communal farmers of Zimbabwe.	40	3	2001
MAINARDI, S.	An econometric analysis of factors affecting tropical and subtropical deforestation.	37	1	1998
ESSA, J.A. & NIEUWOUDT, W.L.	Determinants of hybrid maize seed and fertiliser adoption by emerging farmers in communal areas of KwaZulu-Natal.	40	4	2001
AGBOLA, F.W.	A dynamic adjustment model for South African agriculture: 1965-97.	40	4	2001
BEYERS, L. & HASSAN, R.M.	The structure of South African milk production technology: A parametric approach to supply analysis.	40	4	2001
LYNE, M.C. & GRAHAM, D.H.	The impact of land redistribution on tenure security and agricultural performance in KwaZulu-Natal.	40	4	2001
KALABA, M. & HENNEBERRY, S.R.	The effects of a free trade agreement on South African Agriculture: Competitiveness of fruits in the EU market.	40	4	2001
CONRADIE, B.	Wages and wage elasticities for wine and table grapes in South Africa	44	1	2005
MKHABELA, T.S.	Economic feasibility of using composted feedlot manure on dryland maize	42	1	2003
ALEMU, Z.G., OOSTHUIZEN, K. VAN SCHALKWYK, H.D.	Grain-supply response in Ethiopia: an error-correction approach	42	4	2003
LEAVER, R.	Measuring the supply response function of tobacco in Zimbabwe	43	1	2004
NIEUWOUDT, T.W., NIEUWOUDT, W.L.	The rate of return on R&D in the South African Sugar Industry, 1925-2001	43	3	2004
TALJAARD, P.R., ALEMU, Z.G., VAN SCHALKWYK, H.D.	The demand for meat in South Africa : an almost ideal estimation	43	4	2004
MAHABILE, M.; LYNE, M., PANIN, A.	An empirical analysis of factors affecting the productivity of livestock in southern Botswana	44	1	2005
MEYER, F.H., KIRSTEN, J.F.	Modelling the wheat sector in South Africa	44	2	2005