SOME CONTRIBUTIONS TO FINITE-SAMPLE ANALYSIS IN THREE ECONOMETRIC MODELS

A thesis presented for the degree

of Doctor of Philosophy

in Economics

at

the University of Canterbury

Christchurch, New Zealand

by

J.N. Lye

1990

ABSTRACT

In the standard classical regression model the most commonly used procedures for estimation are based on the Ordinary Least Squares Method, which is justified on the basis of well known finite-sample properties. However, this model consists of a number of assumptions, such as, for example, homoskedastic, serially independent and normally distributed disturbances and nonstochastic regressors. By changing these assumptions in one way or another, different estimating situations are created, in many of which the OLS estimator may have no statistical justification at all. Further, alternative estimation methods have often been justified only on the basis of their asymptotic properties, although in practice economists frequently have to base their statistical analysis on a relatively small number of observations. This suggests that the particular estimator to use in any situation should be chosen on the basis of finite-sample considerations.

The analysis of finite-sample properties of commonly used estimators in three well known Econometric models is the focus of this thesis. In particular the three models considered are: the limited-information simultaneous equations model, the nonnormal linear regression model and the nonnormal limited-information simultaneous equation model. The techniques used include the derivation of the estimators' exact distribution and when this is analytically intractable Monte Carlo methods are employed.

The limited-information simultaneous equation model is analyzed in two stages. First, a useful method of numerically

evaluating many of the commonly used estimators, including the two-stage least squares estimator, is presented. Secondly this method is then used, and combined with Monte Carlo analysis, to compare the distributions of the limited-information maximum likelihood and two-stage least squares estimators in misspecified simultaneous equations models. The result of this comparison indicates the superior performance of the limited-information maximum likelihood estimator over the two-stage least squares estimator in both correctly specified and misspecified simultaneous equations models.

Recently, models with possibly nonnormal distributed disturbances have attracted more attention. For such models, independence and uncorrelatedness of the disturbance terms are not equivalent. Using the nonnormal regression model the statistical consequences of distinguishing between independence and uncorrelatedness are considered when the disturbances are Student-t distributed. The results obtained demonstrate that the distinction between the two assumptions is an important one and the consequences of making the wrong assumption can be serious. Consequently, specification tests are also presented which test for uncorrelatedness versus independence in the elliptically symmetric family.

The nonnormal limited-information simultaneous equation model provides a relatively new area of analysis as there are few published results available on the effects of nonnormal disturbances in the limited- information simultaneous equation model. The objective here is to combine the themes pursued separately in the other two models previously considered. However, to narrow the range of possible models that can be examined, attention is focussed only on the exactly-identified simultaneous equation model. This model has a number of interesting features when the reduced-form disturbances are normally distributed. These features are illustrated and then comparisons are made with the same model when the distribution of the disturbances is widened to include the Student-t family. In this case, as for the nonnormal linear regression model, a distinction needs to be made between independently distributed and jointly distributed disturbances. The consequences of these different assumptions are shown to be important; specification tests relating to this distinction are therefore also presented.

ACKNOWLEDGEMENTS

There are a number of people and organizations whose help and support I would like to acknowledge with gratitude: my supervisor, David Giles, for his overseeing of this thesis, for providing comments on all individual chapters and for his encouragement in the furtherance of my academic career; Dorian Owen, for reading and commenting on numerous drafts of chapters; Aman Ullah, for providing many suggestions and comments that improved the presentation and content of Chapter 5; Ray Byron, for his comments on Chapter 5 in his role as discussant at the second International Postgraduate Conference, Perth, November 1988; Robin Harrison, for help in producing the graphs; Robert Davies, for supplying me with a Fortran version of one of his algorithms that was used in Chapter 6; Ian Coope, for suggesting and providing numerous algorithms used; The University Grants Committee, and The Reserve Bank of New Zealand, for providing financial support. Thanks are also due to The Reserve Bank of New Zealand for providing financial assistance so I could attend the second International Post-graduate Conference, Perth, November 1988, and The Department of Economics, University of Melbourne, for providing financial support so I could return to Canterbury University during Summer 1989 to complete this thesis. I would like to thank Karilyn Smith for her care in preparing the final typed copy, and for her patience in deciphering many pages of untidy manuscript. Finally, I would like to thank my parents for their financial and moral support, throughout all the years in which I worked on this thesis.

CONTENTS

CHAPTER		PAGE

CHAPTER PAGE

LIST OF TABLES

CHAPTER 10 PAGE

CHAPTER 11 PAGE

APPENDIX A

LIST OF FIGURES

CHAPTER 1

INTRODUCTORY COMMENTS

1.1 A GENERAL OVERVIEW

Consider the standard linear multiple regression model that appears in all econometric textbooks (e.g. Johnston (1984), Harvey (1981)):

$$y = X\beta + \epsilon , \qquad (1.1)$$

where $y' = (y_1 \ldots y_N)$, X is an $N \times K$ matrix, $\beta' = (\beta_1 \ldots \beta_K)$ is a vector of unknown parameters and $\epsilon' = (\epsilon_1 \ldots \epsilon_N)$ is a vector of disturbances, and where the following conditions are satisfied:

Condition (i) : X is a nonstochastic matrix of rank $K < N$ and has the property that

$$\lim_{N \Rightarrow \infty} \left(\frac{X'X}{N} \right) = Q,$$

where Q is a finite nonsingular matrix. It is further assumed that there are no variables wrongly included in and/or excluded from the X matrix.

Condition (ii) : ϵ has a multivariate normal distribution with mean 0 and covariance matrix $\sigma^2 I$.

The most commonly used procedures for estimation and inference in this model are based on the Ordinary Least Squares (OLS) principle. This principle is justified on the basis of its well known finite-sample properties which are given in Properties 1.1; for proofs see, for example, Schmidt (1976a, pp.6-31).

Properties 1.1

(i) The least-squares estimator $b = (X'X)^{-1}X'y$, which is also the maximum likelihood estimator, and the associated variance estimator $s^2 = (y-Xb)'(y-Xb)/(N-K)$, are unbiased minimum variance estimators from within the class of all unbiased estimators.

(ii) The joint distribution of b is multivariate normal with mean β and variance covariance matrix $\sigma^2(X'X)^{-1}$, implying that the marginal distribution for an element of the b-vector, say b_j, is normal with mean β_j and variance $\sigma^2(X'X)^{-1}_{jj}$.

(iii) The statistic $(N-K)s^2/\sigma^2$ is distributed as a chi-square random variable with N-K degrees of freedom.

(iv) Under the null hypothesis $\beta_j = 0$, the test statistic $b_j/\sqrt{s^2(X'X)^{-1}_{jj}}$ has a Student-t distribution with N - K degrees of freedom.

However, this model is not sufficient as a basis for modelling many economic data generation processes, simply because in many situations conditions (i) and (ii) do not hold. Consequently, Properties 1.1 are not valid in general and, in particular, the use of OLS techniques may have no statistical justification at all. The relaxation of these conditions has enriched the range of econometric models and has consequently led to the development of a number of estimation and inference techniques which are alternatives to those based on OLS. The introduction of most of these techniques however has only been justified on the basis of their asymptotic properties, asymptotic efficiency and asymptotic normality. However, in practice economists frequently have to base their statistical inferences on a relatively small number of sample observations. This suggests that the choice of the appropriate techniques to use should be based on finite-sample considerations such as those given in Properties 1.1, rather than asymptotic behaviour. However, in general, relatively little is known about these relevant finite-sample considerations.

The objective of this thesis is to extend and develop finite-sample results for various estimators used for estimation and inference in three econometric models. The particular econometric models chosen are well-known extensions of the standard multiple linear regression model when conditions (i) or (ii) or a combination of both conditions are relaxed. Further, each of the econometric models chosen provides a basis for much applied econometric analysis and, in particular, all of the estimators considered are now included in standard and widely-used econometric packages such as SHAZAM and TSP.

The next section describes the three econometric models chosen for investigation and so defines the three main components of this thesis. These models are: the limited-information simultaneous equations model, the nonnormal linear regression model and the nonnormal limited-information simultaneous equations model.

1.2 THE MODELS AND OBJECTIVES

(i) The Limited-Information Simultaneous Equations Model

Econometric models typically consist of sets of equations which incorporate feedback effects from one variable to another. These are known as Simultaneous Equation Models (SEMS). In particular, when the econometrician is interested only in making statistical inferences about the parameters of a single equation of the model, then this is known as "The Limited-Information SEM". Writing this model in the form of (1.1) implies that some of the regressors in X are stochastic and are correlated with the disturbance vector ϵ, in the sense that $(1/N)X'\epsilon$ does not tend to the zero vector as the sample size, N, tends to infinity. Therefore condition (i) is invalidated, and furthermore OLS is an inconsistent estimation technique.

The SEM was first proposed by Haavelmo (1943, 1944, 1947) and this suggestion provided the basis for a research programme undertaken by the Cowles Foundation during the late 1940's and early 1950's. However, the estimators suggested, such as Two Stage Least Squares (TSLS) and the Limited Information Maximum Likelihood estimator (LIML), are rather complicated functions of the underlying random variables, so that the exact distributions are difficult to derive. Nonetheless, the analysis of the exact distributions and their moments began in the early 1960's and in recent years substantial progress has been made for the case when all of the predetermined variables are assumed to be exogenous and the equation is identified by means of zero restrictions (e.g. Nagar (1959); Basmann (1961, 1963, 1974); Mariano (1972, 1973a, 1973b, 1977); Hillier, Kinal and Srivastava (1984); Hillier (1985); Phillips (1980a, 1980b, 1984a, 1984b, 1985); Anderson (1982)).

Although the finite-sample properties of certain test-statistics and variance estimators have received some attention in the literature, most results are concerned with the estimation of the parameters of the structural equation of interest and, in particular, the coefficients of the endogenous regressors. It is this topic that is pursued here.

Traditionally a distinction is made between models in which the structural equation of interest contains only one endogenous regressor, and more than one endogenous regressor. This is because it is only recently that techniques have been developed which allow for the derivation of the exact densities in the case of more than one endogenous regressor, and even then these results are complex and currently not suitable for numerical evaluation. Consequently, most numerical evaluations have concentrated simply on the one endogenous regressor case. One of the themes in this case has been the numerical comparison of the distributions of the LIML and TSLS estimators. In particular the numerical computations of Anderson <u>et al.</u> (1979, 1982) have pointed to the superior performance of the LIML procedure over the TSLS estimator. In this thesis this analysis is extended to the comparison of the distributions of the TSLS and LIML estimators when there are predetermined variables wrongly included in and/or excluded from the model.

The numerical procedures used in this thesis differ from those of Anderson <u>et al.</u> (1979, 1982). In particular, as most of the commonly used estimators, including TSLS, can simply be written as a ratio of quadratic forms in normal variables it is shown how the techniques such as those developed by Imhof (1961) and Davies (1973) can be used to compute the distribution functions. This is an extension of the analysis in Cribbett et al. (1989) which concentrates only on the TSLS estimator. In the case of the LIML estimator, however, the nonparametric density estimator is integrated with a simple Monte-Carlo approach to estimate the density, due to the complexity of numerically evaluating the exact expressions.

(ii) The nonnormal linear regression model

When it is assumed that the error distribution is nonnormal, condition (ii) is invalidated. In the literature a distinction is commonly made on the basis of whether the distribution has a finite or infinite variance.

If the error distribution is assumed to have finite first and second moments then the properties of OLS are well-known. The OLS estimator of β is best linear unbiased (BLUE) and the conventional tests are asymptotically justified in the sense that they have the correct size asymptotically. These results have often been the justification for the use of the least squares estimator under conditions of nonnormality. However, there are two problems with this approach. First, it is well-known that although OLS is BLUE it is, in general, asymptotically inefficient. Consequently there may be nonlinear estimators which have superior finite and asymptotic properties. Secondly, there is a large body of literature (e.g. Mandelbrot (1963a, 1963b, 1966), Fama (1963, 1965, 1970), which suggests that many economic data series, particularly prices in financial and commodity markets, are well represented by a class of distributions with infinite variance. A distribution with an infinite variance has "fat tails" which implies that large values or "outliers" will be relatively frequent. Because the least squares technique minimizes squared deviations, it places relatively heavy weight on outliers, leading to estimates that are extremely sensitive to the presence and values of such observations.

In recent years, to broaden the assumption of nonnormality in the linear regression model, it has often been assumed that the error components follow a joint multivariate elliptically symmetric distribution. Under this assumption it has been shown that the resulting estimators and test statistics possess properties which make them analytically tractable and, furthermore, in many cases, identical to those obtained under the normality assumption. See, for example, Zellner (1976), King (1979, 1980), Singh (1987, 1988).

However, the normal distribution is the only member of the class of multivariate elliptically symmetric distributions where the disturbances are, in fact, independent. Also, it is usually forgotten that the marginal distributions of the disturbance terms under this assumption are identical to those obtained when the disturbances are assumed to be independently and identically distributed (iid) elliptically symmetric. It is these features that lead naturally to the question of the statistical consequences of distinguishing between multivariate and iid elliptically symmetric error distributions and it is this issue that is taken up here.

Kelejian and Prucha (1985) address this problem using asymptotic criteria for the linear regression model and Student-t errors for degrees of freedom greater than 2. This distribution is a particularly important member of the elliptically symmetric class because it is claimed by authors such as Judge et al. (1985) that this distribution may be a reasonable way of modelling tails that are fatter than those of the normal distribution. (see also the recent article by Lange et al. (1989)). The objective here is to extend this analysis by developing properties of the maximum likelihood estimators for the entire Student-t family using finite-sample criteria. Results are obtained assuming the data matrix, X, is nonstochastic.

(iii) The Nonnormal Limited-Information Simultaneous Equations Model

Models (i) and (ii) can be related by simultaneously relaxing both of the conditions associated with the standard linear regression model. This model provides a relatively new area of analysis as there are few published results available on the effects of nonnormal disturbances in the limited- information SEM (e.g. Knight (1985b, 1986), Raj (1980), Donatos (1989)).

The objective here is to combine both of the themes pursued separately in Models (i) and (ii). In particular, in the estimation of the coefficient of the one endogenous regressor in the exactly-identified limited-information SEM, the statistical consequences of distinguishing between multivariate and iid. Student-t error distributions on the LIML and TSLS estimators are examined. Although (because it is exactly-identified) it is a somewhat restrictive model, it is worthy of study because it has a number of interesting features when the errors are normally distributed. In particular, in this case the TSLS, LIML and Least Variance Ratio (LVR) estimators are identical and their distribution is bimodal over part of the parameter space.[1]

1.3 AN OVERVIEW OF THE CHAPTERS

Chapter 2 reviews certain key concepts in probability and statistical inference used in this thesis. It also introduces the notational conventions used.

Chapter 3 reviews an essential tool of analysis that is used throughout this thesis. This is the integration of the nonparametric density estimator with the Monte-Carlo technique. This is a useful technique for approximating many of the density functions considered in the thesis when either the exact distribution is too difficult to derive explicitly or when the exact distribution is known but too complex to be analyzed conveniently. A number of statistical properties of this estimator are discussed. These are all asymptotic properties, but are considered relevant because in the applications considered here, sample size, which is simply the number of replications in the simulation experiment, can be chosen by the investigator.

Chapter 4 discusses the methods used in the simulation experiments. In particular, this includes a discussion of the choice of the number of replications in the simulation experiments, the generation of the random numbers involved, and the algorithms

---

[1] The Least Variance Ratio estimator is the name given to the LIML estimator derived under the assumption of normally distributed errors, when in fact their true distribution is nonnormal.

used to solve the likelihood equations associated with the models considered.

Chapter 5 shows that the exact distribution of a ratio of a bilinear form to a quadratic form in normal variables can be computed using techniques such as those developed by Imhof (1961) and Davies (1973). As many of the commonly used estimators in the limited-information SEM, including TSLS, are of this form, this is a useful technique for the numerical evaluation of their distributions.

Chapter 6 reviews recent relevant finite-sample properties in the literature on the limited-information SEM. It also pursues the theme of comparing the LIML and TSLS distributions, which involves the use of techniques discussed or developed in Chapter 3 and Chapter 5.

Chapter 7 reviews some alternatives to the assumption that the disturbances in the econometric models considered are distributed normally. In particular, the effects of iid nonnormally distributed regression disturbances on the traditional inference and estimation procedures used for normally distributed disturbances are discussed, and a class of alternative estimation techniques collectively labelled "robust estimators" are reviewed. Also in this chapter, the consequences of replacing the normality assumption with the assumption that the regression disturbances follow a multivariate elliptically symmetric distribution are examined. Therefore two types of nonnormally distributed disturbances are reviewed in this chapter, these being, iid nonnormally distributed disturbances and multivariate distributed disturbances. This distinction sets the theme for the remaining chapters. That is, "*an examination of the statistical consequences of distinguishing between the regression disturbances following a multivariate elliptically symmetric distribution and an iid elliptically symmetric distribution*".

Chapters 8 and 9 take up this theme in the nonnormal linear regression model. Chapter 8 considers the "location-scale" model, which is a special case of the linear regression model. It is equivalent to only estimating the intercept term in the linear regression model. Chapter 9 extends the results obtained to the more general model. A distinction is made between the location-scale model and the more general model simply because a number of techniques can be used to examine the problem in the location-scale model that do not generalize to the more general model.

Chapter 10 pursues this theme in the exactly-identified nonnormal limited-information SEM. In particular, the distributions of the TSLS and LIML estimators are compared, since with nonnormal disturbances these two estimation techniques are not necessarily the same.

Chapters 8, 9 and 10 indicate the importance of making the distinction between iid nonnormally distributed disturbances and multivariate nonnormally distributed disturbances. This suggests that it is important to construct appropriate specification tests that make this distinction. This is the topic of Chapter 11.

Finally, Chapter 12 offers some conclusions and presents ideas for future work.

CHAPTER 2

PRELIMINARY DEFINITIONS

2.1 INTRODUCTION

The purpose of this Chapter is to introduce the notational conventions used throughout this thesis.

Section 2.2 defines preliminary mathematical and statistical definitions, such as those given in De Groot (1970), Feller (1966, 1968) and Muirhead (1982). Section 2.3 defines a number of distributions that are used throughout this thesis. These include the multivariate normal, multivariate Student-t, multivariate elliptically-symmetric and Wishart distributions. Finally, Section 2.4 gives a brief note on the layout of the thesis.

2.2 PRELIMINARY MATHEMATICAL AND STATISTICAL DEFINITIONS

(i) Random Variables

A probability space is defined as the combination (Ω,A,P) where, Ω is a set of points, A is a σ-field of subsets of Ω, and P is a probability distribution defined on the elements of A. Furthermore, any set $L \in$ A is known as an event.[1]

---

[1] A σ-field is a set of subsets of Ω which is closed under complementation, countable unions and intersections. A σ-field of interest in the study of probability is the Borel σ-field of subsets of the real line. It is the σ-field generated by the class of all bounded semi-closed intervals of the form (a,b] and is denoted by B. The sets of B are called Borel sets.

A random variable Z is a real valued function from Ω to the real line R which satisfies the condition that for each Borel set B $\in \beta$ on R, the set $Z^{-1}(B) = \{w: Z(w) \in B, w \in \Omega\}$ is an event in **A**.

A collection of random variables $Z_1(w), Z_2(w)\ldots$ on a given pair (Ω, A) will be denoted by $Z_1, Z_2, \ldots$. A random vector is a K-tuple $Z_{\underline{K}} = (Z_1 \ldots Z_K)$ of random variables defined on a given pair (Ω, A).

No distinction will be made between a random variable or vector and the value taken by that random variable or vector.

(ii) Distribution, Probability Density and Characteristic Functions

Associated with a random vector $Z_{\underline{K}}$ on (Ω, A, P) is a distribution function defined on R^K by

$$F_K(t_1 \ldots t_K) = Pr.\left(\{w: Z_1(w) \leq t_1 \ldots Z_K(w) \leq t_K\}\right) \qquad (2.1)$$

for all $t \in R^K$. The joint distribution of $Z_1 \ldots Z_K$ is absolutely continuous if there exists a nonnegative joint probability density function $pdf_K(Z_1 \ldots Z_K)$ such that for every Borel set $B \subset R^K$.

$$F_K(t_1 \ldots t_K) = \int \underset{B}{\cdots} \int pdf_K(Z_1 \ldots Z_K) dZ_1 \ldots dZ_K \ . \qquad (2.2)$$

The characteristic function of a random K-vector $Z_{\underline{K}}$ is defined as

$$\phi_Z(S) = \int \underset{B}{\cdots} \int \exp(isZ'_{\underline{K}})dF, \ S \in R^K, \ i^2 = -1. \qquad (2.3)$$

The characteristic function always exists and no two different distributions yield the same characteristic function so that there is a one-to-one correspondence between characteristic functions and distribution functions.

(iii) Marginal Distribution Functions

The joint distribution of a subset of random variables $Z_1 \ldots Z_p$ of $Z_1 \ldots Z_K$ $(p \leq K)$ is called a marginal distribution. The marginal joint distribution F_p of $Z_1 \ldots Z_p$ is determined from the joint distribution function by the relation

$$F_p(t_1 \ldots t_p) = \Pr(Z_1 \leq t_1 \ldots Z_p \leq t_p) \tag{2.4}$$

$$= \lim F_K(t_1 \ldots t_K) \text{ as } t_j \Rightarrow \infty,\ j = p+1 \ldots K$$

Similarly, the marginal joint probability density function pdf_p of $Z_1 \ldots Z_p$ is determined from the joint probability density function pdf_K $Z_1 \ldots Z_K$ by the relation

$$pdf_p(Z_1 \ldots Z_p) = \int_{R^{K-p}} \cdots \int pdf_K(Z_1 \ldots Z_K)dZ_{p+1} \ldots dZ_K. \tag{2.5}$$

Let G_i denote the marginal univariate distribution function of the random variable Z_i. The random variables $Z_1 \ldots Z_K$ are independent if and only if (iff) their joint distribution function can be factored at every point $(Z_1 \ldots Z_K) \in R^K$ as follows:

$$F_K(t_1 \ldots t_K) = G_1(t_1)G_2(t_2) \ldots G_K(t_K) \tag{2.6}$$

(iv) The Expectation Operator

The expectation $E(Z)$ of any random variable Z with distribution function F is defined as

$$E(Z) = \int_{R^1} Z \; pdf(Z)dZ \tag{2.7}$$

and it exists iff the integral exists. $E(Z)$ is also called the mean of Z or the expected value of Z. For a random vector $Z_{\underline{K}}$ the mean is defined as

$$E(Z_{\underline{K}}) = \left\{E(Z_1)\ldots E(Z_K)\right\} \qquad (2.8)$$

The variance of a random variable Z is given by $E\left[\left(Z-E(Z)\right)^2\right]$ and denoted var(Z). The covariance between random variables Z_1 and Z_2 is defined as $\left[\left(Z_1-E(Z_1)\right)\left(Z_2-E(Z_2)\right)\right]$ and denoted $\text{cov}(Z_i,Z_j)$. Equivalently, this can be expressed as

$$\text{cov}(Z_1,Z_2) = \int_{R^2}\int \left(Z_1-E(Z_1)\right)\left(Z_2-E(Z_2)\right)\text{pdf}(Z_1Z_2)dZ_1dZ_2 \qquad (2.9)$$

For a vector $Z_{\underline{K}}$, the covariance matrix is $\Sigma = (\sigma_{ij})_{K\times K}$, where σ_{ij} = $\text{cov}(Z_i,Z_j)$.

2.3 MULTIVARIATE NORMAL, MULTIVARIATE STUDENT-T, ELLIPTICALLY-SYMMETRIC AND WISHART DISTRIBUTIONS

(i) Multivariate .. Normal and Student-t distributions

A K-dimensional random vector $Z_{\underline{K}}$ has a nonsingular normal distribution with mean $\mu_{\underline{K}}$ and covariance matrix Σ if $Z_{\underline{K}}$ has an absolutely continuous distribution whose probability density function $\text{pdf}(Z_{\underline{K}}|\mu_{\underline{K}},\Sigma)$ is specified at any point $Z_{\underline{K}} \in R^K$ by the equation

$$\text{pdf}(Z_{\underline{K}}|\mu_{\underline{K}},\Sigma) = (2\pi)^{-K/2}|\Sigma|^{-1/2}\exp[-\tfrac{1}{2}(Z_{\underline{K}}-\mu_{\underline{K}})\Sigma^{-1}(Z_{\underline{K}}-\mu_{\underline{K}})'\,]. \qquad (3.1)$$

In (3.1) $\mu_{\underline{K}} = (\mu_1\ldots\mu_K)$ is a K-dimensional vector whose components can be arbitrary real numbers and Σ must be a symmetric and positive definite matrix. This distribution is denoted $N_K(\mu_{\underline{K}},\Sigma)$.

Define the precision matrix T of $N_K(\mu_{\underline{K}},\Sigma)$ to be equal to Σ^{-1}. Suppose that $Y_{\underline{K}}$ is $N_K(\mu_{\underline{K}},\Sigma)$ with precision matrix T and suppose the random variable χ^2 is distributed independently of Y and is chi-square distributed with v degrees of freedom, so that

$$\mathrm{pdf}(\chi^2) = \frac{\exp\left(-\frac{\chi^2}{2}\right)\chi^{2\left(\frac{v-2}{2}\right)}}{2^{v/2}\Gamma(\frac{v}{2})} , \tag{3.2}$$

where $\Gamma(\alpha)$ is the gamma function,

$$\Gamma(\alpha) = \int_0^{\alpha} x^{\alpha-1}\exp(-x)dx, \quad \alpha > 0 \tag{3.3}$$

If the components of $Z_{\underline{K}}$ are defined by the equation,

$$Z_i = Y_i(\frac{\chi^2}{v})^{-\frac{1}{2}} + \mu_i \ , \ i = 1\ldots K, \tag{3.4}$$

then the distribution of $Z_{\underline{K}}$ is multivariate Student-t, with v degrees of freedom, location vector $\mu_{\underline{K}}$ and precision matrix T. It is denoted by $MT_K(\mu_{\underline{K}},T,v)$, and the probability density function of $Z_{\underline{K}} \in R^K$ is,

$$\mathrm{pdf}(Z_{\underline{K}}|\mu_{\underline{K}},T) = \frac{\Gamma\left(\frac{v+K}{2}\right)|T|^{\frac{1}{2}}}{\Gamma\left(\frac{v}{2}\right)(v\pi)^{K/2}}\left[1 + \frac{1}{v}(Z_{\underline{K}}-\mu_{\underline{K}})T(Z_{\underline{K}}-\mu_{\underline{K}})'\right]^{-(\frac{v+K}{2})} . \tag{3.5}$$

For $v > 1$, the mean vector $E(Z_{\underline{K}}) = \mu_{\underline{K}}$ exists, and for $v > 2$ the covariance matrix exists and is equal to $\frac{v}{v-2}\Sigma$.

In both cases the marginal distributions are easy to derive. Suppose that the random vector $Z_{\underline{K}}$ is partitioned in the form,

$$Z_{\underline{K}} = \begin{bmatrix} Z_1^* \\ Z_2^* \end{bmatrix},$$

where the dimension of Z_i^* is K_i $(i = 1,2)$ and $K_1 + K_2 = K$. Also suppose that $\mu_{\underline{K}}$, T and Σ are partitioned as

$$\mu_{\underline{K}} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \qquad T = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix}, \qquad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix},$$

where the dimension of μ_i is K_i $(i = 1,2)$ and the dimension of the sub- matrices T_{ij} and Σ_{ij} is $K_i \times K_j$ $(i,j = 1,2)$. Then, for the normal distribution at any point $Z^*_{1_{\underline{K}_1}} \in R^{K_1}$ the value $pdf_{K_1}(Z_1^*)$ of the marginal probability density function of $Z_{1_{\underline{K}_1}}$ is specified as

$$pdf_{K_1}(Z^*_{1_{\underline{K}}} | \mu_{\underline{K}_1}, T^*) =$$

$$(2\pi)^{-K_1/2} |T^*|^{\frac{1}{2}} \exp\left[-\tfrac{1}{2}(Z^*_{1_{\underline{K}_1}} - \mu_{\underline{K}_1}) T^* (Z^*_{1_{\underline{K}_1}} - \mu_{\underline{K}_1})'\right]. \qquad (3.6)$$

For the multivariate Student-t distribution it is,

$$\mathrm{pdf}_{K_1}\left(Z^*_{1_{\underline{K}_1}} \mid \mu_{\underline{K}_1}, T^*\right) =$$

$$\Gamma\left(\frac{v+K_1}{2}\right) |T^*|^{\frac{1}{2}} \left[1+\frac{1}{v}\left(Z^*_{1_{\underline{K}_1}} - \mu_{\underline{K}_1}\right) T^* \left(Z^*_{1_{\underline{K}_1}} - \mu_{\underline{K}_1}\right)'\right]^{-\left(\frac{v+K_1}{2}\right)} \quad (3.7)$$

where $T^* = T_{11} - T_{12}T_{22}^{-1}T_{21}$. Therefore, in both cases, the marginal distributions are members of the same family as their respective joint distributions.

Further properties of both of these distributions can be found in, for example, De Groot (1970, pp.50-60).

Elliptically Symmetric Distributions

Each of the distributions above belong to the wider family of multivariate elliptically-symmetric distributions. The random vector $Z_{\underline{K}}$ has a multivariate elliptically-symmetric distribution if the characteristic function $\phi_{Z_{\underline{K}}-\mu_{\underline{K}}}(s_{\underline{K}})$ of $(Z_{\underline{K}} - \mu_{\underline{K}})$ is a function of the quadratic form $s_{\underline{K}}\Sigma s_{\underline{K}}'$, (where $s_{\underline{K}}$ is a row vector), such that,

$$\phi_{Z_{\underline{K}}-\mu_{\underline{K}}}(s_{\underline{K}}) = \phi\left(s_{\underline{K}} \Sigma s'_{\underline{K}}\right) = \exp\left(is_{\underline{K}}\mu'_{\underline{K}}\right)\psi\left(s_{\underline{K}} \Sigma s'_{\underline{K}}\right) \quad (3.8)$$

for some function ψ. If it is further assumed that the density function with nonsingular Σ exists, then it is of the form,

$$\mathrm{pdf}(Z_{\underline{K}}) = C_K |T|^{\frac{1}{2}} g\left[\left(Z_{\underline{K}}-\mu_{\underline{K}}\right) T \left(Z_{\underline{K}}-\mu_{\underline{K}}\right)'\right], \quad (3.9)$$

where g is a one-dimensional real-valued function independent of K and C_K is a scalar proportionality constant. This distribution is denoted MES(μ,T) and it has the first two moments, $E(\underline{Z}_K) = \mu_K$ and $Cov(\underline{Z}_K) = \alpha\Sigma$, where $\alpha = -2\psi'(0)$, provided these moments exist. If $\mu_{\underline{K}} = 0$ and $\Sigma = I$ in (3.8) then the multivariate elliptically symmetric distributions are called spherically symmetric distributions. Two properties of these distributions used in this thesis are (see, for example, Muirhead (1982, p.34)):

Properties 3.1

(1) All marginal distributions are elliptical and all marginal density functions of dimension $p \leq K$ have the same functional form.

(2) If $Z_{\underline{K}}$ is $N(\mu_{\underline{K}},\Sigma)$ and Σ is diagonal then the components $Z_1 \ldots Z_K$ of $Z_{\underline{K}}$ are all independent. Within the class of multivariate-elliptically symmetric distributions independence when Σ is diagonal characterizes the normal distribution.

Further properties of these distributions are discussed by authors such as Chmielewski (1981), Kelker (1970), King (1979), Cambanis, Huang and Simons (1981) and Muirhead (1982).

(iii) The Wishart Distribution

The Wishart distribution is used in the derivation of finite-sample properties of common estimators in limited-information simultaneous equations models (as discussed in Chapters 5 and 6). The Wishart distribution is a matrix generalization of the noncentral chi-squared distribution (see, for example, Johnston and Kotz (1972, p.158)). Consider the random $n \times K$ matrix

$$Z = \begin{bmatrix} Z_1' \\ \vdots \\ Z_n' \end{bmatrix} = (Z^{(1)}, Z^{(2)} \ldots Z^{(K)}) ,$$

where the Z_i terms are independent normal random vectors with mean μ_i and covariance matrix Σ. The $K \times K$ matrix $W = Z'Z$, with (i,j)th element $Z^{(i)\prime} Z^{(j)}$, is said to be a Wishart matrix. The elements of W have a non- central Wishart distribution of order K, with n degrees of freedom, covariance matrix Σ and noncentrality parameter $M = \sum_{i=1}^{n} \mu_i \mu_i'$. This is denoted by

$$W \sim W_K(n, \Sigma, M) \tag{3.10}$$

The distribution is said to be central if $M = 0$. The Wishart distribution has properties similar to those of the noncentral chi-squared distribution. In particular, if A and B are symmetric idempotent matrices, then $Z'AZ \sim W_K[q, \Sigma, E(Z)'AE(Z)]$, where q is the rank of A, and $Z'AZ$ and $Z'BZ$ have independent Wishart distributions iff $AB = 0$. Further properties are discussed in Muirhead (1982, pp.441-449); and Johnston and Kotz (1972, pp.158-180).

2.4 NOTE ON LAYOUT AND NOTATION

The purpose of this chapter has been to introduce the basic notational conventions used in this thesis. Other notation used that is not introduced in this chapter is defined when it is required.

The layout of this thesis is as follows. Each chapter is divided into sections. Theorems, Equation Numbers, Properties and Figures within each section of a chapter are denoted by their section number and then in sequence. Therefore, when referenced in other chapters they are denoted by their chapter number first and then their section and sequence number.

CHAPTER 3

KERNEL DENSITY ESTIMATION

3.1 INTRODUCTION

There are a number of techniques that are used to approximate density functions, either when the exact distribution is too difficult to derive explicitly, or when the exact distribution is known but too complex to be analyzed conveniently. For example, the exact sampling distributions of estimators of the unknown coefficients of the endogenous variables in single structural equations, have been shown to depend upon multiple infinite series of zonal-type polynomials, and these present enormous difficulties in numerical work. Phillips (1980a, 1983) has overcome these difficulties by extracting various joint and marginal density approximations using asymptotic expansions.

However, another method which may be used to analyze such distributions is the Monte Carlo method, in which artificial data are generated and from them sampling distributions and moments are estimated. One advantage of this technique is that it can be implemented easily on an extensive range of models and error probability distributions. An extension of this technique which is used in this thesis is the integration of density estimation with the Monte Carlo technique, as suggested by Ullah and Singh (1985). That is, the Monte Carlo approach is used to generate the statistics of interest and then the density of these statistics is estimated using the generated statistics as observations. The objective of this chapter is to briefly review the history and discuss the statistical properties of the Kernel estimator, which is a particular example of a density estimator that is both widely used and thoroughly studied in the statistical literature. The actual Monte Carlo methodology that is used in this thesis is the topic of the next chapter.

The Kernel estimation technique has been reviewed by, for example, Tapia and Thompson (1976), Singh, Ullah and Carter (1987), Wertz (1978), Devroye (1987), Silverman (1986), and Ullah (1988) and the contents of this chapter draw heavily on these reviews. The finite-sample analysis of statistics is the application of the Kernel density estimation technique that will be used throughout this thesis. Recently there has been a great deal of interest in other applications of the technique, such as applying the method to the estimation and testing of econometric models. A review of these applications is beyond the scope of this chapter. However, these applications have been reviewed by Bierens (1986), Singh *et al.* (1987) and Ullah (1988).

In Section 2 the Kernel density estimation technique is defined and its history is briefly reviewed. Section 3 considers the asymptotic properties of this estimator and Section 4 considers the choice of Kernel, window width and sample size. Section 5 concludes this chapter with a simple illustration.

3.2 THE METHOD

Let $X_1, X_2, \ldots, X_{N*}$ be independently and identically distributed observations on a random variable X with probability density function pdf(X). Rosenblatt (1956) and Parzen (1962) developed the Kernel estimator of pdf(X), which is defined as,

$$\hat{pdf}(X) = \frac{1}{N^*h(N^*)} \sum_{j=1}^{N^*} K\left[\frac{X-X_j}{h(N^*)}\right] \qquad (2.1)$$

where $h(N^*)$ is the window width, which is assumed to be a positive function of the sample size, N^*, such that $\lim_{N^*\to\infty} h(N^*) = 0$, and K is the Kernel. If K is everywhere a nonnegative function and satisfies $\int K(x)dx = 1$, then $\hat{pdf}(X)$ will be a probability density function which possesses all of the continuity and differentiability properties of K. Numerous extensions of this estimator have been considered. For example, Breiman et al. (1977) introduced the variable Kernel estimator in which the window width varies across the data points, allowing the tails of the estimator to be smooth while not distorting the central part of the density.

Cacoullos (1966) extended the Kernel estimator to the estimation of multivariate density functions. Let

$$X_i = X\left(X_1^{(i)}\ X_2^{(i)}\ \ldots\ X_m^{(i)}\right) \qquad i = 1 \ldots N^*$$

be a given sample of N^* independent realizations of an m-dimensional random variable $X(X_1\ \ldots\ X_m)$ from a population characterized by a continuous m-variate probability density $f(X_1 \ldots X_m)$. The estimator suggested by Cacoullos is,

$$\hat{pdf}(X) = N^{*-1}h(N^*)^{-m} \sum_{t=1}^{N^*} K\left[\frac{X-X^{(t)}}{h(N^*)}\right], \qquad (2.2)$$

where (as in the univariate case), $h(N^*)$ is the window width, assumed to be a positive function of sample size such that $\lim_{N^*\to\infty} h(N^*) = 0$, and K is the natural generalization of the univariate Kernel. This estimator uses only a single $h(N^*)$ for all m variables, however it has been suggested that this may not be appropriate (see, for example, Ullah (1988, p.634)). On occasions throughout this thesis an estimate of the appropriate marginal density is required, and these can be estimated using the expression in (2.2). Conditional densities can also be estimated; however, the details will not be given here but can be found in Ullah (1988).

Suppose that the vector of realizations is written as,

$$X_i\left(X_1^{(i)}\ X_2^{(i)}\ \ldots\ X_m^{(i)}\right) = X_i\left(Z^{(i)}, y^{(i)}\right)$$

where $Z^{(i)}$ is a $p \times 1$ vector and $y^{(i)}$ is a $q \times 1$ vector such that $p + q = m$. The marginal density of Z_t at Z is

$$\int \hat{pdf}(Z,y)dy\ . \tag{2.3}$$

One example of a joint Kernel K from which marginal densities can be found easily is studied by Epanechnikov (1969), and is given by the equation

$$\hat{pdf}(X) = N*^{-1} \sum_{t=1}^{N*} \prod_{i=1}^{m} \frac{1}{h\ (N*)} K_i\left(\frac{X_i - X_i^{(t)}}{h\ (N*)}\right). \tag{2.4}$$

If each K_i satisfies $\int K_i(x)dx = 1$ and $h_i(N*) = h(N*)$, then (2.3) can be written as,

$$N*^{-1} \sum_{t=1}^{N*} h(N*)^{-q} \prod_{j=1}^{q} K_j\left(\frac{z_j - z_j^{(t)}}{h(N*)}\right). \tag{2.5}$$

In particular, when (2.4) and (2.5) are used in this thesis it is assumed that each K_i has the same form, such as, for example, $K_i = \frac{1}{\sqrt{2\Pi}}\exp\left(-\frac{1}{2}y^2\right)$, the normal Kernel.

3.3 THE ASYMPTOTIC PROPERTIES

The asymptotic properties of the Kernel estimator are particularly relevant for the application of the density estimation technique to the finite-sample analysis of statistics. This is because the sample size, N*, which is the number of replications of the simulation experiment, can be chosen and is bounded only by the limits of the duplication of the random number generator. The objective of this section is to present some of the asymptotic properties of Kernel estimators. In particular, these properties are dependent upon the chosen Kernel, window width and the unknown density. A more extensive review can be found in Ullah (1988, pp.638-642).

To obtain these asymptotic properties, certain regularity conditions are specified on the Kernel, window width and density. The following set of assumptions are taken from Ullah (1988, p.639). Let K be the class of Borel-measurable bounded real valued functions K(x), $x = (x_1 \ldots x_m)$ such that for the:

Kernel

I. (i) $\int K(x)dx = 1$

(ii) $\int |K(x)|dx < \infty$

(iii) $\|x\|^m |K(x)| \Rightarrow 0$ as $\|x\| \Rightarrow \infty$ where $\|.\|$ is the Euclidean norm.

(iv) $\sup|K(x)| < \infty$.

Window Width

II. $h(N^*) \Rightarrow 0$ as $N^* \Rightarrow \infty$

III. $N^*h(N^*)^m \Rightarrow \infty$ as $N^* \Rightarrow \infty$.

Density

IV. $pdf(x)$ is continuous at any point x_0.

Using these assumptions Cacoullos (1966) has shown that if I, II and IV hold,

$$\lim_{N^* \Rightarrow \infty} E\left[\hat{pdf}(x)\right] = pdf(x)$$

which implies pointwise asymptotic unbiasedness, and if I, II, III and IV hold,

$$\hat{pdf}(x) \overset{p}{\Rightarrow} pdf(x) \quad \text{as} \quad N^* \Rightarrow \infty$$

at any point and therefore implies pointwise weak consistency. Other results have also been shown to hold. For example, Deheuvels (1974) develops weaker conditions under which these results hold, and Devroye and Wagner (1976) develop strong consistency results assuming some further conditions.

Each of the properties above are pointwise properties. Some authors (e.g. Bai and Chen (1987)) have obtained results for global properties, using criteria such as those based on the norm Lp, which involve considering conditions under which

$$\left[\int \left|\hat{pdf}(x) - pdf(x)\right|^p dx\right]^{1/p} \Rightarrow 0 \text{ as } N^* \Rightarrow \infty. \qquad (3.1)$$

The last asymptotic property to be discussed is the property of asymptotic normality, which is useful for deriving confidence intervals for $\hat{pdf}(x)$. The results of Parzen (1962) and Cacoullos (1966) imply,

$$\left(N^* h^m(N^*)\right)^{\frac{1}{2}}\left[\hat{pdf}(x) - E\left(\hat{pdf}(x)\right)\right] \sim N\left(0, pdf(x)\int K^2\right) \qquad (3.2)$$

holds. The result given in (3.2) can be achieved if

$\left(N^*h^m(N^*)\right)^{\frac{1}{2}}\text{Bias}\left[\hat{pdf}(x)\right]$ tends to zero asymptotically since,

$$\left(N^*h^m(N^*)\right)^{\frac{1}{2}}\left[\hat{pdf}(x) - pdf(x)\right] = \left(N^*h^m(N^*)\right)\left[\hat{pdf}(x) - E\left(\hat{pdf}(x)\right)\right]$$

$$+ \left(N^*h^m(N^*)\right)^{\frac{1}{2}}\text{Bias}\left[\hat{pdf}(x)\right] \qquad (3.3)$$

Ullah (1988, p.642) shows that $\text{Bias}\left[\hat{pdf}(x)\right]$ is proportional to $h^2(N^*)$. This implies that if $N^*h^{\frac{4+m}{2}}(N^*)$ tends to zero asymptotically then (3.2) holds.

As an example, consider the univariate normal Kernel $\frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}y^2)$, then the 99% asymptotic confidence interval for $\hat{pdf}(x)$ is given by

$$\hat{pdf}(x) \pm 2.58 \left[\frac{\hat{pdf}(x)}{2N^*h\sqrt{\pi}}\right]^{\frac{1}{2}}. \qquad (3.4)$$

3.4 CHOOSING THE KERNEL, WINDOW WIDTH AND SAMPLE SIZE

In the implementation of the Kernel estimator and in the use of the results in the previous section, the selection of h, K and N* is required. Most emphasis in the literature has been given to choosing a suitable window width and Kernel on the basis of minimizing some measure. The usual measures to be taken are approximate bias, mean squared error (MSE), or integrated mean squared error (IMSE) of $\hat{pdf}(x)$ where,

$$\text{IMSE} = \int \text{MSE} = \int E\left[\hat{pdf}(x) - pdf(x)\right]^2 dx. \qquad (4.1)$$

The difference between the two measures MSE and IMSE is that MSE is a measure of the estimator $\hat{f}$ at a single point whereas, IMSE is used as a global accuracy measure of $\hat{f}$ as an estimator of f. The approximations to these measures are obtained using similar methods to Kadane's (1971) small disturbance expansion of estimators, and they can be found in Ullah (1988, p.642). The existence of these approximations require a number of assumptions as given in Ullah (1988, p.641).

The optimal h that minimizes MSE is,

$$h^* = cN^{*-1/(m+4)}; \text{ where,} \tag{4.2}$$

$$c = \left[m\,pdf(x) \left\{ Dpdf(x) \int x^2 K(x)\,dx \right\}^2 \int K^2(x) \right]^{-1/m+4}$$

and for IMSE is,

$$h^* = c^* N^{*-1/(m+4)}; \text{ where,} \tag{4.3}$$

$$c^* = \left[m \left(\int x^2 K(x)\,dx \right)^{-2} \left[\int Dpdf(x)^2 \right]^{-1} \int K^2(x) \right]^{-1/m+4}$$

where Dpdf(X) is the operator $\frac{d^2 pdf(x)}{\partial x \partial x'}$, so that h* converges to 0 as $N^* \Rightarrow \infty$ but only at the rate $N^{*-1/m+4}$.

However, these choices are not in general operational as they depend upon the unknown density. However, suitable operational window widths have been suggested which depend upon the actual estimator $\hat{pdf}(x)$. Simply, in (4.2) and (4.3) above, $\hat{pdf}(x)$ replaces pdf(x), and the iteration process is begun with an initial arbitrary starting value for h. However, the rate of convergence of this estimator may be slow (see, for example, Ullah (1988, p.644)).

There are various other ways of choosing h. The cross-validity approach is one that has often been used. It is also called the modified maximum likelihood method (Duin (1976)), and it involves a completely data- based choice for h. There have been a number of papers that have examined the asymptotic equivalence of the cross-validity choice to MISE (e.g., Hall (1983), Stone (1984)).

Tapia and Thompson (1976) suggest an "interactive" method which is useful mainly in the univariate case. It is recommended that the estimation technique begins with h values that are too large, that is, when the pdf is obviously overly smoothed, and then h is sequentially decreased until overly noisy probability density estimates are obtained. The point where further attempts to improve resolution, by decreasing h lead to noisy estimators is generally fairly sharp and readily observable. Examples of this approach are given by Tapia and Thompson (1976, pp.61-66). Alternatively, they also present an empirical algorithm which iterates according to the algorithm:

$$h_{i+1} = N*^{-1/5}\left[\frac{\int K^2(x)dx}{\int x^2 K(x)dx}\right]^{-1/5}\left[\int |D\hat{pdf}(x)|^2 dx\right]^{-1/5}$$

Other approaches have been suggested and the details are given in Ullah (1988, p.644). A Monte Carlo study of three data-based nonparametric probability density estimators is given by Scott and Factor (1981).

Usually the choice of K will be a symmetric unimodal pdf. Two examples of multivariate kernels are,

$$K(x) = 2\pi^{-m/2}\exp(-\tfrac{1}{2}x'x) \quad , \tag{4.4}$$

the multivariate normal Kernel, and

$$K(x) = \left.\begin{array}{ll} 2^{-1}c_m^{-1}(m+2)(1-x'x) & \text{if } x'x = 1 \\ = 0 & \text{otherwise} \end{array}\right\}, \quad (4.5)$$

where c_m is the volume of the unit m-dimensional sphere. These examples illustrate two different types of Kernels, that is, those with compact or those with non-compact support. Kernels with compact support have two advantages. These are:

- savings in computer time,
- if the density to be estimated has compact support, estimation using a Kernel with noncompact support will always be disturbed by boundary effects (see, for example, Gasser and Müller (1979)).

To obtain the optimal Kernel (4.3) is substituted into (4.1) and IMSE is then minimized. This gives the optimal Kernel given in (4.5) as shown in Epanechnikov (1969).

Davis (1975, 1977) examines the rate at which MSE and IMSE decrease, as sample size increases, for a number of univariate Kernels. Generally though, both the theoretical and the Monte Carlo results have led some researchers to question whether the properties of the Kernel estimator are sensitive to the choice of Kernel. See, for example, Epanechnikov (1969, p.156). However, it is also considered (e.g. Davis (1975)) that if the Kernels are not restricted to be nonnegative, then the degree of approximation may actually improve, although the resulting density estimate may be negative at some points.

Although in many situations the sample size is determined by the availability of data, when the Monte Carlo method is integrated with non- parametric density estimation the investigator can choose the sample size, as it is simply the number of replications in the simulation experiment.

Epanechnikov (1969) gives values of sample size that assure a prescribed level of "minimum relative global error", when the true density is assumed to be multivariate normal and the multivariate normal Kernel is used.[1] This approach is not operational because it depends upon the unknown density. However, using the approximate expressions for MSE and IMSE given by Ullah (1988, p.642), with the estimate $\hat{pdf}(x)$ appropriately replacing pdf(x), then a similar procedure to Epanechnikov (1969) can be performed. However, the properties of this procedure need to be examined. Alternatively, an easy technique to employ is similar to the application of the Kolmogorov-Smirnov statistic in the estimation of the empirical cumulative distribution function. This method is used throughout this thesis and is discussed in the next chapter.

3.5 AN ILLUSTRATION

Epanechnikov (1969) compares various Kernels by calculating the ratio,

$$r = \frac{\int_{-\infty}^{\infty} K^2(y)dy}{\int_{-\infty}^{\infty} K_0^2(y)dy}, \qquad (5.1)$$

where K_0^2 refers to the optimal Kernel given in (4.2) for m = 1.

---

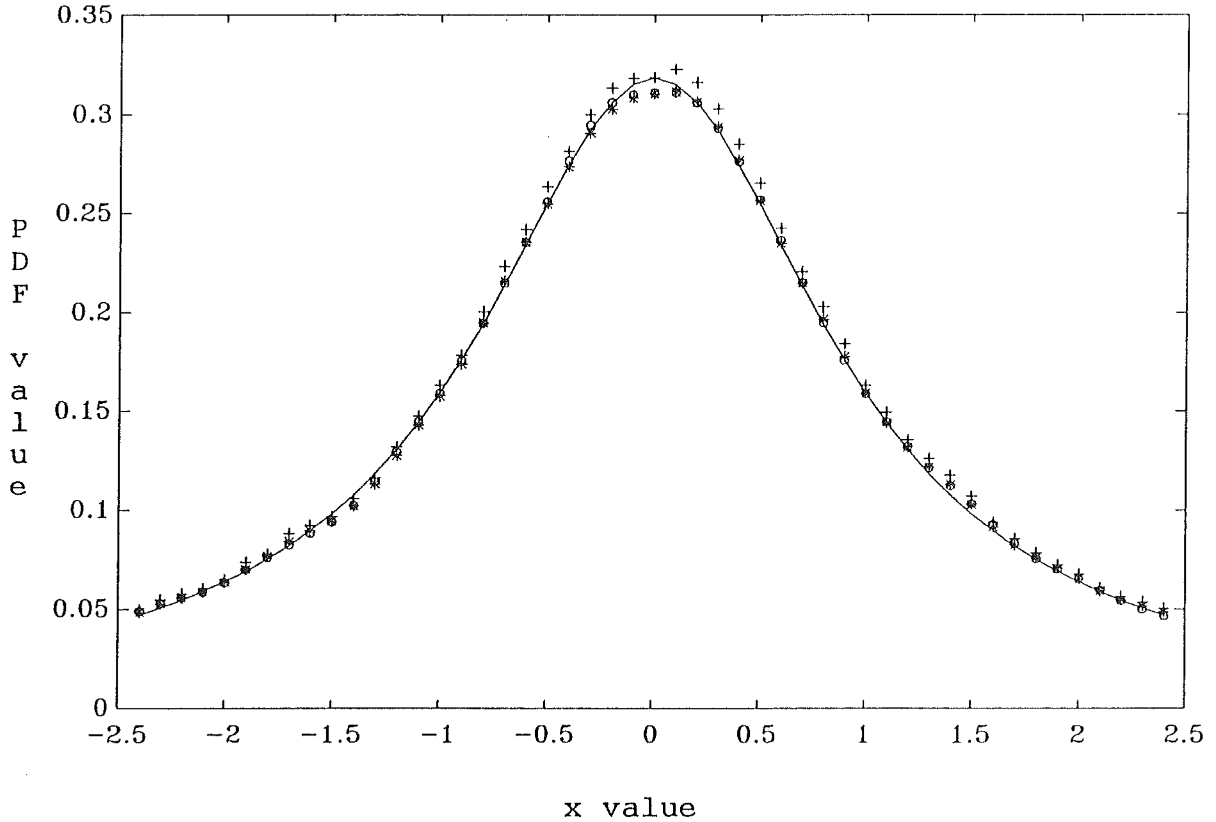
[1] In determining the optimal window width and Kernel, "minimum relative global error" gives the same result as given by minimizing IMSE.

This ratio is used because the optimal Kernel is determined on the basis of minimum IMSE and this is equivalent to minimizing $\int K^2(y)dy$, subject to a number of conditions. For the normal Kernel, given in (4.4), r = 1.051 and for the Laplace Kernel, which is defined by the equation,

$$K(y) = \frac{1}{\sqrt{2}} \exp(\sqrt{2}|y|), \qquad (5.2)$$

r = 1.320. To illustrate the techniques reviewed in the chapter, and to compare the three Kernels mentioned, given the difference in their r values, the standard Cauchy density is estimated. Two sample sizes are chosen, (100,000 and 100 replications), these representing a "large" and "small" sample respectively. The choice of window width is determined using the technique of Tapia and Thompson (1976). Figure 5.1 illustrates the results obtained for 100,000 replications, and given the asymptotic properties presented in Section 3 it is expected that all of the estimated densities will be very similar. In Figure 5.2 when only 100 replications are used, some differences, particularly with the Laplace Kernel, are noticeable, suggesting that in small samples differences do exist between different Kernels. However, these results are only illustrative and the differences obtained with this example may not be generally representative.

FIGURE 5.1 Comparison of Different Kernels for Cauchy Distribution Using 100,000 Replications



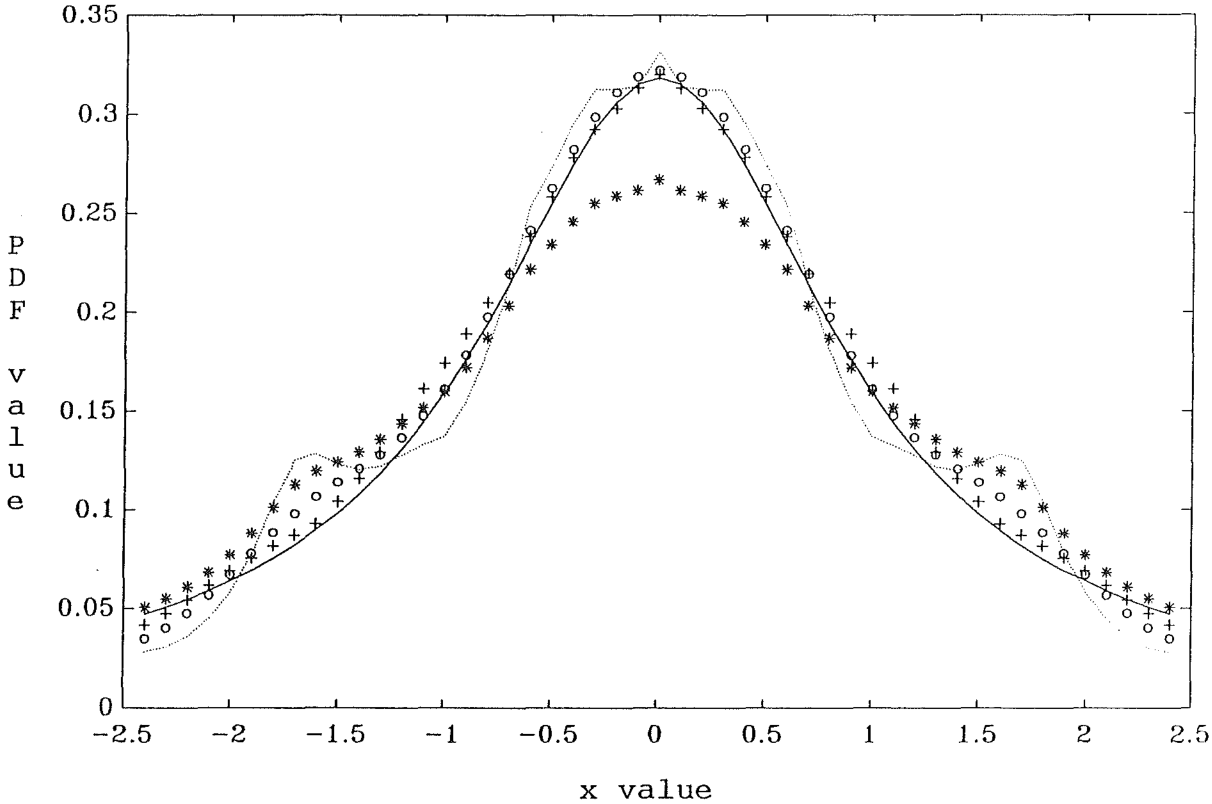
KEY - Standard Cauchy Distribution

° Epanechnikov Kernel; h = 0.03

* Normal Kernel; h = 0.03

+ Laplace Kernel ; h = 0.09

FIGURE 5.2 Comparison of Different Kernels for Cauchy Distribution Using 100 Replications



KEY - Standard Cauchy Distribution

° Normal Kernel; h = 0.60

\+ Epanechnikov Kernel; h = 0.60

* Laplace Kernel; h = 0.60

.. Laplace Kernel; h = 0.40

CHAPTER 4

MONTE CARLO EXPERIMENTS - A DESCRIPTION OF METHODOLOGY

4.1 INTRODUCTION

Each of the Monte Carlo experiments used in this thesis comprises five steps, these being, the choice of the underlying model, the number of replications in the experiment, the generation of appropriate pseudo-random numbers, the estimation of the unknown parameters of interest and, finally, the estimation of the population variance (or when this does not exist some other measure of dispersion such as the Interquartile Range), and/or, the estimation of the probability density function, (or the distribution function, denoted by cdf), of the estimator of interest. Each of the models chosen in Step 1 is discussed in the appropriate chapter, as well as the 'key parameters' on which each experiment is based. In particular, these models are, the LIML estimator with normally distributed disturbances (Chapter 6), the location-scale model (Chapter 8), the multiple regression model (Chapter 9) and the exactly-identified LIML estimator (Chapter 10), each with Student-t disturbances. The topic of this chapter is a description of the remaining steps, that is, Steps 2-5. All computations included in these steps were carried out on a VAX 8350 computer.

As Steps 2 and 5 are not independent they are jointly discussed in Section 2. Section 3 discusses the methods used to generate the pseudo- random numbers and Section 4 outlines the algorithms used to solve the likelihood equations associated with the models given above.

4.2 NUMBER OF REPLICATIONS AND THE ESTIMATION OF DF's, PDF's AND MEASURES OF LOCATION AND DISPERSION

Let $X_1, X_2, \ldots, X_{N*}$ denote a random sample of size N* from a cumulative distribution function DF. Then $Y_1 \leq Y_2 \leq \ldots Y_{N*}$, where Y_1 are the X_1 arranged in order of increasing magnitudes and are defined to be the *order statistics* corresponding to the random sample $X_1, \ldots, X_{N*}$. The unknown DF is estimated using the empirical DF, DF_{N*}, which is a function of order statistics. In particular, DF_{N*}, is defined by,

$$DF_{N*}(x) = (1/N*) * (\text{number of } Y_j \text{ less than or equal to } x). \quad (2.1)$$

The Kolmogorov-Smirnov statistic, D_{N*}, is used to test how well a given set of observations fits some specified DF. It is defined as follows.

$$D_{N*} = \sup_{-\infty<x<\infty} \left| DF_{N*}(x) - DF(x) \right| ,$$

and the exact distribution of D_{N*} has been tabulated for various N* (see e.g. Mood, Graybill and Boes (1986, p.508).

N*, in this thesis, represents the number of replications for the simulation experiments performed. N* is chosen in such a way that on the basis of N* replications we can calculate from the distribution of the Kolmogorov-Smirnov statistic that DF_{N*} is within 0.001 of DF everywhere, with probability more than 0.99, (see e.g. Anderson et al. (1982)). In general, this implies that N^* varies between 60,000 - 90,000 replications.

The integration of the Kernel density estimator (the topic of Chapter 3), with the naive Monte Carlo method is used to obtain the empirical pdf's. In each experiment two Kernels are used, these being the Epanechnikov and normal Kernels (although the final results do not depend on which Kernel is used), and the window width is determined using the technique of Tapia and Thompson (1976), as described in Chapter 3. The number of observations, N*, used in the application of the Kernel estimator is simply the number of replications in the simulation experiment, and is chosen using the bound of estimation, B, associated with the error of estimation. For example, when the normal Kernel and the 99% asymptotic confidence interval are used, as is given in (3.3.4), B is equal to

$$B = 2.58 \left[\frac{\hat{pdf}(x)}{2N^*h(N^*)\pi} \right]^{\frac{1}{2}} .$$

N* is varied until B is less than 0.01 for all points at which the density is estimated. In all of the experiments, N* varies between 60,000 and 90,000 replications. This technique is similar to the use of the Kolmogorov- Smirnov statistic in the estimation of the empirical cumulative distribution function. Given the large number of replications, N*, used the final results do not depend on which kernel is used. This situation is similar to the comparison of different kernels for the Cauchy distribution using a "large sample", as is illustrated in Figure 5.1, in Chapter 3.

The measures of dispersion used include, the median, interquartile range and the variance (if it exists) of the population. These are estimated using the corresponding sample equivalents, (see e.g. Mood, Graybill and Boes (1986, p.75)). The same number of replications used to estimate DF or PDF is used here.

4.3 GENERATION OF RANDOM NUMBERS

The generation of random normal and iid Student-t observations is required in order to obtain the empirical pdf's, cdf's, and dispersion measures, in the analysis of each of the models in Chapters 6, 8, 9 and 10. The analysis of misspecification of error distributions in Chapters 8 and 9, and the analysis of the model in Chapter 11, also require the generation of multivariate Student-t observations. The generation of variates from each of the distributions is based on one or more transformations of uniform random numbers.

Random numbers distributed uniformly on the interval [0,1], denoted U(0,1), are generated using the NAG subroutine G05CAF, which uses a multiplicative congruential method. This generator passes the spectral test which has become the most respected theoretical test of a linear congruential random number generator (Bratley et al. (1983, p.195), Kelejian and Adam (1989, p.3), NAG manual Mark 12 Vol. 6, Algorithm G05CAF). From these variates the following are obtained:

Normal

Standard normal variates, N(0,1), as given by (2.3.1), (with $\Sigma = I$ and $K = 1$), are generated using the NAG Subroutine G05DDF, which is based on Brent's (1974) algorithm. This involves a generalization of Von Neumann's (1951) method of generating random samples from the exponential distribution by comparison of uniform random numbers. These are then converted into normal random variates.

Chi Square

The chi-square distribution, as given by (2.3.2), has a single positive integer parameter v, the degrees of freedom. If $\{Z_i\}$ is a sequence of independent standard normal variates, then

$$X = \sum_{i=1}^{v} Z_i^2$$

has a chi-square distribution with v degrees of freedom. This relationship can be used in the generation of chi-square variates, however, it is thought to be inefficient, except for small v, due to an increasing requirement for normal deviates, as v increases (see, for example, Dagpunar (1988)).

An alternative method, which is considered to be an efficient method for small v, (see, for example, Rubinstein (1981, p.93)), is as follows. If v is even, then X can be computed as,

$$X = -2\ln\left(\prod_{i=1}^{v/2} U_i\right)$$

and if v is odd then

$$X = -2\ln\left(\prod_{i=1}^{v/2-\frac{1}{2}} U_i\right) + Z^2 ,$$

where Z is from N(0,1) and U_i is from U(0,1).

Another approach for generating chi-square variates, (see, for example, Bratley (1963, p.163)), includes making use of the fact that the chi-square distribution is a particular case of a gamma density. This method is particularly useful when v is large.

As most of the focus in this thesis is on small v, the second method is the main method used.

IID Student-t

The standard iid Student-t distribution is defined by (2.3.5) with $\Sigma = I$ and $K = 1$. For degrees of freedom $v < 3$, these variates are generated by the inversion of the distribution function (see, for example, Devroye (1986, p.27)). In particular, for $v = 1$, the Cauchy distribution, standard Cauchy variates are generated as,

$$X = \tan\left(\pi(U-\tfrac{1}{2})\right)$$

and for $v = 2$, the t_2-distribution,

$$X = \sqrt{2}(U-\tfrac{1}{2})/\mathrm{SQRT}\left(U(1-U)\right) ,$$

where U is from U(0,1).

For the rest of the Student-t family, $v \geq 3$, X is generated via a transformation of a symmetric beta variate, (see, for example, Devroye (1986, p.446)). This can be written in terms of independent uniform random numbers U_1, U_2 as,

$$X = \frac{2\sqrt{v}\ \sin\ (2\pi U_1)(1-U_2^{2/v-1})}{(1-\sin^2(2\pi U_1))(1-U_2^{2/v-1})} .$$

This formula is useful as it is valid for all members of the Student-t family with $v \geq 3$. It also does not require the generation of as many random uniform deviates as does the traditional method of generating a t-random variable via its interpretation as a ratio of a standard normal to the square root of an independent normalized chi-square variable.

Multivariate Student-t

For the multivariate Student-t distribution, (2.3.5), with location vector 0 and precision matrix I, the K joint random variates are generated using the relationship (see, for example (2.3.4)),

$$X_i = Z_i\left(\frac{\chi^2}{v}\right)^{-\frac{1}{2}} \qquad i = 1 \ldots K ,$$

where $Z_1 \ldots Z_K$ are K independent standard normal variables and χ^2 is an independent chi-square variable with v degrees of freedom.

For each of the univariate distributions, the methods were tested by estimating the density functions using the Kernel estimator and the generated random variables as observations and comparing the results obtained with the "true pdf", (see, for example, Figure 3.5.1). In the multivariate case, the method was tested against known results, such as, for example, in the linear regression model the t-statistic under the null hypothesis is t-distributed for all v, and the statistic, s^2/σ^2, where s^2 is defined in Properties 1.1.1, is F-distributed with N-K and v degrees of freedom for $v \geq 3$ (see, for example, Zellner (1976)). The results of these tests suggest that the random number generators perform well.

4.4 ESTIMATION OF THE UNKNOWN PARAMETERS OF THE MODELS

The implementation of the simulation experiments performed in this thesis requires the estimation of various parameters of the models involved. The objective of this section is to describe all of the algorithms that are used for this purpose.

Maximum likelihood estimation of the unknown parameters of the models in Chapters 6, 8, 9 and 10 requires the maximization of the appropriate likelihood function or, equivalently, the minimization of the negative of this function, say f(x), where $x \in R^n$.

Quasi-Newton methods for the unconstrained minimization of f(x), $x \in R^n$, are line search algorithms which use the basic iteration

$$x^{(K+1)} = x^{(K)} + \alpha^{(K)}\rho^{(K)}, \quad K = 1,2... \qquad (4.1)$$

to generate a sequence of approximations $\{x^{(K)}, K=2,3...\}$ to a stationary point x* of f(x) from a given starting vector $x^{(1)}$.

A scalar $\alpha^{(K)} > 0$ is usually chosen to reduce the objective function at each iteration so that convergence can be achieved, and this scalar satisfies a descent condition of the form,

$$f(x^{(K+1)}) < f(x^{(K)}) + \rho^{(K)}\alpha^{(K)}\rho^{(K)'}\nabla f\left(x^{(K)}\right),$$

where $\rho \in (0,\frac{1}{2})$, and ∇ is the gradient of f at $x^{(K)}$.

The search direction, $\rho^{(K)} \in R^n$ in (4.1) is determined by solving a system of equations,

$$\beta^{(K)}\rho = - g^{(K)} ,$$

where $\beta^{(K)}$ is a positive-definite approximation to the Hessian matrix of second derivatives $\nabla^2 f(x^{(K)})$ and $g^{(K)} \in R^n$ is the generated vector $\nabla f(x^{(K)})$.

Two algorithms from the Harwell subroutine library are used in this thesis, these being algorithms VAI3AD and VF04AD. Both of these initially choose $\beta^{(1)} = I$ and then use the BFGS formula (Broyden (1970), Fletcher (1970), Goldfard (1970) and Shanno (1970)),

$$\beta^{(K+1)} = \left[\beta - \frac{\beta pp'\beta}{p'\beta p} + \frac{\gamma\gamma'}{\alpha' p'\gamma}\right]^{(K)} ,$$

where $\gamma^{(K)}$ is the vector,

$$\gamma^{(K)} = g^{(K+1)} - g^{(K)} ,$$

to update the matrix $\beta^{(K)} \in R^{n\times n}$. The VAI3AD algorithm, instead of working directly with the matrix β, or its inverse, stores and updates the Choleski factors of β since this enables the search direction ρ to be obtained in $0(n^2)$ operations. This algorithm requires analytical first and second partial derivatives. Algorithm VF04AD, however, uses a conjugate factorization of the approximating Hessian matrix which is useful when gradient information is estimated by finite difference formulae (for further details see, for example, Coope (1987)).

All computations are performed in double precision to 7 decimal places of accuracy. The final results, however, are not dependent upon which algorithm is used in this step.

The other estimation techniques used (Chapter 9) are the trimmed least squares (defined in Chapter 7), and the OLS estimators. To obtain the trimmed least squares estimators, the computation of the θth regression quantile is required. Specifically, for the linear regression model,

$$y_t = x_t'\beta + \epsilon_t , \qquad (4.2)$$

where the ϵ_t are iid distributed with distribution function F, which is symmetrical around zero, and x_t is the t-th row of the nonstochastic matrix of K regressors X, the θth regression quantile, $(0 < \theta < 1)$ is defined as any solution to the minimization problem,

$$\min_{\beta}\left[\sum_{\{t/y_t\geq x_t'\beta\}} \theta|y_t-x_t'\beta| + \sum_{\{t/y_t<x_t'\beta\}} (1-\theta)|y_t-x_t'\beta|\right]. \qquad (4.3)$$

The minimization problem in (4.3) is a linear programming problem which can be solved using the algorithm of Koenker and D'Orey (1987). The alternative modified algorithm of Barrodale and Roberts (1974), (see, for example, Koenker and D'Orey (1987, p.385)) is also used, but there are no differences in the results obtained.

The OLS estimators of β in (4.2) are found using SUBROUTINE ELIM (Gerald and Wheatley (1984, p.144)) which solves a set of linear equations using the Gaussian elimination method.

The solutions of each of the algorithms used were compared with those in the standard Econometric packages TSP and SHAZAM, and were found to give similar results.

CHAPTER 5

THE NUMERICAL CALCULATION OF THE DISTRIBUTION FUNCTION OF A BILINEAR FORM TO A QUADRATIC FORM WITH ECONOMETRIC EXAMPLES

5.1 INTRODUCTION[1]

In many statistical and econometric applications statistics that are the ratio of a bilinear form to a quadratic form are used. The aim of this chapter is to show that the exact distribution of these statistics can be computed using techniques such as those developed by Imhof (1961) and Davies (1973). Numerous examples of the application of this technique will be given, such as Theil's (1961) two-stage least squares (TSLS) and K-class estimators and Nagar's (1962) double K-class estimator of the coefficient of the endogenous regressor in both a correctly specified and misspecified single structural equation. These examples are particularly important because in the last three decades analytical results for the exact density of many of these estimators have been found, as is reviewed in Chapter 6. However, three points can be noted. First, the results have been obtained by alternative techniques. Second, the resulting expressions are complicated and often not suitable for numerical evaluation. Third, the techniques that have been developed for the numerical evaluation of the distribution of one estimator are not easily extended to various other econometric estimators. This then emphasizes the objectives of our approach.

---

[1] The results of this chapter extend the work of Cribbett et al.(1989) on the TSLS estimator.

That is, the use of a single technique for various estimators and easy numerical evaluation.

In Section 2 it is shown that the distribution of a general bilinear form to a quadratic form is essentially the distribution of a quadratic form. In Section 3, various special cases of our main result in Section 2 are given, including estimators of the coefficient of the explanatory endogenous variable in a single structural equation and the estimator of the coefficient of a lagged dependent variable in a dynamic regression model.

5.2 MAIN RESULTS

Consider a class of statistics which are of the form

$$w = (x'A_2x)^{-1}(x'A_1y) \qquad (2.1)$$

where $y = (y_1 \ldots y_N)'$ and $x = (x_1 \ldots x_N)'$ are random column vectors such that the rows of $[y_i, x_i]$ $(i = 1 \ldots N)$ are independently normally distributed, each row having mean $[\mu_{iy}, \mu_{ix}]$ and nonsingular covariance matrix Σ, both A_1 and A_2 are nonstochastic and symmetric matrices and A_2 is assumed to be positive semi-definite.[2] In (2.1), w is the ratio of a bilinear to a quadratic form in normal variables. To obtain the results the following two Lemmas are used:

Lemma 1: The ratio $w = (x'A_2x)^{-1}(x'A_1y)$ can be written as a ratio of quadratic forms,

---

[2] The results in this section also hold for nonsymmetric matrices by noting that $x'Ax = x'\frac{1}{2}(A+A')x$ if A is not symmetric. The positive semi- definiteness of A_2 ensures that the matrix B_2 in (2.2) is positive semi- definite, which is assumed in Lemma 2.

$$w = (z'B_2z)^{-1}(z'B_1z). \qquad (2.2)$$

where z is a 2N × 1 vector distributed as $N_{2N}(\mu_{z_{\underline{2N}}}, \Omega)$,

$$z = \begin{bmatrix} x \\ y \end{bmatrix}, \quad B_1 = \frac{1}{2}\begin{bmatrix} 0 & A_1 \\ A_1 & 0 \end{bmatrix}, \quad B_2 = \begin{bmatrix} 0 & 0 \\ 0 & A_2 \end{bmatrix}$$

$$\mu_z = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \quad \Omega = \left((\Sigma \otimes I_N)\right), \qquad (2.3)$$

The result in the Lemma follows immediately by substituting (2.3) in (2.2).[3]

<u>Lemma 2</u>: The distribution function of w for a given q can be expressed as

$$F(q) = \Pr.\left[(x'A_2x)^{-1}(x'A_1y) \leq q\right]$$

$$= \Pr.\left[(z'B_2z)^{-1}(z'B_1z) \leq q\right]$$

$$= \Pr.\left[z'(B_1 - qB_2)z \leq 0\right]$$

$$= \Pr.\left[z^{*\prime}\Lambda z^{*} \leq 0\right]$$

$$= \Pr.\left[\sum_{j=1}^{2T} \lambda_j z_j^{*2} \leq 0\right]$$

where Λ is a diagonal matrix of eigenvalues of the matrix $\Omega^{\frac{1}{2}}(B_1 - qB_2)\Omega^{\frac{1}{2}}$, P is an orthogonal matrix of corresponding

---

[3] Lemma 1 is as in Cribbett <u>et al.</u> (1989). However, a different transformation is used in <u>Ullah</u> (1985).

eigenvectors and $z^* = P'\Omega^{-\frac{1}{2}}z \sim N_{2N}(P'\Omega^{-\frac{1}{2}}\mu_{\underline{z}_{2N}}, I_{2N})$. Furthermore, z_j^{*2} are independent noncentral chi-square variables each with one degree of freedom and noncentrality parameter $\delta_j^{*2} = [(P'\Omega^{-\frac{1}{2}}\mu)_j]^2$.

The details of Lemma 2 can be found in Koerts and Abrahamse (1971, pp.81-87).

Combining Lemmas 1 and 2 implies that the distribution function of w, which is a ratio of a bilinear form to a quadratic form, reduces to the distribution of a single quadratic form. The distribution of a single quadratic form can be computed easily using techniques such as those of Imhof (1961) and Davies (1973). With these techniques, in order to calculate the distribution function F(q), use is made of the inversion theorem of characteristic functions. This theorem enables a distribution function to be expressed in terms of its characteristic function.

The characteristic function of F(q) is defined as the complex function of the real variable t,

$$\phi(t) = \int_{-\infty}^{\infty} \exp(itx)dF(q).$$

Lévy (1925) proved that a distribution is uniquely determined by its characteristic function. Lévy's (1925) theorem, known as, *"the Uniqueness Theorem of characteristic functions"*, states that if (a-h, a+h) is a continuity interval of the distribution function

$F(q)$,[4] then

$$F(a+h) - F(a-h) = \lim_{T\Rightarrow\infty} \frac{1}{\pi} \int_{-T}^{T} \frac{\sin(ht)}{t} \exp(-ita)\phi(t)d(t). \quad (2.4)$$

From (2.4), it is derived that whenever two distributions have the same $\phi(t)$, the corresponding distribution functions are identical for any interval which is a continuity interval for both distributions. From this result, it then follows that the distributions are identical. However, this result does not give the distribution itself, but the difference $F(a+h) - F(a-h)$. Gil - Peleaz (1951) derived an inversion formula that gives $F(q)$ directly. He showed that for any random variable X with characteristic function $\phi(t)$ we have $Pr.(X \leq q) = F(q)$, where,

$$F(q) = 0.5 + \frac{1}{2\pi} \int_{0}^{\infty} \frac{\exp(itq)\phi(-t) - \exp(-itq)\phi(t)}{it} dt, \quad (2.5)$$

$$= 0.5 + \frac{1}{\pi} \int_{0}^{\infty} \frac{1}{t} I\left[\exp(-itq)\right]\phi(t) \quad dt, \quad (2.6)$$

where I[] denotes the imaginary part of the complex number.

To calculate the distribution function of w in (2.2), which is a ratio of quadratic forms, we know from Lemma 2 that we can calculate the distribution function of the single quadratic

---

[4] An interval (a,b) is called a continuity interval for $F(q)$ when both extremes a and b are continuity points of $F(q)$. That is, $\lim_{q\Rightarrow a} F(q) = F(a)$ and $\lim_{q\Rightarrow b} F(q) = F(b)$. $F(q)$ need not be continuous at every point in (a, b).

form $w^* = z^{*\prime}\Lambda z^*$. Thus, if we have the characteristic function of w^* we can use the formulae (2.5) and (2.6) noting in this case $q = 0$ and $\phi(t)$ is the characteristic function of w^*. That is, letting $X = z^{*\prime}\Lambda z^*$ and $q = 0$ in (2.5) and (2.6) we have,

$$F(q) = \Pr.\left[(x'A_2x)^{-1}(x'A_1y) \le q\right]$$

$$= \Pr.\left[z^{*\prime}\Lambda z^* \le 0\right],$$

where,

$$\Pr.\left[z^{*\prime}\Lambda z^* \le 0\right] = 0.5 + \frac{1}{2\pi}\int_0^\infty \frac{\phi(-t) - \phi(t)}{it}\, dt,$$

$$= 0.5 + \frac{1}{\pi}\int_0^\infty \frac{1}{t}\, I\left[\phi(t)\right] dt. \tag{2.7}$$

To find the characteristic function of w^* we note from Lemma 2 that,

$$\Pr.\left[z^{*\prime}\Lambda z^* \le 0\right] = \Pr.\left[\sum_{j=1}^{2T} \lambda_j z_j^{*2} \le 0\right].$$

Since z_j^{*2} are independent noncentral chi - square random variables, each with one degree of freedom and noncentrality parameter δ_j^{*2}, it is well known that the characteristic function of z_j^{*2} equals,

$$h(t) = (1 - 2it)^{-0.5} \exp\left[i\,\frac{\delta_j^{*2} t}{1 - 2it}\right]$$

(see e.g. Rao (1965, p.147). Furthermore, as the characteristic function of $\lambda_j z_j^{*2}$ equals $h_j(\lambda_j t)$, and also using the rule that the characteristic function of a sum of independent random variables equals the product of the individual characteristic functions, (see e.g. Lukacs and Laha (1964, p.21)), then,

$$\phi(t) = \prod_i (1 - 2i\lambda_j t)^{-0.5} \exp(i \Sigma \frac{\delta_j^{*2}\lambda_j t}{1 - 2i\lambda_j t}).$$

This implies that (2.7) is equal to,

$$\Pr.\left[z^{*\prime}\Lambda z^* \leq 0\right] =$$

$$0.5 + \frac{1}{\pi}\int_0^{\infty} \frac{1}{t} I\left[\prod_i(1-2i\lambda_j t)^{-0.5}\exp(i\Sigma\frac{\delta_j^{*2}\lambda_j t}{1 - 2i\lambda_j t})\right]dt. \quad (2.8)$$

Imhof (1961), (see also Koerts and Abrahamse (1971, pp. 78 - 80)), expresses I[ ] in known quantities, showing that (2.8) may be written as,

$$\Pr.\left[z^{*\prime}\Lambda z^* \leq 0\right] = 0.5 + \frac{1}{\pi}\int_0^{\infty} \frac{\sin \epsilon(u)}{u\gamma(u)} du,$$

where,

$$\epsilon(u) = 0.5 \Sigma \left[\tan^{-1}(\lambda_j u) + \delta_j^{*2}\lambda_j u(1 + \lambda_j^2 u^2)^{-1}\right] - 0.5qu$$

$$\gamma(u) = \Pi(1 + \lambda_j^2 u^2)^{0.25} \exp\left[(0.5 \Sigma(\delta_j^{*2}\lambda_j^2 u^2)/(1 + \lambda_j^2 u^2)\right].$$

Hence $\Pr.\left[z^{*\prime}\Lambda z^{*} \le 0\right]$ can be calculated by numerical integration.

In numerical work, the integration is carried out on a finite range only, say $0 \le u \le U$. Therefore, the degree of approximation will depend on two types of error, as well as the usual rounding - off errors. These are, the error arising from using an approximate rule to compute the integral, and secondly a truncation error,

$$t_u = \frac{1}{\pi}\int_0^\infty \frac{\sin \epsilon(u)}{u\gamma(u)}\, du.$$

Davies (1980) and Koerts and Abrahamse (1971) program the techniques of Davies (1973) and Imhof (1961) respectively, for the numerical inversion of (2.8)[5]. Farebrother (1984) has shown that the Davies (1980) routine achieves the desired level of accuracy more rapidly than the Koerts and Abrahamse (1971) routine. Another advantage of the Davies routine is that both the truncation and the numerical integration errors are controlled with guaranteed accuracy. For the numerical implementation of either of these techniques, the eigenvalues and corresponding eigenvectors of the matrix $\Omega^{\frac{1}{2}}(B_1-qB_2)\Omega^{\frac{1}{2}}$ are needed. These can always be obtained numerically and in some cases, as shown in Section 3, can be found analytically.[6]

---

[5] A Fortran version of Davies (1980) is used in this thesis which was supplied to the author by Robert Davies.

[6] The method used to find the eigenvalues and noncentrality parameters is illustrated in the appendix. It is similar to the methods of Anderson and Sawa (1973).

5.3 SPECIAL CASES

In Section 2 it was shown that the exact distribution of the ratio of a general bilinear to quadratic form can be obtained by using techniques such as those proposed by Imhof (1961) and Davies (1973). This result provides a simple method of obtaining the exact distribution of various econometric estimators and test statistics by using a single algorithm. For illustrative purposes five examples are considered:

5.3.i Double K-class estimator.

5.3.ii Reciprocal Double K-class estimator.

5.3.iii Misspecification Analysis.

5.3.iv Ratio of normal variables.

5.3.v Other Cases.

The objective here is to show that the exact distribution of each of these cases can be evaluated by using just one algorithm. The detailed analysis of each case is, however, beyond the scope of this chapter.

5.3.i Double K-class Estimator

We consider the distribution of the Double K-class estimator of the structural parameter β in an equation,

$$y_1^* = y_2^*\beta + X_1\gamma_1 + u, \qquad (3.1)$$

where y_1^* and y_2^* are N-component vectors of observations on the endogenous variables, X_1 is a $N \times G_1$ matrix of observations on exogenous variables, β is a scalar parameter, γ_1 is a G_1-component vector of parameters and u is a N-component vector of structural disturbances. The reduced-form of the system of structural equations includes,

$$(y_1^* \; y_2^*) = (X_1 X_2)\begin{bmatrix} \pi_{11}^* & \pi_{12}^* \\ \pi_{21}^* & \pi_{22}^* \end{bmatrix} + (v_1 v_2) = X\Pi^* + v \;, \qquad (3.2)$$

where X_2 is a $N \times G_2$ matrix of observations on G_2 exogenous variables that are excluded from (3.1), π_{11}^* and π_{12}^* are G_1-component vectors, π_{21}^* and π_{22}^* are G_2-component vectors, of reduced-form coefficients and $(v_1 v_2)$ is a $N \times 2$ matrix of reduced-form disturbances.

ASSUMPTION 1: The rows of (v_1, v_2) are independently normally distributed, each row having mean 0 and non-singular covariance matrix.

$$\Omega = \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} .$$

As $u = v_1 - \beta v_2$, the variance of each component of u is,

$$\sigma^2 = w_{11} - 2\beta w_{12} + \beta^2 w_{22}.$$

ASSUMPTION 2: The $N \times G$ $(G = G_1 + G_2)$ matrix X of exogenous variables is of rank G $(< N)$.

ASSUMPTION 3: The matrix $(\pi_{21}^* \; \pi_{22}^*)$ is of rank one and π_{22}^* has at least one non-zero component so that (3.1) is identified.

For any matrix D of full column rank let $P_D = D(D'D)^{-1}D'$ and $\bar{P}_D = I - D(D'D)^{-1}D'$. Then Nagar's (1962) Double K-class (DK) estimator with non- stochastic parameters K_1 and K_2 is

$$\hat{\beta}_{DK} = (y_2^{*\prime} A_1 y_2^*)^{-1}(y_2^{*\prime} A_2 y_1^*) \qquad (3.3)$$

where $A_j = K_j(P_x - P_{x_1}) + (1-K_j)\bar{P}_{x_1}$ $j = 1,2$. This class of estimators provides considerable appeal as a summary statement of several commonly used estimators. In particular, when $K_1 = K_2 = K$, (3.3) is Theil's K-class estimator. Also, if $K_1 = 1 - G(N - G - 3)^{-1}$ and $K_2 = 1 - G(1 + g)^{-1}(N - G - 3)^{-1}$ for a chosen g, we get Zellner's (1986) extended MELO estimator.

There exist transformations of the variables and parameters of the model given by (3.1) and (3.2) which transform it into one in which $\Omega = I_2$, the canonical form of the model. These transformations are given in Anderson and Sawa (1973) for example, and are

$$y_1 = w_{22}^{\frac{1}{2}}|\Omega|^{-\frac{1}{2}}\left[y_1^* - (w_{12}/w_{22})y_2^*\right] \quad , \quad y_2 = w_{22}^{-\frac{1}{2}}y_2^*.$$

The canonical form of the model depends on six key parameters. These are, the noncentrality parameter

$$\delta^2 = \frac{\pi_{22}^{*\prime} X_2\bar{P}_{x_1}X_2\pi_{22}^*}{w_{22}}, \tag{3.4}$$

the standard structural coefficient

$$\alpha = w_{22}|\Omega|^{-\frac{1}{2}}\left(\beta - \frac{w_{12}}{w_{22}}\right) \tag{3.5}$$

the number of excluded exogenous variables G_2, and the parameters N - G, K_1 and K_2. The corresponding form of the canonical model is

$$\Pi = \begin{bmatrix} w_{22}^{\frac{1}{2}}|\Omega|^{-\frac{1}{2}}\left(\pi_{11}^* - \frac{w_{12}}{w_{22}}\pi_{12}^*\right) & w_{22}^{\frac{1}{2}}\pi_{12}^* \\ w_{22}^{\frac{1}{2}}|\Omega|^{-\frac{1}{2}}\left(\pi_{21}^* - \frac{w_{12}}{w_{22}}\pi_{12}^*\right) & w_{22}^{\frac{1}{2}}\pi_{22}^* \end{bmatrix}. \tag{3.6}$$

The DK estimator of α in (3.5) is

$$\hat{\alpha}_{DK} = (y_2' A_1 y_2)^{-1}(y_2' A_2 y_1). \tag{3.7}$$

Applying Lemma 1 in Section 2 we can write (3.7) as

$$\hat{\alpha}_{DK} = (z' B_2 z)^{-1}(z' B_1 z)$$

where z is a $2N \times 1$ vector distributed $N_{2N}(\mathrm{Vec}(X\Pi), I_{2N})$ and B_1 and B_2 are symmetric matrices such that

$$z = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \qquad B_1 = \frac{1}{2}\begin{bmatrix} 0 & A_2 \\ A_2 & 0 \end{bmatrix}, \qquad B_2 = \begin{bmatrix} 0 & 0 \\ 0 & A_1 \end{bmatrix},$$

Further, using Lemma 2,

$$\Pr.(\hat{\alpha}_{DK} \le q) = \Pr.(z'(B_1 - qB_2)z \le 0) \tag{3.8}$$

where

$$z'(B_1 - qB_2)z = \sum_{r=1}^{2} \lambda_r \chi_r^2(G, \delta_r^2) + \sum_{r=3}^{4} \lambda_r \chi_r^2(N-G, 0). \tag{3.9}$$

The λ_r in (3.9) are the non-zero eigenvalues of the matrix $(B_1 - qB_2)$ such that

$$\lambda_1 = -\frac{1}{2}\left(q - (1+q^2)^{\frac{1}{2}}\right), \quad \lambda_2 = -\frac{1}{2}\left(q + (1+q^2)^{\frac{1}{2}}\right). \tag{3.10}$$

both with multiplicity G_2 and

$$\lambda_3 = -\frac{1}{2}\left(q(1-K_1) - (q^2(1-K_1)^2 + (1-K_2)^2)^{\frac{1}{2}}\right)$$
$$\lambda_4 = -\frac{1}{2}\left(q(1-K_1) + (q^2(1-K_1)^2 + (1-K_2)^2)^{\frac{1}{2}}\right) \tag{3.11}$$

both with multiplicity N - G. χ_1^2 and χ_2^2 in (3.9) are noncentral chi-square variables with G_2 degrees of freedom and noncentrality parameters δ_1^2 and δ_2^2 respectively, where

$$\delta_1^2 = \frac{\delta^2}{2}\left\{1 + \alpha^2 + (2\alpha - q + q\alpha^2)(1 + q^2)^{-\frac{1}{2}}\right\} \tag{3.12}$$

$$\delta_2^2 = \frac{\delta^2}{2}\left\{1 + \alpha^2 - (2\alpha - q + q\alpha^2)(1 + q^2)^{-\frac{1}{2}}\right\} \tag{3.13}$$

and χ_3^2 and χ_4^2 are each central chi-square variables with N - G degrees of freedom.

5.3.ii Reciprocal Double K-Class Estimator

It is well known that the DK estimator of β in (3.1) is not invariant to normalization. That is, we could apply DK to estimate $1/\beta$ in,

$$y_2^* = y_1^* \frac{1}{\beta} + X_1\left(\frac{-\gamma_1}{\beta}\right) + \left(\frac{-u_1}{\beta}\right) , \tag{3.14}$$

and then take the reciprocal of this statistic as an estimate of β. This will be called RDK and it yields a different estimation technique to DK for $G_2 > 1$. For a special case, Reciprocal Two Stage Least Squares ($K_1 = K_2 = 1$), Anderson and Sawa (1977) compare the reciprocal and direct procedures using approximate asymptotic expansions. However, for RDK in general a version of the Lemmas in Section 2 can be applied to find the exact distribution function. Using the canonical form of the model, the RDK with nonstochastic K_1 and K_2 is,

$$\hat{\alpha}_{RDK} = (y_1' A_2 y_2)^{-1}(y_1' A_1 y_1) , \tag{3.15}$$

which is the ratio of a quadratic to bilinear form. However, this can be written as a ratio of quadratic forms,

$$\hat{\alpha}_{RDK} = (z' B_2^* z)^{-1}(z' B_1^* z) \tag{3.16}$$

where z is a (2N × 1) vector distributed $N_{2N}(\text{vec}(X\Pi), I_{2N})$ and B_1^*

and B_2^* are symmetric matrices such that

$$z = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \qquad B_1^* = \begin{bmatrix} A_1 & 0 \\ 0 & 0 \end{bmatrix}, \qquad B_2^* = \frac{1}{2}\begin{bmatrix} 0 & A_2 \\ A_2 & 0 \end{bmatrix}.$$

Applying Lemma 2 we can write,

$$\Pr.(\hat{\alpha}_{RDK} \le q) = \Pr.\left(z'(B_1^* - qB_2^*)\, z \le 0\right) \tag{3.17}$$

where

$$z'(B_1^* - qB_2^*)z = \sum_{r=1}^{2} \lambda_{rR}\chi_r^2(G_2\ \delta_{rR}^2) + \sum_{r=3}^{4} \lambda_{rR}\chi_r^2(N-G,0) \tag{3.18}$$

The λ_{rR} are the non-zero eigenvalues of the matrix $(B_1^* - qB_2^*)$ such that

$$\lambda_{1R} = \frac{1}{2}\left(1 + (1+q^2)^{\frac{1}{2}}\right), \quad \lambda_{2R} = \frac{1}{2}\left(1 - (1+q^2)^{\frac{1}{2}}\right) \tag{3.19}$$

both with multiplicity G_2, and

$$\lambda_{3R} = \frac{1}{2}\left\{(1-K_1) + \left((1-K_1)^2 + (1-K_2)^2q^2\right)^{\frac{1}{2}}\right\} \tag{3.20}$$

$$\lambda_{4R} = \frac{1}{2}\left\{(1-K_1) - \left((1-K_1)^2 + (1-K_2)^2q^2\right)^{\frac{1}{2}}\right\} \tag{3.21}$$

both with multiplicity N - G. χ_1^2 and χ_2^2 are noncentral chi-square variables with G_2 degrees of freedom and noncentrality parameters.

$$\delta_{1R}^2 = \frac{\delta^2}{2}\left[\left(q^2 - 2\alpha q(1+(1+q^2)^{\frac{1}{2}}) + \alpha^2\left(1+(1+q^2)^{\frac{1}{2}}\right)^2\right)\left(1+(1+q^2)^{\frac{1}{2}}+q^2\right)^{-1}\right] \tag{3.22}$$

$$\delta_{2R}^2 = \frac{\delta^2}{2}\left[\left(q^2 + 2\alpha q(1+(1+q^2)^{\frac{1}{2}}-1) + \alpha^2\left(1-(1+q^2)^{\frac{1}{2}}\right)^2\right)\left(1-(1+q^2)^{\frac{1}{2}}+q^2\right)^{-1}\right] \tag{3.23}$$

for $q \neq 0$, and

$$\delta^2_{1R} = \alpha^2\delta^2$$

for q = 0 since $\lambda_2 = 0$. χ^2_3 and χ^2_4 are central chi-square variables with N - G degrees of freedom and δ^2 and α are defined in (3.4) and (3.5).

5.3.iii Misspecification Analysis

Rhodes and Westbrook (1981) derive the exact probability density function of the TSLS estimator when exogenous variables are wrongly excluded from the equation being estimated, but not from the system. The analysis of this type of misspecification of the DK and RDK estimators can easily be analyzed using the Lemmas of Section 2.

Suppose that the correctly specified pair of equations in the simultaneous equations model are

$$y_1^* = y_2^*\beta + X_1\gamma_1 + X_2\gamma_2 + u_1 \qquad (3.24)$$

$$y_2^* = y_2^*\lambda + X_1\gamma_3 + X_2\gamma_4 + u_2 \; . \qquad (3.25)$$

If X_1 is partitioned as

$$X_1 = (X_1^+ \; X_1^{++}) \; ,$$

where X_1^+ is $N \times G_1^+$ and X_1^{++} is $N \times G_1^{++}(G_1 = G_1^+ + G_1^{++})$ then (3.24) may be written as

$$y_1^* = \beta y_2^* + X_1^+\gamma_1^+ + X_1^{++}\gamma_1^{++} + X_2\gamma_2 + u_1 \qquad (3.26)$$

where γ_1 has been partitioned so as to conform with X_1. Let the correct specification for identification of (3.26) be

$$\gamma_2 = \underline{0} \; .$$

Misspecification occurs when the actual specification of (3.26) asserts

$$\gamma_2 = \underline{0} \qquad \text{and} \qquad \gamma_1^{++} = \underline{0}\ . \tag{3.27}$$

The DK estimator with this type of misspecification is

$$\hat{\beta}_{DK}^{M} = (y_2' A_1^M y_2)^{-1}(y_2' A_2^M y_1) \tag{3.28}$$

where $A_j^M = K_j(P_x - P_{x_1^+}) + (1-K_j)\bar{P}_{x_1^+}$ $j = 1,2$. Using the canonical form of the model and applying Lemma 1 in Section 2 we can write

$$\hat{\alpha}_{DK}^{M} = (z' B_2^M z)^{-1} (z' B_1^M z) \tag{3.29}$$

where z is a 2N × 1 vector distributed $N_{2N}(\mathrm{vec}(X\Pi), I_{2N})$ and B_1^M and B_2^M are symmetric matrices such that

$$z = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \qquad B_1^M = \frac{1}{2}\begin{bmatrix} 0 & A_2^M \\ A_2^M & 0 \end{bmatrix}, \qquad B_2^M = \begin{bmatrix} 0 & 0 \\ 0 & A_1^M \end{bmatrix},$$

Applying Lemma 2 we can write

$$\Pr.(\hat{\alpha}_{DK}^{M} \le q) = \Pr.\left(z'(B_1^M - qB_2^M)\, z \le 0\right) \tag{3.30}$$

$$z'(B_1^M - qB_2^M)z = \sum_{r=1}^{2} \lambda_r \chi_r^2(G_1^{++} + G_2, \delta_2^{2+}) + \sum_{r=3}^{4} \lambda_r \chi_r^2(N-G, 0) \tag{3.31}$$

The λ_r in (3.31) have previously been defined in (3.10) and (3.11). χ_1^2 and χ_2^2 are noncentral chi-square variables with $G_1^{++} + G_2$ degrees of freedom and noncentrality parameters,

$$\delta_1^{2+} = \frac{1}{2}\left((1+q^2+q(1+q^2)^{\frac{1}{2}}\right)^{-1}\left[\bar{\pi}_{12}^{++\prime}\,\bar{\pi}_{12}^{++} + \left(q+(1+q^2)^{\frac{1}{2}}\right)\left\{\bar{\pi}_{11}^{++\prime}\,\bar{\pi}_{12}^{++}\right.\right.$$

$$\left.\left. + \left(q+(1+q^2)^{\frac{1}{2}}\right)\bar{\pi}_{11}^{++\prime}\,\bar{\pi}_{11}^{++}\right\}\right] + \delta_1^2 \tag{3.32}$$

$$\delta_2^{2+} = \frac{1}{2}\left((1+q^2-q(1+q^2)^{\frac{1}{2}}\right)^{-1}\left[\bar{\pi}_{12}^{++\prime}\,\bar{\pi}_{12}^{++}+\left(q-(1+q^2)^{\frac{1}{2}}\right)\left\{\bar{\pi}_{11}^{++\prime}\,\bar{\pi}_{12}^{++}\right.\right.$$

$$\left.\left.+\left(q-(1+q^2)^{\frac{1}{2}}\right)\bar{\pi}_{11}^{++\prime}\,\bar{\pi}_{11}^{++}\right\}\right] + \delta_2^2 \qquad (3.33)$$

where δ_1^2 and δ_2^2 are defined in (3.12) and (3.13). The $\bar{\pi}_{ij}$, j = 1,2 are the relevant components of the matrix,

$$\begin{bmatrix} (X_1^{+\prime}\,X_1^{+}) & (X_1^{+\prime}\,X_1^{+})^{\frac{1}{2}}X_1^{+\prime}\,X_3 \\ 0 & X_3'\;\bar{P}_{x_1}\;X_3 \end{bmatrix}\Pi \ , \qquad (3.34)$$

where Π is defined in (3.6) and X_3 is a matrix containing (X_1^{++},X_2). χ_3^2 and χ_4^2 are central chi-square variables with N - G degrees of freedom.

The RDK with this type of misspecification is

$$\hat{\beta}^M_{RDK} = (y_1'A_2^M y_2)^{-1}(y_1'A_1^M y_1) \ , \qquad (3.35)$$

where A_j^M for j = 1,2 has been defined in (3.28). Using the canonical form of the model, and applying Lemma 1 in Section 2 we have

$$\hat{\alpha}^M_{RDK} = (z'B_2^{*M}z)^{-1}(z'B_1^{*M}z) \qquad (3.36)$$

where z is a (2N × 1) vector distributed $N_{2N}(\text{vec}(X\Pi),I_{2N})$ and B_1^{*M} and B_2^{*M} are symmetric matrices such that

$$z = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \qquad B_1^{*M} = \begin{bmatrix} A_1^M & 0 \\ 0 & 0 \end{bmatrix}, \qquad B_2^{*M} = \frac{1}{2}\begin{bmatrix} 0 & A_2^M \\ A_2^M & 0 \end{bmatrix}.$$

Applying Lemma 2 we can write

$$\Pr.(\hat{\alpha}^{M}_{RDK} \le q) = \Pr.(z'(B^{*M}_{1} - qB^{*M}_{2})z \le 0) \qquad (3.37)$$

where

$$(z'(B^{*M}_{1} - qB^{*M}_{2})z = \sum_{r=1}^{2} \lambda_{rR}\chi^2_r(G^{++}_1 + G_2, \delta^{2+}_{rR})$$

$$+ \sum_{r=3}^{4} \lambda_{rR}\chi^2_r(N - G, 0) \quad . \qquad (3.38)$$

The λ_{rR} in (3.37) have previously been defined in (3.19), (3.20) and (3.21). χ^2_1 and χ^2_2 are noncentral chi-square variables with G^{++}_1 and G_2 degrees of freedom and noncentrality parameters,

$$\delta^{2+}_{1R} = \tfrac{1}{2}\left((1+ \frac{1}{q^2} + \frac{1}{q^2}(1+q^2)^{\frac{1}{2}}\right)^{-\frac{1}{2}}\left[\bar{\pi}^{++'}_{12}\,\bar{\pi}^{++}_{12} + \frac{1}{q}\left(1+(1+q^2)^{\frac{1}{2}}\right)\right.$$

$$\left.\times \left\{\bar{\pi}^{++'}_{11}\,\bar{\pi}^{++}_{12} + \frac{1}{q}\left(1+(1+q^2)^{\frac{1}{2}}\right)\bar{\pi}^{++'}_{11}\,\bar{\pi}^{++}_{11}\right\}\right] + \delta^2_{1R} \quad . \qquad (3.39)$$

$$\delta^{2+}_{2R} = \tfrac{1}{2}\left((1+ \frac{1}{q^2} - \frac{1}{q^2}(1+q^2)^{\frac{1}{2}}\right)^{-\frac{1}{2}}\left[\bar{\pi}^{++'}_{12}\,\bar{\pi}^{++}_{12} - \frac{1}{q}\left(1-(1+q^2)^{\frac{1}{2}}\right)\right.$$

$$\left.\times \left\{\bar{\pi}^{++'}_{11}\,\bar{\pi}^{++}_{12} - \frac{1}{q}\left(1-(1+q^2)^{\frac{1}{2}}\right)\bar{\pi}^{++'}_{11}\,\bar{\pi}^{++}_{11}\right\}\right] + \delta^2_{2R} \quad . \qquad (3.40)$$

for $q \neq 0$ and where δ^2_{1R} and δ^2_{2R} are defined in (3.22) and (3.23). For $q = 0$, since $\lambda_2 = 0$ we just have,

$$\delta^2_{1R} = \bar{\pi}^{++'}_{11}\,\bar{\pi}^{++}_{11} + \bar{\pi}^{++'}_{12}\,\bar{\pi}^{++}_{12} + \alpha^2\delta^2 \qquad (3.41)$$

where δ^2 is defined in (3.4). χ^2_3 and χ^2_4 are central chi-square variables with N - G degrees of freedom.

5.3.iv Ratio of Normal Variables

Another special case of interest is when $A_1 = A_2 = A$ in (2.1) and A is an $m \times m$ positive semidefinite matrix of rank 1 which can be written as aa', where a is an $m \times 1$ vector. Therefore w in (2.1) becomes the ratio of normal variables $(a'x)^{-1}(a'y)$ and its distribution can be studied using Lemmas 1 and 2 in Section 2. The distribution of such ratios has been studied by, for example, Geary (1930), Fieller (1932) and Marsaglia (1965).

A particular example of this ratio is the TSLS (or RTSLS) estimator when $G_2 = 1$. Substituting $K_1 = K_2 = 1$ in (3.7) the estimator is,

$$\hat{\alpha} = \frac{X_2' \bar{P}_{X_1} y_1}{X_2' \bar{P}_X y_2} . \tag{3.42}$$

To evaluate the distribution function of $\hat{\alpha}$ the relevant eigenvalues and noncentrality parameters are, λ_1, λ_2, δ_1^2 and δ_2^2 as given in (3.10), (3.12) and (3.13) corresponding to $K_1 = K_2 = 1$.

To calculate the distribution function of a ratio $(a'x)^{-1}(a'y)$ in general, we note that it suffices to consider the distribution function of

$$w' = (c + p_2)^{-1}(b + p_1) \tag{3.43}$$

for b,c nonnegative constants and p_1, p_2 independent standard normal variables.[7] In this case

$$\Pr.(w' \leq q) = \Pr.\left[\sum_{r=1}^{2} \lambda_r \chi_r^2(1, \delta_r^2) \leq 0 \right], \tag{3.44}$$

where λ_1, λ_2 are given in (3.10) corresponding to $K_1 = K_2 = 1$ and

---

[7] If w is the ratio of two arbitrary normal random variables which may be correlated or not, then there exist constants c_1 and c_2 such that $c_1 + c_2 w$ have the same distribution as w′ where $w' = (b + p_1)(c + p_2)^{-1}$ for c, b nonnegative constants and p_1, p_2 are independent standard normal variables.

χ_1^2 and χ_2^2 are noncentral chi-square variables with 1 degree of freedom and non-centrality parameters,

$$\delta_1^2 = \frac{1}{2}\left(1+q^2+q(1+q^2)^{\frac{1}{2}}\right)^{-1}\left[c^2+\left(q+(1+q^2)^{\frac{1}{2}}\right)\left\{bc+\left(q+(1+q^2)^{\frac{1}{2}}\right)b^2\right\}\right] \qquad (3.45)$$

$$\delta_2^2 = \frac{1}{2}\left(1+q^2-q(1+q^2)^{\frac{1}{2}}\right)^{-1}\left[c^2+\left(q-(1+q^2)^{\frac{1}{2}}\right)\left\{bc-\left(q-(1+q^2)^{\frac{1}{2}}\right)b^2\right\}\right] \qquad (3.46)$$

5.3.v Other Cases

All of the previous examples have concentrated on the application of the main result presented in Section 2 to the evaluation of the distribution function of various estimators in the limited information linear simultaneous equations model. However, other examples do exist, such as the evaluation of the distribution functions of estimators in macro models with expectations (e.g. see Ullah (1985)) and the test statistic of a set of restrictions in the general linear model. Recently, Knight (1985a) used a technique developed by Davis (1976) to obtain the characteristic function of a quadratic form with a nonnormal error process characterized by an Edgeworth or Gram - Charlier series expansion, (see e.g. Knight (1985a, p.232) or Peters (1989, p.283). Therefore, using the notation of Lemma 2, where z* is now a vector of iid Edgeworth variables with mean μ_{z*} and covariance I, we have,

$$F(q) = \Pr.\left[\,(z'B_2z)^{-1}\,(z'B_1z) \le q)\,\right],$$

$$= \Pr.\left[\,z'(B_1 - qB_2)z \le q\,\right],$$

$$= \Pr.\left[z^{*\prime}\Lambda z^{*} \qquad \le 0\right],$$

$$= 0.5 + \frac{1}{2\pi}\int_0^\infty \frac{\phi(-t) - \phi(t)}{it}\, dt. \qquad (3.47)$$

From Equation (2.4) of Knight (1985a, p.234) we have,

$$\phi(t) = |I - 2it\Lambda|^{-1/2}\exp(1/2\mu' S\mu)$$

$$\Bigg\{ 1 + \frac{K_3}{6}\left[3 \sum_j (\Sigma\mu_k S_{kj}) S_{jj} + \sum_j(\Sigma\, \mu_k S_{kj})^3\right] +$$

$$\frac{K_4}{24}\left[3\sum_j S_{jj}^2 + 6\sum_j(\Sigma\mu_k S_{jj})^2 S_{jj} + \sum_j(\Sigma\mu_k S_{kj})^4\right] +$$

$$\frac{K_3^2}{72}\Big[9\sum_i\sum_j S_{ii}S_{jj}S_{ij} + 6\sum_i\sum_j S_{ij}^3 + 18\sum_i\sum_j (\Sigma\mu_k S_{ki})(\Sigma\mu_l S_{lj})S_{ij}^2 +$$

$$18\sum_i\sum_j (\Sigma\mu_k S_{ki})2S_{ij}S_{ij} + 9\sum_i\sum_j (\Sigma\mu_k S_{ki})(\mu_l S_{li})S_{ii}S_{jj} +$$

$$6\sum_i\sum_j (\Sigma\mu_k S_{ki})^3(\Sigma\mu_l S_{lj})S_{jj} + 9\sum_i\sum_j (\Sigma\mu_k S_{ki})^2(\Sigma\mu_l S_{lj})^2 S_{ij} +$$

$$\sum_i\sum_j (\Sigma\mu_k S_{ki})^3(\Sigma\mu_l S_{lj})^3\Big] \ldots \Bigg\},$$

where $\Omega^* = (I-2it\Lambda)$, $S = \Omega^{*-1} - I$ and K_r $(r=0,1,2 \ldots)$, are the standard cumulants, where in particular it is assumed that $K_3 \neq 0$ and $K_4 \neq 3$. The calculation of P and Λ are as for the normality assumption and the distribution function given in (3.47) can be numerically obtained by using appropriate algorithms in standard packages such as NAG and IMSL. Consequently, all of the previous examples can now be extended to include this type of nonnormality. Note however, that there are various limitations with this type of nonnormality. In particular, all of the z_i^*'s are assumed to have the same 3rd and 4th cumulants, and various restrictions are needed on the values of K_3 and K_4 to obtain a positive probability density function, these being $K_3 \leq 0.6$ and $0.3 \leq K_4 \leq 4.0$.

Another example is the evaluation of the distribution function of the Ordinary Least Squares (OLS) estimator of the coefficient of the lagged endogenous variable in dynamic models.[8] This example will be considered in some detail assuming both a correctly-specified and misspecified equation structure.

Suppose that the correctly specified equation is,

$$y = \alpha y_{-1} + X_1\beta_1 + X_2\beta_2 + u \tag{3.48}$$

where y and y_{-1} (the subscript referring to a one-period lagged value of y) are $n \times 1$ random vectors ($n = N - 1$), X_1 and X_2 are nonstochastic matrices of order $n \times K_1$ and $n \times K_2$ respectively, and α satisfies $|\alpha| < 1$. Assuming that the vector u is normally distributed with $E(u) = 0$, $E(uu') = \Omega_u$, then so is y with $E(y) = \mu$ where the t-th element of μ is,

$$\mu_t = \left(\frac{1}{1-\alpha L}\right)(X_{1t}\beta_1 + X_{2t}\beta_2) \ , \tag{3.49}$$

L being the Lag operator, and the variance-covariance matrix Ω_y is determined by the specification of Ω_u.[9] The OLS estimator for α is,

---

[8] Sawa (1978) evaluates the exact mean and variance of the least squares estimator of the stationary first-order autoregressive coefficient using the moments of a ratio of quadratic forms.

[9] In the Adaptive Expectation Model, for example, it is assumed that $u_t = v_t - \alpha v_{t-1}$ so that $\Omega_y = \sigma^2 I$.

$$\hat{\alpha} = (y'_{-1}My_{-1})^{-1}(y'_{-1}My) = (z'N_1z)^{-1}(z'Nz) \tag{3.50}$$

where $M = I - X(X\ X)^{-1}X'$, (using the notation $X = (X_1,X_2)$), an idempotent matrix of rank $n - (K_1+K_2)$, $z = [y_1 \ \dots \ y_T]$, $N = \frac{1}{2}(D_1'MD_2 + D_2'MD_1)$ and $N_1 = D_1'MD_1$ with $D_1 = [I_n,0]$ and $D_2 = [0,I_n]$, that is, identity matrices bordered by one column of zeroes.

However, if, for example, the exogenous variables X_2 are erroneously excluded from (3.48), then the misspecified equation is,

$$y = \alpha y_{-1} + X_1\beta_1 + u\ , \tag{3.51}$$

and the OLS estimator of α is

$$\hat{\alpha}^M = (y'_{-1}M^My_{-1})^{-1}(y'_{-1}M^My) = (z'N_1^Mz)^{-1}(z'N^Mz)\ , \tag{3.52}$$

where $M = I - X_1(X_1'X_1)^{-1}X_1'$, an idempotent matrix of $n - K_1$, $N^M = \frac{1}{2}(D_1'M^MD_2 + D_2'M^MD)$, $N_1 = D_1'M^MD_1$ and z, D_1 and D_2 are defined as above.

Both of these estimators, $\hat{\alpha}$, $\hat{\alpha}^M$, are ratios of quadratic forms so that the main result of Section 2 is applicable. However, in this example the eigenvalues and eigenvectors of the matrices $\Omega_y^{\frac{1}{2}}(N-qN_1)\Omega_y^{\frac{1}{2}}$ and $\Omega_y^{\frac{1}{2}}(N^M-qN_1^M)\Omega_y^{\frac{1}{2}}$ need to be found numerically. An illustration is given in Lye (1988).

CHAPTER 6

THE LIML AND TSLS ESTIMATORS

6.1 INTRODUCTION

The study of simultaneity did not become a dominant research program by the Cowles Foundation until Haavelmo (1944) recognized it as a unified approach to demand systems (Wright (1934)), Tinbergen's (1930) macroeconometric models and Frisch's (1933, 1934) confluent systems (see, for example, Epstein (1987)). The Cowles Foundation developed the theory of simultaneity as a multiple equation problem in Fisher's (1925) likelihood framework and, in particular, a distinction was made between limited-information and full-information SEM's. In the limited-information SEM, attention focusses on just one particular equation. In this case the investigator is not prepared to specify fully the equations of the rest of the system, but recognizes the necessity to develop special techniques that acknowledge the endogeneity of some of the regressors. This chapter looks at some finite-sample properties of the two most common estimators in limited-information SEM's, these being the LIML and TSLS estimators.

Section 2 of this chapter defines the two estimators. Recently, the finite-sample properties of these single-equation methods have been investigated extensively, and Section 3 reviews some of these studies. This review is divided up into three parts, these being moment results, properties of the exact distributions and misspecification analysis. Sections 4 and 5 present some new results on the comparison of the two estimators based on their finite-sample distributions when exogenous variables are wrongly excluded from the equation of interest but not from the system. Finally, Section 6 presents some conclusions and suggestions for future work. Throughout this chapter the canonical form of the limited- information SEM and the same notation will be used as is given in Section 3 of Chapter 5.

6.2 THE ESTIMATORS

The TSLS estimator is defined in Chapter 5, (5.3.7), as a member of the double K-class estimator family with nonstochastic parameters $K_1 = K_2 = 1$. That is, if we consider again the structural equation (5.3.1) in canonical form,

$$y_1 = y_2\alpha + X_1\gamma + u = H_1\delta_1 + u \ , \tag{2.1}$$

then the TSLS estimator is equal to,

$$\hat{\alpha} = \frac{y_2'\left(P_X - P_{X_1}\right)y_1}{y_2'\left(P_X - P_{X_1}\right)y_2} \ ,$$

where $P_D = D(D'D)^{-1}D'$ for any matrix D of full column rank. To define the LIML estimator, rewrite (2.1) as

$$Y_\Delta\beta_\Delta + X_1\gamma + u = 0 \ , \tag{2.2}$$

where $Y_\Delta = (y_1, y_2)$ and $\beta_\Delta = (-1, \alpha)'$.[1] The LIML estimator of β_Δ is the estimator obtained by maximizing the joint likelihood function of Y_Δ subject to the constraint $\Pi_2^0\gamma^0 = 0$, where $\Pi_2^0 = (\pi_{21} \ \pi_{22})$ and

---

[1] Equation (2.1) and (5.3.1) contain one endogenous regressor only, as this structural equation is the main focus of Chapters 5 and 6. However, the expressions given for the TSLS (see 5.3.7) and LIML (see 2.4) are also relevant for G(> 1) endogenous regressors by assuming the sizes of y_2 and α are respectively $(N \times G)$ and $(G \times 1)$.

$\gamma^0 = (1 \ \gamma)'$. However, when the reduced-form disturbances are multivariately normally distributed, the resulting estimator is identical to the LVR estimator of $\hat{\beta}_\Delta$. This estimator minimizes the variance ratio,

$$\ell = \frac{\beta'_\Delta Y'_\Delta \bar{P}_{x_1} Y_\Delta \beta_\Delta}{\beta'_\Delta Y'_\Delta \bar{P}_x Y_\Delta \beta_\Delta} , \tag{2.3}$$

where $\bar{P}_D = I - P_D$, and therefore is the solution to the equation,

$$(Y'_\Delta \bar{P}_{x_1} Y_\Delta - \hat{\ell} Y'_\Delta \bar{P}_x Y_\Delta)\hat{\beta}_\Delta = 0 , \tag{2.4}$$

where ℓ is the smallest root of the determinental equation,

$$|Y'_\Delta \bar{P}_{x_1} Y_\Delta - \hat{\ell} Y'_\Delta \bar{P}_x Y_\Delta| = 0 . \tag{2.5}$$

Some normalization rule must be imposed if the solution to (2.4) is to be unique, however, which normalization rule is imposed is of no consequence. This estimator is also a member of the double K-class estimator family with stochastic parameters $K_1 = K_2 = \hat{\ell}$.

Both the TSLS and LIML estimators are instrumental variable estimators (see, for example, Bowden and Turkington (1984, pp.110-113)), and this interpretation is useful for comparing the two estimation techniques. Using the second expression in (2.1) the LIML estimator of δ_1 is,

$$\tilde{\delta}_1 = (\tilde{H}'_1 \ H_1)^{-1}\tilde{H}_1 y_1 , \tag{2.6}$$

which is an instrumental variable estimator where the matrix of instruments is given by $\tilde{H}_1 = [\tilde{y}_2, X_1]$ where

$$\tilde{y}_2 = (X_1 \ X_2)\begin{pmatrix} \tilde{\pi}_{12} \\ \tilde{\pi}_{22} \end{pmatrix} ,$$

$H_1 = [y_2, X_1]$ and $\tilde{\pi}_{12}$, $\tilde{\pi}_{22}$ are the maximum likelihood estimators of the respective population coefficients subject to the restriction

$\Pi_2^0 \gamma_2^0 = 0$. The TSLS estimator of δ_1 is,

$$\hat{\delta}_1 = (\hat{H}_1' H_1)^{-1} \hat{H}_1 y_1 ,$$

where in this case the matrix of instruments $\hat{H}_1$ is given by,

$$\hat{H}_1 = (X_1, X_2) \begin{pmatrix} \hat{\pi}_{12} \\ \hat{\pi}_{22} \end{pmatrix} = P_x y_2 ,$$

where $\hat{\pi}_{12}$, $\hat{\pi}_{22}$ are the OLS estimators of the corresponding population parameters. Therefore, in forming the instruments the LIML estimator takes account of the overidentification restrictions whereas the TSLS estimator does not use this information.

However, both estimators have the same asymptotic distribution, that is,

$$\sqrt{N}(\tilde{\delta}_1 - \delta_1),\ \sqrt{N}(\hat{\delta}_1 - \delta_1) \overset{D}{\Rightarrow} MN(0,V)$$

where $V = \text{plim}\ (H_1' P_x H_1/N)^{-1}$. Further, both estimators are BAN.

6.3 FINITE SAMPLE PROPERTIES OF THE TSLS AND LIML ESTIMATORS: A REVIEW

Throughout this chapter it is assumed that the structural equation of interest is identified by means of zero restrictions; the sample size is greater than the number of exogenous variables and all of the predetermined variables are assumed to be exogenous. In these equations, one further distinction is made between the case of one and more than one endogenous regressor, due to the complexity of deriving finite-sample analytical results for the latter case. Equations of this type have been of interest for many years (see, for example, Haavelmo (1947), Bergstrom (1962) and Basmann (1961, 1963)).

This review concentrates on the finite-sample properties of the estimators of the coefficients on the endogenous regressors.[2] As both the TSLS and LIML estimators are complicated functions of the underlying random variables, their exact distributions are difficult to derive. Consequently, their use was first justified on the basis of large sample criteria, such as consistency and asymptotic efficiency. However, in the early 1960's the analysis of the exact distributions and moments of these estimators began, and since this time substantial progress has been made. Although these estimators are asymptotically equivalent, recent research has shown that their finite-sample properties are substantially different, and these differences are the focus of this review. In particular, three areas are considered, these being, moment results, exact distributions and misspecification analysis.

(i) Moment Results

The necessary and sufficient condition for the TSLS estimator to have finite absolute moments of positive order is that the order of the moments must be less than or equal to the degree of overidentification. This result was shown for special cases by Basmann (1961), Richardson (1968) and Sawa (1969) and extended more generally by Kinal (1980), and Hillier, Kinal and Srivastava (1984). The LIML estimator, however, has no positive finite moments of any order, as shown by Mariano and Sawa (1972) for the case of one endogenous regressor and more generally in an unpublished paper by Sargan (1970), and also in Phillips (1984a).

These moment results imply that the LIML estimator is inadmissible under a strictly quadratic loss function. However,

[2] Some results also exist for the estimators of the coefficients of the exogenous regressors, see for example Phillips (1984b).

this does not mean that the LIML estimator should be dropped in favour of the TSLS estimator, since, for example, we could be comparing a Cauchy distribution with high concentration about the true parameter value and a normal distribution with a finite but very large variance. Consequently, the two estimators have been compared using measures other than those that depend upon the existence of moments. One such measure, for example, is the "Probability of Concentration around the true parameter value" (see, for example, Rao (1981)), and in using such measures knowledge of the finite-sample distribution is important.[3]

(ii) Finite-Sample Distribution

In the case of one endogenous regressor, Richardson (1968) and Sawa (1969) derived the density of the TSLS estimator and Mariano and Sawa (1972) gave the density of the LIML estimator. Phillips (1980a, 1984a, 1985) extended both of these results to the case where there is an arbitrary number of endogenous regressors. However, the expressions for the densities involve complicated functions making general comparisons difficult, and numerical computations to date have concentrated only on the one endogenous regressor case. This is because the general expressions involve zonal-type polynomials which converge slowly and so are not yet suitable for numerical evaluation.

Anderson et al. (1979, 1982) give tables of the distribution functions of the two estimators when there is only one endogenous regressor. The method used to evaluate the distribution function of

---

[3] The Probability of Concentration for a particular estimator $\hat{\theta}$ of θ is defined as, $\Pr.(|\hat{\theta}-\theta| < \tau)$, for some τ. In other words, it considers the concentration of $\hat{\theta}$ around θ in a particular neighbourhood of θ.

the TSLS estimator is similar to the method explained in Chapter 5; however, it is not as general (see, for example, Cribbett et al. (1989)). To obtain the corresponding tables for the LIML estimator a simulation method is used. Comparisons of the performance of the estimators based on these tables indicate two major differences in their finite-sample distributions. These are:

- The distribution of the LIML estimator is essentially median- unbiased whereas the distribution of the TSLS estimator is badly distorted except for small α and/or large noncentrality parameters.

- The approach to its asymptotic distribution is very slow for the TSLS estimator and very rapid for the LIML estimator, so that even though the moments of the LIML estimator are not finite, the normal distribution is a very good approximation to the actual distribution.

Hillier (1988) considers that the differences in the performances of the two estimators are a result of the dependence of the TSLS estimator on the normalization rule, whereas the LIML estimator is invariant to this.

To analyze the sampling behaviour of the estimators when there is an arbitrary number of endogenous regressors, asymptotic approximations to the exact formulae have been used. In particular, Phillips (1983) applies the method of extracting marginal density approximations using the multivariate version of the Laplace formula to the instrumental variables estimators, which includes the TSLS estimator. Some features emerge from the numerical computations of these approximations, such as:

- as the number of endogenous regressors increases, the marginal distribution concentrates more slowly as N tends to infinity.

- the marginal distribution displays more bias as the degree of overidentification increases.

- the true values of the coefficients of the other endogenous regressors in the equation can affect the Probability of Concentration around the true parameter value of the marginal distribution of the estimator of the coefficient of the endogenous variable of interest.

These features are illustrated in Phillips (1980, pp.872-876; 1983, pp.13-19). No corresponding computations exist for the LIML estimator. However, Anderson et al. (1986) compare a number of estimators on the basis of their mean-squared errors and their Probability of Concentration around the true parameter value. These measures are computed by means of asymptotic expansions of their distributions when the disturbance variance tends to zero and, alternatively, when the sample size increases indefinitely. In particular, from these comparisons, it is recommended that the TSLS estimator should not be used in practice, and several modifications of the LIML estimator are given that are asymptotically admissible in the large-sample asymptotic theory. That is, they are third-order efficient. The particular choice of modification depends on criteria such as asymptotic mean-unbiasedness (e.g. Fuller (1977)) or asymptotic median unbiasedness (e.g. the LIML-estimator itself).

The combination of these results indicates the superior performance of the LIML estimator over TSLS, and therefore a long-standing issue over the choice of a single-equation estimator in a correctly specified SEM has been resolved.

(iii) Misspecification Analysis

Since typically in applied econometric studies economic theory provides some guidance but falls short of specifying the precise form of structural relationship, the possibilities for misspecification in SEM's are numerous. Although this area of analysis has not received a great deal of attention in the literature (see, for example, the comments of Taylor (1983) and Zellner (1979)), there have been contributions from Fisher (1961, 1966, 1967), Hale *et al.* (1980), Mariano and Ramage (1978, 1983) and Rhodes and Westbrook (1981, 1983), Knight (1982) and Skeels (1988).[4]

Fisher (1961, 1966, 1967) compares the large-sample asymptotic behaviour of the TSLS and LIML estimators in the presence of misspecification consisting of exclusion of relevant variables in a single equation. His principal result is that neither TSLS nor LIML dominates the other for all possible values of the specification error according to his criterion, which amounts to a weighted sum of squares of the large-sample asymptotic bias.

Hale *et al.* (1980) examine the effects of misspecification on the exact sampling moments of the K-class estimator family for nonstochastic K, and this family includes the TSLS estimator.

---

[4] See Phillips (1982, p. 503) for the correction of a minor error in Theorem 2.1 of Rhodes and Westbrook (1981).

Exact expressions and large concentration parameter asymptotic expansions are presented and analyzed for the bias and MSE of the K-class estimators in the case of one endogenous regressor. In particular, when relevant exogenous variables are omitted from the estimated equation but not from the system, the entire K-class for nonstochastic K between 0 and 1 is dominated in terms of large concentration parameter asymptotic MSE by either TSLS or OLS. In a similar study, Mariano and Ramage (1978) included the LIML estimator, which was also found to be dominated by either OLS or TSLS with respect to asymptotic MSE. Mariano and Ramage (1980) consider other types of misspecification including the omission of relevant endogenous variables and the misclassification of endogenous regressors as exogenous.

Knight (1982) gives an alternative derivation to that of Hale et al. (1980) of the effects of misspecification on both the OLS and TSLS estimators. Skeels (1988) examines the finite - sample properties of a class of instrumental variable estimators, (including OLS and TSLS but excluding LIML), when the system of equations, and in particular the equation being estimated, are misspecified by the incorrect exclusion of exogenous variables.

Rhodes and Westbrook (1981) compute the exact density function of the OLS and TSLS estimators, when exogenous variables are wrongly excluded from the equation being estimated and when there is only one endogenous regressor. The misspecified distributions are compared with the correctly specified ones on the basis of density function concentration (that is, the length of 90% probability intervals), and location around the true parameter value (that is, the midpoint of the probability interval). It is concluded that the effect of misspecification on estimator performance is ambiguous. In particular, for the TSLS estimator the following is concluded:

- the lengths of the TSLS probability intervals may increase or decrease under misspecification errors.

- the deviation of the midpoint of the probability intervals from the true parameter value may increase or decrease and may even change sign.

Overall, they conclude that under misspecification OLS may indeed be the superior estimation technique. However, no similar analysis exists for the LIML estimator, although Rhodes and Westbrook (1983) have considered some specific examples from which no general conclusions can be drawn. In the rest of this chapter, Rhodes and Westbrook's (1983) analysis is extended to include the LIML estimator, so that further comparisons between the TSLS and LIML estimators can be made.

6.4 THE KEY PARAMETERS IN THE MISSPECIFIED CANONICAL DISTRIBUTIONS

Let the structural equation of interest be written as,

$$y_1 = y_2\alpha + X_1^+\gamma_1^+ + X_1^{++}\gamma_1^{++} + X_2\gamma_2 + u_1, \tag{4.1}$$

where X_1^+ is $N*G_1^+$, X_1^{++} is $N*G_1^{++}$, $(G_1^+ + G_1^{++} = G_1)$, X_2 is $N*G_2$ and γ_1 has been partitioned so as to conform with X_1. Let the correct specification after identification of (4.1) be $\gamma_2 = \underline{0}$. Misspecification occurs when the actual specification of (4.1) asserts $\gamma_2 = \underline{0}$, and $\gamma_1^{++} = \underline{0}$. The TSLS estimator with this type of misspecification is,

$$\hat{\alpha}^{M} = (y_2'(P_x - P_{x_1^+})y_2)^{-1}(y_2'(P_x - P_{x_1^+})y_1). \tag{4.2}$$

In Chapter 5, Section 3, it is shown that (4.2) can be written as a ratio of quadratic forms, that is,

$$\hat{\alpha}^{M} = (z' B_2^M z)^{-1} (z' B_1^M z)$$

where z is a 2N * 1 normally distributed vector, and B_1^M and B_2^M are symmetric matrices such that

$$z = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \quad B_1^M = \frac{1}{2}\begin{bmatrix} 0 & (P_x - P_{x_1^+}) \\ (P_x - P_{x_1^+}) & 0 \end{bmatrix}, \quad B_2^M = \begin{bmatrix} 0 & 0 \\ 0 & (P_x - P_{x_1^+}) \end{bmatrix}.$$

This implies that the distribution function can be calculated as follows:

$$\Pr.(\hat{\alpha}^{M} \le q) = \Pr.\left(z'(B_1^M - qB_2^M)\, z \le 0\right)$$

$$z'(B_1^M - qB_2^M)z = \sum_{r=1}^{2} \lambda_r \chi_r^2 (G_1^{++} + G_2, \delta_2^{2+}).$$

The λ_r are the non-zero eigenvalues of the matrix $(B_1^M - qB_2^M)$ defined as,

$$\lambda_1 = \frac{-1}{2}\left[q - (1 + q^2)^{\frac{1}{2}}\right], \qquad \lambda_2 = \frac{-1}{2}\left[q + (1 + q^2)^{\frac{1}{2}}\right],$$

both with multiplicity $G_1^{++} + G_2$. χ_1^2 and χ_2^2 are noncentral chi-square variables with $G_1^{++} + G_2$ degrees of freedom and noncentrality parameters,

$$\delta_1^{2+} = \frac{1}{2}\left((1+q^2+q(1+q^2)^{\frac{1}{2}}\right)^{-1}\left[\bar{\pi}_{12}^{++'}\bar{\pi}_{12}^{++}+\left(q+(1+q^2)^{\frac{1}{2}}\right)\left\{\bar{\pi}_{11}^{++'}\bar{\pi}_{12}^{++}\right.\right.$$

$$\left.\left.+\left(q+(1+q^2)^{\frac{1}{2}}\right)\bar{\pi}_{11}^{++'}\bar{\pi}_{11}^{++}\right\}\right] + \delta_1^2$$

(4.3)

$$\delta_2^{2+} = \frac{1}{2}\left((1+q^2-q(1+q^2)^{\frac{1}{2}}\right)^{-1}\left[\bar{\pi}_{12}^{++'}\bar{\pi}_{12}^{++}+\left(q-(1+q^2)^{\frac{1}{2}}\right)\left\{\bar{\pi}_{11}^{++'}\bar{\pi}_{12}^{++}\right.\right.$$

$$\left.\left.+\left(q-(1+q^2)^{\frac{1}{2}}\right)\bar{\pi}_{11}^{++'}\bar{\pi}_{11}^{++}\right\}\right] + \delta_2^2$$

where δ_1^2, δ_2^2 and δ^2 are equal to,

$$\delta_1^2 = \frac{\delta^2}{2}\left[\ 1 + \alpha^2 + (2\alpha - q + q\alpha^2)(1 + q^2)^{-1/2}\right]$$

$$\delta_2^2 = \frac{\delta^2}{2}\left[\ 1 + \alpha^2 - (2\alpha - q + q\alpha^2)(1 + q^2)^{-1/2}\right] \qquad (4.4)$$

$$\delta^2 = \frac{\pi_{22}^{*'}\ X'(I - P_{X_1})X_2\pi_{22}^{*}}{w_{22}},$$

where π_{22}^* and w_{22} are defined in Chapter 5, Section 3,(see Assumption 1 and (5.3.6)). The $\bar{\pi}_{1j}$, j = 1,2 are the relevant components of the matrix,

$$\begin{bmatrix} (X_1^{+'}X_1^{+}) & (X_1^{+'}X_1^{+})^{\frac{1}{2}}X_1^{+'}X_3 \\ 0 & X_3'\ \bar{P}_{X_1}\ X_3 \end{bmatrix}\Pi ,$$

where X_3 is a matrix containing (X_1^{++}, X_2). Given this information, the exact distribution function of $\hat{\alpha}^M$ can be calculated using the techniques such as those developed by Imhof (1961) and Davies (1973, 1980), as described in Chapter 5. Furthermore, from (4.3) and (4.4) the "key parameters" of this distribution are given by the non-centrality parameter, δ^2, the true parameter value α, the degrees of freedom parameter $G_1^{++} + G_2$ and the parameters $\bar{\pi}_{11}^{++\prime}$, $\bar{\pi}_{11}^{++}$, $\bar{\pi}_{11}^{++\prime}$, $\bar{\pi}_{12}^{++}$ and $\bar{\pi}_{12}^{++\prime}$, $\bar{\pi}_{12}^{++}$.

The derivation of the LIML estimator of the endogenous regressor in the correctly specified model (5.3.1) begins with the joint distribution of two independent Wishart matrices (see Chapter 2), which will be denoted W and S. In particular, $W = Y_\Delta'\left(P_X - P_{X_1}\right)Y_\Delta$ (where X_1, X are given in (5.3.2) and the Projection matrix P_D is defined in (5.3.3)), is a noncentral Wishart matrix with degrees of freedom G_2, covariance matrix I and noncentrality parameter $\bar{M}$ where,

$$\bar{M} = \delta^2 \begin{bmatrix} \alpha^2 & \alpha \\ \alpha & 1 \end{bmatrix} ,$$

and $S = Y_\Delta'(I - P_X)Y_\Delta$ is a central Wishart statistic with N - G degrees of freedom and covariance matrix I. When the equation is misspecified as in (5.3.26) however, the distribution of W changes, although the distribution of S remains the same. In this case $W^M = Y_\Delta'\left(P_X - P_{X_1^+}\right)Y_\Delta$ (the superscript M representing the misspecified Wishart matrix, and X_1^+ is as defined in (5.3.26)) is a noncentral Wishart with degrees of freedom parameter $G_1^{++} + G_2$, covariance matrix I and noncentrality parameter,

$$M^* = \begin{bmatrix} \bar{\pi}_{11}^{++\prime}\,\bar{\pi}_{11}^{++} & \bar{\pi}_{11}^{++\prime}\,\bar{\pi}_{12}^{++} \\ \bar{\pi}_{12}^{++\prime}\,\bar{\pi}_{11}^{++} & \bar{\pi}_{12}^{++\prime}\,\bar{\pi}_{12}^{++} \end{bmatrix} + \bar{M} .$$

However, the matrices W^M and S remain independent since $\left(P_X - \bar{P}_{X_1^+}\right) \times (I - P_X) = 0$ (see Chapter 2), so that the results of Hillier (1987) can be used to obtain the analytical expression for the distribution function of the LIML estimator subject to this type of misspecification. However, for the purposes of the simulation experiment performed in this chapter, the key parameters have already been identified, and are the parameters in the Wishart distributions of S and W^M. These parameters are N - G, $G_2 + G_1^{++}$, α, δ^2, $\bar{\pi}_{11}^{++\prime}$, $\bar{\pi}_{11}^{++}$, $\bar{\pi}_{12}^{++\prime}$, $\bar{\pi}_{11}^{++}$ and $\bar{\pi}_{12}^{++\prime}$, $\bar{\pi}_{12}^{++}$. Therefore, as in the correctly specified case, the only Key parameter that differs from those that affect the TSLS distribution function is N - G.

6.5 PROPERTIES OF THE MISSPECIFIED DISTRIBUTIONS

When the structural equation of interest is misspecified by the exclusion of relevant exogenous variables, such as in (5.3.26), the density functions of the TSLS and LIML estimators of the endogenous regressor coefficients contain a number of key parameters in addition to those that affect the densities in the correctly specified model. These parameters are: $\bar{\pi}_{11}^{++\prime}$, $\bar{\pi}_{11}^{++}$ and $\bar{\pi}_{12}^{++\prime}$, $\bar{\pi}_{12}^{++}$ which are both non - negative, ≥ 0, and $\bar{\pi}_{11}^{++\prime}$, $\bar{\pi}_{12}^{++} \gtreqless 0$ and, in addition, the degrees of freedom parameter increases to include the number of wrongly excluded exogenous variables. The effect of these parameters on the density functions is examined in this section.

In particular, the influence of misspecification upon the density functions is analyzed for a number of parameter constellations by numerically evaluating the exact distribution of the TSLS estimator as descibed above and in Chapter 5, and simulating the distribution of the LIML estimator via the integration of the Kernel density estimator with the naive Monte Carlo method. The kernel estimate at point X is equal to,

$$\hat{pdf}(X) = \frac{1}{N^*h(N^*)}\sum_j k\left[\frac{X - X_j}{h(N^*)}\right], \qquad (5.1)$$

where k[.] is the standard N(0,1) density. The window width h(N*) is chosen using the interactive approach of Tapia and Thompson (1978). In all cases this approach led to the use of a window width between 0.02 and 0.09. N* is simply the number of replications in the simulation experiment, and is chosen using the bound of estimation. For example, the results of Parzen (1962) and Cacoullos (1966) imply that,

$$\left(N^*h^m(N^*)\right)^{\frac{1}{2}}\left[\hat{pdf}(x) - E\left(\hat{pdf}(x)\right)\right] \sim N\left(0, pdf(x)\int K^2\right). \qquad (5.2)$$

The result in (5.2) can be achieved if $\left(N^*h^m(N^*)\right)^{\frac{1}{2}}Bias\left[\hat{pdf}(x)\right]$ tends to zero asymptotically since,

$$\left(N^*h^m(N^*)\right)^{\frac{1}{2}}\left[\hat{pdf}(x) - pdf(x)\right] = \left(N^*h^m(N^*)\right)\left[\hat{pdf}(x) - E\left(\hat{pdf}(x)\right)\right]$$

$$+ \left(N^*h^m(N^*)\right)^{\frac{1}{2}}Bias\left[\hat{pdf}(x)\right].$$

Ullah (1988, p.642) shows that $Bias\left[\hat{pdf}(x)\right]$ is proportional to $h^2(N^*)$. This implies that if $N^*h^{(4+m)/2}(N^*)$ tends to zero asymptotically then (5.2) holds. Therefore, for the normal kernel, $\frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}y^2)$, the 99% asymptotic confidence interval for $\hat{pdf}(X)$ is given by,

$$\hat{pdf}(X) \pm 2.58 \left[\frac{\hat{pdf}(X)}{2N^* h\sqrt{\pi}}\right]^{\frac{1}{2}},$$

so that B is given by,

$$B = 2.58\left[\frac{\hat{pdf}(X)}{2N^* h(N^*)\pi}\right]^{\frac{1}{2}}$$

N* is varied until B is less than 0.01 for all points at which the density is estimated. In all experiments, N* varies between 60,000 and 90,000 replications[5]. The input of X_j in (5.1) involves numerically maximizing the likelihood function to obtain the LIML estimator of α. Two algorithms from the Harwell Subroutine library are used, these being algorithms VAI3AD and VF04AD, which both use the BFGS formula, (Broydon (1970), Fletcher (1970), Goldfard (1970) and Shanno (1970)). All computations are performed in double precision to 7 decimal places of accuracy. The final results, however, are not dependent upon which algorithm is used in this step. Furthermore, the solutions of each of the algorithms used were compared with those in the standard econometric packages TSP and SHAZAM, and were found to give similar results. Random numbers distributed uniformly on the interval [0,1], denoted U, are generated using the NAG subroutine GOFCAF, which uses a multiplicative congruential method. Standard normal variates, N(0,1), were generated using the NAG subroutine G05DDF, which is

---

[5]Empirical densities were also computed using the Epanechnikov (1969) kernel. However, given the number of replications used, the results proved not to depend on which kernel is used. This situation is similar to the comparison of different kernels for the Cauchy distribution using a "large sample", as is illustrated in Figure 5.1 in Chapter 3.

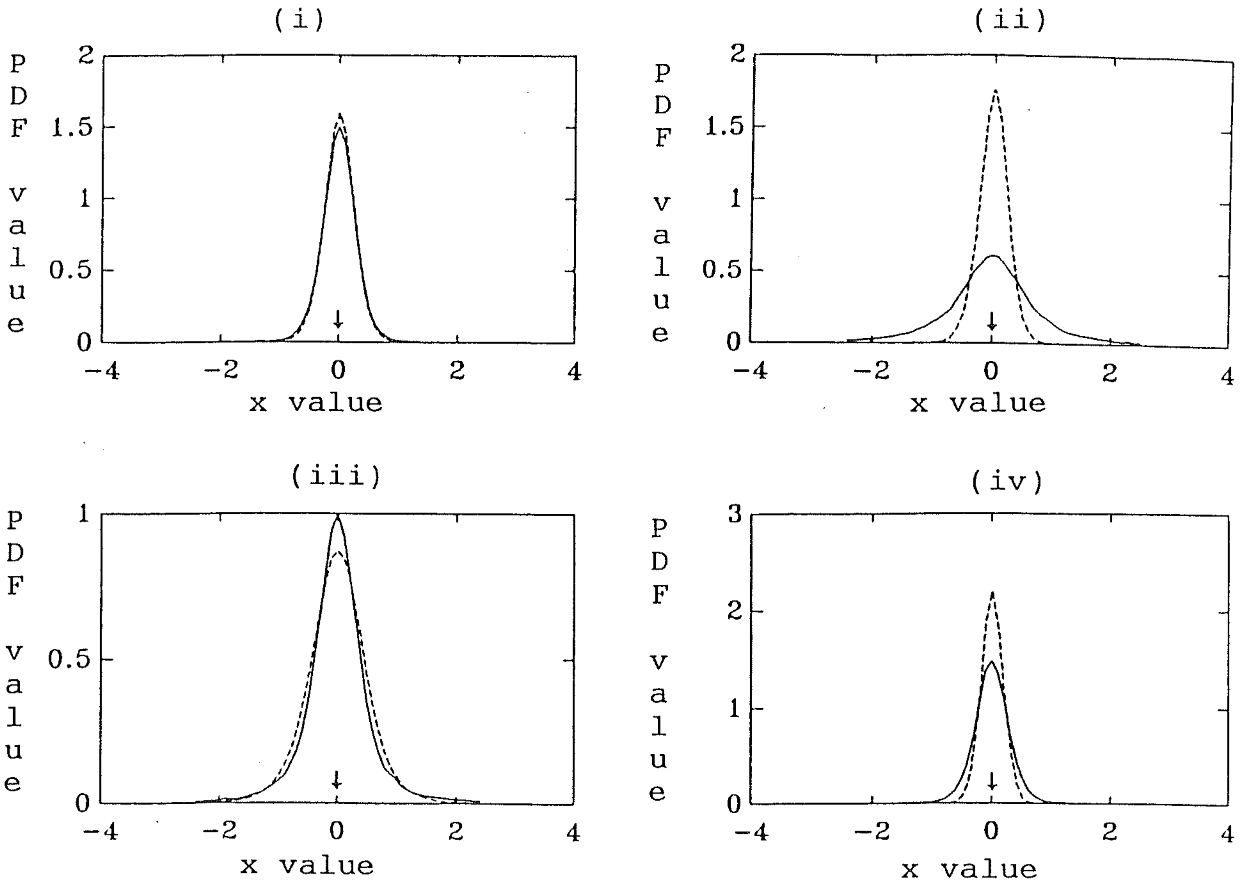
based on Brent's (1974) algorithm. Further details are given in Chapter 4.

A selection of these densities is presented in Figures 5.1-5.3. The median and interquartile range (IQR) are also computed and these are used as summary measures of the influence of misspecification on the location and concentration of the density functions. A selection of these values is presented in Table 5.1. These were calculated exactly for the TSLS estimator using the Davies (1980) routine and were estimated for the LIML estimator as described in Mood, Graybill and Boes (1986, p.75). The same number of replications used to estimate the PDF is used here. In each of these computations it is necessary only to consider the parameter space defined by $\alpha \geq 0$, as the respective densities for $\alpha < 0$ are simply the mirror images of their corresponding positive counterpart.

The analysis begins with the case of one wrongly excluded exogenous variable and the effect of misspecification on the location and IQR of the density functions. In this case in determining the effect of the parameters $\bar{\pi}_{11}^{++\prime}$, $\bar{\pi}_{11}^{++}$, $\bar{\pi}_{12}^{++\prime}$, $\bar{\pi}_{11}^{++}$, $\bar{\pi}_{12}^{++\prime}$, $\bar{\pi}_{12}^{++}$ on the misspecified distributions, it is sufficient to discuss the effects of $\bar{\pi}_{11}^{++}$ and $\bar{\pi}_{12}^{++}$ only. From Table 5.1 the following comments can be made. For the TSLS estimator the conclusions are similar to those of Rhodes and Westbrook (1981).

- the IQR can increase or decrease in comparison to the correctly specified model. Increases in the IQR are associated with increases in $|\bar{\pi}_{11}^{++}|$, while decreases are associated with increases in $|\bar{\pi}_{12}^{++}|$. Therefore, these two parameters exert opposing influences, although in general, misspecification is associated

FIGURE 5.1 Densities for the LIML and TSLS Estimators when $\alpha = 0.0, \delta^2 = 16$



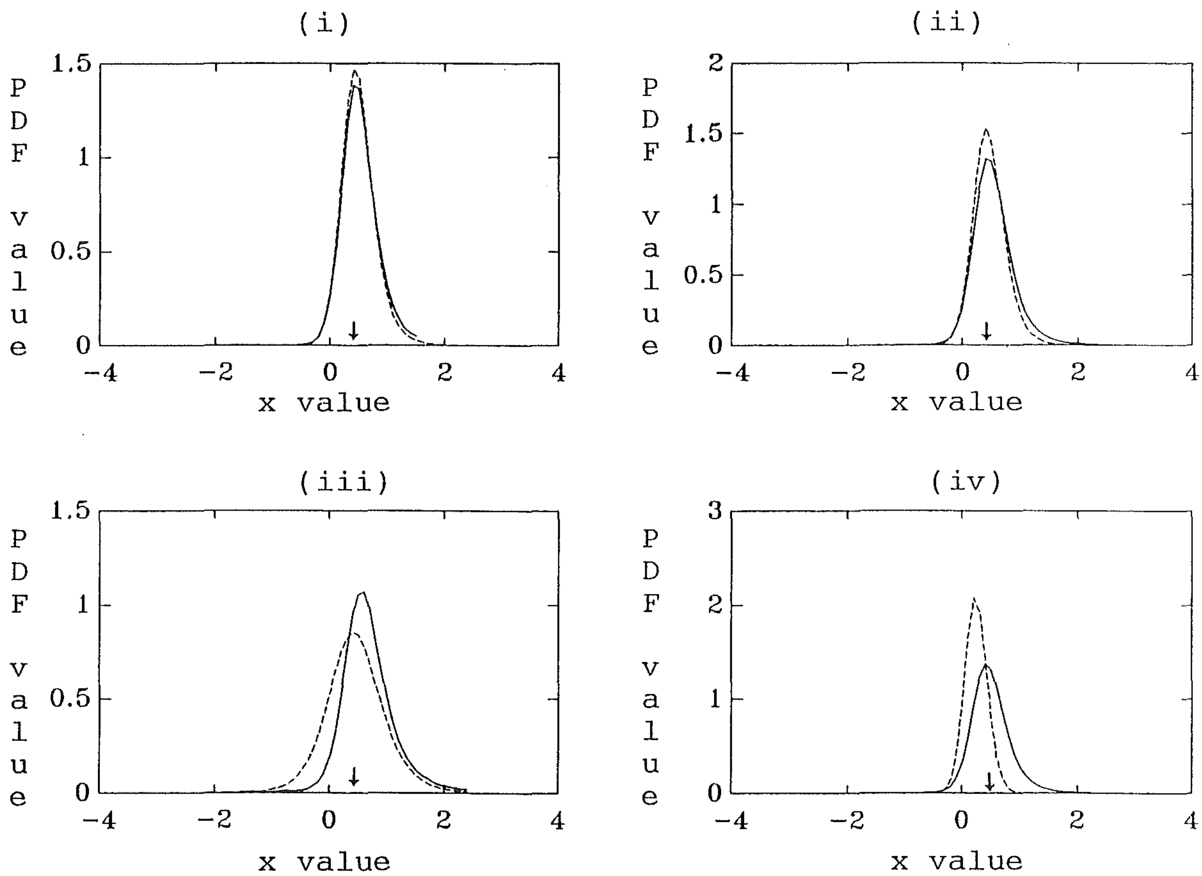
KEY ↓ True Parameter Value

\- LIML Estimator

-- TSLS Estimator

FIGURE 5.2 Densities for the LIML and TSLS Estimators

When $\alpha = 0.5$, $\delta^2 = 16$



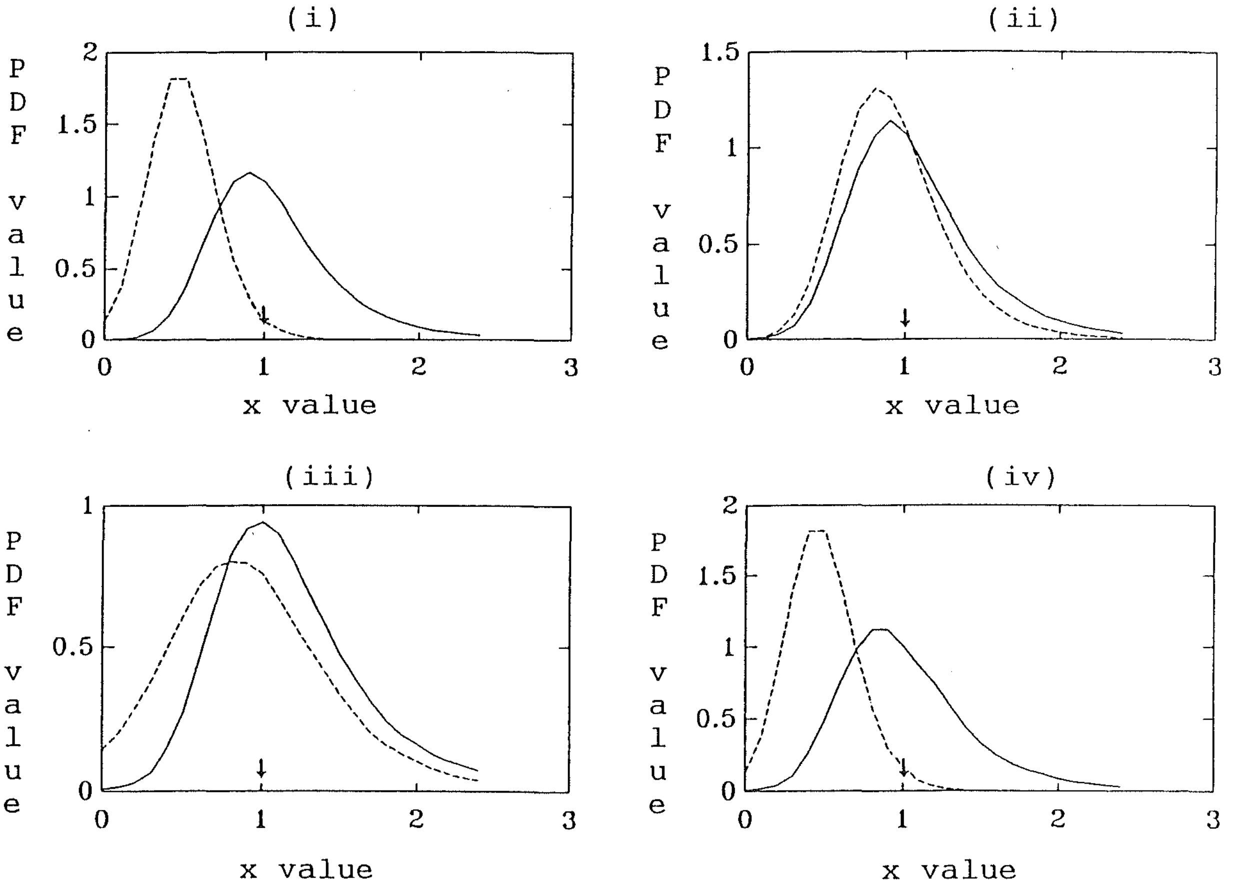
KEY ↓ True Parameter Value

– LIML Estimator

-- TSLS Estimator

FIGURE 5.3 Densities for the LIML and TSLS Estimators

When $\alpha = 1.0,\ \delta^2 = 16$



KEY ↓ True Parameter Value

– LIML Estimator

-- TSLS Estimator

Table 5.1: Median and IQR for LIML and TSLS Estimators

	Median for $\bar{\pi}_{11}^{++}$ =				IQR for $\bar{\pi}_{11}^{++}$ =			
$\bar{\pi}^{++}$	0	2	4	7	0	2	4	7
	LIML $\alpha = 0.0$ $\delta^2 = 16$ $N - G = 20$ $G_2 = 2$ $G_1^{++} = 1$							
0	0.0	0.0	0.0	0.0	0.374	0.400	0.4029	0.4680
2	0.0	0.011	0.0239	0.0441	0.370	0.381	0.4001	0.4600
4	0.0	0.0216	0.0434	0.0830	0.369	0.374	0.3930	0.4491
	correctly specified Median = 0.0 IQR = 0.42							
	TSLS $\alpha = 0.0$ $\delta^2 = 16$ $G_2 = 2$ $G_1^{++} = 1$							
0	0.0	0.0	0.0	0.0	0.3200	0.354	0.444	0.624
2	0.0	0.091	0.182	0.318	0.2900	0.316	0.382	0.5230
4	0.0	0.117	0.235	0.412	0.2320	0.246	0.282	0.36399
	correctly specified Median = 0.0 IQR = 0.333							
	LIML $\alpha = 0.5$ $\delta^2 = 5$ $N - G = 10$ $G_1^{++} = 1$ $G_2 = 2$							
0	0.4778	0.5022	0.5666	0.8230	0.8452	0.8956	1.0712	1.9363
2	0.4542	0.5142	0.6089	0.9021	0.8174	0.8362	0.7445	1.7909
4	0.4056	0.4892	0.5973	0.8537	0.7531	0.9430	0.7988	1.4788
	correctly specified Median = 0.4834 IQR = 0.7634							
	TSLS $\alpha = 0.5$ $\delta^2 = 5$ $G_1^{++} = 1$ $G_2 = 2$							
0	0.338	0.338	0.338	0.338	0.5440	0.6690	0.9630	1.50
2	0.223	0.403	0.586	0.8999	0.4299	0.4910	0.6467	0.961
4	0.109	0.282	0.456	0.717	0.291	0.3130	0.3730	0.501
	correctly specified Median 0.397 IQR = 0.60199							
	LIML $\alpha = 0.5$ $\delta^2 = 16$ $N - G = 10$ $G_2 = 2$ $G_1^{++} = 1$							
0	0.5003	0.5037	0.5302	0.6075	0.4147	0.4246	0.4644	0.5865
2	0.4886	0.5083	0.5435	0.6338	0.4106	0.4138	0.4403	0.5308
4	0.4676	0.4967	0.5405	0.6371	0.3903	0.3953	0.4118	0.4760
	correctly specified Median 0.4935 IQR = 0.3966							
	TSLS $\alpha = 0.5$ $\delta^2 = 16$ $G_2 = 2$ $G_1^{++} = 1$							
0	0.4419	0.4419	0.4419	0.4419	0.3480	0.3800	0.464	0.638
2	0.363	0.4530	0.5440	0.6810	0.3110	0.3450	0.396	0.531
4	0.235	0.353	0.470	0.6470	0.2460	0.257	0.292	0.370
	correctly specified Median = 0.4960 IQR = 0.3640							

Table 5.1 continued.

$\bar{\pi}_{12}^{++}$	Median for $\bar{\pi}_{11}^{++}$ = 0	2	4	7	IQR for $\bar{\pi}_{11}^{++}$ = 0	2	4	7
	LIML α = 0.5 δ^2 = 25 N - G = 10 G_2 = 2 G_1^{++} = 1							
0	0.5007	0.5055	0.5197	0.5683	0.3206	0.3301	0.3507	0.4077
2	0.4949	0.5077	0.5001	0.5846	0.32296	0.3252	0.3158	0.3810
4	0.4814	0.5288	0.5274	0.5899	0.3188	0.3394	0.3235	0.3572
	correctly specified Median = 0.4985 IQR = 0.2971							
	TSLS α = 0.5 δ^2 = 25 N - G = 10 G_2 = 2 G_1^{++} = 1							
0	0.46199	0.46199	0.46199	0.46199	0.28699	0.30399	0.35099	0.45799
2	0.4030	0.46699	0.53199	0.62799	0.2634	0.27799	0.31699	0.40399
4	0.291	0.384	0.47599	0.61599	0.221	0.22899	0.25299	0.311
	correctly specified Median = 0.48099 IQR = 0.29499							
	LIML α = 1.0 δ^2 = 16 N - G = 10 G_1^{++} = 1.0 G_2 = 2							
0	0.991	1.0073	1.0372	1.1234	0.5157	0.5257	0.5512	0.6299
2	0.9884	0.9989	1.0272	1.1051	0.5175	0.5125	0.5237	0.5729
4	0.9601	0.9711	0.9989	1.072	0.5108	0.4942	0.4948	0.5205
-2	NA	0.9993	1.0302	1.1207	NA	0.5396	0.5817	0.6962
-4	NA	0.9699	0.9991	1.084	NA	0.5493	0.6074	0.6432
	correctly specified Median = 0.999 IQR = 0.4952							
	LIML α = 1.0 δ^2 = 16 N - G = 20 G_1^{++} = 1.0 G_2 = 2							
0	1.0024	1.0068	1.0231	1.0686	0.5132	0.5196	0.5496	0.5718
2	0.9965	1.0012	1.0169	1.0626	0.5152	0.5114	0.5188	0.5442
4	0.9799	0.9852	1.0016	1.0444	0.5113	0.5026	0.5028	0.5185
	correctly specified Median = 1.0003 IQR = 0.4899							
	TSLS α = 1.0 δ^2 = 16 G_1^{++} = 1.0 G_2 = 2							
0	0.885	0.885	0.885	0.885	0.420	0.450	0.530	0.680
2	0.730	0.815	0.912	1.045	0.370	0.378	0.440	0.565
4	0.471	0.588	0.702	0.886	0.280	0.289	0.325	0.395
-2	NA	0.635	0.544	0.410	NA	0.392	0.451	0.577
-4	NA	0.353	0.236	0.059	NA	0.292	0.327	0.399
	correctly specified Median = 0.930 IQR = 0.451							

Table 5.1: Median and IQR for LIML and TSLS Estimators

$\bar{\pi}^{++}$	Median for $\bar{\pi}^{++}_{11}$ = 0	2	4	7	IQR for $\bar{\pi}^{++}_{11}$ = 0	2	4	7
	LIML α = 0.0 δ^2 = 16 N - G = 20 G_2 = 2 G_1^{++} = 1							
0	0.0	0.0	0.0	0.0	0.374	0.400	0.4029	0.4680
2	0.0	0.011	0.0239	0.0441	0.370	0.381	0.4001	0.4600
4	0.0	0.0216	0.0434	0.0830	0.369	0.374	0.3930	0.4491
	correctly specified Median = 0.0 IQR = 0.42							
	TSLS α = 0.0 δ^2 = 16 G_2 = 2 G_1^{++} = 1							
0	0.0	0.0	0.0	0.0	0.3200	0.354	0.444	0.624
2	0.0	0.091	0.182	0.318	0.2900	0.316	0.382	0.5230
4	0.0	0.117	0.235	0.412	0.2320	0.246	0.282	0.36399
	correctly specified Median = 0.0 IQR = 0.333							
	LIML α = 0.5 δ^2 = 5 N - G = 10 G_1^{++} = 1 G_2 = 2							
0	0.4778	0.5022	0.5666	0.8230	0.8452	0.8956	1.0712	1.9363
2	0.4542	0.5142	0.6089	0.9021	0.8174	0.8362	0.7445	1.7909
4	0.4056	0.4892	0.5973	0.8537	0.7531	0.9430	0.7988	1.4788
	correctly specified Median = 0.4834 IQR = 0.7634							
	TSLS α = 0.5 δ^2 = 5 G_1^{++} = 1 G_2 = 2							
0	0.338	0.338	0.338	0.338	0.5440	0.6690	0.9630	1.50
2	0.223	0.403	0.586	0.8999	0.4299	0.4910	0.6467	0.961
4	0.109	0.282	0.456	0.717	0.291	0.3130	0.3730	0.501
	correctly specified Median 0.397 IQR = 0.60199							
	LIML α = 0.5 δ^2 = 16 N - G = 10 G_2 = 2 G_1^{++} = 1							
0	0.5003	0.5037	0.5302	0.6075	0.4147	0.4246	0.4644	0.5865
2	0.4886	0.5083	0.5435	0.6338	0.4106	0.4138	0.4403	0.5308
4	0.4676	0.4967	0.5405	0.6371	0.3903	0.3953	0.4118	0.4760
	correctly specified Median 0.4935 IQR = 0.3966							
	TSLS α = 0.5 δ^2 = 16 G_2 = 2 G_1^{++} = 1							
0	0.4419	0.4419	0.4419	0.4419	0.3480	0.3800	0.464	0.638
2	0.363	0.4530	0.5440	0.6810	0.3110	0.3[illegible]50	0.396	0.531
4	0.235	0.353	0.470	0.6470	0.2460	0.257	0.292	0.370
	correctly specified Median = 0.4960 IQR = 0.3640							

with decreases in concentration as $|\overline{\pi}_{11}^{++}|$ has to be considerably larger with respect to $|\overline{\pi}_{12}^{++}|$ before the IQR increases.

- the parameters $|\overline{\pi}_{11}^{++}|$ and $|\overline{\pi}_{12}^{++}|$ also exert opposing influences on the median. In particular, the value of the median increases as $|\overline{\pi}_{11}^{++}|$ increases and decreases as $|\overline{\pi}_{12}^{++}|$ increases. In the correctly specified model, the TSLS estimator is badly median under-biased (see, for example, Anderson et al. (1979)). In the misspecified model, the TSLS estimator may become median over-biased, unbiased or remain under-biased, depending on the values of $|\overline{\pi}_{11}^{++}|$ and $|\overline{\pi}_{12}^{++}|$.

Although both $|\overline{\pi}_{11}^{++}|$ and $|\overline{\pi}_{12}^{++}|$ exert similar influences on the misspecified density of the LIML estimator, the extent to which this density is affected by these parameters differs. In particular,

- in the correctly specified model the LIML estimator is essentially median-unbiased (see, for example, Anderson et al. (1982)), so that increases in $|\overline{\pi}_{11}^{++}|$ imply the estimator becomes median over-biased and increases in $|\overline{\pi}_{12}^{++}|$ imply the estimator becomes median under-biased. However, even for very large values of $|\overline{\pi}_{11}^{++}|$ and $|\overline{\pi}_{12}^{++}|$, the density becomes only mildly median-biased.

- the IQR for the misspecified densities only moderately differs from that of the correctly specified model, although in general it increases as $|\overline{\pi}_{11}^{++}|$ exceeds the value of $|\overline{\pi}_{12}^{++}|$.

These comments consider the absolute effects of $\overline{\pi}_{11}^{++}$ and $\overline{\pi}_{12}^{++}$, however, it is also interesting to note the differences that occur when $\overline{\pi}_{11}^{++}\,\overline{\pi}_{12}^{++} < 0$ rather than $\overline{\pi}_{11}^{++}\,\overline{\pi}_{12}^{++} > 0$. These are illustrated in Table 5.1 for $\alpha = 1.0$, $\delta^2 = 16$. In particular, for the LIML estimator, although the value of the median remains similar in both cases, the value of the IQR is much larger for $\bar{\pi}_{11}^{++} \bar{\pi}_{12}^{++} < 0$. The opposite occurs for the TSLS estimator. In this case, while the value of the IQR is similar in both cases, when $\bar{\pi}_{11}^{++} \bar{\pi}_{12}^{++} < 0$ the value of the median is substantially smaller compared to when $\bar{\pi}_{11}^{++} \bar{\pi}_{12}^{++} > 0$.

In the correctly specified model, the degrees of freedom parameter is equal to G_2, however under this type of misspecification this increases to $G_2 + G_1^{++}$. The effect of this increase is isolated by considering the parameter values $\bar{\pi}_{11}^{++} = \bar{\pi}_{12}^{++} = 0$ in Table 5.1. In particular, the IQR falls (and concentration increases) for the LIML estimator. However, for the TSLS estimator, the median in general decreases in comparison to the correctly specified model, so that the pdf becomes more concentrated around the wrong parameter value. For the LIML estimator, the median remains essentially unbiased, so that although dispersion increases, the pdf is concentrated around the true parameter value.

Figures 5.1-5.3 illustrate a selection of pdf's for the TSLS and LIML estimators, showing a subset of the range of behaviours represented in Table 5.1, and clearly displaying a number of comparisons between the two alternative estimation techniques. Each figure is divided up into four plots, these corresponding to:

(i) the correctly specified model;

(ii) $\bar{\pi}_{11}^{++} = \bar{\pi}_{12}^{++} = 0$;

(iii) $\bar{\pi}_{11}^{++} = 0$ $\bar{\pi}_{12}^{++} = 4$; and

(iv) $\bar{\pi}_{11}^{++} = 7$ $\bar{\pi}_{12}^{++} = 0$.

Further, in each figure, the true parameter value is increased, while the other key parameter values are kept equal to $\delta^2 = 16$, N - G = 10, $G_2 = 1$ and $G_1^{++} = 1$.

In the plots of the correctly specified models for the small parameter values, $\alpha = 0.0$ and $\alpha = 0.5$ (Figures 5.1-5.2), the distributions of the LIML and TSLS estimators are almost identical. Therefore, plots (i), (ii), (iii) and (iv) in each of these figures illustrates not only the effects of misspecification on each estimator but also how the effects differ between the two estimation techniques. In particular, the TSLS estimator appears to be more sensitive to misspecification as is easily seen by comparing the location and spread between the plots.

In Figure 5.3, when the true parameter value is equal to 1, even in the correctly specified model the two estimation techniques are clearly distinguishable. However, the differences between the two become even more apparent in the plots corresponding to the misspecified models. Once again, the TSLS estimator is clearly more affected by misspecification as the LIML estimator maintains a similar shape as that in the correctly specified case.

Similar results to those above are also reported in Table 5.2, where in this case both the number of correctly excluded, G_2, and incorrectly excluded, G_1^{++}, variables are increased corresponding to the true parameter value $\alpha = 1.0$ and noncentrality parameter $\delta^2 = 25$.

Consequently then, the results presented in this section suggest that although the TSLS and LIML distributions are affected in similar ways under this type of misspecification, the LIML estimator tends to be more robust, particularly in the location of its distribution in relation to the true parameter value.

Table 5.2: Median and IQR for LIML and TSLS Estimators

	Median for $\bar{\pi}_{11}^{++}$ =			IQR for $\bar{\pi}_{11}^{++}$ =		
$\bar{\pi}_{12}^{++}$	0	2	4	0	2	4
	LIML α = 1.0 δ^2 = 25 N - G = 10 G_2 = 3 G_1^{++} = 5					
0	0.99744	0.97639	0.91383	0.44929	0.44856	0.43939
2	0.97764	0.99823	1.0597	0.44903	0.43591	0.45986
4	0.91522	0.93990	0.99784	0.43971	0.40784	0.40588
-2	NA	1.0002	1.0747	NA	0.49438	0.60225
	correctly specified Median = 1.0019 IQR = 0.40167					
	TSLS α = 1.0 δ^2 = 25 G_2 = 3 G_1^{++} = 5					
0	0.77599	0.77899	0.77899	0.63699	0.34900	0.4800
2	0.39100	0.67199	0.86499	0.19699	0.24800	0.31400
4	0.22400	0.40100	0.58099	0.14100	0.14999	0.17500
-2	NA	0.28800	0.09600	NA	0.2500	0.32600
	correctly specified Median = 0.92399 IQR = 0.40107					
	LIML α = 1.0 δ^2 = 25 N - G = 10 G_2 = 6 G_1^{++} = 5					
0	1.0791	1.0387	0.98313	0.47246	0.37936	0.40969
2	1.0169	1.0273	1.0567	0.38684	0.37532	0.35851
4	0.98313	0.99453	1.0258	0.40969	0.38216	0.39651
-2	NA	1.0282	1.0588	NA	0.39738	0.40324
	correctly specified Median = 1.0003 IQR = 0.4252					
	TSLS α = 1.0 δ^2 = 25 G_2 = 6 G_1^{++} = 5					
0	0.70899	0.70999	0.71399	0.27200	0.32400	0.441
2	0.45299	0.63499	0.81599	0.21499	0.23700	0.299
4	0.21700	0.39100	0.56499	0.13800	0.14699	0.172
-2	NA	0.04400	0.09100	NA	0.14900	0.311
	correctly specified Median = 0.82899 IQR = 31400					

6.6 SOME FINAL COMMENTS

In a correctly specified SEM, the LIML estimator is considered to be a superior estimation technique to the TSLS estimator, as it is essentially median-unbiased whereas the distribution of the TSLS estimator is, in general, badly distorted.

The numerical results presented in this chapter extend the comparison of the two estimation techniques to the case when the structural equation of interest is misspecified by the exclusion of relevant exogenous variables (which are, however, not excluded from the system as a whole). The key parameters of the distribution are identified, and are shown to affect the distributions in a similar way. However, in general, the LIML estimator is more robust as, although it tends to be more dispersed than the TSLS estimator it is, in general, better located around the true parameter value.

The numerical results presented here, combined with those of Anderson et al. (1979, 1982), are also applicable to the analysis of other types of misspecification, specifically the inclusion of irrelevant exogenous variables and a combination of inclusion of irrelevant and exclusion of relevant exogenous variables, from the structural equation of interest. This is because in the case of the inclusion of irrelevant exogenous variables, only the degrees of freedom parameter in the Wishart matrix, W, (as in Section 6.4) is affected. This is easily seen by applying a similar argument to that given in Section 6.4.

CHAPTER 7

EXTENSIONS OF THE NORMALITY ASSUMPTION: A REVIEW

7.1 THE NORMAL ASSUMPTION

The early statistical researchers, in particular De Moivre (1733), regarded the normal distribution only as a convenient approximation to the binomial distribution, and it was not until the nineteenth century that appreciation of its broader theoretical importance spread with the work of Gauss (1809) and Laplace (1812). Gauss and Laplace were both led to the rediscovery of the normal distribution through their work on the theory of errors of observations. Laplace in particular gave the first statement (although incomplete) of the general theorem, now well known under the title of the Central Limit Theorem.[1] Today in the majority of cases the distribution of disturbances in econometric equations is assumed to be normal. This is by and large true for single equation and simultaneous equation models, both linear and nonlinear. A number of reasons have been suggested for the dominance and popularity of the normality assumption in this context. Two of the most frequently used arguments are the following:

---

[1] This only briefly summarizes the major developments in the derivation of the normal distribution. Considerable attention has been paid to its historical development by authors such as Johnston and Kotz (1970), Cramér (1946) and Stigler (1986).

- Haavelmo (1944) argues that the random disturbance terms of econometric models can be considered to be the sum of a large number of independent small elementary random shocks, and therefore will be approximately normal by virtue of central limit theorem considerations.

- In at least the simplest models, such as the classical linear regression model, the assumption of normality implies that the maximum likelihood and the BLU least squares estimators coincide. Further, a large collection of finite sampling distributions are analytically tractable and consequently have been extensively studied.

However, the widespread use of the normality assumption does not mean it has escaped criticism, and comments such as the following have frequently appeared in the literature:

-*"everyone believes in the Gaussian law of errors, the experimenters because they think it is a mathematical theorem, the mathematicians because they think it is an experimental fact"* - Lippmann (quoted by Poincaré (1912)).

- *"normality is a myth, there never was and never will be a normal distribution"* - Geary (1947, p.209).

- *"Practical statisticians have tended to disregard nonnormality, partly for lack of an adequate body of mathematical theory to which an appeal can be made, partly because they think it is too much trouble, and partly because of a hazy tradition that all mathematical ills arising from nonnormality will be cured by sufficiently large numbers. This last idea presumably stems from limit theorems, or rumors or inaccurate recollection of them"* - Hotelling (1961, p.319).

- *"however, it is rather puzzling that investigators who are generally loathe to adopt informative priors about the systematic structure of the models about which theoretical considerations and part empirical evidence should provide substantive evidence, should find themselves so well informed about the unobservable constituents of their model's unobservable errors to argue that they satisfy a Lindeberg condition"* - Koenker and Bassett (1978, p.34).

There is a large body of empirical literature (e.g. Mandelbrot (1963a, 1963b, 1966, 1967, 1969) and Fama (1963, 1965, 1970)) which suggests that many economic time series, particularly prices in financial and commodity markets, are well represented by nonnormal distributions, especially those with infinite variance. Another example of econometric models in which errors are nonnormal is the study of frontier production and cost function models (e.g. Schmidt (1976b), Waldman (1982)).

Even the asymptotic justification of the normal distribution has been questioned. Bartels (1977) argues that limit-theorem arguments in the context of economic statistics are just as likely to lead to a nonnormal stable distribution as to a normal distribution, so that limiting arguments cannot guarantee that a variable will be normal.

Consequently, there has been a substantial interest in alternatives to the normality assumption. This interest has essentially followed two different directions. The first direction considers iid nonnormal disturbances. This has led to the derivation and use of estimators other than Gaussian-type estimators. The next two sections examine properties of estimators and test statistics optimal in a Gaussian sense under iid nonnormal conditions and considers an important class of alternative methods given under the rubric of robust regression. With the use of the normality assumption the terms independence and uncorrelatedness are equivalent. However, with nonnormally distributed disturbances they are not. Therefore, the second direction broadens the assumption of normality by assuming the disturbances of the regression model follow a joint multivariate elliptical distribution (as defined in Chapter 2). The results that have been obtained under this assumption are reviewed in Section 4. The final section of this chapter combines both of the directions by considering the importance of distinguishing between independence and uncorrelatedness in nonnormal models. This section sets the theme of the remaining chapters in this thesis.

7.2 NONNORMAL IID DISTURBANCES - THE EFFECT ON GAUSSIAN-TYPE STATISTICS

"What are the effects of nonnormality on the traditional normal procedures?"

The objective of this section is to answer this question by drawing together numerous results which have been published on the properties of Gaussian-type statistics under the regime of nonnormal disturbances. Attention is given only to symmetric nonnormal parent populations.

First consider the relaxation of condition (ii) in Section (ii) of Chapter 1. That is, consider the usual linear regression model with non- normal errors. If the errors are assumed to be independently and identically distributed with zero mean and finite variance σ^2, then the following properties hold for the ordinary least-squares based statistics:

Properties 2.1:

(i) $\hat{\beta} = (X'X)^{-1}X'y$ is unbiased, consistent, BLUE and has covariance matrix $\sigma^2(X'X)^{-1}$.

(ii) $s^2 = \frac{1}{N-K}(y-X\hat{\beta})'(y-X\hat{\beta})$ is unbiased and consistent.

(iii) $\hat{\beta}$ does not have a normal distribution and $\frac{(N-K)s^2}{\sigma^2}$ does not have a chi-squared distribution.

(iv) The usual t- and F-tests are not in general valid, however,

$$\sqrt{N}(\hat{\beta}-\beta) \overset{(d)}{\Rightarrow} N(0,\sigma^2 Q^{-1}). \tag{2.1}$$

where $Q = \lim_{N\Rightarrow\infty} \frac{X'X}{N}$. Further, under the null hypothesis $R\beta=r$, where R is a $1 \times K$ known vector and r a known scalar then

$$\frac{\sqrt{N}(R\hat{\beta}-r)}{s\sqrt{R\left(\frac{X'X}{N}\right)^{-1}R'}} \overset{(d)}{\Rightarrow} N(0,1) \ , \tag{2.2}$$

A proof of these properties is given by Schmidt (1976a, pp.55-60).

If the existence of a finite mean and variance of the errors is not assumed then Properties 2.1 do not in general hold. There are many examples of distributions without any finite moments (e.g. Cauchy) or finite mean only (e.g. t_2) and it is believed that these types of distributions are representative of many economic data series, particularly prices in financial and commodity markets (e.g. Fama (1963, 1965, 1970)). These distributions have "fat tails" implying that large values or "outliers" will be relatively frequent. Because the least squares technique minimizes squared deviations it places a relatively heavy weight on these outliers, and their presence can lead to estimates that are extremely sensitive to the presence and values of such outliers. For example, it is well known that the mean of a sample of n values from the standard Cauchy distribution is the same as that of a single observation so consequently even the moments of the distribution of the mean do not exist. A further feature with this class of distributions is that the t- and F-test do not have the usual asymptotic justification as described above in Properties 2.1. For example, Logan et al. (1973), (see also Phillips and Hajivassiliou (1987)) examine the asymptotic distribution and density of the t-statistic in the location model[2] when the observations are a sample from a symmetric stable distribution with index α, where $0 < \alpha \leq 2$. They conclude that:

- the tails of the distribution are Gaussian-like at $\pm \infty$
- if $0 < \alpha < 1$ then the density has infinite singularities at $|1 \pm x|^{-\alpha}$
- for $1 < \alpha < 2$ there are finite "bumps" in the density at ± 1. These disappear as α approaches 2 as the distribution converges to the standard Gaussian density.

This then illustrates the importance of distinguishing between the existence and non-existence of the first two moments of the error distribution, even in the consideration of asymptotic properties.

The finite-sample properties of the t- and F-tests under various moment assumptions have also received much attention in the literature. For the t-statistic, interest has focussed primarily on the location model and the following remarks briefly sketch the main results to illustrate the magnitudes of the differences when compared with classical results. The studies were pioneered by Pearson and Adyantaya (1929) with some empirical investigations on the size and power of the t-test. Many articles soon followed in

---

[2] Using Equation (1.2.1) the location model is equivalent to $y = \beta + \epsilon$.

which theoretical investigations were carried out using the first four terms of Edgeworth series expansions by authors such as Bartlett (1935), Gayen (1949) and Srivastava (1958), and similarly using Laguerre polynomials and gamma density functions by Tiku (1971). It was concluded from these studies that if there is only a moderate departure from the normal distribution then the effect on the properties of the t-test is negligible. Bondesson (1983) establishes that if the distribution function has finite moments of all orders and if the t-test statistic is distributed as central t with N-1 degrees of freedom under the null hypothesis for all sample sizes $N \geq 2$, then the distribution function is normal.

However, these results depend on the existence of moments of the parent population up to a certain order. Unfortunately for many parent populations such as the Student-t with small degrees of freedom, these procedures either fail or do not work well. Yuen and Murthy (1974) perform a Monte Carlo experiment to determine percentage points for the t-statistic when the parent population is Student-t with $v \geq 3$. They suggest the following approximation:

$$\tilde{t} = t_{N-1} \frac{(1-2.08-1.18 \log \text{size } \%)}{Nv}$$

where $\tilde{t}$ is the t-statistic for a parent Student-t family with $v \geq 3$, t_{N-1} is the classical normal t-statistic, and $\log \alpha\%$ assumes the values 0, 0.7, 1.0 corresponding to size $\alpha\%$ of the test equal to 1, 5, 10 respectively. This implies that the $\tilde{t}$ statistic is conservative, that is, the size of the test is smaller than it would be under normality, whenever the value of the ratio is less than 1. More generally, this result is believed to hold for all long-tailed parent distributions (see, for example Cressie (1980), Johnston (1978) and Efron (1969)). However, Benjamini (1983) claims that this is too broad a statement and using various criteria for long-tailedness proves that the t-test is conservative but only for large enough critical values.

Results on the performance of the F-test in the linear regression model (e.g. Schrader and Hettmansperger (1980)) suggest that it is moderately robust with respect to the size of the test but loses power rapidly even in the presence of small departures from the normality assumption of the errors.

The model of interest in this thesis when both conditions (i) and (ii) from Section 2, Chapter 1 are relaxed is the limited information linear simultaneous equations model. In this case the asymptotic distribution of estimators such as OLS, TSLS and the LVR are well known to be normal under certain conditions (e.g. Theil (1971, p.505), Bowden and Turkington (1984, p.26)). However, although the finite-sample properties of these estimators have attracted a great deal of attention in recent years, there are few published results available on the effect of nonnormal disturbances. Knight (1985b, 1986) analyzes the effect of nonnormal disturbances on the moments and distribution of OLS and TSLS estimators by applying results of Davis (1976). Although it is concluded that nonnormality has little effect, the analysis is very limited in the sense that all common nonnormal distributions are excluded. Therefore it is only valid for very small departures from normality. Raj (1980) considers four alternative forms of two parameter normal and nonnormal error distributions and reports on a Monte Carlo study of the small-sample properties of estimators including OLS and TSLS. On the basis of 1,000 replications of sample size 20, in two experiments on an overidentified model, it is found that the small-sample rankings of the estimators of both structural coefficients and forecasts of endogenous variables, according to parametric and nonparametric measures[3] of bias, dispersion and dispersion including bias do not change for any of the four error distributions. This study, too, has its limitations, particularly in the error distributions which are chosen so as to satisfy the existence of the first two moments. A similar study was also carried out by Donatos (1989) and reached similar conclusions.

This section has reviewed several properties of the traditional Gaussian-type statistics in both the general linear and linear simultaneous equations model under a variety of nonnormal distributions. Therefore the question posed at the beginning of this section can now be answered. In the linear regression model the answer is clear-cut. In this model two results, namely Properties (i) and (ii) are often used to justify the use of least-squares statistics under conditions of nonnormality. However, these properties require the existence of the first two moments of the parent distribution, and if this condition is not met the least-squares statistics can have vastly different properties. Even if they do hold, the finite sample properties of the usual inference procedures may be substantially different from those under the classical assumptions. Further, the class of linear estimators tends to be drastically restrictive as its members generally are asymptotically inefficient relative to many non-linear estimators. Although very little analysis has been carried out on the linear simultaneous equations model, similar comments can be made. In particular, the traditional estimators such as OLS, TSLS and the LVR are in general asymptotically

---

[3] Nonparametric measures were used as there exist few results on the existence of moments for nonnormal situations.

inefficient. Consequently, these results have led to the development of other estimation techniques, which is the topic of the next section.

7.3 ROBUST ESTIMATION TECHNIQUES

Are there estimators that are not much worse than least squares when the disturbances are normal but considerably better for nonnormal distributions?

Judge et al. (1985) recommend that if a priori information about the likely form of a nonnormal distribution exists, then because of its known desirable asymptotic properties maximum likelihood estimation should be used. Otherwise, a robust estimation technique should be used. Robust estimators are independent of a distributional assumption on the errors of the model and are "robust" in the sense that they are reasonably efficient irrespective of the form of the underlying distribution. These estimation techniques have been used since the nineteenth century especially in astronomical calculations (see, for example, Stigler (1973)). Although initially most attention was focussed on the location model, recently there have been developments which are relevant for both the linear regression and linear simultaneous equations models. The rest of this section will only briefly outline the three major classes of estimators in this area as there are excellent reviews in the literature. These include Mosteller and Tukey (1977), Huber (1972, 1973, 1977, 1981), Bickel (1976), Koenker (1982) and Koenker and Bassett (1978, 1982). The three classes of estimators are the M, L and R estimators.

The M estimators are also known as "maximum-likelihood like" estimators. Suppose the errors of the model are iid with pdf $f(e_t)$ and are symmetrically distributed about zero. The first order condition of the log likelihood equation for the unknown vector β is,

$$\sum_{t=1}^{N} \frac{x_t' f'(y_t - x_t'\beta)}{f(y_t - x_t'\beta)} = 0 \tag{3.1}$$

However, if the density function is unknown then this equation cannot be solved. An M-estimator is found by replacing f'/f with another function ψ. For robust estimation this function is chosen so that outliers are weighted less heavily than in the least squares solution. The scale factor σ has also been introduced into this estimation technique. One method is to solve the equation

$$\sum_{t=1}^{N} x_t' \psi\left(\frac{y_t - x_t'\beta}{\hat{\sigma}}\right) = 0 \tag{3.2}$$

where $\hat{\sigma}$ is a robust scale estimator of σ. Another approach, however, is to set up a "pseudo maximum-likelihood estimator" for both β and σ (see, for example, Huber (1981)).

The asymptotic properties of M-estimators have been investigated by authors such as Huber (1973, 1981) and Yohai and Maronna (1979). If, in addition to some mild conditions on ψ and F, it is assumed that

(i) $\int_{-\infty}^{\infty} \psi(e_t) dF = 0$

(ii) $\int_{-\infty}^{\infty} \psi(e_t)^2 dF < \infty$

(iii) $\lim N^{-1}(X'X) = Q$ is positive definite,

then the corresponding M-estimator, say $\hat{B}_M$, is consistent and

$$\sqrt{N}(\hat{\beta}_M - \beta) \overset{d}{\Rightarrow} N\left(0, \sigma^2(\psi, F) Q^{-1}\right) \tag{3.3}$$

where

$$\sigma^2(\psi,F) = \frac{\int_{-\infty}^{\infty} \psi(e_t)^2 dF}{\left[\int_{-\infty}^{\infty} \psi'(e_t) dF\right]^2} .$$

In the location model L-estimators are simply estimators which involve linear combinations of order statistics, where the order statistics are defined as the observations ordered, in ascending order. Equivalently L-estimators can be regarded as linear functions of the sample quantiles. The definition of L-estimators on the basis of sample quantiles has been extended to the linear regression model by Koenker and Bassett (1978). This definition was used because the usual concept of order statistics is no longer adequate in the regression model, because what constitutes an appropriate ordering depends on the vector β.

In the linear regression model the θth sample quantile, $0 < \theta < 1$, is defined as any solution to the minimization problem:

$$\min_{\beta} \left[\sum_{(t/y_t \geq x_t'\beta)} \theta |y_t - x_t'\beta| + \sum_{(t/y < x_t'\beta)} (1-\theta) |y_t - x_t'\beta| \right] . \tag{3.4}$$

Koenker and Bassett (1978) have established a number of properties of the estimators that are solutions to (3.4).

The rth-trimmed mean estimator in the location model is defined as

$$\sum_{i=r+1}^{N-r} \frac{y_{(i)}}{N-2r} \tag{3.5}$$

where $y_{(i)}$ are the order statistics. Ruppert and Carroll (1980) and Koenker (1987) discuss alternative estimators based on sample quantiles say $\hat{\beta}_{TLS}$ which, asymptotically, behave similarly to the rth-trimmed mean estimator. That is, under appropriate conditions they show that

$$\sqrt{N}(\hat{\beta}_{TLS}-\beta) \stackrel{d}{\to} N\left(0, \sigma^2(rF)Q^{-1}\right) \tag{3.6}$$

where $\sigma^2(r,F)$ denotes the asymptotic variance of the corresponding rth- trimmed mean from a population with distribution F.

When $\theta = 0.5$ in (3.4), the corresponding estimator is defined as the least absolute deviations estimator. This estimator is also a $\hat{\beta}_{TLS}$ estimator and in this case (3.6) (Koenker and Bassett (1982)) reduces to

$$\sqrt{N}(\hat{\beta}_{LAD}-\beta) \stackrel{d}{\to} N\left(0, [2f(0)]^{-2}Q^{-1}\right) \tag{3.7}$$

where f(0) is the value of the density at the median. This implies that the least absolute deviations estimator is asymptotically more efficient than OLS for all error distributions where the median is superior to the mean as an estimator of location. Amemiya (1982) and Powell (1983) have extended this estimator to simultaneous equation models.

R-estimators, proposed by Jaeckel (1972), are based on a ranking of the residuals in linear models. He wrote the regression model as

$$y = \beta_1 + X\beta^* + u , \tag{3.8}$$

where X is the usual regressor matrix except for the column of 1's and β^* is the usual coefficient vector except for the intercept term. Jaeckel estimator maximizes

$$D(y-Xb^*) = \sum_{t=1}^{N} \left[R_t - \frac{N+1}{2}\right](y_t - x_t' b^*) , \tag{3.9}$$

where R_t = rank $(y_t - x_t' b^*)$. Jaeckel proves that D is a nonnegative, continuous and convex function of b* and that his estimator is asymptotically normal with mean β^* and covariance matrix,

$$\tau^2\left\{X'\left[I-\frac{11'}{N}\right]X\right\}^{-1} \tag{3.10}$$

where τ^2 is $\frac{1}{12}\left[\int f^2(u)du\right]^{-2}$ and 1 is a column of 1's.

A number of Monte Carlo experiments have compared the performance of robust estimators to the OLS estimator in a variety of nonnormal iid distributions. Using the location model Andrews et al. (1972) reports a Monte Carlo study of 68 robust estimators. Their study shows that the performance of the sample mean is clearly inferior for heavy tailed distributions. Similar studies have been carried out by, for example, Hill and Holland (1977), Forsythe (1972), Koenker (1987) and Ruppert and Carroll (1980), for the regression model. These studies have indicated that the particular choice of robust estimator to use depends upon the assumed distribution. Therefore, Amemiya (1985, p.75) concludes that in choosing an appropriate robust estimator, a preliminary study is required to narrow the range of distributions that the given data are supposed to follow.

7.4 MULTIVARIATE ELLIPTICALLY SYMMETRIC DISTRIBUTIONS

In recent years in the econometric literature, to broaden the assumption of nonnormality in the linear regression model, it is assumed that the error components follow a joint multivariate elliptical distribution, as defined in Chapter 2; see also Muirhead (1982). The objective of this section is to review the results in the literature that have been obtained under this assumption.

Thomas (1970) looks at the univariate general linear model $y = X\beta+\epsilon$, where X is a nonstochastic design matrix and ϵ has a spherically symmetric distribution. He shows that the usual t- and

F-statistics used for β have unchanged null distributions for this wider class of spherically symmetric laws and he also gives expressions for the non-null distributions. Zellner (1976) considers the problem for multivariate Student-t errors, a special case, and shows that $\hat{\beta}$ is a maximum likelihood estimator for β and, furthermore, that $\hat{\beta}$ is a maximum likelihood estimator for β for all likelihood functions which are monotonically decreasing functions of $(y-X\beta)'(y-X\beta)$. He further adds that if second moments exist, then $\hat{\beta}$ is a minimum variance unbiased estimator. He also presents the corresponding Bayesian analysis. With a diffuse prior probability density function it is found that the joint posterior distribution for the regression coefficients is in precisely the same multivariate Student-t form as arise from the usual normal model. However, the posterior distribution for the scale parameter is in the form of an F-distribution whereas in the normal model it has an inverted gamma density. He also presents a natural conjugate prior. Extensions to this result have been considered by Jammalamadaka, Chib and Tiwari (1987, 1988).

The methods used to obtain results for the univariate regression model have been mechanized routinely to give the distribution theory for the multivariate regression model. For example, Sutradhar and Ali (1986) consider the multivariate regression model defined by,

$$Y = \beta X + \epsilon \quad , \tag{4.1}$$

where Y is the (p×N) matrix of dependent variables, β is a (p×k) matrix of unknown parameters to be estimated, X is a nonstochastic matrix of order (k×N) and $\epsilon = (\epsilon_1 \dots \epsilon_j \dots \epsilon_N)$, an error variable where $\epsilon_j = (\epsilon_{1j} \dots \epsilon_{pj})$ each with covariance matrix Σ. The probability density function of ϵ is given by the multivariate-t distribution,

$$\mathrm{pdf}(\epsilon) = \frac{(v-2)^{\frac{v}{2}}\Gamma\left(\frac{v+Np}{2}\right)}{\pi^{\frac{Np}{2}}\Gamma\left(\frac{v}{2}\right)} |\Sigma|^{-\frac{N}{2}} \left\{(v-2) + \sum_{j=1}^{N} \epsilon_j' \Sigma^{-1} \epsilon_j\right\}^{-\left(\frac{v+Np}{2}\right)}. \quad (4.2)$$

They show that the ordinary least squares estimator of β is unbiased and weakly consistent. Using the method of moments, they also consider the consistent estimation of v, for $v > 4$. Singh (1987) also considers this problem in the univariate regression model.

Andrews and Phillips (1987) consider optimal median-unbiased estimation in a linear regression model with the distribution of the errors lying in a subclass of the elliptically symmetric distributions. The generalized least squares estimator is shown to be best for any loss function that is nondecreasing as the magnitude of underestimation or overestimation increases. For the same loss functions, a restricted generalized least squares estimator is shown to be best when the estimator is known to lie in an interval. The class of error distributions that is considered are rotated variance mixtures of multivariate normal distributions.

The properties of a number of statistical tests have also been examined by, for example, King (1979, 1980), Ullah and Phillips (1986), Sutradhar (1988), Ullah and Zinde-Walsh (1984, 1985, 1987) and Anderson, Fang and Hsu (1986).

King (1979, 1980) establishes the result that statistics which are invariant to the scale of the disturbances have the same small sample distributions as they do under normality.

In the special case of multivariate-t errors, Ullah and Phillips (1986) analyze the distribution of the F-ratio for testing a set of linear restrictions and in particular derive its non-null density function. Sutradhar (1988) also examines this problem and calculates the power of the test for a particular set of exogenous variables. In this case the power of the test depends upon the degrees of freedom parameter, which is assumed known.

In a series of papers, Ullah and Zinde-Walsh (1984, 1985, 1987) consider the F, Likelihood-Ratio (LR), Lagrangian Multiplier (LM), Wald (W) and Rao-Score (RS) tests for testing a set of linear restrictions. They describe these statistics as being numerically robust over a class of error distributions if their values are independent of the specific error distribution from that class, and inferentially robust if no matter which error distribution from that class of distributions is considered the test statistics remain unchanged. Using these criteria, they show that if the error disturbance is assumed to be spherically normally distributed, F and LR are numerically robust against the class of all monotonically decreasing continuous spherical distributions, but RS and W are not. However, all these statistics are inferentially robust over this class so that the test conclusions reached under the assumption of normality will not be overturned if the error distribution is spherical. They also extend these results based on the assumption of spherical normality against the general class of elliptical error distributions. In particular, they obtain conditions for numerical robustness for the class of covariance matrices often used in econometrics such as in autoregressive, moving average and heteroskedastic models. Their investigations show that for these covariance matrices the numerical robustness of test statistics under consideration is rare and they develop bounds for critical values which ensure that the conclusions based on the usual tests are not affected by a particular class of distributions.

Anderson, Fang and Hsu (1986) obtain likelihood ratio criteria for a class of null hypotheses for monotonically-decreasing continuous elliptically contoured distributions.

The topic of spherical matrix distributions and a multivariate model has been studied by many authors, among whom are Dawid (1977), Fraser and Ng (1980), Jensen and Good (1981), Kariya (1981), Eaton (1983) and Sutradhar and Ali (1986).

Stein (1955) shows that in higher dimensional problems, the sample mean of a multivariate normal distribution is inadmissible against expected squared error loss. This result was extended and analyzed for the vector of regression coefficients when the disturbances are distributed normally by James and Stein (1961) and Brown (1966) and they show the inadmissibility of the OLS estimator for greater than two regressors. Because of this deficiency, Stein-type improved estimators have been developed (see Judge _et al._ (1985, p.82)). Recently, several authors have extended the analysis to include spherically symmetric disturbances. These include, Strawderman (1974), Berger (1975), Brandwein and Strawderman (1978, 1980), Brandwein (1979), Judge _et al._ (1985). and Judge and Yancey (1986). Judge and Yancey (1986, p.271) conclude that, "in general, the risk characteristics for traditional Stein-like estimators for the nonnormal errors were found to be the same as for the normal case". The Stein-type estimators have been shown to be another type of pretest estimator for combining the unrestricted and restricted least-squares estimator (see, for example, Judge _et al._ (1985, p.86)). Giles

(1990) has derived some results on pretesting with models whose errors are assumed to be normally distributed but, in fact, follow a spherically symmetric distribution.

Knight (1986) considers the compound normal distributions which are contained within the elliptically symmetric class, in a simultaneous equation framework. In particular he establishes the result that the OLS and TSLS estimators in the leading case are robust to this class of non- normal distributions. This implies that the estimators possess the same moment results as under the normality assumption. Using the techniques of Ullah and Phillips (1986) and Giles (1990) the results of Chapter 5 could be extended to consider the distribution of these estimators in the general case under the assumption of multivariate-t errors.

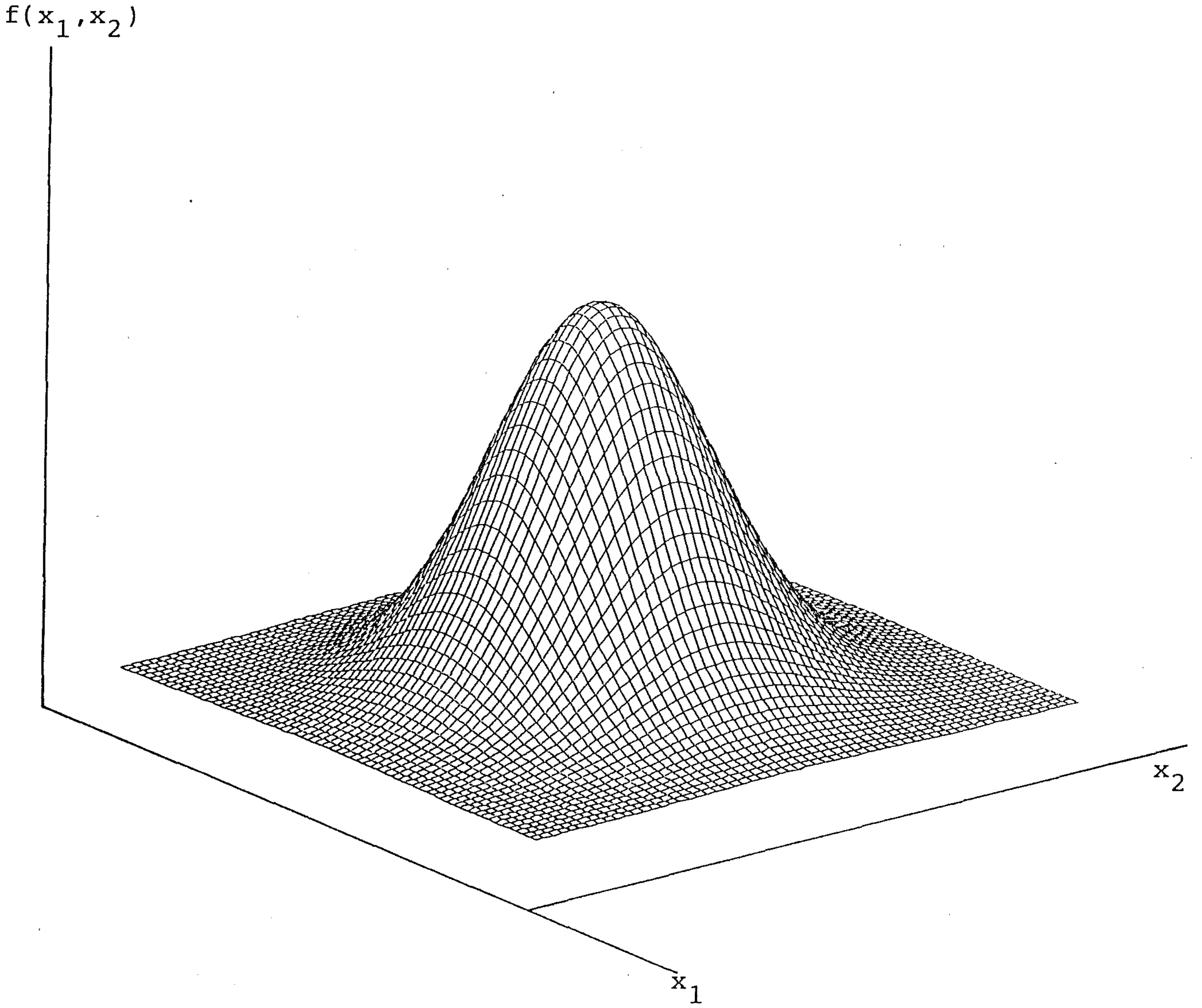
The results reviewed in this section suggest that by replacing the normality assumption with the assumption that the regression disturbances follow a multivariate elliptically symmetric distribution in the linear regression model, the resulting distributions possess properties which make them analytically tractable and, further, in many cases identical to those obtained under the normality assumption. However, the marginal distributions of the disturbance terms from this assumption are identical to those when it is assumed the disturbances are distributed identically and independently elliptically symmetric and, in this case, the results of Sections 2 and 3 are applicable. The differences in the results reviewed in Sections 2, 3 and 4 suggest it is important to distinguish between multivariate and iid elliptically symmetric distributed disturbances. The importance of this distinction is discussed further in the next section.

7.5 JOINTLY DISTRIBUTED VERSUS INDEPENDENT DISTURBANCES

It is well known that within the class of elliptically symmetric distributions, independence when the covariance matrix is diagonal, characterizes the normal distribution (see Chapter 2). The bivariate normal distribution with covariance matrix and location vector zero is illustrated in Figure 5.1. The corresponding bivariate joint-Cauchy distribution is illustrated in Figure 5.2. This distribution has a "bell-shape" similar to that of the bivariate normal distribution. However, the independent-Cauchy distribution, as given in Figure 5.3, has a rather different shape, especially in the tails. These features are also reflected in the reviews of Sections 2, 3 and 4 of this Chapter. In particular, Section 4 illustrates the robustness of many Gaussian statistics when disturbances are distributed multivariate elliptically symmetric. However, the properties of these statistics when the disturbances are independently distributed (Section 2), has led to the development of a wide range of alternative methods (Section 3).

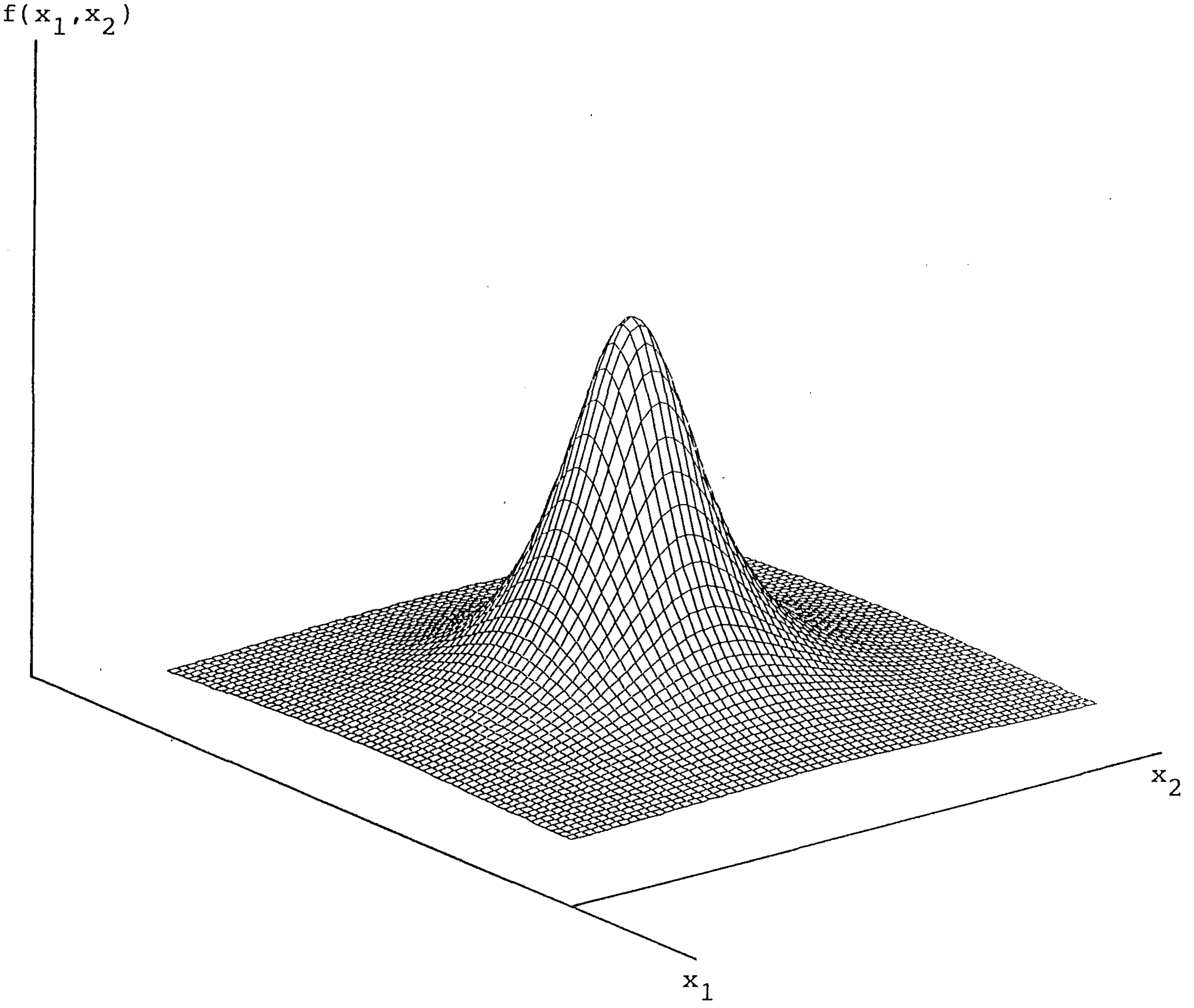
Consequently, when it is assumed the disturbances are nonnormally distributed, it is important to distinguish between "jointly-distributed" and "independently-distributed" disturbances, as they lead to quite different estimation and inference techniques. This problem is similar to distinguishing between "heteroskedastic versus homoskedastic disturbances" or "autocorrelated versus serial-independent disturbances". However, while it is standard in virtually every econometric textbook to study the implications of misspecifying "heteroskedastic and homoskedastic disturbances" or "autocorrelated and serial independent disturbances", it would seem that the article by

FIGURE 5.1 BIVARIATE SURFACE FOR SPHERICAL NORMAL DISTRIBUTION



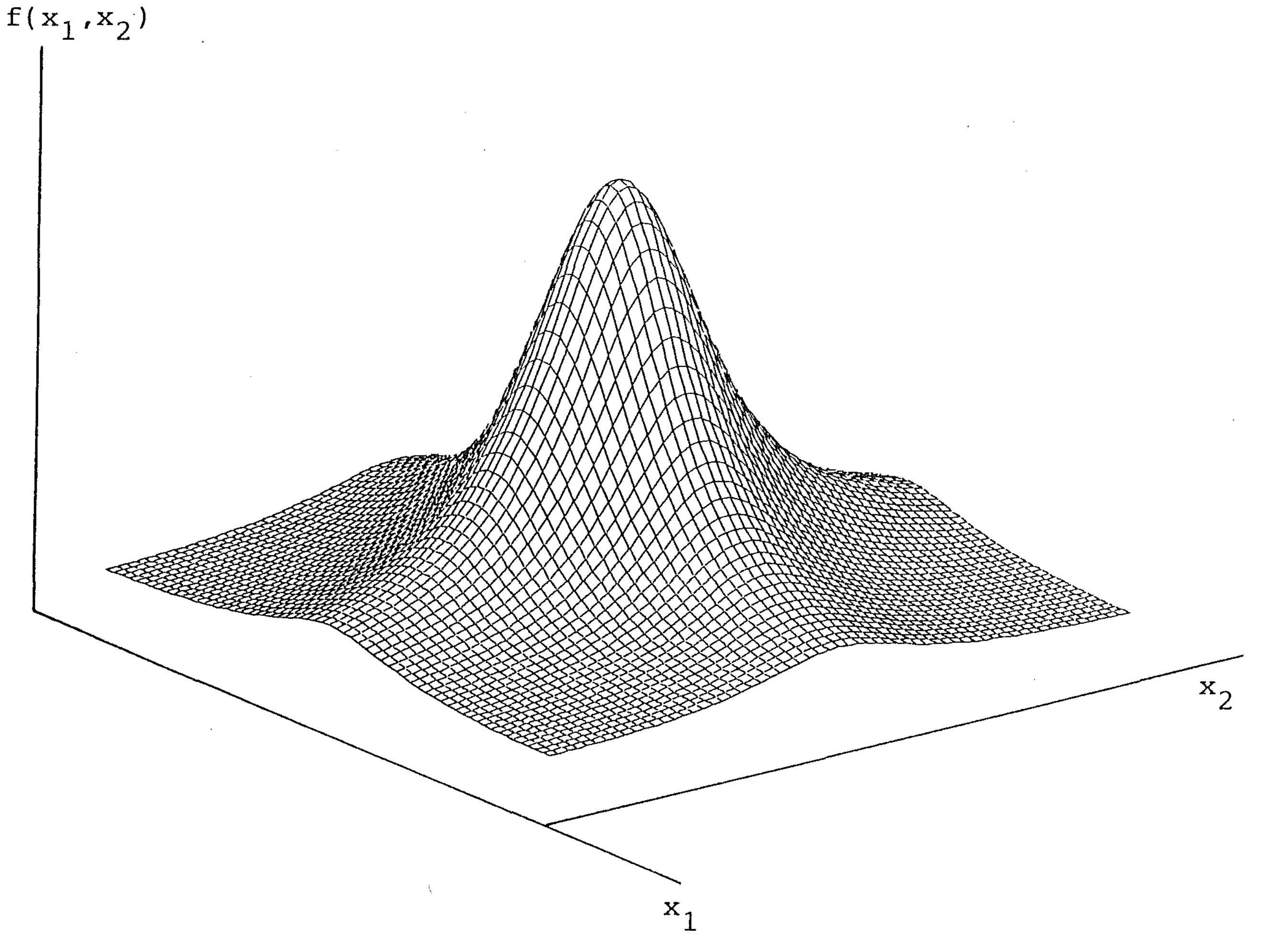
$$f(x_1,x_2) = \frac{1}{2\pi} \exp(-\tfrac{1}{2}\{x_1^2 + x_2^2\})$$

FIGURE 5.2 BIVARIATE SURFACE FOR JOINT SPHERICAL CAUCHY DISTRIBUTION



$$f(x_1,x_2) = \frac{1}{2\pi}(1 + \{x_1^2 + x_2^2\})^{-3/2}$$

FIGURE 5.3 BIVARIATE SURFACE FOR INDEPENDENT SPHERICAL CAUCHY DISTRIBUTION



$$f(x_1,x_2) = \frac{1}{\pi^2}(1 + x_1^2)^{-1}(1 + x_2^2)^{-1}$$

Kelejian and Prucha (1985) is the only one which attempts to address this issue for misspecifying "jointly-distributed" and "independently-distributed disturbances". In this paper they consider this issue using the Student-t distribution. This is an important nonnormal distribution as it is considered that it is a reasonable way of modelling tails that are fatter than those of the normal distribution, (see e.g. Jeffreys (1961)), and this is relevant for many economic data series such as prices in financial and commodity markets, (see e.g. Judge et al. (1985, p.825), and the recent paper by Lange et al. (1989)). In particular, Kelejian and Prucha (1985) compare the asymptotic properties of the maximum likelihood estimators of the linear regression model, when the disturbances are assumed either to be distributed multivariate Student-t with $v \geq 3$ (uncorrelated disturbances) or iid Student-t with $v \geq 3$ (independent disturbances). In this example, if the disturbances are assumed to be independent when they are only uncorrelated, and the regression parameters are correspondingly estimated, the estimator of the variance-covariance matrix is inconsistent. On the other hand, if the disturbances are independent, but they are only assumed to be uncorrelated, efficiency is lost and inferences are based on an incorrect large sample distribution. Further, the efficiency loss is substantial for certain parameter values.

The objective of the next three chapters is to extend this analysis to finite-sample differences between the two alternative assumptions for the entire Student-t family (i.e. $v \geq 1$). Given the extent of the existing literature in the dependent case, most attention is given to developing properties of maximum likelihood statistics in the independent case. These include, for example, the estimation of variance (if it exists), the shape of the distribution function of the statistics and their relationship to the robust estimators of Section 3.[4] Once these properties are established, we examine the statistical consequences of using the maximum likelihood estimator associated with one assumption, when in fact the other assumption is true. Further, because properties of the maximum likelihood estimator for iid Student-t disturbances are compared with a number of robust estimators, we can also see the statistical consequences of making one error assumption over the other when a more general robust estimator is used. Chapter 7 concentrates solely on the location model. Chapters 8 and 9 extend these results to both the linear regression model and the exactly-identified limited-information SEM.

Since the distinction between the two assumptions is important, specification tests need to be developed to make this distinction. This topic is also discussed in the following chapters. In particular, tests are developed which make this distinction in the elliptically-symmetric class of distributions and which use existing tests for normality. The properties of such tests are illustrated for the Student-t family.

---

[4] The comparison with the robust estimators is only carried out for the location-scale and linear regression models. This is because the theory of general robust estimation techniques is not well developed in limited- information SEMs, (see e.g. Powell (1983)).

CHAPTER 8

THE LOCATION/SCALE MODEL WITH STUDENT-t OBSERVATIONS

8.1 INTRODUCTION

The topic of the next three chapters is the statistical comparison of the maximum likelihood estimators of the unknown parameters in the linear regression and limited-information SEM's, when it is assumed the disturbances are distributed either as iid Student-t or multivariate Student-t. This problem is similar to the comparison of alternative assumptions in econometric models, such as autocorrelation versus serial independence, or heteroskedasticity versus homoskedasticity, which are standard analyses in all econometric textbooks.

The analysis begins in this chapter with the location-scale model, which is the simplest case of the linear regression model. This refers to the estimation of location μ, and scale σ, in the model,

$$y_i = \mu + u_i \quad , \qquad i = 1 \ldots N \tag{1.1}$$

where if it is assumed that $u_1 \ldots u_N$ have a multivariate Student-t distribution then

$$\mathrm{pdf}(u_1 \ldots u_N | v, 0_N, \sigma_N^2) = \frac{\Gamma\left(\frac{v+N}{2}\right)}{\Gamma\left(\frac{v}{2}\right)(v\pi)^{N/2}\sigma^N}\left[1+\frac{1}{v\sigma^2}\,(u_i^2+\ldots+u_N^2)\right]^{-\left(\frac{N+v}{2}\right)} \tag{1.2}$$

or, alternatively, if it is assumed that the elements of u_i are iid distributed as,

$$\mathrm{pdf}(u_i|v,0_N,\sigma_I^2) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)(v\pi)^{1/2}\sigma}\left[1+\frac{1}{v\sigma^2}u_i^2\right]^{-\left(\frac{v+1}{2}\right)} \tag{1.3}$$

then the joint distribution of the disturbances is,

$$\mathrm{pdf}(u_1 \ldots u_N|v,0,\sigma^2) = \left[\frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)\sqrt{v\pi}\,\sigma}\right]^N \prod_{i=1}^{N}\left[1+\frac{1}{v\sigma^2}u_i^2\right]^{-\left(\frac{v+1}{2}\right)}. \tag{1.4}$$

Section 4 of Chapter 7 reviews finite-sample properties of the maximum likelihood estimators when the joint distribution of the disturbances is given by (1.2). These properties include:

Properties 1.1

(i) $\hat{\mu}_{OLS}$, the sample mean, is the maximum likelihood estimator of μ.

(ii) $(\hat{\mu}_{OLS} - \mu)/\sigma$ is distributed $MT_1(0,1/N,v)$.

(iii) $\hat{\mu}_{OLS}$ is the MVB estimator and therefore also the BLUE when $v \geq 3$.

(iv) $\hat{\mu}_{OLS}$ is median-unbiased, is at least as concentrated about μ as any other median-unbiased linear estimators and is "best" for any monotone loss function (that is, any loss function that is non- decreasing as the magnitude of underestimation or overestimation increases), for all v.

(v) For $v \geq 3$, an unbiased estimator of scale is,

$$\hat{\sigma} = \frac{1}{B(N,v)}\left[\Sigma(y_i-\hat{\mu})^2\right]^{\frac{1}{2}},$$

where B(N,v) is an adjustment factor, which depends upon N and v.

However, when the disturbances are jointly distributed as (1.4) the corresponding properties of the maximum likelihood estimators are not known. In this chapter they are developed by using properties of order statistics. Using order statistics, Lloyd (1952) derives the exact generalised BLU estimators of μ and σ. These estimators are compared with the maximum likelihood estimators and from this comparison similar properties, such as those in Properties 1.1 for multivariate Student-t disturbances, are developed. Once these properties are developed, the statistical consequences of making one error assumption over the other are discussed.

Section 2 reviews Lloyd's (1952) BLU estimators, using the order statistics of the sample. Section 3 considers properties of the maximum likelihood estimators with independent Student-t observations. This section is divided into a number of parts, including discussions of the numerical maximization of the log-likelihood function, the asymptotic distribution and the finite-sample distribution of the maximum likelihood estimators. Section 4 considers the statistical consequences of making one error assumption over the other in the location-scale model; and Section 5 concludes with some final comments.

8.2 LLOYD'S BEST LINEAR UNBIASED ESTIMATORS

Suppose the u_i in the location-scale model of (1.1) are iid such that

$$\mathrm{pdf}(u_i) = \frac{1}{\sigma}\,\mathrm{pdf}\left(\frac{y_i-\mu}{\sigma}\right) \quad , \quad \sigma > 0.$$

For this family of distributions, using order statistics, Lloyd (1952) obtained the unbiased and minimum variance estimators of location and scale, within the L-class of robust-estimators (Chapter 7). These estimators are defined as follows. Suppose the "ordered" location model is,

$$y_{(i)} = \mu + \sigma u_{(i)} ,$$

where (i) denotes the order statistics of the sample (that is, the observations of the sample are arranged in ascending order), then using formulae for the means, variances and covariances of order statistics as given by David (1970, pp.25-30),

$$E\left(u_{(i)}\right) = \alpha_i \qquad ; \qquad Var\left(u_{(i)}, u_{(j)}\right) = \beta_{ij}$$

Since $u_{(i)} = \left(y_{(i)} - \mu\right)/\sigma$, this implies that

$$E\left(y_{(i)}\right) = \mu + \sigma\alpha_i \; ; \qquad Var\left(y_{(i)}, y_{(j)}\right) = \sigma^2\beta_{ij}$$

Lloyd's BLU estimator is obtained by rewriting (1.1) as,

$$y_{(i)} = \mu + \sigma\alpha_i + \sigma u^{\alpha}_{(i)} , \qquad (2.1)$$

so that, $E(u^{\alpha}_{(i)}) = 0$, and $var(u^{\alpha}_{(i)}, u^{\alpha}_{(j)}) = \beta_{ij}$, and then applying Generalized Least Squares to (2.1) to obtain,

$$\hat{\beta} = (X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}Y) ,$$

where

$$\hat{\beta} = \begin{bmatrix} \hat{\mu}_{LB} \\ \hat{\sigma}_{LB} \end{bmatrix} , \; \Omega = \left[\beta_{ij}\right] , \; X = \left[1, \alpha\right] , \; \alpha = \left[\alpha_i\right] , \; Y = \left[y_{(i)}\right]$$

and 1 is a column of 1's. For symmetrical parent populations the formulae become,

$$\hat{\mu}_{LB} = \frac{1'\Omega^{-1}Y}{1'\Omega^{-1}1} , \quad \hat{\sigma}_{LB} = \frac{\alpha'\Omega^{-1}Y}{\alpha'\Omega^{-1}\alpha} , \qquad (2.2)$$

with variances,

$$\text{var}(\hat{\mu}_{LB}) = \frac{\sigma^2}{1'\Omega^{-1}1}, \quad \text{var}(\sigma^2_{LB}) = \frac{\sigma^2}{1'\Omega^{-1}\alpha}. \qquad (2.3)$$

This is shown in David (1970, p.104). In particular, the BLUE for μ corresponds to the sample mean iff $1'\Omega^{-1} = 1$, or equivalently, iff all of the rows of the covariance matrix add to unity. Bondesson (1976) proves that the sample mean is BLUE iff the underlying distribution is either normal or the gamma distribution.

For the Student-t distribution with iid observations, a number of results for the calculation of the BLU estimators in (2.2) are used in the following sections. These results are:

(i) For sample sizes less than 20, these estimators can be calculated when $v > 2$ using the means, variances and covariances of the order statistics calculated by Tiku and Kumra (1985).

(ii) Jung (1962) considers the asymptotic distribution of these estimators when $v > 2$. In particular, he shows them to be consistent, asymptotically normally distributed and asymptotically efficient.

(iii) For the Cauchy distribution (for which the means, variances and covariances of the order statistics are calculated by Barnett (1966)), and the t_2-distribution, some care is needed in obtaining the BLU estimators, as the extreme order statistics have infinite variances. However, in this case, the standard expressions (2.2) are used, by assuming the coefficients of the extreme order statistics are zero. For the Cauchy distribution, these are the first and last two order statistics of the sample, and for the t_2-distribution, the first and last order statistic. Asymptotically these estimators are consistent and asymptotically normally distributed, since as $N \to \infty$, the order statistics $Y_{(r_1)} \ldots Y_{(r_K)}$, for $r_j = P_j N$ with $0 < P_j < 1$, $j = 1, \ldots, K$, are asymptotically multivariate normal (see, for example, Cox and Hinkley (1974, p.469)).

8.3 MAXIMUM LIKELIHOOD ESTIMATORS FOR INDEPENDENT STUDENT-t OBSERVATIONS

This section is divided into four parts. The first part defines the maximum likelihood estimators for both μ and σ, and since analytical expressions do not exist for these estimators in general, the numerical maximization of the likelihood function is also discussed. Parts 2 and 3 concentrate on the distribution function of the maximum likelihood estimators. Part 2 considers the asymptotic distribution and Part 3 develops properties of the finite-sample distribution. Finally, Part 4 summarizes the results of this section in a form similar to Properties 1.1.

(a) Definition

Let $u_1 \ldots u_N$ be a random sample with joint distribution function given by (1.4). The log-likelihood function is given by,

$$\mathcal{L} = \text{constant} + (-N+v+1) \log \sigma - \left(\frac{v+1}{2}\right) \sum_{i=1}^{N} \log\left[v\sigma^2+(y_i-\mu)^2\right]. \qquad (3.1)$$

If v is specified and both μ and σ are assumed unknown then the first order conditions for the maximum likelihood estimators are given by the equations,

$$\frac{\partial \mathcal{L}}{\partial \mu} = \sum_{i=1}^{N} \frac{y_i - \mu}{v\sigma^2 + (y_i - \mu)^2} = 0 \tag{3.2}$$

$$\frac{\partial \mathcal{L}}{\partial \sigma} = \frac{-N+v+1}{\sigma} - \frac{v+1}{2} \sum_{i=1}^{N} \frac{2v\sigma}{v\sigma^2 - (y_i - \mu)^2} = 0. \tag{3.3}$$

Ferguson (1978) finds closed-form expressions for the solutions to these equations for the Cauchy distribution when the sample size is 3 or 4. However, in general, this is not possible and the maximum likelihood estimators must be obtained by numerical methods. Copas (1975) shows that the joint likelihood function for the Cauchy distribution has exactly one point of maxima and at most one stationary point. This result has been extended to the t_v-distribution in general by Gabrielson (1982). This implies that the maximized likelihood function for given degrees of freedom is unimodal and that numerical maximization of (3.1) produces the global maximum likelihood estimators, and these will be denoted $\hat{\mu}_{ML}$ and $\hat{\sigma}_{ML}$.[1]

(b) The Asymptotic Distribution

While the existing literature contains results on the asymptotic distributions of some specific members of the Student-t family (for example, Haas, Bain and Antle (1970) and Norden (1972) for the Cauchy distribution and Kelejian and Prucha (1985) for $v \geq 3$), none of these authors considers the Student-t family as a whole. However, it is easy to generalise their results as in the following theorem:

---

[1] Alternatively, although not considered in this thesis, σ can be assumed to be known in (3.1). In this case, Barnett (1966) shows that (3.1), for the Cauchy distribution, will often have multiple roots. This argument can be extended to the t_v-distribution in general.

Theorem 3.1

There exist solutions $\hat{\mu}_{ML}$ and $\hat{\sigma}_{ML}$ of the likelihood equations (3.2) and (3.3) such that

(i) $\hat{\mu}_{ML}$ and $\hat{\sigma}_{ML}$ are consistent estimators of μ and σ respectively.

(ii) $\sqrt{N}\left(\hat{\mu}_{ML}-\mu, \hat{\sigma}_{ML}-\sigma\right)$ is asymptotically bivariate normal with vector mean zero and covariance matrix $\begin{bmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{bmatrix}^{-1}$ where $I_{ij} = -E\left[\frac{\partial^2 \mathscr{L}}{\partial\theta_i \, \partial\theta_j}\right]$, that is, the Cramér - Rao Lower Bound, (CRLB). In this case we have,

$$I_{11} = \frac{1}{\sigma^2} \int_{-\infty}^{\infty} \left[\frac{pdf'(y)}{pdf(y)}\right]^2 pdf(y)\,dy = \left(\frac{v+1}{v+3}\right)\frac{1}{\sigma^2}$$

$$I_{22} = \frac{2}{\sigma^2\left(1+\frac{4}{v}+\frac{3}{v^2}\right)} \left[1+\frac{1}{v}\right],$$

and $I_{12} = I_{21} = 0$.

(iii) $\hat{\mu}_{ML}$ and $\hat{\sigma}_{ML}$ are asymptotically efficient in the sense that

$$\sqrt{N}\left(\hat{\mu}_{ML}-\mu\right) \xrightarrow{D} N\left(0, I_{11}^{-1}\right) \tag{3.4}$$

$$\sqrt{N}\left(\hat{\sigma}_{ML}-\sigma\right) \xrightarrow{D} N\left(0, I_{22}^{-1}\right)$$

Proof:

The proof of the theorem follows by considering the combination of the following two points:

(1) There exist linear combinations of order statistics that are estimators of μ and σ, as given in Section 2, which are consistent and asymptotically normally distributed.

(2) If $\hat{\theta}$ is a consistent estimator of θ such that $\sqrt{N}(\hat{\theta}-\theta)$ has a proper limit distribution, the second round estimator is asymptotically normally distributed and asymptotically efficient. Details of the argument on which this result is based are given in Appendix B.

Therefore, by beginning the numerical maximization process with the estimators given in (1), since the likelihood function is unimodal, the resulting estimators are the maximum likelihood estimators and from (2) are consistent, asymptotically normally distributed and asymptotically efficient.

(c) Finite-Sample Distribution

In this section members of the Student-t distribution are divided into two cases, those where the variance of the disturbances is finite ($v > 2$), and those where it is infinite ($v \leq 2$). In each of these cases the variances and probability density functions of the standardized maximum likelihood and Lloyd's BLU estimators of location, $(\hat{\mu}-\mu)/\sigma$, and scale, $\hat{\sigma}/\sigma$, are estimated for various sample sizes and degrees of freedom, as Antle and Bain (1969) have shown that these distributions depend only on sample size. Figures 3.1 - 3.4 and a number of entries in Tables 3.1 - 3.2 are based on the results of Monte - Carlo experiments. Details of these experiments were given in Chapter 4, but they will be briefly outlined here for completeness. Empirical variances of $(\hat{\mu}_{ML} - \mu)/\sigma$, (Table 3.1), and empirical biases and variances of $\hat{\sigma}_{ML}/\sigma$ are estimated using 40,000 - 60,000 replications. Empirical densities of, $(\hat{\mu}_{ML} - \mu)/\sigma$, $(\hat{\mu}_{LB} - \mu)/\sigma$, (Figures 3.1 - 3.2, 3.4 - 3.6), and $\hat{\sigma}_{ML}/\sigma$, $\hat{\sigma}_{LB}/\sigma$, (Figure 3.3), were estimated via the integration of the kernel density estimator with the naive Monte Carlo method. The kernel estimate at point X is equal to,

$$\hat{pdf}(X) = \frac{1}{N{*}h(N{*})} \sum_j k\left[\frac{X - X_j}{h(N{*})}\right] \tag{3.5}$$

where k[.] is the standard N(0,1) density. The window width h(N*) is chosen using the interactive approach of Tapia and Thompson (1978). In all cases this approach led to the use of a window width between 0.02 and 0.09. N* is simply the number of replications in the simulation experiment, and is chosen using the bound of estimation. For example, the results of Parzen (1962) and Cacoullos (1966) imply,

$$\left(N{*}h^m(N{*})\right)^{\frac{1}{2}}\left[\hat{pdf}(x) - E\left(\hat{pdf}(x)\right)\right] \sim N\left(0, pdf(x)\int K^2\right) \tag{3.6}$$

holds. The result given in (3.6) can be achieved if $\left(N{*}h^m(N{*})\right)^{\frac{1}{2}}Bias\left[\hat{pdf}(x)\right]$ tends to zero asymptotically since,

$$\left(N{*}h^m(N{*})\right)^{\frac{1}{2}}\left[\hat{pdf}(x) - pdf(x)\right] = \left(N{*}h^m(N{*})\right)\left[\hat{pdf}(x) - E\left(\hat{pdf}(x)\right)\right] + \left(N{*}h^m(N{*})\right)^{\frac{1}{2}}Bias\left[\hat{pdf}(x)\right].$$

Ullah (1988, p.642) shows that $Bias\left[\hat{pdf}(x)\right]$ is proportional to $h^2(N{*})$. This implies that if $N{*}h^{(4+m)/2}(N{*})$ tends to zero asymptotically then (3.6) holds. Therefore, for the normal kernel

$\frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}y^2)$, the 99% asymptotic confidence interval for $\hat{pdf}(X)$ is given by,

$$\hat{pdf}(X) \pm 2.58 \left[\frac{\hat{pdf}(X)}{2N^*h\sqrt{\pi}}\right]^{\frac{1}{2}},$$

so that B is given by,

$$B = 2.58\left[\frac{\hat{pdf}(X)}{2N^*h(N^*)\pi}\right]^{\frac{1}{2}}$$

N* is varied until B is less than 0.01 for all points at which the density is estimated. In all experiments, N* varies between 60,000 and 90,000 replications[2]. The input of X_j in (3.5) involves numerically maximizing the likelihood function (3.1) to obtain the maximum likelihood estimators and calculating (2.2) to obtain the BLU estimators. Two algorithms from the Harwell Subroutine library are used, these being algorithms VAI3AD and VF04AD, which both use the BFGS formula, (Broydon (1970), Fletcher (1970), Goldfard (1970) and Shanno (1970)). All computations are performed in double precision to 7 decimal places of accuracy. The final results, however, are not dependent upon which algorithm is used in this step. Furthermore, the solutions of each of the algorithms used were compared with those in the standard Econometric packages TSP and SHAZAM, and were found to give similar results. Random numbers

---

[2]Empirical densities were also computed using the Epanechnikov (1969) kernel. However, given the number of replications used, the results proved not to depend on which kernel is used. This situation is similar to the comparison of different kernels for the Cauchy distribution using a "large sample", as is illustrated in Figure 5.1 in Chapter 3.

distributed uniformly on the interval [0,1], denoted U, are generated using the NAG subroutine G05CAF, which uses a multiplicative congruential method. Standard iid Student-t variates, for degrees of freedom $v < 3$, are generated by the inversion of the distribution function (see, for example, Devroye (1986, p.27)). In particular, for $v = 1$, the Cauchy distribution, standard Cauchy variates are generated as,

$$X = \tan\left(\pi(U-\tfrac{1}{2})\right)$$

and for $v = 2$, the t_2-distribution,

$$X = \sqrt{2}(U-\tfrac{1}{2})/\mathrm{SQRT}\left(U(1-U)\right) ,$$

where U is from U(0,1). For the rest of the Student-t family, $v \geq 3$, X is generated via a transformation of a symmetric beta variate, (see, for example, Devroye (1986, p.446)). This can be written in terms of independent uniform random numbers U_1, U_2 as,

$$X = \frac{2\sqrt{v}\ \sin\ (2\pi U_1)(1-U_2^{2/v-1})}{(1-\sin^2(2\pi U_1))(1-U_2^{2/v-1})} .$$

This formula is useful as it is valid for all members of the Student-t family with $v \geq 3$. Also it does not require the generation of as many random uniform deviates as does the traditional method of generating a t-random variable via its interpretation as a ratio of a standard normal to the square root of an independent normalized chi-square variable. Further details of the Monte Carlo methodology are given in Chapter 4.

Finite Variance $v > 2$

Table 3.1 gives the estimated variances corresponding to various v and N for the standardized location estimator. Also given in this table is the Cramér-Rao lower bound (CRLB) as given in (3.4). However, this bound is attainable only asymptotically due to the joint application of the results of Koopman (1936) and Pitman (1936). The results of Koopman (1936, p.408) imply that a pair of jointly sufficient statistics for the unknown parameters exist only for the normal distribution ($v = \infty$). Consequently, the Cramér-Rao lower bound (CRLB) is not attainable in finite-samples due to the joint results of Koopman (1936) and Pitman (1936) and summarized in, for example, Theorem 9 of Dhrymes (1970) "if an... unbiased MVB estimator of θ exists, pdf ($Y_1 \ldots Y_N$) admits a set of jointly sufficient statistics for its parameters..."

However, as indicated in Table 3.1 the empirical variances are well approximated by applying the asymptotic theory for small v and N, for example $v = 3$ and 5 and $N = 20$. Consequently, in these cases the maximum likelihood estimator is the MVB estimator. More generally though, a relationship between the maximum likelihood estimator and Lloyd's BLU estimator of location can be established and also given in Table 3.1 are the known finite-sample variances of Lloyd's estimator as defined in (2.3). These variances are the same as the empirical variances of the maximum likelihood estimator to at least two decimal places. Figures 3.1 and 3.2 compare the estimated densities for the two estimators for different v and N. These results indicate that the maximum likelihood estimator of location can be regarded as the BLU estimator since their estimated densities are indistinguishable from one another.

TABLE 3.1: The Variance of $\hat{\mu}_{LB}$, Empirical Variance of $\hat{\mu}_{ML}$, the CRLB and the Degrees of Freedom Parameter, γ, in the Student-t Approximation for the Distribution of $\hat{\mu}_{LB}$ and $\hat{\mu}_{ML}$.

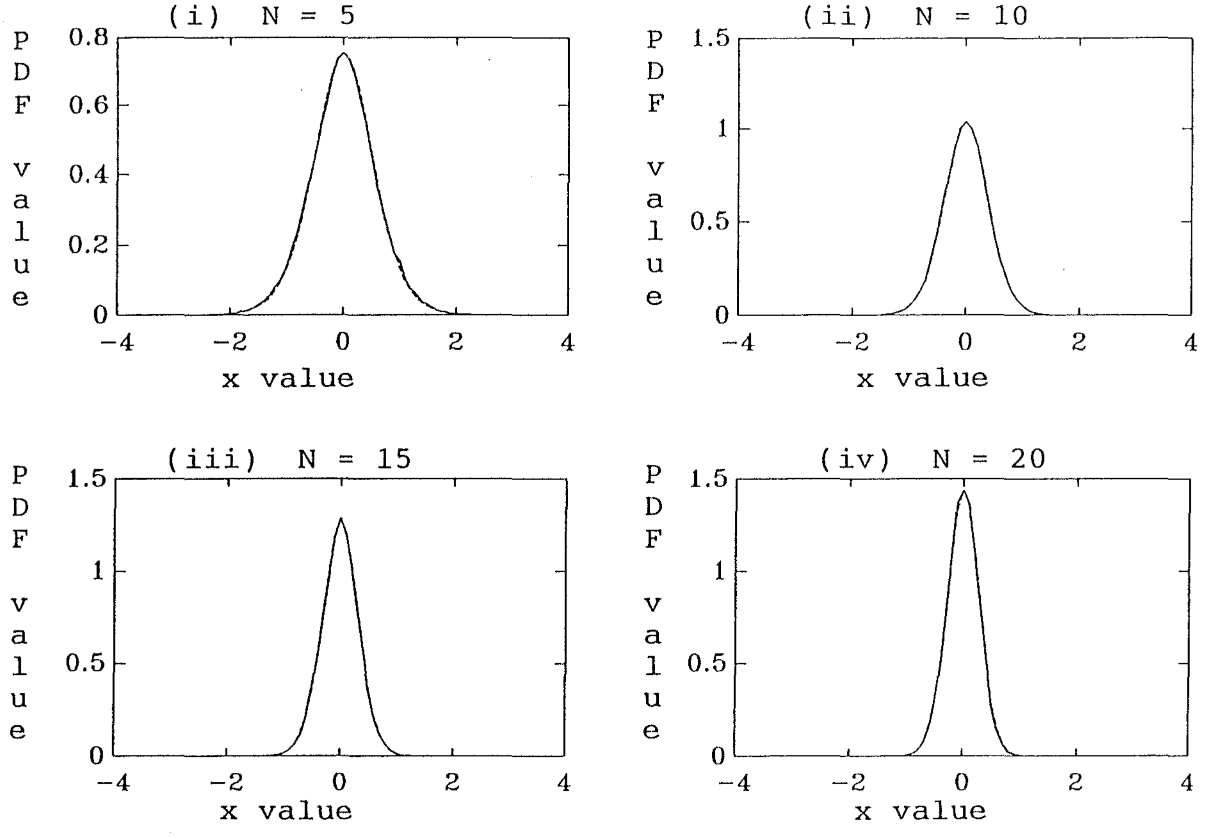
N	5	10	20	25	30
v=1					
Blue	1.2213	0.3263	0.1820	0.1261	0.0713
Empirical Variance	1.0160	0.29004	0.16961	0.11884	0.0683
CRLB	0.4000	0.2000	0.1300	0.1000	0.0667
γ	3	6	8	11	16
v=2					
Blue	*	*	*	*	*
Empirical Variance	0.4606	0.1952	0.1233	0.0900	0.0581
CRLB	0.3333	0.1667	0.1111	0.0833	0.0556
γ	6	11	16	21	∞
v=3					
Blue	0.3599	0.1634	0.1060	0.0782	0.0541
Empirical Variance	0.3571	0.1634	0.1057	0.0783	0.0540
CRLB	0.3000	0.1000	0.1000	0.0750	0.0533
γ	10	15	20	∞	∞

Table 3.1 continued

N	5	10	20	25	30
v=5					
Blue	0.2916	0.1399	0.0920	0.0683	0.0441
Empirical Variance	0.28952	0.1393	0.0916	0.0682	0.0441
CRLB	0.2667	0.1333	0.0889	0.0667	0.0440
γ	24	∞	∞	∞	∞
v=10					
Blue					
Empirical Variance		Essentially equal to the CRLB			
CRLB	0.2364	0.1182	0.0788	0.0591	0.0400
γ	∞	∞	∞	∞	∞
v=19					
Blue	0.2221	Essentially equal to the CRLB			
Empirical Variance	0.2211				
CRLB	0.2200	0.1100	0.0733	0.0550	0.0367
γ	∞	∞	∞	∞	∞

* Order statistics not available

FIGURE 3.1 Comparison of Maximum Likelihood, BLUE and Student-t Approximation for the Standardized Location Parameter Corresponding to v = 3 and Different N

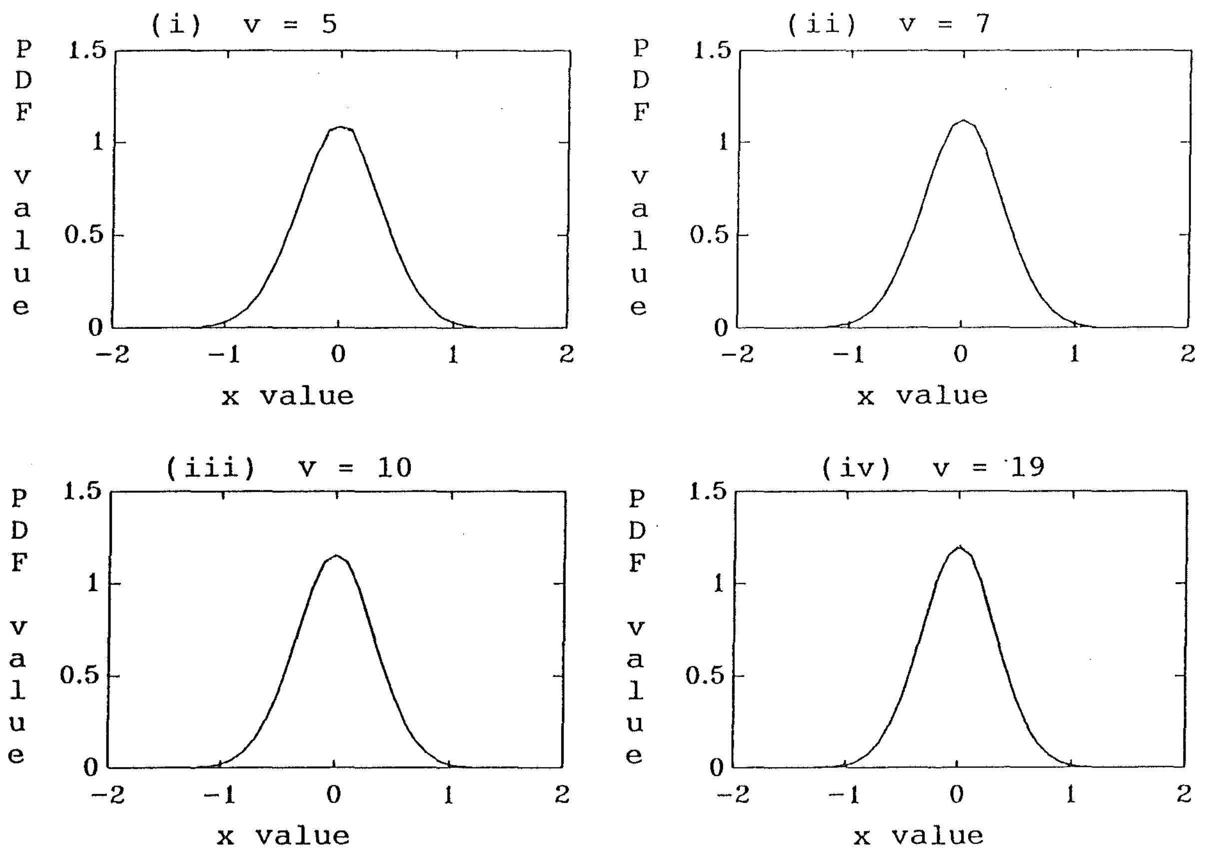


KEY - Empirical Distribution of Maximum Likelihood Estimator

-- Student-t Approximation

: Empirical Distribution of BLU Estimator

FIGURE 3.2 Comparison of Maximum Likelihood, BLUE, and Student-t Approximation for the Standardized Location Parameter Corresponding to Different v and N = 10.



KEY - Empirical Distribution of Maximum Likelihood Estimator

-- Student-t Approximation

: Empirical Distribution of BLU Estimator

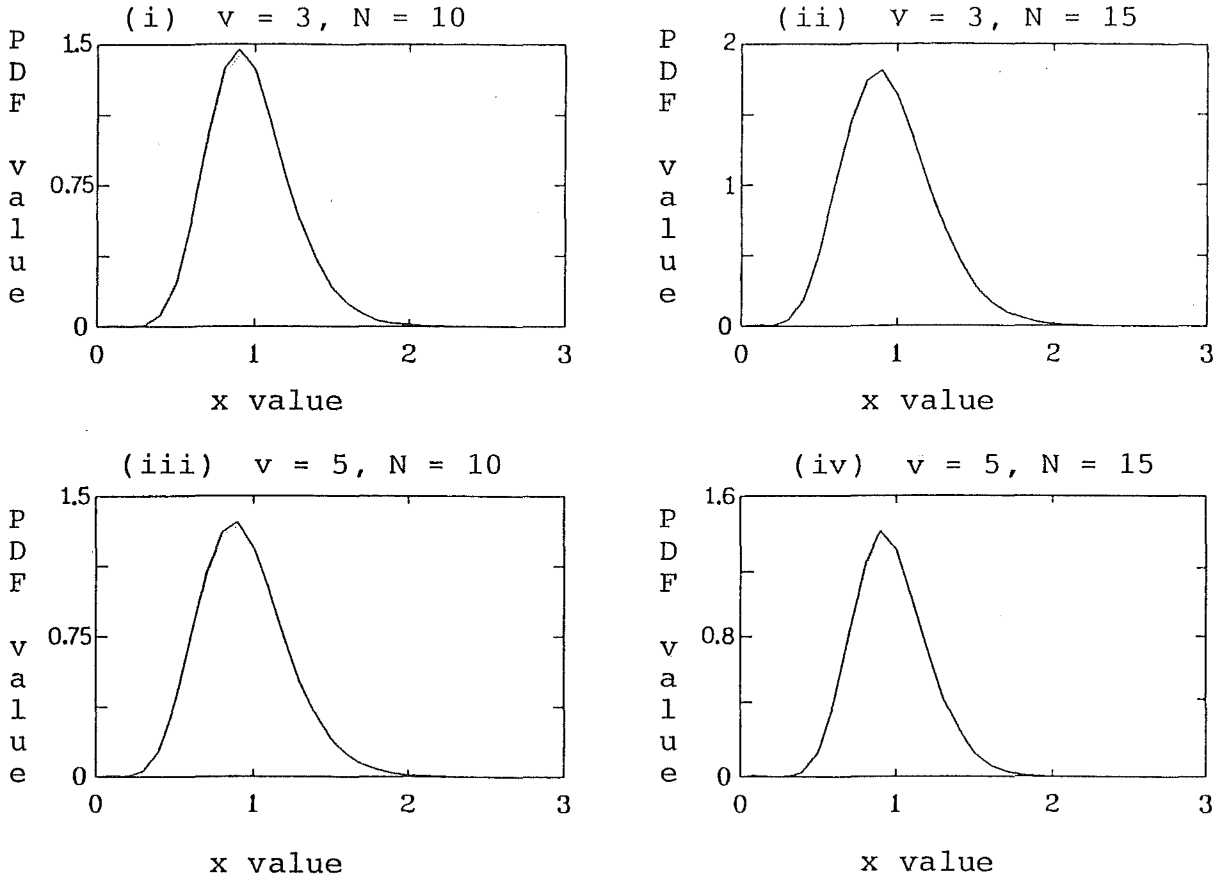
The distribution of the standardized location estimator is normally distributed as either v and/or N tend to infinity. However, for both small v and N the empirical densities indicate they have tails that are fatter than those of the normal distribution, which suggests approximating them by the Student-t distribution. The degrees of freedom parameter in this approximation, γ, is given in Table 3.1, and is chosen by matching the empirical variance with the variance obtained from the Student-t distribution with different degrees of freedom. The value of γ obtained again illustrates the closeness between the finite-sample and asymptotic approximations. Some examples of the approximation are illustrated in Figures 3.1 and 3.2.

Similar properties can be derived for the standardized maximum likelihood estimator of the scale parameter. Table 3.2 reports the bias of this estimator and its variance once the estimator has been adjusted for bias. As for the location parameter, the variance of this estimator corrected for bias will be above the CRLB (which is also given in Table 3.2), although the empirical variances are well approximated by the asymptotic variances at least for sample sizes greater than 20, and as v increases for even smaller sample sizes. Therefore in these cases the adjusted for bias maximum likelihood estimator of scale is the minimum variance estimator. More generally though, the maximum likelihood estimator is closely related to the BLU estimator of scale as given in Section 2. Table 3.2 gives the known finite-sample variance of this estimator, as defined in (2.3), and it is equal to the empirical variance of the maximum likelihood estimator adjusted for bias to at least 2 decimal places. Figure

TABLE 3.2: The Bias of $\hat{\sigma}_{ML}$, the Variance of $\hat{\sigma}_{LB}$ and the Variance of $\hat{\sigma}_{ML}$ Adjusted for Bias.

	N	10	15	20
v = 1	Bias($\hat{\sigma}_{ML}$)	0.0	0.0	0.0
	var($\hat{\sigma}_{ML}$)$_{adj}$	0.2896	0.1678	0.1175
	var($\hat{\sigma}_{LB}$)	0.4158	0.2070	0.1375
	CRLB	0.200	0.1330	0.1000
v = 3	Bias($\hat{\sigma}_{ML}$)	-0.0404	-0.0270	-0.0202
	var($\hat{\sigma}_{ML}$)$_{adj}$	0.1155	0.0753	0.0554
	var($\hat{\sigma}_{LB}$)	0.1231	0.0764	0.0552
	CRLB	0.1000	0.0667	0.0500
v = 5	Bias($\hat{\sigma}_{ML}$)	-0.0509	-0.0340	-0.0256
	var($\hat{\sigma}_{ML}$)$_{adj}$	0.0898	0.0579	0.0419
	var($\hat{\sigma}_{LB}$)	0.0913	0.0583	0.0420
	CRLB	0.800	0.0533	0.0400
v = 10	Bias($\hat{\sigma}_{ML}$)	-0.0612	-0.0408	-0.0302
	var($\hat{\sigma}_{ML}$)$_{adj}$	0.0709	0.0456	0.0329
	var($\hat{\sigma}_{LB}$)	0.0715	0.0464	0.0338
	CRLB	0.540	0.0433	0.0325
v = 19	Bias($\hat{\sigma}_{ML}$)	-0.0684	-0.0455	-0.0336
	var($\hat{\sigma}_{ML}$)$_{adj}$	0.0641	0.0386	0.0302
	var($\hat{\sigma}_{B}$)	0.0640	0.0390	0.0301
	CRLB	0.0579	0.0386	0.0289

FIGURE 3.3 Comparison of Maximum Likelihood and BLU Estimators for the Standardized Scale Parameter



KEY - Empirical Distribution of Maximum Likelihood Estimator

-- Empirical Distribution of BLU Estimator

3.3 compares the empirical densities of some cases corresponding to different v and N. These figures indicate that the maximum likelihood estimator for scale, adjusted for bias, is essentially the BLU estimator, as their densities are identical.

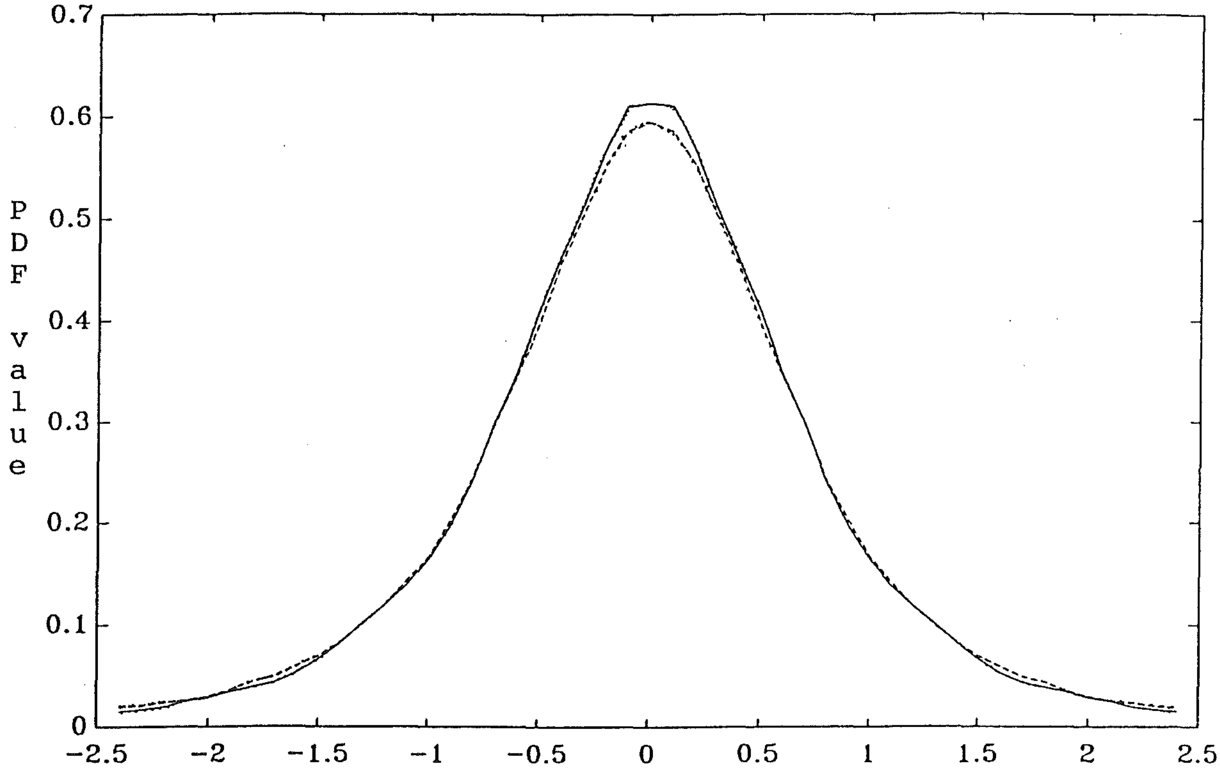
Infinite Variance: $v \leq 2$

The empirical variances for the standardized maximum likelihood estimator of the location parameter are given in Table 3.1. By applying the same argument as above, these variances will be greater than the CRLB. The sample size at which the asymptotic variance approximates the empirical variances is much larger than in the finite-variance case. For example, for both $v = 1,2$, the sample size needs to be greater than 30.

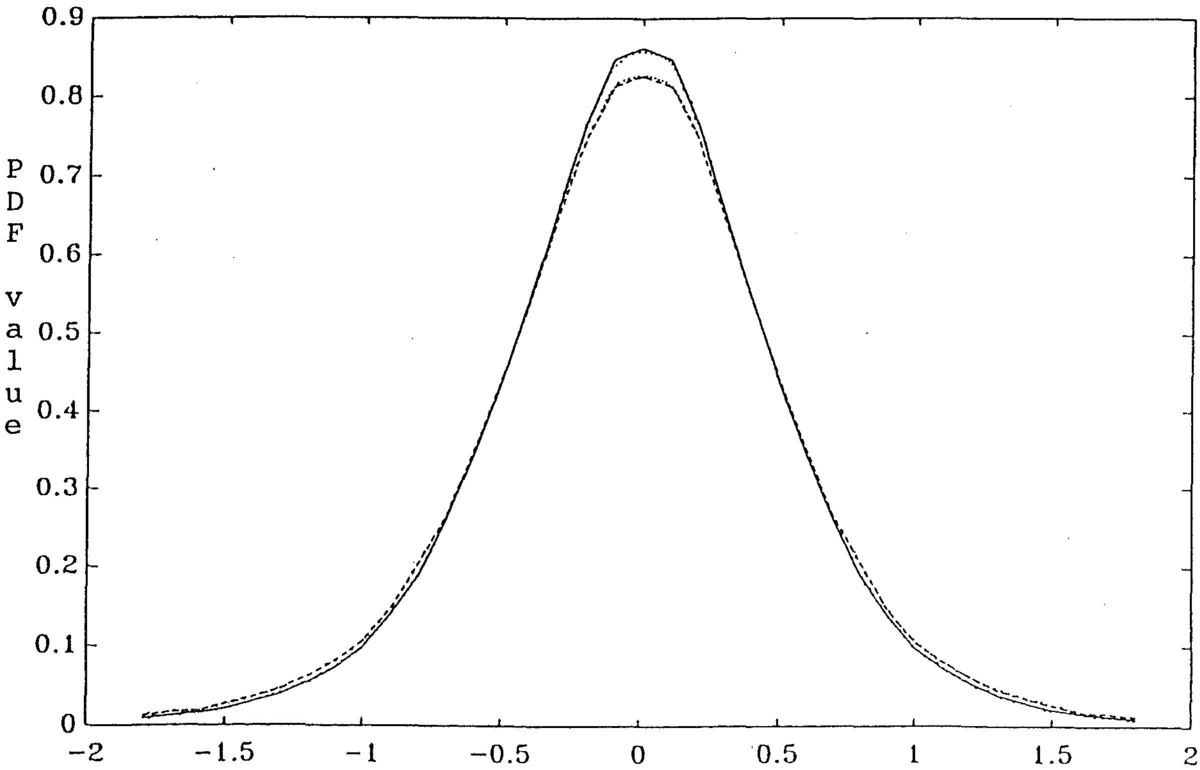
There are some differences between the BLU and maximum likelihood estimators. This is illustrated for the Cauchy distribution ($v = 1$). Table 3.1 gives the known finite sample variances of the BLU estimator, as defined in (2.3). These variances are greater than the empirical variances of the maximum likelihood estimator, and the two converge only asymptotically. Similarly, Figure 3.4 illustrates the empirical densities for two sample sizes, $N = 5$ and 10. These differences imply that for the infinite variance distributions, the maximum likelihood estimator is a nonlinear function of the "*ordered*" sample observations.

The empirical densities can be approximated by Student-t distributions, although the degrees of freedom parameter, γ, (as given in Table 3.1), in this approximation is much smaller than in the finite-variance case. Some cases are illustrated in Figures 3.4, 3.5 and 3.6 for different sample sizes. This approximation also indicates that the only difference between the BLU and maximum

FIGURE 3.4 Comparison of Maximum Likelihood, BLUE and Student-t Approximation for the Standardized Location Parameter Corresponding to v = 1 and N = 5 and 10



(i) N = 5



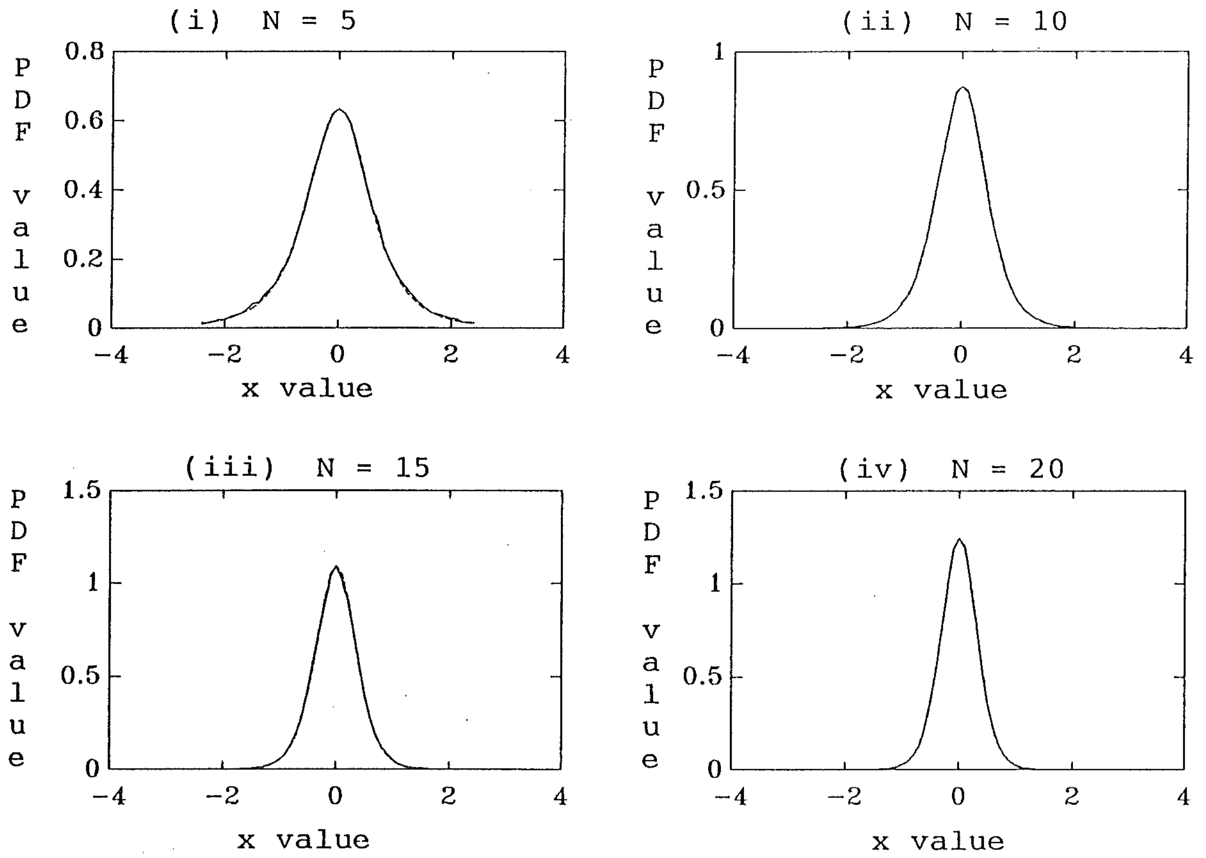
(ii) N = 10

KEY - Empirical Distribution of Maximum Likelihood Estimator

-- Student-t Approximation

: Empirical Distribution of BLU Estimator

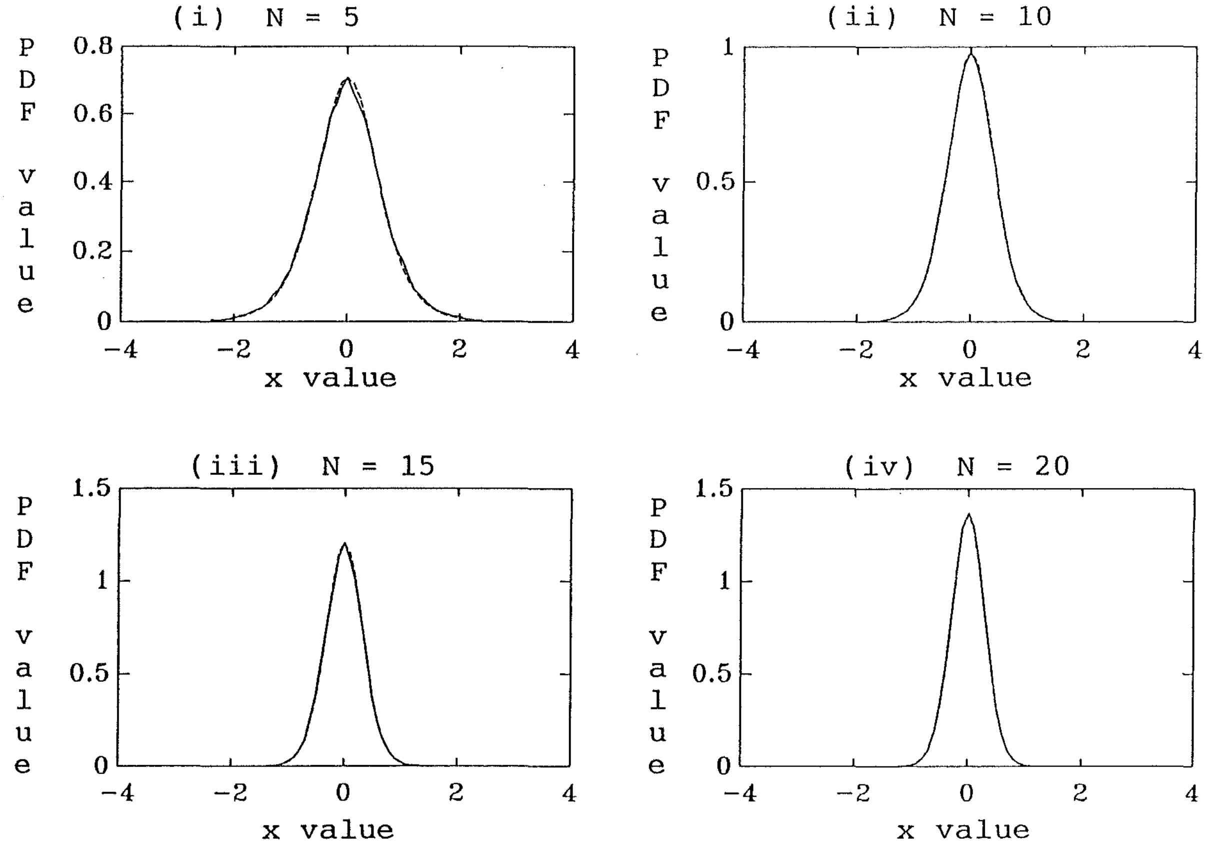
FIGURE 3.5 Comparison of Maximum Likelihood and Student-t Approximation for the Standardized Location Parameter Corresponding to ν = 1 and different N



KEY - Empirical Distribution of Maximum Likelihood Estimator

-- Student-t Approximation

FIGURE 3.6 Comparison of Maximum Likelihood and Student-t Approximation for the Standardized Location Parameter Corresponding to v = 2 and different N



KEY - Empirical Distribution of Maximum Likelihood Estimator

-- Student-t Approximation

likelihood estimators is different variances. This is because the same degrees of freedom parameter in this approximation is suitable for both estimators.

Similar properties hold for the standardized scale parameter. In particular, the maximum likelihood estimator of this parameter is more efficient in small samples than the corresponding BLU estimator. This is illustrated for the Cauchy distribution in Table 3.2. Further, for these sample sizes the variance of the maximum likelihood estimator is not approximated by the CRLB. Finally, it is interesting to note that for the Cauchy distribution, the maximum likelihood estimator is unbiased. This estimator only becomes biased as v increases.

(d) Summary of Properties

To complete this section, the results obtained are summarized in a form similar to Properties 1.1

Properties 3.1

(i) $\hat{\mu}_{ML}$ and $\hat{\sigma}_{ML}$, the joint solutions to (3.2) and (3.3) are the maximum likelihood estimators of μ and σ respectively.

(ii) The distribution of $(\hat{\mu}_{ML}-\mu)/\sigma$ can be approximated by a Student-t distribution with γ degrees of freedom, where γ is given in Table 3.1. As sample size tends to infinity, so too does γ.

(iii) For $v > 2$, $\hat{\mu}_{ML}$ has the same distribution as Lloyd's BLU estimator, $\hat{\mu}$. Even for small v and N, these estimators can be regarded as being the MVB estimator, since their variance is well approximated by the CRLB.

(iv) For $v \leq 2$, $\hat{\mu}_{ML}$ is a nonlinear estimator that is more efficient than $\hat{\mu}_{LB}$.

(v) For $v > 2$, the distribution of $\hat{\sigma}_{ML}$, adjusted for bias, can be approximated by Lloyd's BLU estimator $\hat{\sigma}_{LB}$. For $v \leq 2$, $\hat{\sigma}_{ML}$ adjusted for bias, has smaller variance than $\hat{\sigma}_{LB}$.

8.4 JOINT VERSUS IID STUDENT-t DISTURBANCES

To assess the importance of developing specification tests to distinguish between the assumption of jointness versus independence, it is necessary first to consider the consequences of misspecification. Therefore, the topic of this section is the statistical analysis of the properties of the appropriate estimator to use under one assumption, when the alternative assumption is actually correct.

Throughout this section, the superscripts I and D will be used to denote whether the standardized estimators are being used when the disturbances are iid Student-t (I) or multivariate Student-t (D). As in Section 3 it will be assumed that the estimators have been appropriately standardized, that is, they are in the form, $(\hat{\mu} - \mu)/\sigma$ and $(\hat{\sigma}/\sigma)$.

Finite Variance v > 2

The Location Parameter

Consider the case in which the disturbance terms are assumed to be independent, but are only uncorrelated. Then $\hat{\mu}_{ML}^{(I)}$ would be taken as the maximum likelihood estimator, with a distribution function that is identical to $\hat{\mu}_{LB}^{(I)}$, and can be approximated by a Student-t distribution with γ degrees of freedom (Table 3.1) and variance,

$$\text{var}\left(\hat{\mu}_{M}^{(I)}\right)=\text{var}\left(\hat{\mu}_{LB}^{(I)}\right)=(1'B1)^{-1}\left(1'B\text{var}\left[u_{(i)}\right]B\right)(1'B1)^{-1}=(1'B1)^{-1}, \quad (4.1)$$

where B is the inverse of the covariance matrix of $\left[u_{(i)}\right]$ and 1 is column vector of 1's. However, Figure 4.1 illustrates a number of empirical densities for $\hat{\mu}_{ML}^{(D)}$ and $\hat{\mu}_{ML}^{(D)}$, for different v and N, and as illustrated, these estimators are unbiased but are Student-t distributed with v degrees of freedom, where $v < \gamma$.[3] These densities are estimated via the integration of the kernel density estimator with the naive Monte - Carlo method. Details of this approach is given in Section 3, although in this case, multivariate Student - t variates are generated using the relationship (see, for example (2.3.4)),

$$X_i = Z_i\left(\frac{\chi^2}{v}\right)^{-\frac{1}{2}} \qquad i = 1 \ldots K ,$$

where $Z_1 \ldots Z_K$ are K independent standard normal variables and χ^2 is an independent chi-square variable with v degrees of freedom. The "correct" BLU and maximum likelihood estimator is $\hat{\mu}_{OLS}^{(D)}$ with variance,

$$\text{var}\left(\hat{\mu}_{OLS}^{(D)}\right) = \frac{v}{v-2}\,\frac{1}{N} \,. \qquad (4.2)$$

whereas, the actual variance of $\hat{\mu}_{ML}^{(D)}$ and $\hat{\mu}_{LB}^{(D)}$ is,

$$\text{var}\left(\hat{\mu}_{ML}^{(D)}\right) = \text{var}\left(\hat{\mu}_{LB}^{(D)}\right) = (1'B1)^{-1}\left(1'B\text{var}\left[u_{(i)}\right]B1\right)(1'B1)^{-1}$$

$$= (1'B1)^{-1}(1'B\Omega{*}B_1)(1'B1)^{-1} \quad , \qquad (4.3)$$

---

[3] The unbiasedness follows from properties of symmetrical parent distributions (see David (1970, p.105)). The distribution follows from the dependent structure.

where Ω^* is the covariance matrix of $\left[u_{(i)}\right]$ for the multivariate Student-t sample. A selection of the values of the variances in (4.1), (4.2) and (4.3) are given in Table 4.1 for different v and N. Given the range of v and N covered in Table 4.1, a comparison of these variances indicates that in general:

(1) $\text{var}\left(\hat{\mu}_{ML}^{(D)}\right) = \text{var}\left(\hat{\mu}_{LB}^{(D)}\right) \geq \text{var}\left(\hat{\mu}_{OLS}^{(D)}\right)$,

so that $\hat{\mu}_{ML}^{(D)}$ and $\hat{\mu}_{LB}^{(D)}$ are inefficient with respect to $\hat{\mu}_{OLS}^{(D)}$, except as $v \to \infty$, in which case all of the estimators are equivalent, and

(2) $\text{var}\left(\hat{\mu}_{ML}^{(D)}\right) = \text{var}\left(\hat{\mu}_{LB}^{(D)}\right) \leq \text{var}\left(\hat{\mu}_{ML}^{(I)}\right) = \text{var}\left(\hat{\mu}_{LB}^{(I)}\right)$,

so that the actual variance of $\hat{\mu}_{ML}^{(D)}$ and $\hat{\mu}_{LB}^{(D)}$ is substantially less than the assumed variance for small v.

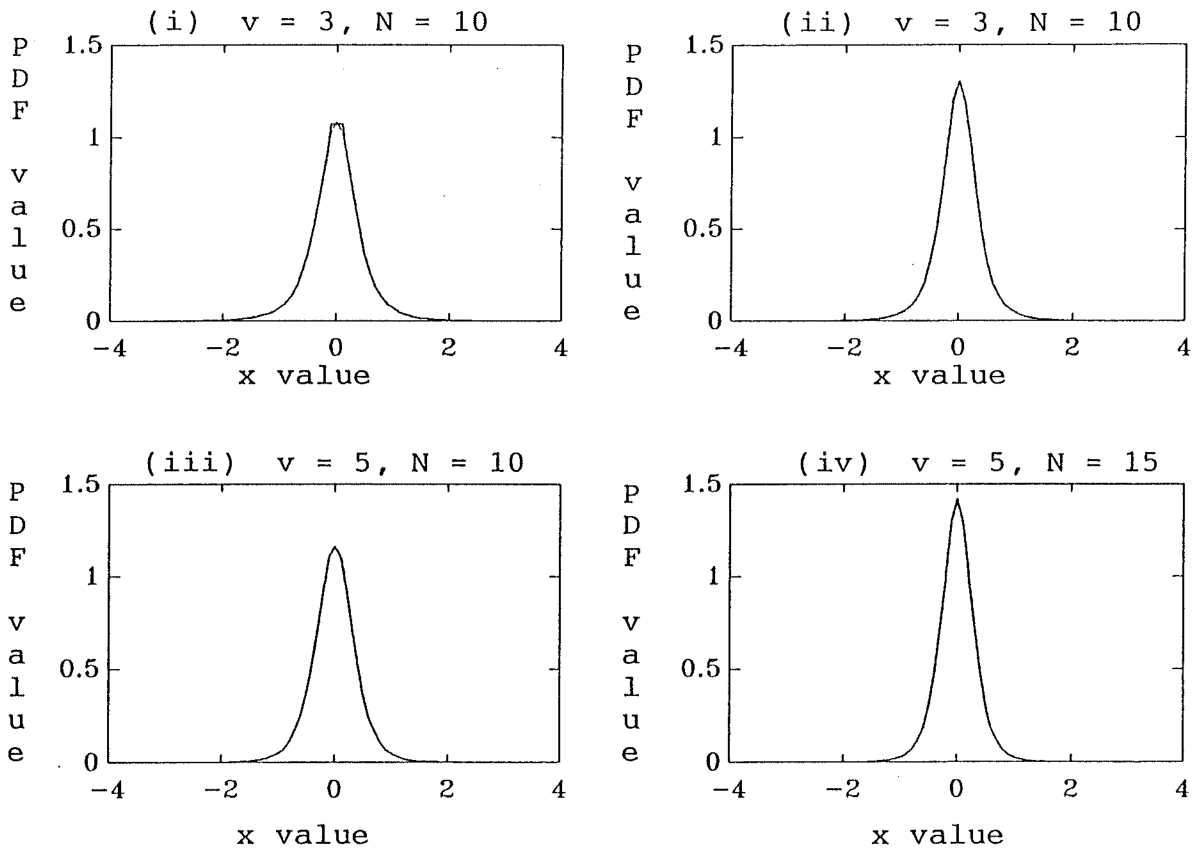
Therefore, if the disturbances are assumed to be independently distributed but are only uncorrelated, an inefficient estimator will be used which will be assumed to have a "thinner-tailed" distribution with a smaller variance than its actual distribution. Consequently, the estimator will be thought of as being more precise than it actually is. Furthermore, inferences will be based on the use of the wrong distribution, although the implications of this are beyond the scope of this thesis.

Consider, on the other hand, the case in which the disturbance terms are assumed to be uncorrelated only, but are in fact independent. Then $\hat{\mu}_{OLS}^{(D)}$ would be used as the appropriate maximum likelihood estimator, assumed to be distributed Student-t with v degrees of freedom. However, Figure 4.2 illustrates empirical densities, generated via the integration of the kernel

TABLE 4.1: The Actual and Assumed Variances of $\hat{\mu}^{(D)}_{ML}$, $\hat{\mu}^{(I)}_{OLS}$.

v	N	5	10	15	20
3	$var(\hat{\mu}^{I}_{LB})=var(\hat{\mu}^{I}_{ML})$	0.3599	0.1634	0.1061	0.0786
	$var(\hat{\mu}^{D}_{LB})=var(\hat{\mu}^{D}_{ML})$	0.65613	0.3595	0.2131	0.1660
	$var(\hat{\mu}^{I}_{OLS})=var(\hat{\mu}^{I}_{OLS})$	0.6000	0.3000	0.2000	0.1500
5	$var(\hat{\mu}^{I}_{LB})=var(\hat{\mu}^{I}_{ML})$	0.2916	0.1399	0.0916	0.0682
	$var(\hat{\mu}^{D}_{LB})=var(\hat{\mu}^{D}_{ML})$	0.25433	0.1787	0.1210	0.0868
	$var(\hat{\mu}^{I}_{OLS})=var(\hat{\mu}^{I}_{OLS})$	0.3355	0.1667	0.1111	0.0833
10	$var(\hat{\mu}^{I}_{LB})=var(\hat{\mu}^{I}_{ML})$	0.2434	0.1182	0.0788	0.0591
	$var(\hat{\mu}^{D}_{LB})=var(\hat{\mu}^{D}_{ML})$	0.2620	0.1340	0.0860	0.0651
	$var(\hat{\mu}^{I}_{OLS})=var(\hat{\mu}^{I}_{OLS})$	0.2500	0.1250	0.0833	0.0625
19	$var(\hat{\mu}^{I}_{LB})=var(\hat{\mu}^{I}_{ML})$	0.2221	0.1100	0.0733	0.0550
	$var(\hat{\mu}^{D}_{LB})=var(\hat{\mu}^{D}_{ML})$	0.2250	0.1126	0.0748	0.0561
	$var(\hat{\mu}^{I}_{OLS})=var(\hat{\mu}^{I}_{OLS})$	0.2235	0.1118	0.0745	0.0559

FIGURE 4.1 Distributions of $\hat{\mu}_{ML}^{(D)}$ and $\hat{\mu}_{LB}^{(D)}$ when the Disturbances are Uncorrelated



KEY - Empirical Distribution of Maximum Likelihood Estimator

-- Empirical Distribution of BLU Estimator

: Student-t Approximation (i) t_3

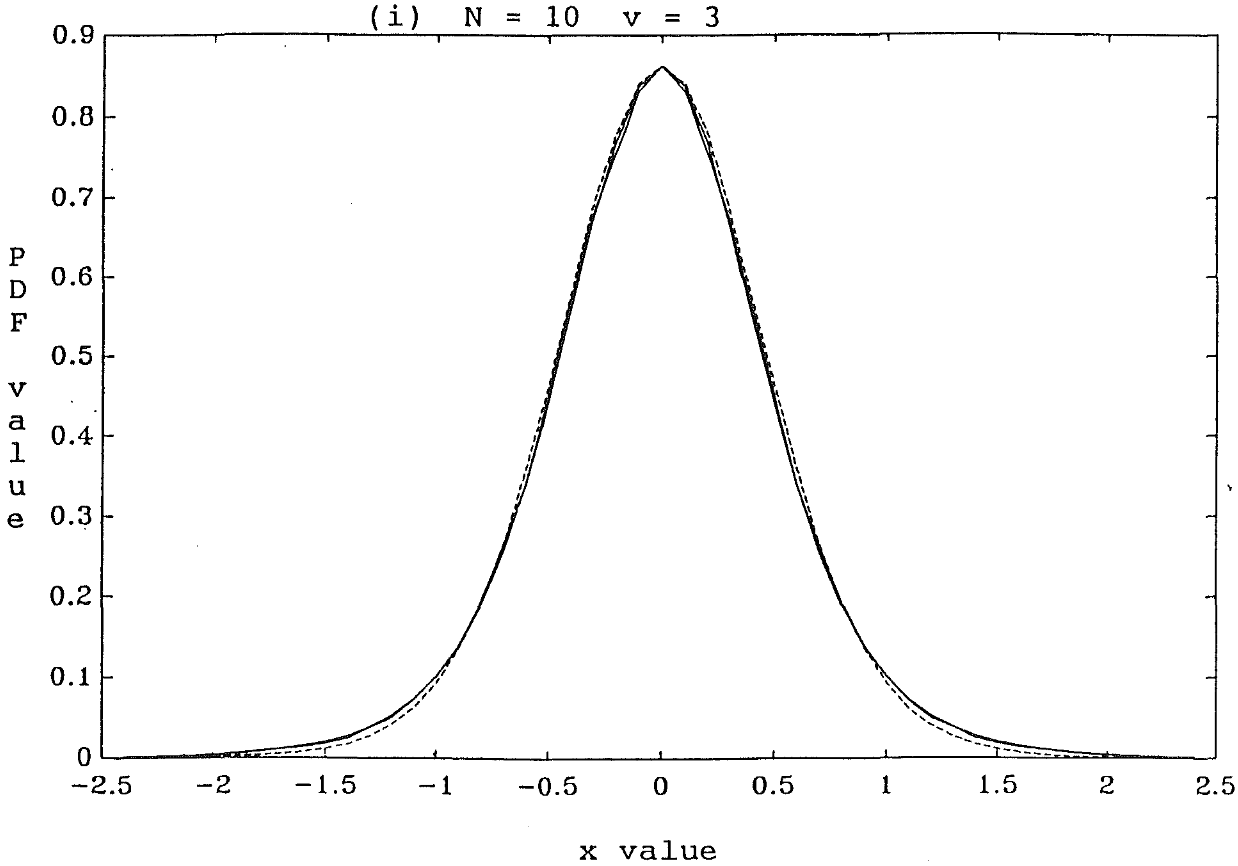
(ii) t_3

(iii) t_5

(iv) t_5

FIGURE 4.2 The Distribution of $\hat{\mu}^{(I)}_{OLS}$ for IID Student-t Disturbances

(NB Only v = 3 is shown here, as this illustrates well the feature that the distribution of $\hat{\mu}^{(I)}_{OLS}$ is fatter tailed than $\hat{\mu}^{(I)}_{ML}$.)

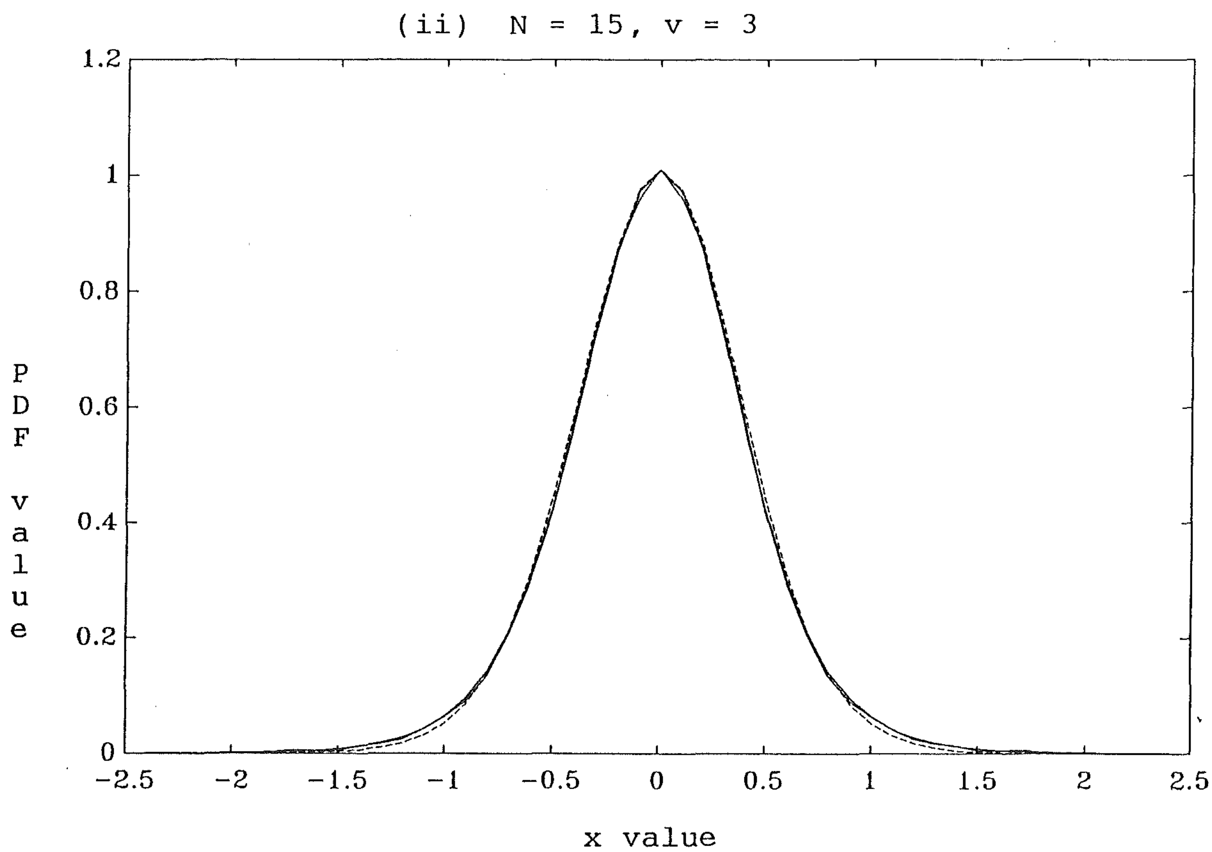


KEY - Empirical Density

- Student-t Approximation with 6 degrees of freedom

-- Student-t Approximation with 15 degrees of freedom (which is the same approximation used for $\hat{\mu}^{(I)}_{ML}$).

FIGURE 4.2(ii)



KEY - Empirical Density

- Student-t Approximation with 14 degrees of freedom

-- Student-t Approximation with 20 degrees of freedom (which is the same approximation used for $\hat{\mu}_{ML}^{(I)}$).

method with the naive Monte - Carlo method as described in Section 3, for the unbiased $\hat{\mu}^{(I)}_{OLS}$ estimator and the Student-t approximation to this distribution with γ^* degrees of freedom.[4] In this case $\gamma^* > v$, so that the assumed distribution is "fatter-tailed than the actual distribution. Further, the actual distribution of $\hat{\mu}^{(I)}_{OLS}$ is "fatter-tailed" than the distribution of the correct maximum likelihood estimator, $\hat{\mu}^{(I)}_{ML}$, since $\gamma^* < \gamma$. As well, the selection of variances given in Table 4.1 illustrates that in general,

$$(3) \qquad \mathrm{var}\left(\hat{\mu}^{(I)}_{OLS}\right) = \mathrm{var}\left(\hat{\mu}^{(D)}_{OLS}\right) \leq \mathrm{var}\left(\hat{\mu}^{(I)}_{LB}\right) = \mathrm{var}\left(\hat{\mu}^{(I)}_{ML}\right).$$

A comparison of all of these features indicates they can be substantial for small v.

Therefore, if the disturbances are assumed to be uncorrelated, when in fact they are independently distributed, an inefficient estimator with a "fatter-tailed" distribution than the "correct maximum likelihood estimator" will be used. Consequently, there is more probability of obtaining outliers. Furthermore, the distribution of this estimator will be assumed to have a "fatter-tailed" distribution than its "actual distribution", and this will in turn have consequences for inference. However, these consequences are beyond the scope of this thesis.

The Scale Parameter

When the disturbance terms are assumed to be independent, but are only uncorrelated, $\hat{\sigma}^{(D)}_{ML}$ is used to estimate the scale parameter, σ, instead of $\hat{\sigma}^{(D)}_{OLS}$, the "correct" maximum likelihood

---

[4] The unbiasedness follows from properties of symmetrical parent distributions (see David (1970), p.105).

TABLE 4.2: The Bias of $\hat{\sigma}^{(D)}_{ML}$ and $\hat{\sigma}^{(I)}_{OLS}$.

N	5	10	15	50
v=3				
Bias$\left(\hat{\sigma}^{D}_{ML}\right)$	0.0269	0.0783	0.0913	0.0929
Bias$\left(\hat{\sigma}^{I}_{OLS}\right)$	0.1036	0.1228	0.1378	0.1847
v=5				
Bias$\left(\hat{\sigma}^{D}_{ML}\right)$	0.0004	0.0023	0.0118	0.0156
Bias$\left(\hat{\sigma}^{I}_{OLS}\right)$	0.0426	0.0566	0.0595	0.07612

TABLE 4.3: Median-Bias of Adjusted $\hat{\sigma}^{(D)}_{ML}$ and $\hat{\sigma}^{(I)}_{OLS}$ Estimators for the Cauchy Distribution.

N	5	10	15	50	100
$\hat{\sigma}^{(D)}_{ML}$	-0.4013	-0.1402	-0.13375	-0.09666	-0.0959
$\hat{\sigma}^{(I)}_{OLS}$	0.88476	1.58122	2.0929	4.6639	6.9358

estimator.[5] On the other hand, when the disturbance terms are assumed to be uncorrelated only, but are in fact independent, the estimator $\hat{\sigma}^{(I)}_{OLS}$ is used instead of $\sigma^{(I)}_{ML}$. In both cases then, the estimator is biased. This bias is illustrated in Table 4.2, where the entries in this Table are based on the results of Monte - Carlo experiments using 40,000 - 60,000 replications. In particular, we see that the bias increases with N and decreases with v. Consequently, the estimated standard deviation of the unstandardized location parameter, $\hat{\sigma}\left(\mathrm{var}(\hat{\mu}^{s})\right)^{\frac{1}{2}}$, is also biased (where s denotes the standardized parameter). This bias is greater when the disturbance terms are assumed to be uncorrelated but are in fact independent.

Infinite Variance $v \leq 2$

For the infinite variance distributions the statistical consequences of inappropriately using the least squares or robust iid Student-t maximum likelihood estimators are even more serious. First, consider the inappropriate use of the OLS estimator (that is, when the disturbances are wrongly assumed to be multivariate Student-t). Because the least squares technique minimizes squared deviations, it places relatively heavy weight on outliers, so that least squares estimates are extremely sensitive to the presence and values of such observations. For iid infinite-variance distributions, "outliers" occur frequently since these distributions have "fat tails". Consequently, in repeated samples, the least squares estimates vary more than in the finite-variance case.

---

[5] The $\hat{\sigma}_{ML}$ discussed in this section is assumed to have an equivalent distribution to $\hat{\sigma}_{LB}$, see Section 3(c).

Andrews and Phillips (1986) discuss the inappropriate use of the robust and Student-t maximum likelihood estimators (that is, when the disturbances are wrongly assumed to be independent Student-t). In particular, they show that the least squares estimator is strictly preferred to the robust technique.

To consider these statistical consequences in more detail, the rest of this section will assume that the disturbances are Cauchy distributed.

Location Parameter

When the disturbances are assumed to be multivariate Student-t distributed, $\hat{\mu}_{OLS}^{(I)}$ is assumed to be the appropriate maximum likelihood estimator. However, if the disturbances are actually iid Student-t distributed, then the distribution of $\hat{\mu}_{OLS}^{(I)}$ is standard Cauchy, although its assumed distribution is Cauchy with scale factor 1/N. Consequently, for large N, it will be assumed $\hat{\mu}_{OLS}$ is very concentrated around zero, when in fact it has the same distribution as that of a single standardized observation (see Kendall and Stuart (1969, p.248)).

Alternatively, if the disturbances are assumed to be iid Student-t, $\hat{\mu}_{ML}^{(I)}$ is assumed to be the appropriate maximum likelihood estimator; the distribution of which will be taken to be approximately Student-t, with at least the first finite two moments for $N \geq 5$. Furthermore, it will be assumed that the limiting distribution is normal. However, if the disturbances are actually multivariate Student-t, the distribution of $\hat{\mu}_{ML}^{(D)}$ is Cauchy (even asymptotically), with no finite moments.

Scale Parameter

To consider the effects of misspecification on the scale parameter a simulation experiment is first used to transform $\hat{\sigma}^{(I)}_{ML}$ and $\hat{\sigma}^{(D)}_{OLS}$ into median- unbiased estimators. Then, the "median-bias" of the resulting estimators, under the appropriate type of misspecification, is calculated using 40,000 - 60,000 replications in a simple Monte - Carlo experiment. This bias is reported in Table 4.3. The results indicate that the adjusted estimator, $\hat{\sigma}_{OLS}$, is extremely sensitive to misspecification whereas the adjusted estimator $\hat{\sigma}_{ML}$, although "median-biased" is more robust.

8.5 SOME FINAL COMMENTS

Recently, models with nonnormal disturbances have attracted substantial attention (see Chapter 7). However, in such models a distinction needs to be made between multivariate distributed disturbances and iid distributed disturbances. This section has concentrated on the importance of making this distinction in the location-scale model with Student-t disturbances. In this section, small sample properties of the standardized maximum likelihood estimators of the location and scale parameters when the disturbances are distributed iid Student-t, are developed. In the literature (see, for example, Chapter 7) attention has been given to the properties of the distributions assuming multivariate Student-t disturbances. The results obtained demonstrate that the distinction between the two assumptions is an important one and the consequences of making the wrong assumption is serious, especially for small v.

Therefore, it must also be important to develop specification tests as a way of choosing between the alternative assumptions. However, before this topic is discussed, the results obtained here for the location-scale model are generalized to the multiple regression model and the exactly- identified SEM.

CHAPTER 9

THE GENERAL LINEAR REGRESSION MODEL WITH STUDENT-t DISTURBANCES

9.1 INTRODUCTION

This chapter considers the statistical comparison of the maximum likelihood estimator of the unknown β in the linear regression model (1.1.1), when it is assumed that the disturbances are distributed either as iid Student-t, or multivariate Student-t. This extends the results obtained for the location-scale model considered in Chapter 8.

In Section 2, the results of Zellner (1976) are used to develop finite-sample properties for the maximum likelihood estimator for multivariate Student-t disturbances. These properties are easily seen to be a simple generalization of those obtained for the location-scale model. In Section 3, similar properties are developed for the maximum likelihood estimator for independent Student-t disturbances. However, these properties are not a simple generalization of those obtained for the location-scale model. This is mainly because order statistics were used to develop these properties in the location-scale model. However, in the general linear regression model the usual concept of order statistics is no longer adequate, because what constitutes an appropriate ordering depends on the vector of unknown regression coefficients β. Section 4 considers the statistical consequences of making one error assumption when in fact the other assumption is valid and Section 5 concludes with some final comments.

9.2 PROPERTIES OF MAXIMUM LIKELIHOOD ESTIMATORS WITH DEPENDENT STUDENT-T ERRORS

If it is assumed that $\epsilon_1 \ldots, \epsilon_N$ are multivariate-t distributed disturbances as in (8.1.2), (i.e. with precision matrix $\sigma^2 I$), then the likelihood function for the linear regression model,

$$y = X\beta + \epsilon \, , \tag{2.1}$$

where, $y' = (y_1, \ldots, y_N)$, X is an N * K matrix of nonstochastic regressors, (K assumed to be greater than 1), $\beta' = (\beta_1, \ldots, \beta_K)$ is a vector of unknown parameters, is given by,

$$\mathcal{L}(y|\beta, v, \sigma) = \left[g(v)/(\sigma^2)^{N/2}\right]\left\{v + \left[(N-K)s^2 + (\beta-b)'X'X(\beta-b)\right]/\sigma^2\right\}^{-(N+v)/2} \tag{2.2}$$

where,

$$g(v) = \frac{v^{v/2}\Gamma[(v+N)/2]}{\pi^{N/2}\Gamma(v/2)},$$

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}\exp(-x)dx, \quad \alpha > 0 \, ,$$

$$b = (X'X)^{-1}X'y \, ,$$

$$s^2 = (y-Xb)'(y-Xb)/(N-K) \, .$$

In this case the disturbances are homoskedastic but not serially independent. It is easily seen from (2.2), (see also Zellner (1976, p.401) or Chapter 7) that b and s^2 are sufficient statistics and further, that b is the maximum likelihood estimator of β. Furthermore, from the review given in Chapter 7, we have the following set of properties:

PROPERTIES 1.1

1. b is the maximum likelihood estimator of β.

2. When $v > 2$, b is the minimum variance unbiased estimator and is therefore also BLUE, with covariance matrix $\left(v/(v-2)\right)\sigma^2(X'X)^{-1}$.

3. For all v, b is the optimal median-unbiased estimator for any loss function that is nondecreasing as the magnitude of underestimation or overestimation increases.

4. Assuming $\Sigma = \lim(X'X)/N$ is finite and nonsingular, then the limiting distribution of $N^{1/2}(b-\beta)$ is multivariate Student-t with a location vector of zeros and characteristic matrix $\sigma^2\Sigma^{-1}$. This also describes the finite-sample distribution.

These properties are easily seen to be straight generalizations of Properties 8.1.1 for the location-scale model. In the next section, the corresponding properties are developed for the maximum likelihood estimator of β in the linear regression model when the disturbances are distributed iid Student-t. However, in this case there are distinguishing features between the location-scale and general linear regression model.

9.3 PROPERTIES OF MAXIMUM LIKELIHOOD ESTIMATORS WITH INDEPENDENT STUDENT-t ERRORS

In this case it is assumed that $\epsilon_1,\ldots,\epsilon_N$ are homoskedastic and serially-independent iid Student-t distributed as in (8.1.4), so that the density of $\epsilon = (\epsilon_1,\ldots,\epsilon_N)$ is given by,

$$\text{pdf}(\epsilon|v,\sigma) = \text{pdf}(\epsilon_1|v,\sigma) \ldots \text{pdf}(\epsilon_N|v,\sigma) .$$

Throughout this section we will concentrate only on developing properties of β assuming σ^2 is unknown. Consequently, the likelihood function for the regression model (2.1) is given by,

$$\mathcal{L}(y|\beta,v,\sigma) = \text{constant} - N\log(\sigma)$$

$$- \left(\frac{v+1}{2}\right)\sum_i \log\left[1 + \left\{\frac{Y_i - X_i\beta}{\sqrt{v}\sigma}\right\}^2\right] . \qquad (3.1)$$

where X_i refers to the ith row of the matrix X in (2.1). As in the location-scale model of Chapter 8, the OLS estimator is not the maximum likelihood estimator; although the OLS estimator is BLUE for $v > 2$, it is asymptotically inefficient.

It is not possible to give a closed-form expression for the maximum likelihood estimator of β, say $\hat{\beta}_{ML}$, so $\hat{\beta}_{ML}$ is obtained via the numerical optimization of (3.1). However, unlike the location-scale model, the likelihood function is in general multimodal, as shown by Gabrielson (1982), since for all v and all linear models with $K > 1$, there exist, with probability greater than zero, data such that the joint likelihood function for both $\beta_1,\ldots,\beta_K$ and σ^2 is multimodal. Therefore, because of the multimodality of the likelihood function, it is important to have appropriate initial starting values for the unknown parameters β and σ^2. These are obtained, for example, using Amemiya (1985, p.138) who states that if $\hat{\theta}_1$ is a consistent estimator of $\hat{\theta}_0$ such that $\sqrt{N}(\hat{\theta}_1 - \theta_0)$ has a proper limit distribution, the second round estimator $\hat{\theta}_2$ has the same asymptotic distribution as a consistent root of the likelihood equation, and so too does the final converged root $\hat{\theta}_c$. Details of the argument on which this result is based are given in Appendix B. The actual first round estimators used in this Chapter are given in the discussion of the asymptotic distribution of $\hat{\beta}_{ML}$.

The multimodality of the likelihood function (3.1) is one distinguishing feature between the linear regression model and the location- scale model. Another difference between the two models arises as a result of the definition of order statistics. In the location-scale model, finite- sample properties of $\hat{\beta}_{ML}$ (which corresponds to K = 1 in (2.1)), are developed for v > 2, by showing a relationship with Lloyd's BLUE, which is the BLU estimator among the class of L-estimators. However, when the more general linear model is considered, the usual concept of order statistics is no longer adequate, because what constitutes an appropriate ordering depends on the vector β. Consequently, there is no generalization of Lloyd's BLU estimator. However, there have been generalizations of some of the estimators contained in the L-class, such as generalizations of the trimmed- mean estimator, which will be denoted as a class of estimators by $\hat{\beta}_{TLS}$. In the finite-sample analysis of $\hat{\beta}_{ML}$, the mean square error (MSE) of some members of $\hat{\beta}_{TLS}$ are compared with the corresponding MSE of $\hat{\beta}_{ML}$. The objective of this comparison is to determine if there is a generalized relationship between the maximum-likelihood and L-class estimators in the linear regression model.

The rest of this section is divided up into four parts. The first part discusses the asymptotic distribution of the maximum likelihood estimator, and the second part develops properties of the finite-sample distribution of $\hat{\beta}_{ML}$. The third part summarizes all of the properties of $\hat{\beta}_{ML}$ obtained, and the fourth part offers some overall comments.

(i) Asymptotic Distribution of $\hat{\beta}_{ML}$

Although in the likelihood function (3.1) it is assumed that both β and σ are unknown parameters, in the discussion in this section we are interested only in developing the asymptotic distribution of $\hat{\beta}_{ML}$, as the properties of this estimator are the focus of this chapter. Therefore, for the purposes of this section it will be assumed that σ is known in (3.1), since from Lehmann (1983, p.438), the asymptotic efficiency in this case is the same as if σ is assumed unknown, because the distribution of ϵ is symmetric.

Kelejian and Prucha (1985) consider the limiting distribution of $\hat{\beta}_{ML}$ corresponding to $v > 2$. In particular, they show that,

$$\sqrt{N}(\hat{\beta}_{ML}-\beta) \overset{D}{\to} N\left[0,\left\{\frac{v+3}{v+1}\right\}\sigma^2\Sigma^{-1}\right] \quad , \; v > 2 \qquad (3.2)$$

where $\Sigma_X = \lim_{N\Rightarrow\infty}(X'X)/N$. However, as for the location scale model, it is relatively easy to show that this result holds for all v. This result is proved in Theorem 3.1.

Theorem 3.1

There exists a solution $\hat{\beta}_{ML}$ to the likelihood (3.1) such that (3.2) holds for all v.

Proof

The proof of the theorem follows by considering the combination of the following two points:

1. There exist estimators of β which are consistent and asymptotically normally distributed. The estimators used in

this chapter are OLS for $v > 2$ and $\hat{\beta}_{LAD}$ as described in (7.3.4) with asymptotic distribution in (7.3.7) for $v \leq 2$. These two estimators are used because, not only do they satisfy the requirements of consistency and asymptotic normality, but they also proved to be efficient in terms of the number of iterations required to obtain a maximum of (3.1).[1]

2. If $\hat{\theta}$ is a consistent estimator of θ such that $\sqrt{N}(\hat{\theta}-\theta)$ has a proper limit distribution, the second round estimator is asymptotically normally distributed and asymptotically efficient. The details of the argument this result is based on is given in Appendix B.

Therefore, if the numerical maximization of (3.1) begins with the estimators in (1), this implies from (2) that the resulting converged root of the likelihood equation corresponding to (3.1), $\hat{\beta}_{ML}$, will be asymptotically normally distributed and asymptotically efficient.

(ii) <u>Properties of the Finite-Sample Distribution</u>

In developing properties of the finite-sample distribution of $\hat{\beta}_{ML}$, we consider the standardized maximum likelihood estimators, that is, $(\hat{\beta}_{ML}-\beta)/\sigma$. This is because Antle and Bain (1969) show that these statistics depend only on the sample size N.

To develop the finite-sample properties a number of results are obtained. In particular, mean square errors (MSE's) are

---

[1] In the actual numerical computation of (3.1), we also need to supply an initial starting value for σ. For $v < 2$, s is used, while for $v \geq 2$, we find the residuals from the least absolute regression and then take the median of these residuals as our starting value (see, for example, Judge (1985, p.831)).

estimated for the linear regression model with sample sizes N = 20, 50 and numerous values of K.[2] These MSE's are reported in Table 3.1, (N = 20) and Table 3.2 (N = 50). Each of the entries in the tables are based on at least 40,000 replications for K < 10. However, for K > 10, since the convergence of the likelihood equation (2.2) is very slow, the number of replications is decreased substantially, and often less than 10,000 replications are used. This number of replications was chosen on the basis of available computer processor time. The MSE's require iid Student - t variates to be generated. For degrees of freedom $v < 3$, these are generated by the inversion of the distribution function (see, for example, Devroye (1986, p.27)). In particular, for $v = 1$, the Cauchy distribution, standard Cauchy variates are generated as,

$$X = \tan\left(\pi(U-\tfrac{1}{2})\right)$$

and for $v = 2$, the t_2-distribution,

$$X = \sqrt{2}(U-\tfrac{1}{2})/\mathrm{SQRT}\left(U(1-U)\right) \quad ,$$

where U is from U(0,1), generated using the NAG subroutine G0FCAF, which uses a multiplicative congruential method. For the rest of the Student-t family, $v \geq 3$, X is generated via a transformation of a symmetric beta variate, (see, for example, Devroye (1986, p.446)). This can be written in terms of independent uniform random numbers U_1, U_2 as,

---

[2] All of the estimators in this chapter are unbiased. This holds because the error distributions assumed are symmetrical, (see David (1970, p.105)). Hence, MSE is simply the sum of the individual variances of each of the estimated coefficients of (2.1).

$$X = \frac{2\sqrt{v}\ \sin(2\pi U_1)(1-U_2^{2/v-1})}{(1-\sin^2(2\pi U_1))(1-U_2^{2/v-1})} .$$

This formula is useful as it is valid for all members of the Student-t family with $v \geq 3$. It also does not require the generation of as many random uniform deviates as does the traditional method of generating a t-random variable via its interpretation as a ratio of a standard normal to the square root of an independent normalized chi-square variable. The design matrix X is generated as a column of ones with remaining entries drawn as iid realizations from the N(0,1) distribution generated using NAG subroutine G05DDF which is based on Brent's (1974) algorithm. In the preliminary analysis, numerous other X matrices were used, but the results given in Table 3.1 and Table 3.2 illustrate the general results obtained. Also reported in these tables is the corresponding asymptotic MSE, which is calculated on the basis of the formula given in (3.2).

As well as estimating the MSE's, a number of pdf's for $(\hat{\beta}_{ML_i}-\beta)/\sigma$, $(i = 1,\ldots,K)$ are estimated using the procedures described in Chapters 3 and 4. Therefore, we only briefly detail them here for completeness. Empirical densities are estimated via the integration of the kernel density estimator with the naive Monte-Carlo method. The kernel estimate at point X is equal to,

$$\hat{pdf}(X) = \frac{1}{N^*h(N^*)}\sum_j k\left[\frac{X - X_j}{h(N^*)}\right], \qquad (3.3)$$

where k[.] is the standard N(0,1) density. The window width h(N*) is chosen using the interactive approach of Tapia and Thompson

(1978). In all cases this approach led to the use of a window width between 0.02 and 0.09. N* is simply the number of replications in the simulation experiment, and is chosen using the bound of estimation. For example, the results of Parzen (1962) and Cacoullos (1966) imply,

$$\left(N^*h^m(N^*)\right)^{\frac{1}{2}}\left[\hat{pdf}(x) - E\left(\hat{pdf}(x)\right)\right] \sim N\left(0, pdf(x)\int K^2\right). \quad (3.4)$$

holds. The result given in (3.4) can be achieved if $\left(N^*h^m(N^*)\right)^{\frac{1}{2}}Bias\left[\hat{pdf}(x)\right]$ tends to zero asymptotically since,

$$\left(N^*h^m(N^*)\right)^{\frac{1}{2}}\left[\hat{pdf}(x) - pdf(x)\right] = \left(N^*h^m(N^*)\right)\left[\hat{pdf}(x) - E\left(\hat{pdf}(x)\right)\right] + \left(N^*h^m(N^*)\right)^{\frac{1}{2}}Bias\left[\hat{pdf}(x)\right].$$

Ullah (1988, p.642) shows that $Bias\left[\hat{pdf}(x)\right]$ is proportional to $h^2(N^*)$. This implies that if $N^*h^{(4+m)/2}(N^*)$ tends to zero asymptotically then (3.4) holds. Therefore, for the normal kernel $\frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}y^2)$, the 99% asymptotic confidence interval for $\hat{pdf}(X)$ is given by,

$$\hat{pdf}(X) \pm 2.58 \left[\frac{\hat{pdf}(X)}{2N^*h\sqrt{\pi}}\right]^{\frac{1}{2}},$$

so that B is given by,

$$B = 2.58\left[\frac{\hat{pdf}(X)}{2N^*h(N^*)\pi}\right]^{\frac{1}{2}}.$$

N* is varied until B is less than 0.01 for all points at which the density is estimated. In all experiments, N* varies between 60,000 and 90,000 replications[3]. The input of X_j in (3.3) involves numerically maximizing the likelihood function (3.1). Two algorithms from the Harwell Subroutine library are used, these being algorithms VAI3AD and VF04AD, which both use the BFGS formula, (Broydon (1970), Fletcher (1970), Goldfard (1970) and Shanno (1970)). All computations are performed in double precision to 7 decimal places of accuracy. The final results, however, are not dependent upon which algorithm is used in this step. Furthermore, the solutions of each of the algorithms used were compared with those in the standard econometric packages TSP and SHAZAM, and were found to give similar results. Standard iid Student-t variates, are generated as described above. Further details of the Monte Carlo methodology are given in Chapter 4.

Empirical densities are illustrated in Figures 3.1 and 3.2 for one particular i (as similar results are obtained for the others), for $v = 3$ and $v = 10$ respectively, with $N = 20$ and $K = 2, 5, 10$ and 12. In Figure 3.3, empirical densities are also illustrated for $v = 1$, $N = 20$ and $K = 2, 10$. In each of these figures the empirical densities are compared with the corresponding appropriate asymptotic distribution.

Finally, various MSE's are estimated using at least 40,000 replications in a simple Monte-Carlo experiment. These MSE's are

---

[3]Empirical densities were also computed using the Epanechnikov (1969) kernel. However, given the number of replications used, the results proved not to depend on which kernel is used. This situation is similar to the comparison of different kernels for the Cauchy distribution using a "large sample", as is illustrated in Figure 5.1 in Chapter 3.

reported in Table 3.3 for a number of trimmed-mean robust estimators, $\hat{\beta}_{TLS}$. All of the estimators considered are based on solutions to (7.3.4), and are obtained numerically using the subroutine of Koenker and D'Orey (1987). This subroutine involves linear programming techniques. In particular we have,

1. $\hat{\beta}_{LAD}$, which corresponds to the solution of (7.3.4) when $\theta = 0.5$. This estimator has asymptotic distribution given by (7.3.7).

2. $\hat{\beta}_{TLS1}$, requires the calculation of a preliminary estimate, $\hat{\beta}_0$. $\hat{\beta}_0$ is obtained as the average of the θ and $(1-\theta)$ regression quantiles. These regression quantiles are obtained as solutions to (7.3.4). Then the residuals from $\hat{\beta}_0$ are calculated and the observations corresponding to the $[N\theta]$ smallest and $[N\theta]$ largest residuals are removed.[4] $\hat{\beta}_{TLS1}$ is defined as the least squares estimate calculated from the remaining observations and has asymptotic distribution given in (7.3.6).

3. $\hat{\beta}_{TLS2}$, also requires the calculation of a preliminary estimate. The regression quantiles obtained as solutions to (7.3.4) for $0 < \theta < 0.5$ are calculated corresponding to θ (denote by $\hat{\beta}(\theta)$), and $(1-\theta)$, (denoted by $\hat{\beta}(1-\theta)$). Then, any observation whose residual from $\hat{\beta}(\theta)$ is

---

[4] The notation $[\alpha]$ denotes the greatest integer not exceeding α.

Table 3.1: Empirical and Asymptotic MSE's for $\hat{\beta}_{ML}$ for N = 20

	K = 2	K = 5	K = 10	K = 12	K = 14
Empirical MSE $\hat{\beta}_{ML}$ v=1	0.5370	1.7750	16.337	23.4105	99.5194
Asymptotic MSE $\hat{\beta}_{ML}$	0.3566	0.7300	3.3794	2.9824	5.6424
Empirical MSE $\hat{\beta}_{ML}$ v=2	0.3466	0.9085	5.9611	5.6886	13.0976
Asymptotic MSE $\hat{\beta}_{ML}$	0.2972	0.6083	2.8162	2.4853	4.7020
Empirical MSE $\hat{\beta}_{ML}$ v=3	0.2934	0.6933	4.4360	4.6078	10.1311
Asymptotic MSE $\hat{\beta}_{ML}$	0.2675	0.5475	2.5346	2.2368	4.2318
Empirical MSE $\hat{\beta}_{ML}$ v=5	0.2474	0.5376	0.28608	2.7820	6.2487
Asymptotic MSE $\hat{\beta}_{ML}$	0.2377	0.4867	2.2529	1.9883	3.7616
Empirical MSE $\hat{\beta}_{ML}$ v=10	0.2129	0.4449	2.1309	1.9013	3.6882
Asymptotic MSE $\hat{\beta}_{ML}$	0.2107	0.4314	1.9970	1.7623	3.3341

Table 3.2: Empirical and Asymptotic MSE's for $\hat{\beta}_{ML}$ for N = 50

		K = 3	K = 5	K = 10	K = 30
Empirical MSE $\hat{\beta}_{ML}$	v=1	0.19674	0.3681	1.1868	∞^*
Asymptotic MSE $\hat{\beta}_{ML}$		0.1516	0.2444	0.6216	4.3796
Empirical MSE $\hat{\beta}_{ML}$	v=2	0.14104	0.24233	0.68398	∞^*
Asymptotic MSE $\hat{\beta}_{ML}$		0.1263	0.2037	0.5180	3.6497
Empirical MSE $\hat{\beta}_{ML}$	v=3	0.12045	0.20207	0.5440	7.2081
Asymptotic MSE $\hat{\beta}_{ML}$		0.1137	0.1833	0.4662	3.2847
Empirical MSE $\hat{\beta}_{ML}$	v=5	0.10335	0.17069	0.4547	3.9620
Asymptotic MSE $\hat{\beta}_{ML}$		0.1011	0.1629	0.3673	2.9197
Empirical MSE $\hat{\beta}_{ML}$	v=10	0.08996	0.14679	0.03807	2.9536
Asymptotic MSE $\hat{\beta}_{ML}$		0.0896	0.1444	0.3673	2.5897

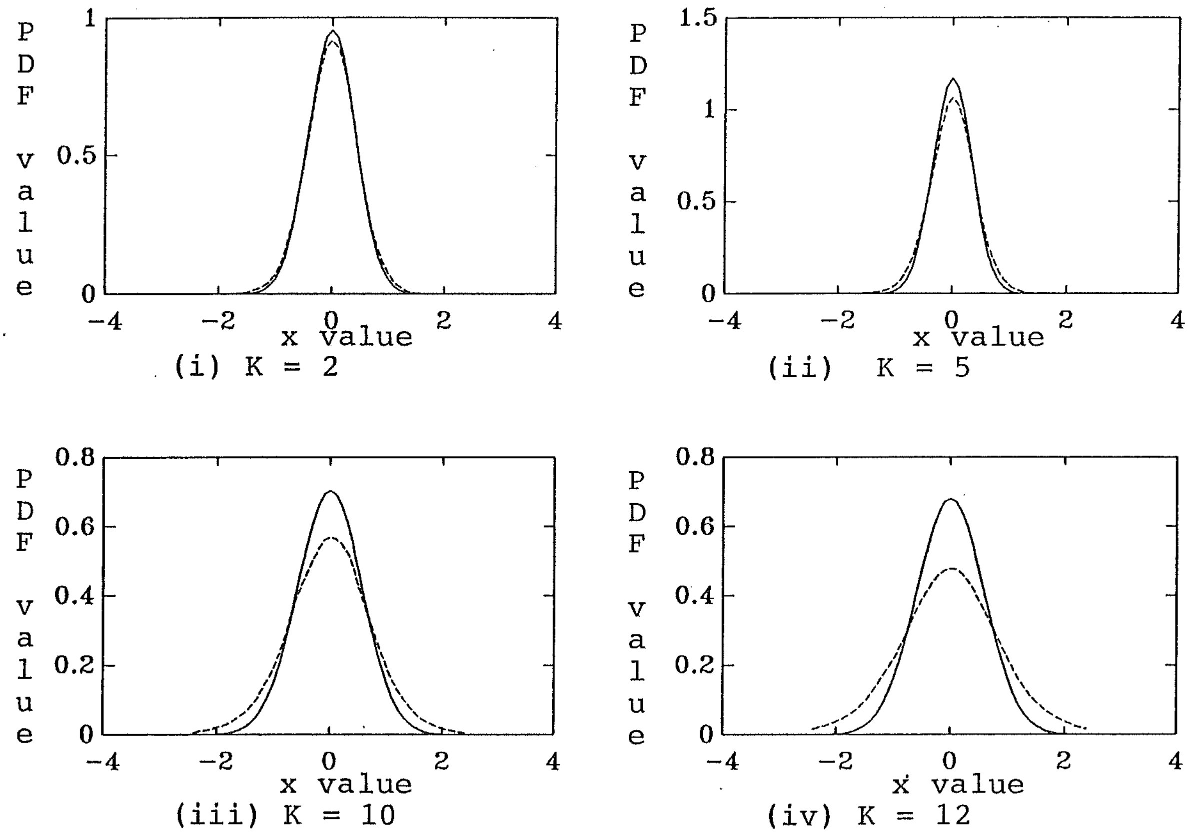
*Conjectured on the basis of empirical results.

Table 3.3: Empirical MSE's for $\hat{\beta}_{TLS}$ and Actual MSE's for b for N = 20

		K = 2	K = 5	K = 10
Empirical MSE $\hat{\beta}_{LAD}$	v=1	0.6958	2.4039	36.3327
Empirical MSE $\hat{\beta}_{TR1}$		0.9575	2.8250	∞^*
Empirical MSE $\hat{\beta}_{TR2}$		1.4144	10.1941	∞^*
+Actual MSE b		∞	∞	∞
Empirical MSE $\hat{\beta}_{LAD}$	v=2	0.4491	1.1811	6.6794
Empirical MSE $\hat{\beta}_{TR1}$		0.5205	2.3078	158.40
Empirical MSE $\hat{\beta}_{TR2}$		0.4454	1.2337	29.3768
+Actual MSE b		∞	∞	∞
Empirical MSE $\hat{\beta}_{LAD}$	v=3	0.3709	0.8375	4.5484
Empirical MSE $\hat{\beta}_{TR1}$		0.3376	0.9487	9.8426
Empirical MSE $\hat{\beta}_{TR2}$		0.3360	0.8370	3.5280
+Actual MSE b		0.5349	0.9880	5.0691
Empirical MSE $\hat{\beta}_{LAD}$	v=5	0.3543	0.7032	3.4456
Empirical MSE $\hat{\beta}_{TR1}$		0.2864	0.7675	6.9117
Empirical MSE $\hat{\beta}_{TR2}$		0.2763	0.6428	2.8998
+Actual MSE b		0.2972	0.6085	2.8162
Empirical MSE $\hat{\beta}_{LAD}$	v=10	0.3049	0.6223	2.8603
Empirical MSE $\hat{\beta}_{TR1}$		0.2553	0.6633	5.6639
Empirical MSE $\hat{\beta}_{TR2}$		0.2416	0.5388	2.2101
+Actual MSE b		0.2229	0.4563	2.1121

*Conjectured on the basis of empirical results.
+For a comparison between β_{ML} and b when K is greater than 10 we can note that the actual MSE's of b for K=14, v=3,5, and 10 are 8.4656, 4.702 and 3.5265 respectively.

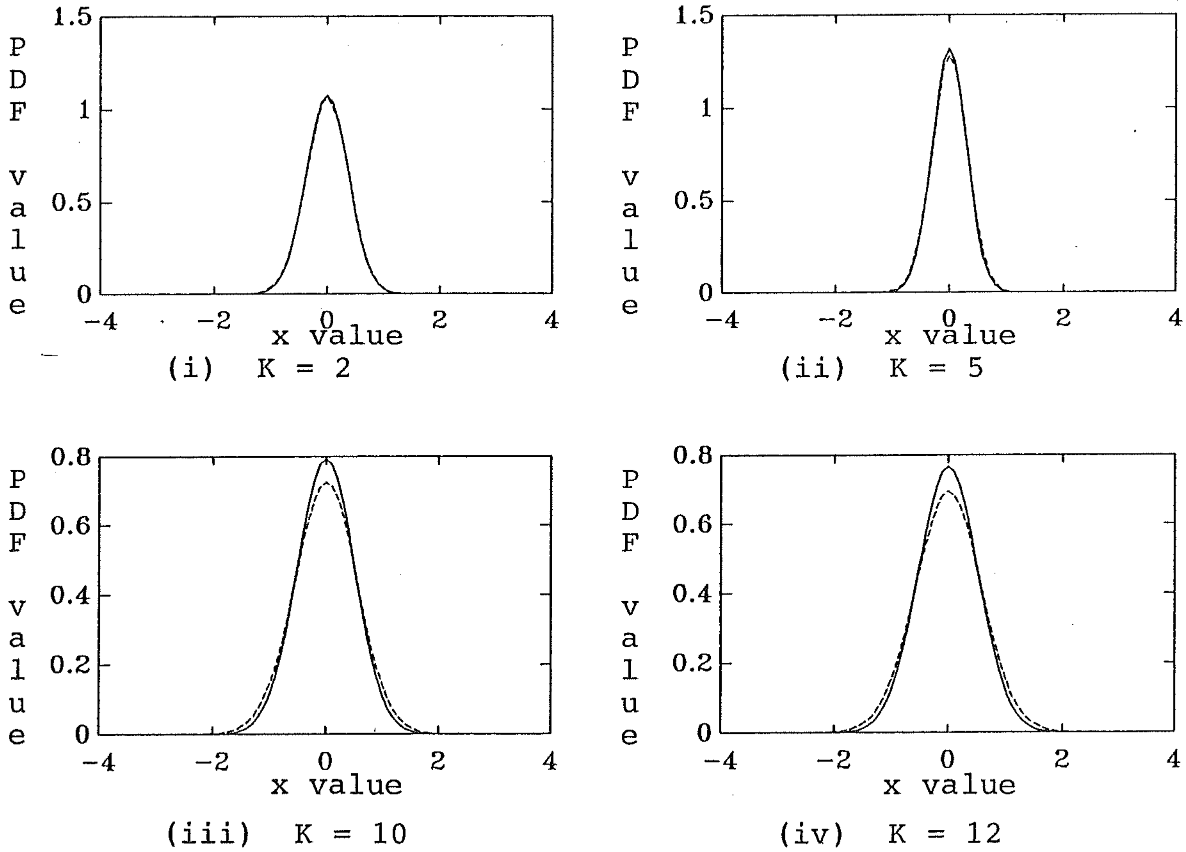
FIGURE 3.1 Comparison of the Finite-Sample Distribution of $\hat{\beta}_{ML_i}$ with its Asymptotic Distribution for v = 3, N = 20.



KEY -- Empirical distribution of $\hat{\beta}_{ML_i}$

- Asymptotic distribution of $\hat{\beta}_{ML_i}$

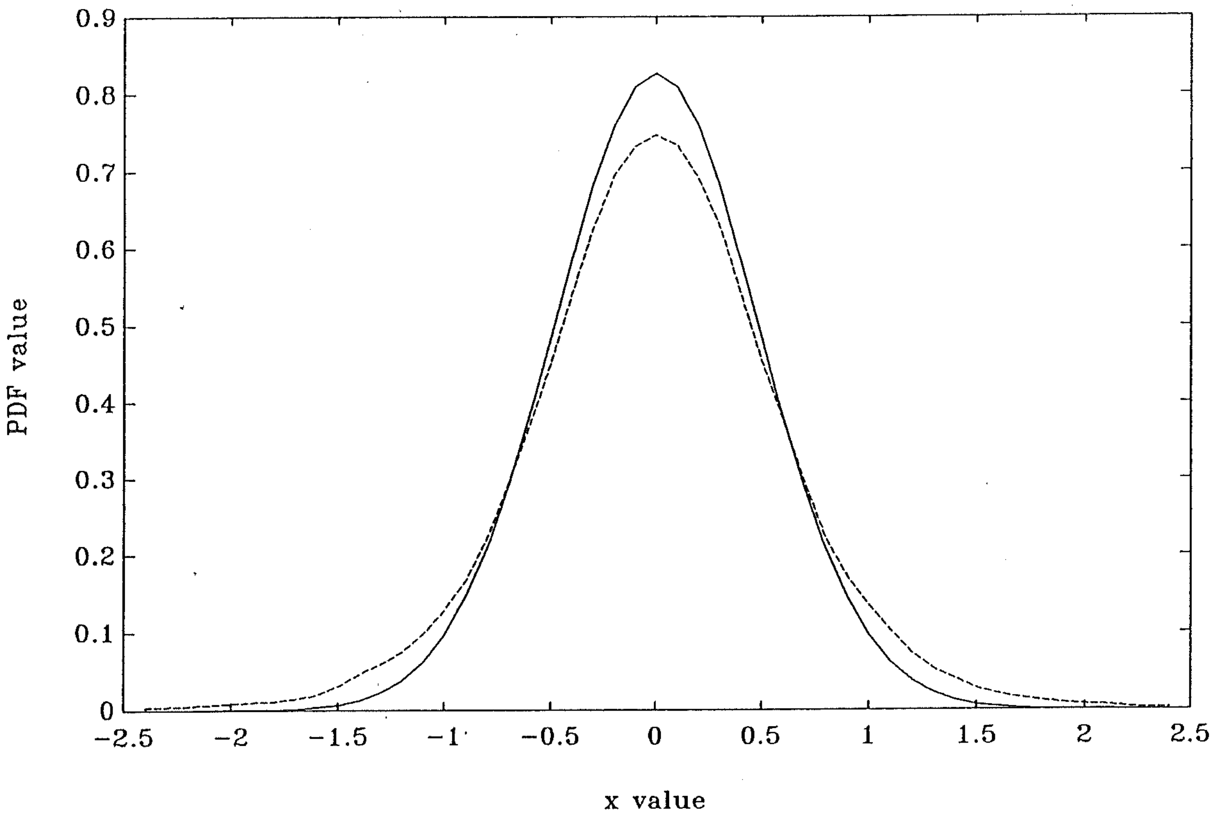
FIGURE 3.2 Comparison of the Finite-Sample Distribution of $\hat{\beta}_{ML_i}$ with its Asymptotic Distribution for v = 10, N = 20



KEY -- Empirical distribution of $\hat{\beta}_{ML_i}$

- Asymptotic distribution of $\hat{\beta}_{ML_i}$

FIGURE 3.3 Comparison of the Finite-Sample Distribution of $\hat{\beta}_{ML_i}$ with its Asymptotic Distribution for v = 1, N = 20



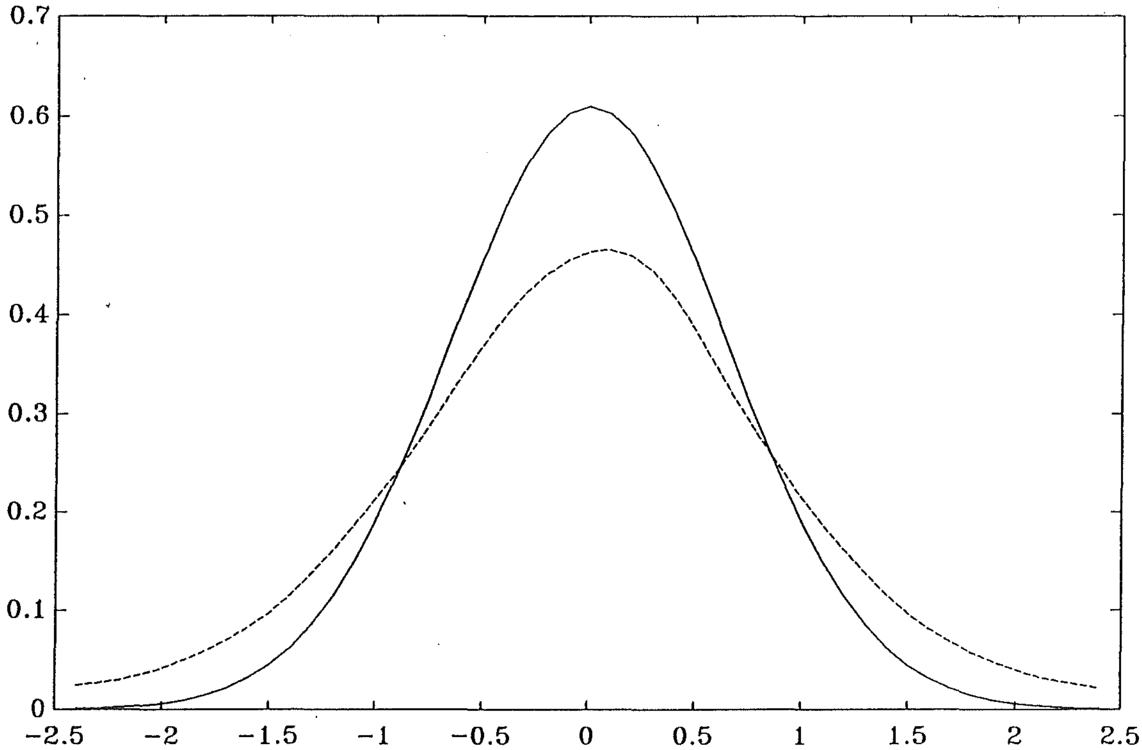
(i) K = 2

KEY

-- Empirical Distribution of $\hat{\beta}_{ML_i}$

- Asymptotic Distribution of $\hat{\beta}_{ML_i}$

FIGURE 3.3 (ii) K = 10



KEY

-- Empirical Distribution of $\hat{\beta}_{ML_i}$

– Asymptotic Distribution of $\hat{\beta}_{ML_i}$

negative or whose residual from $\hat{\beta}(1-\theta)$ is positive is removed from the sample. $\hat{\beta}_{TLS2}$ is defined as the least squares estimate calculated from the remaining observations, and has asymptotic distribution given in (7.3.6).

In particular, both $\hat{\beta}_{TLS1}$ and $\hat{\beta}_{TLS2}$ are calculated assuming $\theta = 0.2$. This θ is chosen so as to represent a "slightly" trimmed estimator, whereas $\hat{\beta}_{LAD}$ represents a "drastic" robust estimator (see, for example, Amemiya (1985, p.75).

The discussion of all of these results is divided into two parts. The finite-sample distribution of $(\hat{\beta}_{ML}-\beta)/\sigma$ is, first compared with its corresponding asymptotic distribution and secondly, compared with the results obtained for the finite-sample distribution for each of the $(\hat{\beta}_{TLS}-\beta)/\sigma$ considered. The discussion is also broken down into finite- variance ($v > 2$) and infinite-variance ($v \leq 2$) distributions.

(i) Comparison With Limiting Distribution

It is important to make comparisons between the finite-sample distribution and the limiting distribution, as the limiting distribution is often used as an approximation to the finite-sample distribution.

Finite Variance: $v > 2$

From the results reported in Table 3.1 then, the following general comments can be made. For small models, the asymptotic MSE is a good approximation to the actual MSE. This was also true for the individual variances, although they are not reported here.

However, as the number of regressors increases for a fixed N, the asymptotic MSE considerably understates the actual MSE. These features are also illustrated in Figures 3.1 and 3.2, where for larger sized models the empirical pdf is much "fatter- tailed" than the corresponding asymptotic distribution.

These results suggest that for small K relative to N the asymptotic distribution can be used to approximate the finite-sample distribution. However, for large K relative to N, some other approximation is needed, perhaps based on the Student-t distribution which has fatter-tails than the normal distribution. However, this approximation is not pursued here as the results obtained in the comparison with other estimators suggest that the maximum likelihood estimator may not be the appropriate estimator to use in this case.

Infinite Variance: $v \leq 2$

From the results reported in Table 3.1, we can see that the asymptotic MSE understates the actual MSE considerably, even in models where K is small relative to N. This is also shown in Figure 3.3 for the Cauchy distribution.

Consequently, even for moderately-sized N and small K, the asymptotic distribution should not be used to approximate the finite-sample distribution. For the infinite-variance distributions, some other approximation, perhaps based on the Student-t distribution, should be used instead.

(ii) Comparison With Other Estimators

Finite Variance: $v > 2$

From Table 3.3 we can see that the performance of the members of $\hat{\beta}_{TLS}$ chosen deteriorates rapidly as K increases for fixed N. In comparison with these estimators, $\hat{\beta}_{ML}$ can clearly be seen to be superior on the basis of MSE. However, when $\hat{\beta}_{ML}$ is compared with b, the OLS estimator, this superiority holds only for small and moderately-sized K. In this case, we have an interesting example of an asymptotic inefficient estimator having superior finite-sample performance, at least over some of the parameter space.[5]

Infinite Variance: $v \leq 2$

As in the case when $v > 2$, the performance of all of the $\hat{\beta}_{TLS}$ estimators deteriorates rapidly for moderately-sized K. In particular, the MSE appears to approach infinity as it does for b. While for small and moderately-sized K $\hat{\beta}_{ML}$ is superior to these estimators on the basis of MSE, it too has an infinite MSE for large K. Therefore, while for moderately- sized values of K, $\hat{\beta}_{ML}$ is substantially superior to the other estimators considered for large values of K, all of the estimators seem to have infinite MSE as does b, so in this case, on the basis of MSE, they are indistinguishable.

---

[5] Although the results reported here for $\hat{\beta}_{ML}$ correspond to the converged root of (3.1) with b as the initial starting value, similar results were obtained with other starting values, for example $\hat{\beta}_{LAD}$.

(iii) Summary of Properties

The results obtained on the finite-sample distribution of $\hat{\beta}_{ML}$ are now summarized in Properties 2.1.

Properties 2.1

1. $\hat{\beta}_{ML}$ is the maximum likelihood estimator of β. It is found via the numerical optimization of the likelihood function (3.1).
2. When $v > 2$, $\hat{\beta}_{ML}$ is superior to a wide number of robust estimators and b on the basis of MSE, for moderately-sized K and fixed N. However, as K increases, the performance of $\hat{\beta}_{ML}$ deteriorates rapidly, and $\hat{\beta}_{ML}$ becomes inefficient relative to b, where b is BLUE but asymptotically inefficient.
3. For $v \geq 2$, $\hat{\beta}_{ML}$ is superior to a wide range of robust estimators and b for moderately-sized K and fixed N. However, for large K corresponding to fixed N, all of these estimators have infinite MSE, so that on the basis of this measure it is impossible to distinguish between them.
4. Assuming $\Sigma = \lim_{N \Rightarrow \infty}(X'X)$ is finite and nonsingular, there exists a solution, β_{ML} to (3.1) such that, $\sqrt{N}(\hat{\beta}_{ML}-\beta)/\sigma$ is multivariate normal with a mean vector of zeros and covariance matrix $\sigma^2\left[\frac{v+3}{v+1}\right]\Sigma^{-1}$. This limiting distribution is only useful as an approximation to the finite-sample distribution in linear regression models where K is small and $v \geq 2$.

(iv) Overall Comments

A comparison of Properties 1.1 and 2.1 suggest substantial differences between the maximum likelihood estimators b and $\hat{\beta}_{ML}$, corresponding to joint and independent Student-t disturbances respectively. Therefore, it is important to consider the consequences of misspecifying "joint versus independent" disturbances. This analysis is carried out in the next section.

9.4 JOINT VERSUS IID STUDENT-T DISTURBANCES

In this section we consider the statistical consequences of misspecifying "jointly-distributed" and "independently-distributed" disturbances. Such an analysis will assess the importance of developing specification tests to make this distinction in the linear regression model.

Throughout this section, the superscripts I and D will be used to denote whether the standardized estimators are being used when the disturbances are iid Student-t, (I), and multivariate Student-t, (D). As in Section 3, it will be assumed that $\hat{\beta}_{ML}$ has been appropriately standardized, that is, it is written as, $(\hat{\beta}_{ML_i} - \beta_i)/\sigma$ as Antle and Bain (1969) show that these statistics depend only on the sample size N.

Finite Variance: $v > 2$

Consider the case in which the disturbance terms are assumed to be independent, but are only uncorrelated. In this case, $\hat{\beta}_{ML}$ will be assumed to be the correct maximum likelihood estimator to use. Although this estimator is unbiased, there are a number of consequences as a result of using this estimator, rather than the correct maximum likelihood estimator, the OLS estimator, b.[6]

In Table 4.1, empirical MSE's are reported for $\hat{\beta}_{ML}^{(D)}$ corresponding to N = 20, K = 2, 3, 5, 10. These MSE's are estimated as described in the previous section, although in this case multivariate Student - t variates, say X_i, need to be generated. These are generated using the relationship (see, for example (2.3.4)),

$$X_i = Z_i \left(\frac{\chi^2}{v} \right)^{-\frac{1}{2}} \qquad i = 1 \ldots K ,$$

where $Z_1 \ldots Z_K$ are K independent standard normal variables and χ^2 is an independent chi-square variable with v degrees of freedom. The chi-square and standard normal variables are generated as described in Section 3. In comparison with the actual MSE's for b, which are also given in Table 4.1, we can see that while $\hat{\beta}_{ML}^{(D)}$ is robust for small models, it becomes increasingly inefficient.

Further, the large-sample distribution of $\hat{\beta}_{ML}^{(D)}$ will incorrectly be assumed to be given by (3.2). There are two implications associated with this. First, the asymptotic variances associated with (3.2) will be used to approximate the actual variances for each $\hat{\beta}_{ML}^{(D)}$, (i = 1,...,K). Some examples of the use of this approximation are given in Table 4.2 for $(\hat{\beta}_{ML_i}^{(D)} - \beta)/\sigma$ (i = 1,...,K), for N = 20, K = 2, v = 3, 5, 10. These examples are illustrative of a more general comparison, from which it can be concluded that the use of this incorrect approximation results in

---

[6] The unbiasedness follows from properties of symmetrical parent distributions (see David (1970, p.105)).

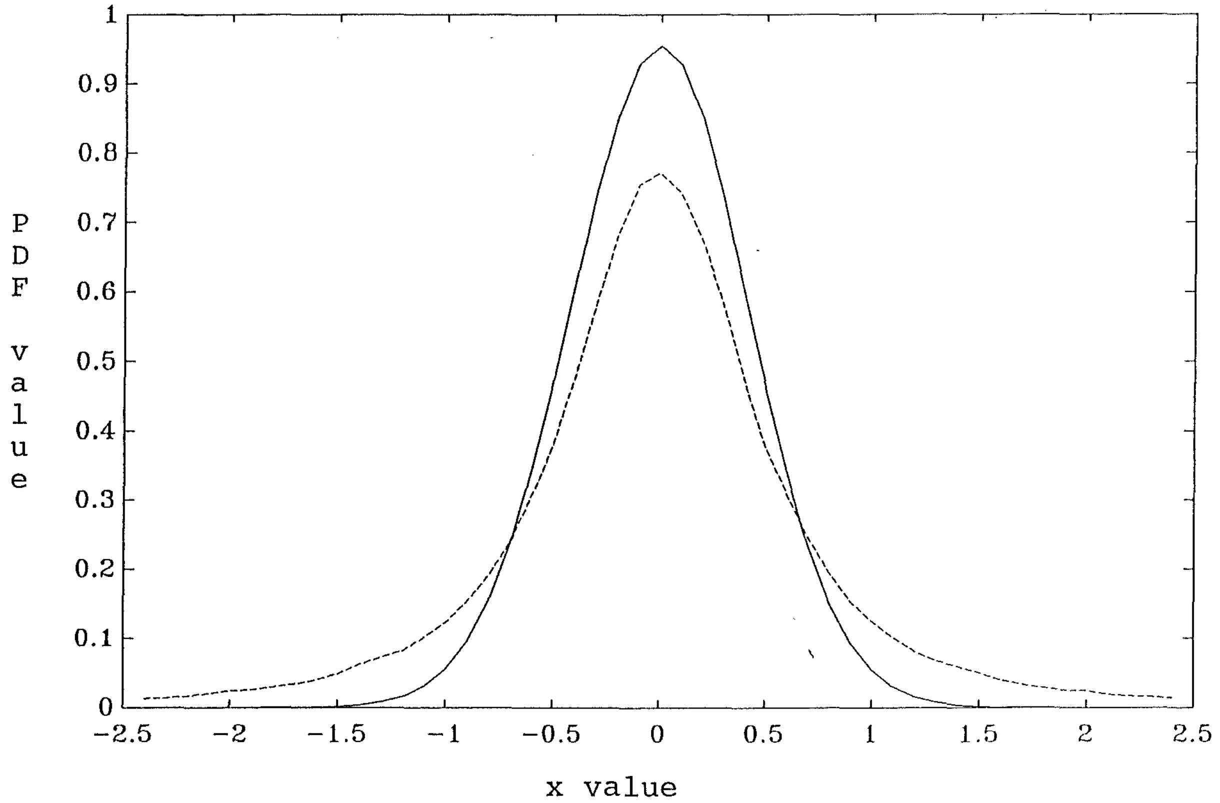
Table 4.1: Comparison of MSE's for $\hat{\beta}_{ML}^{(D)}$ and b for N = 20

		K = 2	K = 3	K = 5	K = 10
Empirical MSE $\hat{\beta}_{ML}$	v = 3	0.6207	1.0773	1.3099	7.1583
MSE b		0.5349	0.2609	0.9885	5.0691
Empirical MSE $\hat{\beta}_{ML}$	v = 5	0.3202	0.5803	0.6706	2.4783
MSE b		0.2972	0.5338	0.6083	2.8162
Empirical MSE $\hat{\beta}_{ML}$	v = 10	0.2270	0.4142	0.4684	2.006
MSE b		0.2229	0.4003	0.4563	2.1121

Table 4.2: Comparison of Empirical Variances for $\hat{\beta}_{ML}^{(D)}$ with Asymptotic Variance of $\hat{\beta}_{ML}^{(I)}$ for K = 2, N = 20

		Empirical MSE $\hat{\beta}_{ML_i}^{(D)}$	Asymptotic MSE $\hat{\beta}_{ML_i}^{(I)}$
v = 3	i = 1	0.39876	0.17475
	i = 2	0.22195	0.0927
v = 5	i = 1	0.20884	0.1553
	i = 2	0.11131	0.0824
v = 10	i = 1	0.14801	0.13768
	i = 2	0.07900	0.07303

FIGURE 4.1 Comparison of the distribution of $\hat{\beta}_{ML}^{(D)}$ with its Incorrectly Assumed Asymptotic Distribution $\hat{\beta}_{ML}^{(I)}$ for v = 3, N = 20



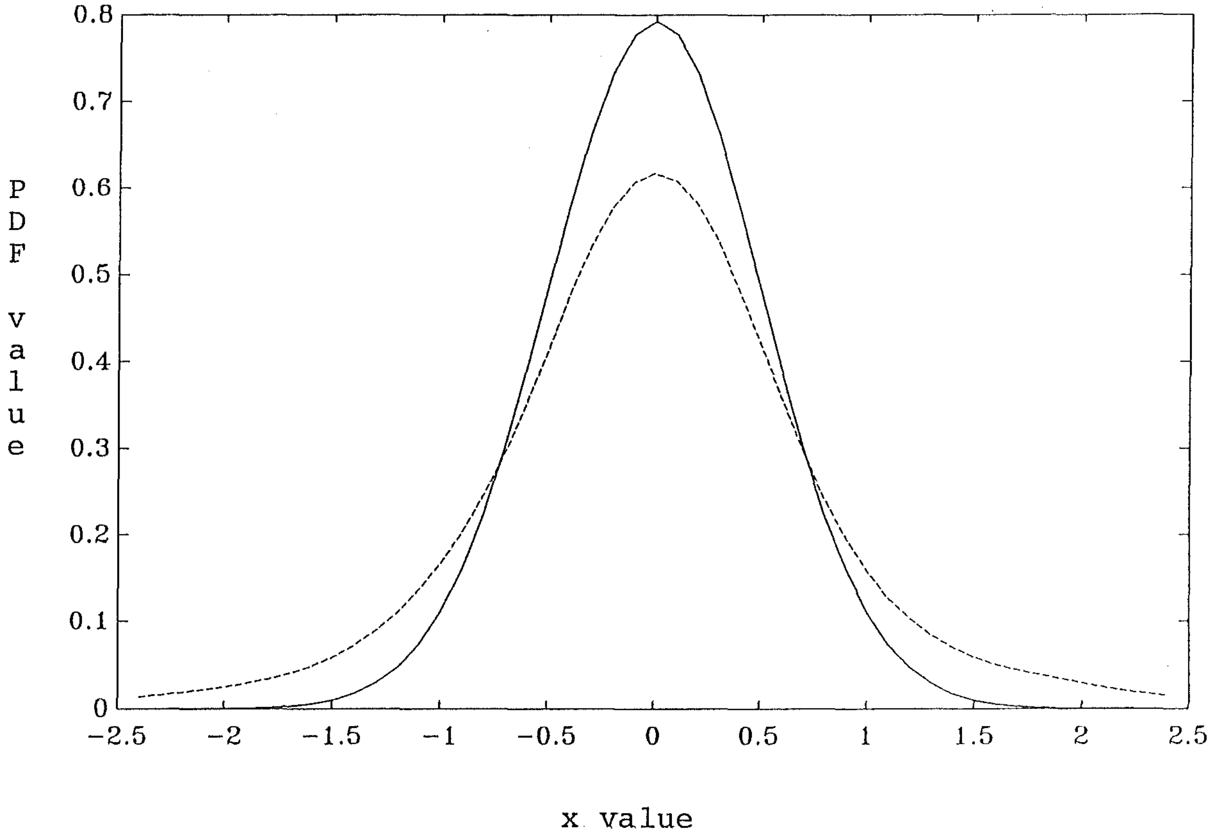
(i) K = 2

KEY

-- Empirical Distribution of $\hat{\beta}_{ML}^{(D)}$

- Incorrectly Assumed Asymptotic Distribution given by $\hat{\beta}_{ML}^{(I)}$

FIGURE 4.1 (ii) K = 10



KEY

-- Empirical Distribution of $\hat{\beta}_{ML}^{(D)}$

- Incorrectly Assumed Asymptotic Distribution given by $\hat{\beta}_{ML}^{(I)}$

$\hat{\beta}_{ML}^{(D)}$ being considered to be substantially more precise than it actually is. Secondly, the multivariate normal distribution will be used to approximate the finite- sample distribution of $\hat{\beta}_{ML}$, which is actually distributed multivariate Student-t with v degrees of freedom. The effect of this is illustrated for a particular $(\hat{\beta}_{ML_i}-\beta)/\sigma$ corresponding to v = 3, N = 20, K = 2, 10 in Figure 4.1 (i) and (ii) respectively. This density was estimated via the integration of the kernel method with the naive Monte-Carlo method as described in Section 3, and using multivariate - t random numbers as described above. This figure emphasizes that the use of the wrong limiting distribution implies that one is much more confident that the estimator is located around the true parameter value than one should be. These results hold for all sample sizes N, as the wrong asymptotic distribution is used even asymptotically.

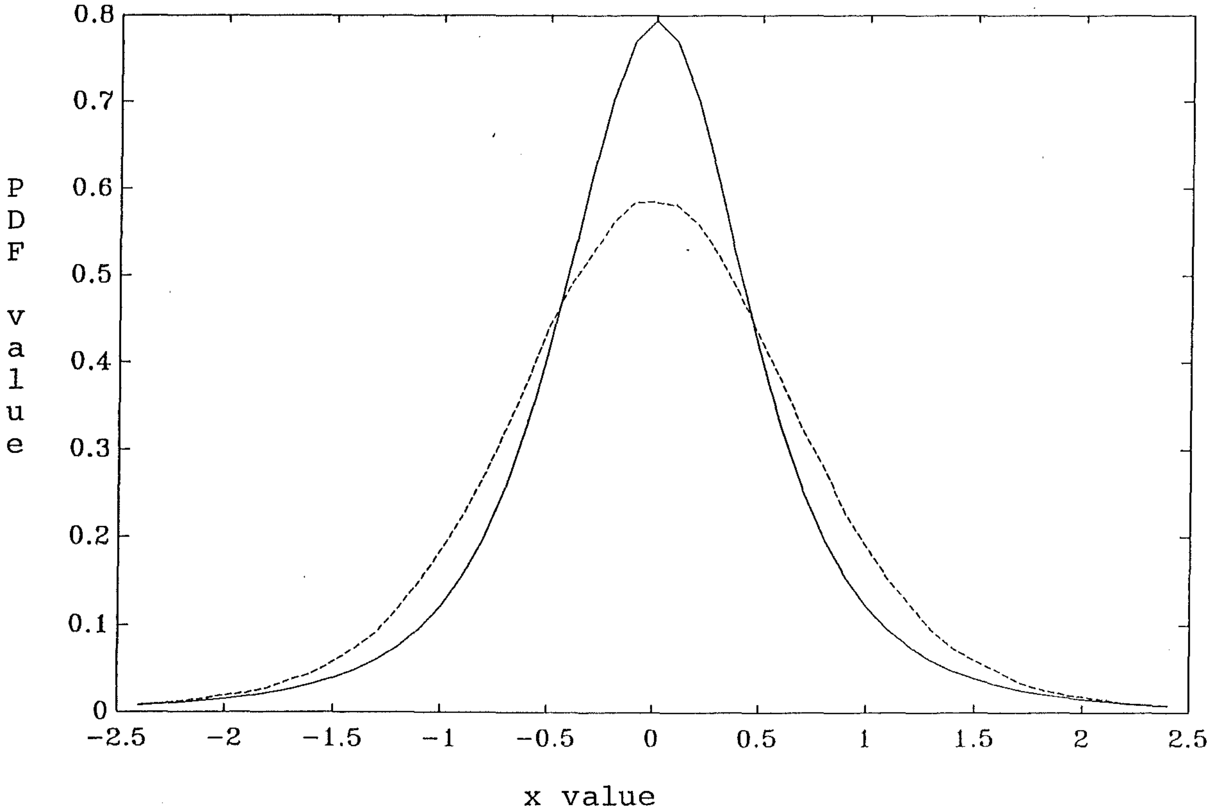
Consequently, when the disturbances are assumed to be independent, but are only uncorrelated, an inefficient estimator is used, and the inefficiency increases as K increases. Further, the wrong limit distribution is used as an approximation to the finite-sample distribution, which results in one assuming that the estimator is more located around the true parameter value than it actually is.

Consider now the case in which the disturbances are assumed to be jointly-distributed, but they are actually independently distributed. Then the OLS estimator, b, will be taken to be the correct maximum likelihood estimator to use. Although this estimator is unbiased, there are a number of consequences of using

Table 4.3: Empirical MSE's for $\hat{\beta}^{(I)}_{ML}$ and Actual MSE's for $b^{(I)}$ for N = 20

		K = 2	K = 5	K = 10
Empirical MSE $\hat{\beta}^{(I)}_{ML}$	v = 1	0.5370	1.7750	16.337
Actual MSE $b^{(I)}$		∞	∞	∞
Empirical MSE $\hat{\beta}^{(I)}_{ML}$	v = 2	0.3466	0.9085	5.9611
Actual MSE $b^{(I)}$		∞	∞	∞
Empirical MSE $\hat{\beta}^{(I)}_{ML}$	v = 3	0.2934	0.6933	4.4360
Actual MSE $b^{(I)}$		0.5349	0.9880	5.0691
Empirical MSE $\hat{\beta}^{(I)}_{ML}$	v = 5	0.2474	0.5376	0.28608
Actual MSE $b^{(I)}$		0.2972	0.6083	2.8162
Empirical MSE $\hat{\beta}^{(I)}_{ML}$	v = 10	0.2129	0.4449	2.1309
Actual MSE $b^{(I)}$		0.2229	0.4563	2.1121

FIGURE 4.2 Comparison of the Distribution of $b^{(I)}$ with its Incorrectly Assumed Asymptotic Distribution $b^{(D)}$ for v = 3, N = 20



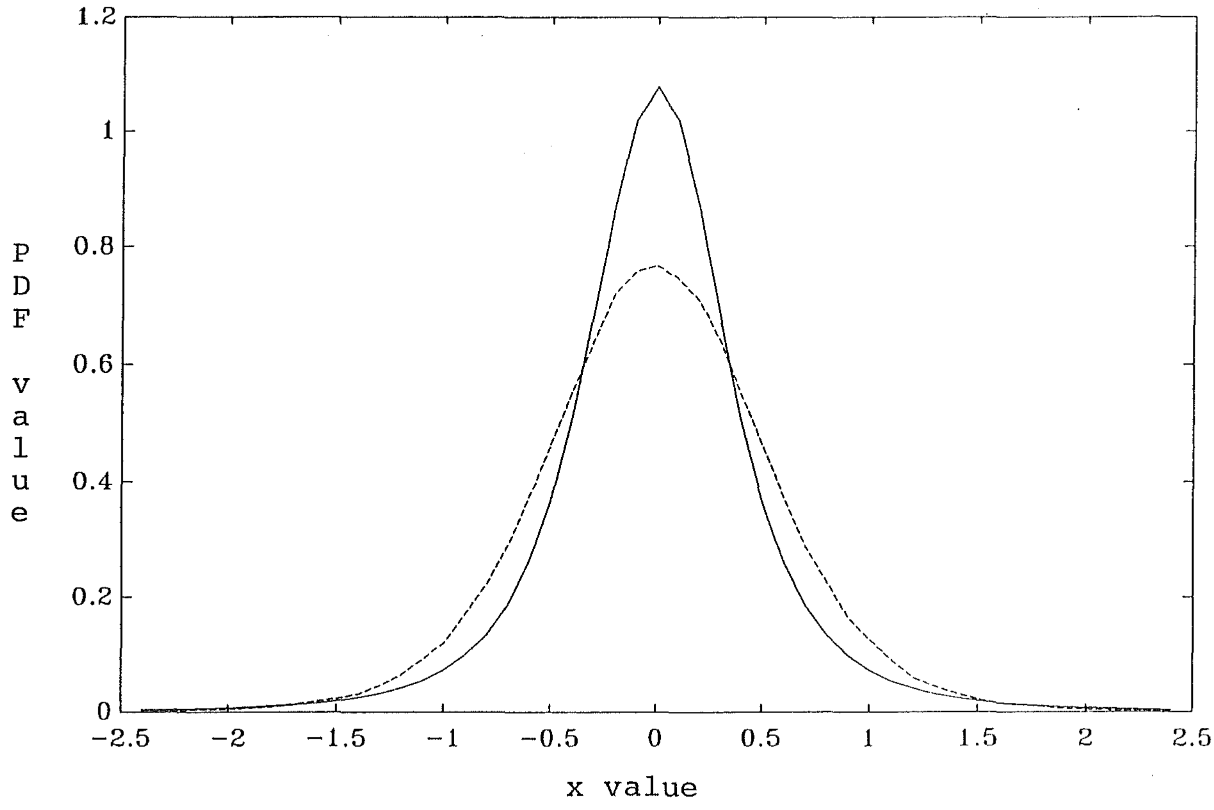
(i) K = 2

KEY

-- Empirical Distribution of $b^{(I)}$

\- Incorrectly Assumed Asymptotic Distribution given by $b^{(D)}$

FIGURE 4.2 (ii) K = 10



KEY

-- Empirical Distribution of $b^{(I)}$

- Incorrectly Assumed Asymptotic Distribution given by $b^{(D)}$

this estimator rather than the correct maximum likelihood estimator, $\hat{\beta}_{ML}$.[7]

In Table 4.3, actual MSE's are given for $b^{(I)}$, and empirically estimated MSE's for $\hat{\beta}^{(I)}_{ML}$, (these were taken from the relevant entries in Table 3.2), corresponding to N = 20 and numerous values of K. A comparison of these illustrates that for small K, b is substantially inefficient, but becomes more robust as K increases.

Further, while for each individual b_i, (i = 1,...,K), the correct variance will be estimated, the finite-sample distribution will be approximated by the distribution of $b_i^{(D)}$, which is Student-t with v degrees of freedom. The effect of this is illustrated in Figure 4.2, for a particular i (only one i is illustrated as the results are similar for the others), for v = 3, N = 20, K = 2 (i) and K = 10 (ii). Again, this density is estimated via the integration of the kernel density method with the naive Monte - Carlo method as described in Section 3, using iid student - t random variables. In this case however, the subroutine ELIM from Gerald and Wheatley (1984, p.144) is used to obtain the OLS inputs for (3.3). This subroutine solves a set of linear equations using the Gaussian elimination method. We can see that the actual distribution of $b_i^{(I)}$ has thinner tails than the incorrectly assumed distribution, which is to be expected as the actual distribution limits to the normal distribution. However, for the central part of the distribution, the use of the wrong

---

[7] The unbiasedness of b follows from properties of symmetrical parent distributions (see David (1970, p.105)).

limit distribution implies that one is much more confident that the estimator is located around the true parameter value.

Consequently, when the disturbances are assumed to be uncorrelated but are actually independent, for small K for fixed N, a substantially inefficient estimator will be used, although this estimator becomes robust as K increases. However, the wrong limit distribution is used to approximate the finite-sample distribution, which results in one assuming that the estimator is more located around the true parameter value than it actually is, at least for the central part of the distribution.

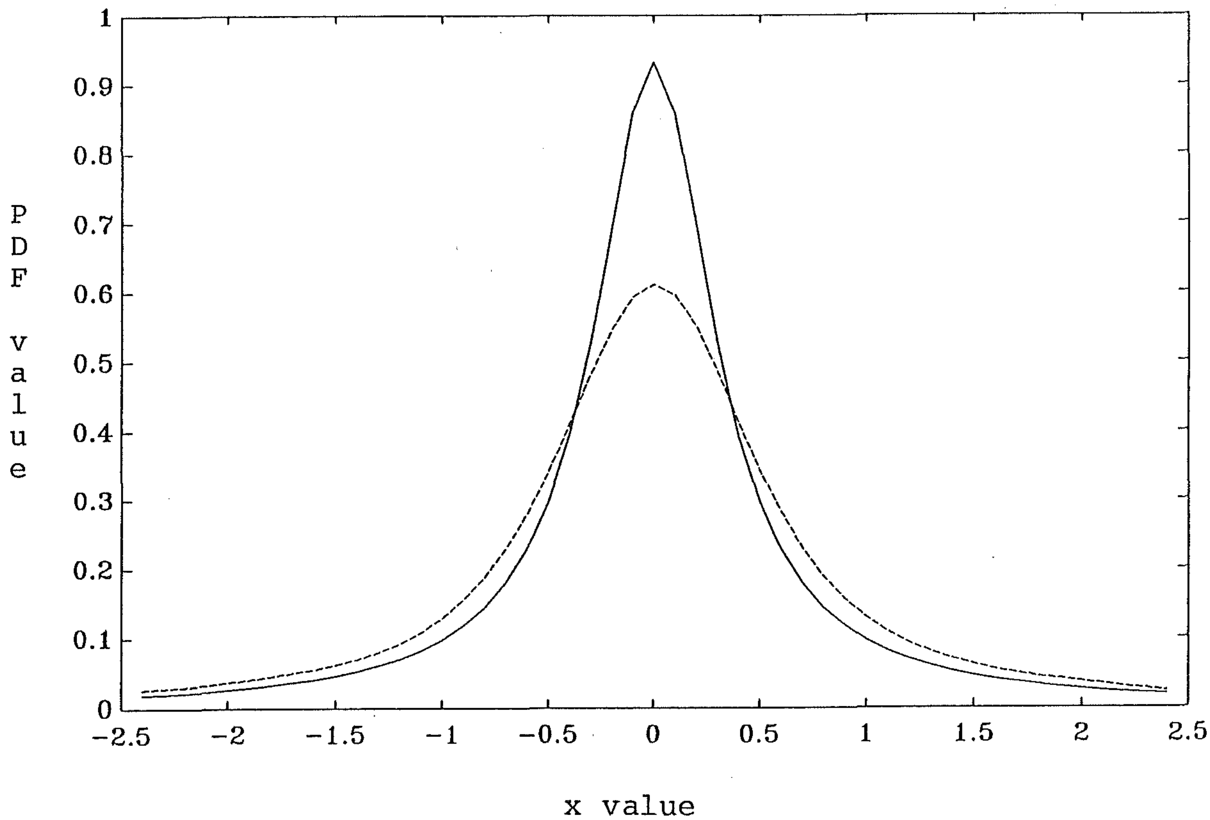
Infinite Variance: $v \leq 2$

As in the location-scale model, for the infinite variance distributions the statistical consequences of inappropriately using b or $\hat{\beta}_{ML}$ are even more serious. To illustrate this, we will assume the disturbances are Cauchy distributed.

When the disturbances are assumed to be multivariate Student-t distributed, b is assumed to be the correct maximum likelihood estimator to use. However, if the disturbances are actually iid Student-t distributed, then the correct maximum likelihood estimator is $\hat{\beta}_{ML}$. In Table 4.3,(where the appropriate values have been taken from Table 3.2 for $\hat{\beta}_{ML}$), we see that on the basis of MSE this results in a particularly inefficient estimator being used, as b has infinite MSE, whereas $\hat{\beta}_{ML}$, at least for moderately-sized regression models, has finite MSE. Further, the distribution of $b_i^{(I)}$ is standard Cauchy with scale $\Sigma|[X(X'X)^{-1}X']_i|$ for i = i,...,K and where $\Sigma|[X(X'X)^{-1}X']_i|$ denotes the sum of the absolute terms of the ith row of the matrix $[X(X'X)X']$ (see, for example, Johnston and Kotz (1970, p.157)). However, because the disturbances are assumed to be multivariate Student-t, it will be incorrectly assumed that the finite-sample distribution of $b_i^{(I)}$ is Cauchy with scale $\left((X'X)_i^{-1}/N\right)$, where $(X'X)_i^{-1}$ is the ith diagonal term of the matrix $(X'X)^{-1}$. Some examples of the consequences of this incorrect assumption are illustrated in Figure 4.3, (i) K = 2, and (ii) K = 10. In particular we can see from these figures that the estimator is thought to be substantially more located around the true parameter value than it actually is, especially as K increases.

Alternatively, if the disturbances are assumed to be iid Student-t, $\hat{\beta}_{ML}$ is assumed to be the correct maximum likelihood estimator, with asymptotic distribution given by (3.2). However, if the disturbances are actually multivariate Student-t distributed, then the finite-sample distribution of $\hat{\beta}_{ML}$ is Cauchy. Consequently, the normal distribution with finite variance, which is the assumed asymptotic distribution in (3.2), will be used to approximate the Cauchy distribution, which has infinite variance. Some examples of this incorrect approximation are illustrated in Figure 4.4, (i) and (ii). These densities were estimated via an integration of the kernel and naive Monte-Carlo methods as described in Section 3. From the figures we can see that the estimator is assumed to be more concentrated around the true parameter value than it actually is.

FIGURE 4.3 Comparison of the distribution of $\hat{\beta}_{ML}^{(D)}$ with its Incorrectly Assumed Asymptotic Distribution $\hat{\beta}_{ML}^{(I)}$ for v = 1, N = 20



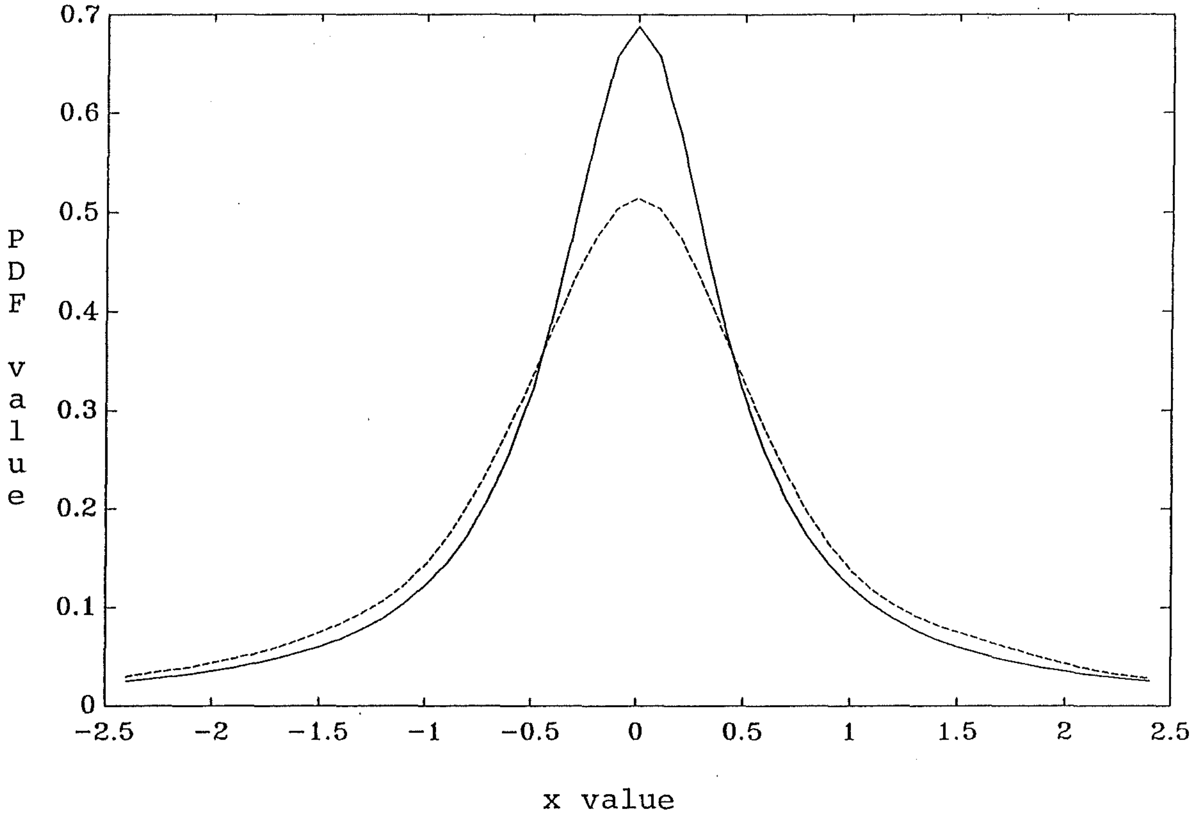
(i) K = 2

KEY

-- Empirical Distribution of $\hat{\beta}_{ML}^{(D)}$

\- Incorrectly Assumed Asymptotic Distribution given by $\hat{\beta}_{ML}^{(I)}$

FIGURE 4.3 (ii) K = 10

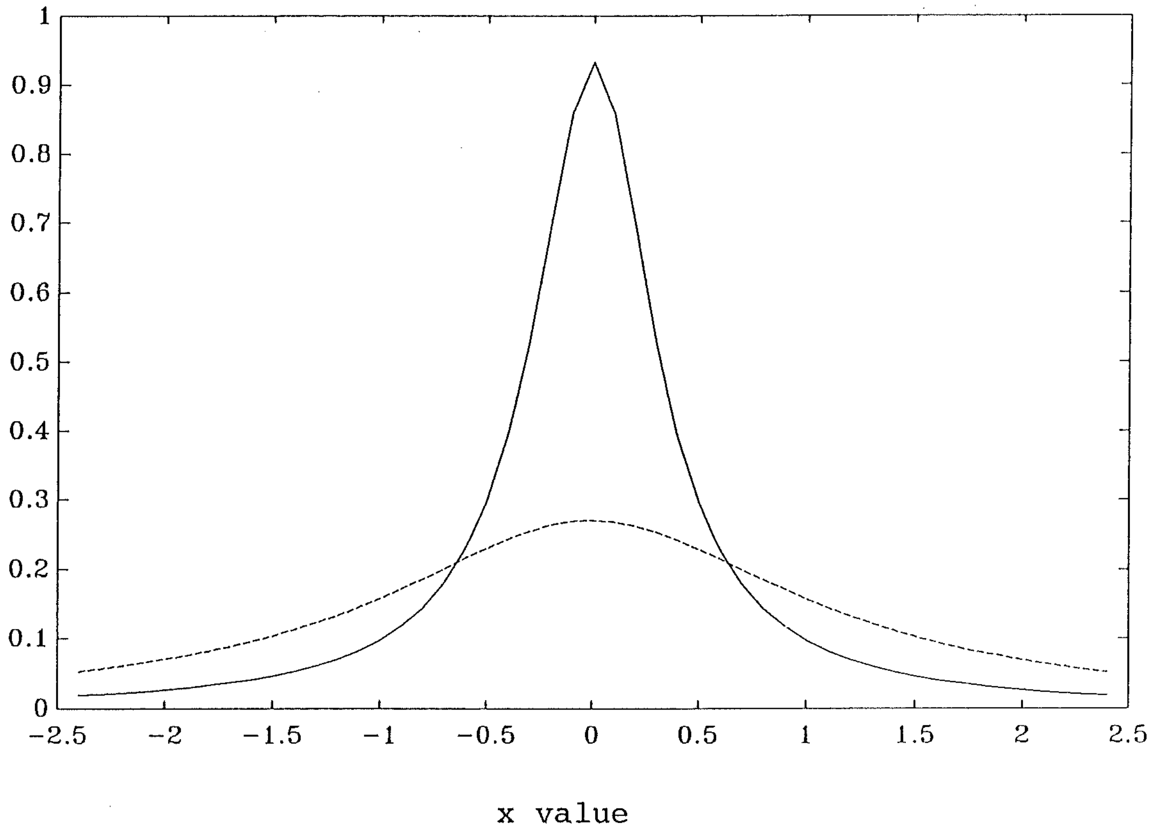


KEY

-- Empirical Distribution of $\hat{\beta}_{ML}^{(D)}$

- Incorrectly Assumed Asymptotic Distribution given by $\hat{\beta}_{ML}^{(I)}$

FIGURE 4.4 Comparison of the Distribution of $b^{(I)}$ with its Incorrectly Assumed Asymptotic Distribution $b^{(D)}$ for $v = 1$, $N = 20$



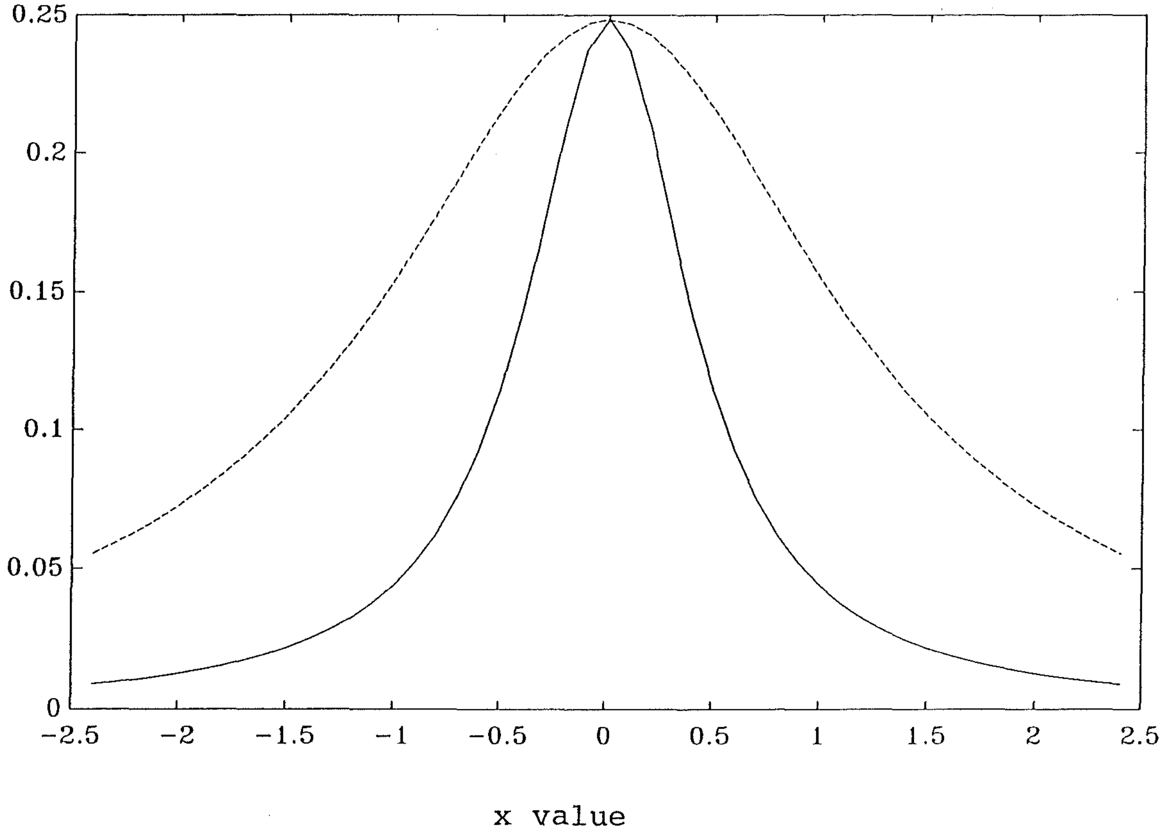
(i) K = 2

KEY

-- Empirical Distribution of $b^{(I)}$

\- Incorrectly Assumped Asymptotic Distribution given by $b^{(D)}$

FIGURE 4.4 (ii) K = 10



KEY

-- Empirical Distribution of $b^{(I)}$

- Incorrectly Assumed Asymptotic Distribution given by $b^{(D)}$

9.5 SOME FINAL COMMENTS

In nonnormal models a distinction needs to be made between multivariate distributed disturbances and iid distributed disturbances. In this chapter the importance of this distinction in finite-samples, has been illustrated for the maximum likelihood estimators of the regression coefficients in the general linear regression model, when the disturbances are Student-t distributed. This extends the results of Chapter 8, in which the location-scale model was assumed.

Properties of the maximum likelihood estimator of the regression coefficients when the disturbances are multivariate Student-t distributed, (i.e., the OLS estimator, (b), are well known, and more importantly, they are a simple generalization of those given for the location-scale model. However, similar properties for the maximum likelihood estimator of the regression coefficients, when the disturbances are iid Student-t (i.e., the robust estimator, ($\hat{\beta}_{ML}$), are not known and so are developed in this chapter; these properties are shown not to be a simple generalization of those given in the location-scale model. These properties are then used to consider the implications of misspecification. That is, to consider the implications of using the maximum likelihood estimator associated with one assumption, when in fact the other error assumption is correct. Although these implications depend on the number of regressors in the model, in general the consequences of making the wrong assumption are serious, with respect to the efficiency of the resulting estimator, and the use of the wrong limit distribution to approximate the finite-sample distribution.

However, before specification tests are developed to test for this distinction, we first discuss the implications of "jointness versus independence" in the nonnormal limited-information SEM. This is the topic of the next chapter.

CHAPTER 10

THE NONNORMAL LIMITED-INFORMATION SIMULTANEOUS EQUATIONS MODEL

10.1 INTRODUCTION

The nonnormal limited-information SEM provides a relatively new area of analysis as there are few published results on the effects of nonnormal disturbances in the limited-information SEM (e.g. Knight (1985b, 1986), Raj (1980), Donatos (1989)). However, the objective in this chapter is simply to combine the themes pursued in this thesis for the limited-information SEM and the nonnormal linear regression model.

To narrow the range of possible models to consider, attention is focussed only on the exactly-identified SEM. This model, although somewhat restrictive, is worthy of study because the finite-sample distribution of the estimator of the coefficient of the endogenous regressor, has a number of interesting features when the errors are normally distributed. In particular in this chapter, finite-sample properties of the LIMLK estimator of the coefficient of the one endogenous regressor in the exactly-identified SEM are considered. The LIMLK estimator is the LIML estimator assuming the covariance matrix of the reduced-form disturbances is known. It is considered here because in the cases when the LIMLK estimator is not equivalent to the LIML estimator, it is numerically easy to compute, and it is considered that the distribution of the two estimators will have similar features. We begin by first considering properties of the finite-sample distribution of this estimator when the reduced-form disturbances are normally distributed. In this case the LIMLK estimator reduces to LIML and TSLS, and a number of interesting properties of the resulting distribution are illustrated. Next these properties are examined when the assumption of normally-distributed reduced-form disturbances is widened to Student-t disturbances. In this case two assumptions are considered, these being, jointly-distributed Student-t reduced-form disturbances and iid Student-t reduced-form disturbances. Finally, the statistical consequences of distinguishing between these two assumptions are considered to determine how important it is to make this distinction by applying appropriate specification tests.

There are two sections in this chapter. Section 2 discusses the properties of the LIMLK estimator. Part (i) of this section assumes normally-distributed reduced-form disturbances, and Parts (ii) and (iii) concentrate on Student-t distributed reduced-form disturbances. Section 3 considers the statistical consequences of misspecifying the jointness versus iid Student-t distributed reduced-form disturbances.

10.2 EXACTLY-IDENTIFIED LIMITED-INFORMATION SEM

(i) Normally distributed disturbances

In the exactly-identified SEM with normally-distributed reduced-form disturbances, the TSLS and LIMLK estimators reduce to indirect least squares (ILS). Using the notation of Chapter 5, this estimator takes the form,

$$\hat{\alpha} = [X_2'y_2]^{-1}[X_2'y_1] \qquad (2.1)$$

where, in particular, X_2 is of dimension (N*1). In this case where there are only two-included endogenous variables, (2.1) reduces to a ratio of normal variables.

Ratios of normal random variates of the form,

$$z = (c_2+b)^{-1}(c_1+a) \qquad (2.2)$$

where a, b are nonnegative constants, and c_1 and c_2 are independent standard normal variables, have been studied by authors such as Geary (1930), Fieller (1932) and Marsaglia (1965). These studies are also relevant for the ILS estimator when there are only two included endogenous variables in the structural equation of interest, as in this case (2.1) can be written in the form of (2.2), where $a = \alpha\pi_{22}$ and $b = \pi_{22}$.

Geary (1930) gives the distribution of z when $a = b = 0$. This distribution can easily be seen to be the Cauchy distribution and further, as Phillips (1982, p.64) notes, it provides the leading term in the multiple series expansion of the more general case $a \neq 0$, $b \neq 0$. Therefore, the ILS estimator possesses no moments of finite-order, which implies that, in general, its distribution will have "fat tails". Fieller (1932) gives the following expression for the pdf of z in this case.

$$pdf(z) = \frac{1}{\pi}\frac{1}{1+z^2}\exp\left(-\frac{1}{2}(a^2+b^2)\right)$$

$$+ \exp\left(-\frac{1}{2}\frac{(a-bz)^2}{1+z^2}\right)\frac{b+az}{(1+z^2)^{3/2}}\int_0^{p*}\exp\left(-\frac{1}{2}y^2\right)dy, \qquad (2.3)$$

where, $p* = \frac{(b+az)}{(1+z^2)^{1/2}}$ This pdf depends only on the parameter values a and b, which for the ILS estimator correspond to δ^2 and α, as is given in Chapter 5, equations (5.3.4) and (5.3.5) respectively. Phillips (1982, p.63 eqn. 3.78) gives the form of (2.3) for the ILS estimator. Marsaglia (1965) gives the limiting distribution of (2.3) and for the ILS estimator (see also Anderson (1982, p.1015)) this is equal to,

$$\sqrt{\frac{\delta^2}{1+\alpha^2}}\,(\hat{\alpha}-\alpha) \rightarrow N(0,1) \quad \text{as} \quad \delta^2 \rightarrow \infty . \qquad (2.4)$$

Table 2.1 presents a number of points of the DF of the normalized ILS estimator for different δ^2 and for two parameter values α = 0.5 and 5.33.[1] These points are calculated using the method given in Chapter 5, Section 3, and are useful in determining the approach of the density of the normalized ILS estimator to the standardized normal distribution. Although this depends upon the

---

[1] Only points on the right hand side of the distribution are presented as the approach on the left hand side was very similar for the δ^2 chosen.

Table 2.1: Points of the Distribution Function of $\hat{\alpha}$ in the Exactly-Identified Limited-Information SEM with Normally-Distributed Reduced-form Disturbances, α = 0.5 and 5.33 and Various δ^2.

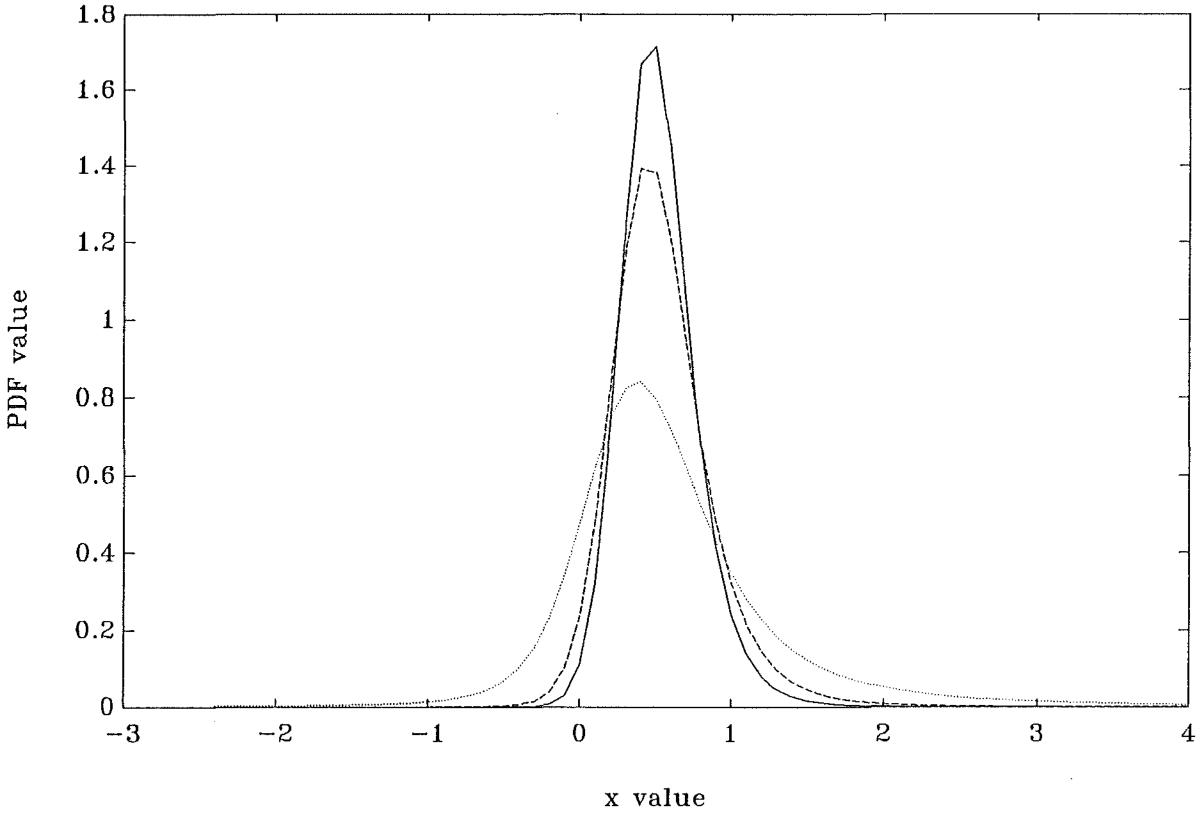
α = 0.5

	δ^2=25	δ^2=100	δ^2=1000	Normal
60%	0.26265	0.25933	0.25688	0.255466
70%	0.55739	0.54289	0.53267	0.5301033
80%	0.92617	0.87962	0.85496	0.8469008
90%	1.4998	1.3803	1.3132	1.282630
95%	2.0543	1.8055	1.6941	1.645
99%	3.3184	2.6833	2.4289	2.326

α = 5.33

	δ^2=25	δ^2=100	δ^2=1000	Normal
60%	0.26879	0.26227	0.25780	0.255466
70%	0.59083	0.55777	0.53832	0.5301033
80%	1.0179	0.92562	0.87049	0.8469008
90%	1.7242	1.4717	1.3406	1.282630
95%	2.4489	1.9691	1.7404	1.645
99%	4.3095	3.0101	2.5024	2.326

FIGURE 2.1 Distributions of Maximum Likelihood Estimator in Exactly-Identified SEM with Normally Distributed Reduced-form Disturbances Corresponding to $\alpha = 0.5$



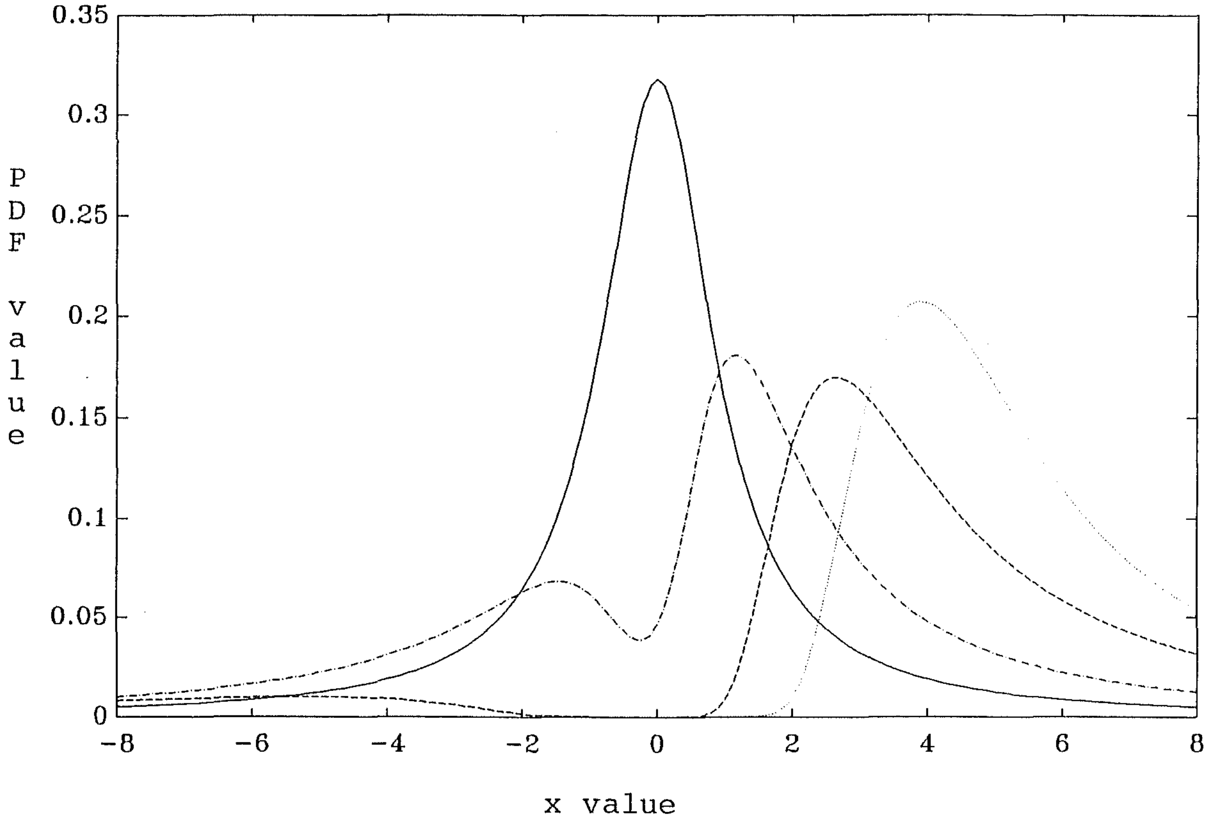
KEY

: $\delta^2 = 5.0$

-- $\delta^2 = 15.0$

– δ^2 – 23.0

FIGURE 2.2 Distributions of Maximum Likelihood Estimator in Exactly-Identified SEM with Normally Distributed Reduced-form Disturbances Corresponding to $\alpha = 5.33$



KEY

- $-$ $\delta^2 = 0.0001$
- $-\cdot$ $\delta^2 = 0.1410$
- $--$ $\delta^2 = 1.0$
- $\cdot\cdot$ $\delta^2 = 4.0$

values of α and δ^2, in general, even though the moments of the ILS estimator are not finite, the standard normal distribution is a good approximation to its actual distribution. Hence the usual methods with asymptotic standard deviation give reasonable inference.

Marsaglia (1965) also presents an interesting numerical analysis from which it is concluded that the density of z is unimodal or bimodal, according to the value a takes, and in particular, when a $\gtrsim$ 2.33 (~ implies this result is based on asymptotic behaviour see, for example, Marsaglia (1965, p.197)), the density is bimodal, although one of the modes may be insignificant. Applying this result to the ILS estimator means that its distribution will be bimodal when,

$$\alpha\pi_{22} \geq 2.255 = \delta^2 \geq \frac{5.09}{\alpha^2} \qquad (2.5)$$

Consequently, for example, as $\alpha \to \infty$ the distribution will be bimodal for all values of the noncentrality parameter whereas as $\alpha \to 0$, the distribution should always be unimodal.

To determine the significance of the bimodality, several densities for the ILS estimator are illustrated in Figures 2.1 and 2.2 for $\alpha = 0.5$ and 5.33 respectively. These values are chosen to represent a small and large value of α respectively. Furthermore, they are calculated by first finding the points of the DF using the method in Chapter 5, Section 3, and then the pdf is obtained via numerical differentiation. For $\alpha = 0.5$, the distribution of the ILS estimator can be considered to be unimodal whereas for $\alpha = 5.33$, prominent bimodality occurs, but is only a feature of the distribution for very small δ^2. In general, the distribution locates around the true parameter value very quickly.

Consequently, even though the density in (2.3) is an interesting example in which maximum likelihood estimators may have bimodal distributions for certain parameter values, the feature is only prominent for relative large α values and very small noncentrality parameters and, in general, the distribution of the ILS estimator tends to be well approximated by the normal distribution in finite-samples, even though the ILS estimator has no moments of finite order.

Next, the assumption of normally-distributed reduced-form disturbances is widened to Student-t disturbances and properties of the LIMLK estimator are developed.

(ii) Dependent Student-t Disturbances

In this case, again assuming there are two-included endogenous variables in the structural equation of interest, as in (5.3.1), the distribution of the reduced-form disturbances (v_1, v_2) is given by,

$$\text{pdf}(v_{11} \ldots v_{1N} v_{21} \ldots v_{2N}) =$$

$$\frac{\Gamma\left(\frac{v+2N}{2}\right)}{\Gamma\left(\frac{v}{2}\right)(v\pi)^N}\left[1+\frac{1}{v}(v_{11}^2+\ldots+v_{1N}^2+v_{21}^2+\ldots+v_{2N}^2)\right]^{-\left(\frac{v+2N}{2}\right)}, \quad (2.6)$$

and it is easily shown that the LIMLK estimator takes the same form as when the disturbances are normally distributed. This can be shown quite simply in the exactly-identified SEM by using the relationship between the reduced- form and structural parameters. In particular we have,

$$\alpha = \pi_{12}/\pi_{22} \quad . \tag{2.7}$$

Using the invariance property of maximum likelihood, this implies that the maximum likelihood estimator of α is,

$$\hat{\alpha} = \hat{\pi}_{12}/\hat{\pi}_{22} \quad . \tag{2.8}$$

The results of Sutradhar and Ali (1986) show that the maximum likelihood estimators of the reduced-form parameters are OLS, and this is all that is needed to establish the result that $\hat{\alpha}$ is the same as for normally- distributed (ν_1,ν_2). Consequently, (2.8) takes the form,

$$\hat{\alpha} = (X_2'y_2)^{-1}(X_2'y_1) \quad , \tag{2.9}$$

which is a ratio of correlated bivariate Student-t variables.

In a similar manner to Marsaglia (1965), Press (1969) considers the distribution of the ratio,

$$z^* = (c_1+b)^{-1}(c_2+a) \quad ,$$

where c_1 and c_2 have a bivariate-t distribution, and this ratio, of course, includes (2.9) with $a = \alpha\Pi_{22}$ and $b = \Pi_{22}$. In particular, he gives the following expression for the pdf of z^*.

$$\text{pdf } (z^*) = \frac{k_1}{1+z^{*2}} \left\{1 + \frac{k_2 q}{q^{*v+1}}\left[2F_{v+1}\left(\frac{q\sqrt{v+1}}{q^*}\right) - 1\right]\right\}, \tag{2.10}$$

where

$$K_1 = \frac{1}{\pi\left(1+\frac{a^2+b^2}{v}\right)^{v/2}}, \quad K_2 = \frac{\pi^{\frac{1}{2}} v^{(v+2)/2}\Gamma\left(\frac{v+1}{2}\right)}{2\Gamma\left(\frac{v+2}{2}\right)\left(1+\frac{a^2+b^2}{v}\right)^{-v/2}}$$

$$q = \frac{-az^*+b}{(1+z^{*2})^{\frac{1}{2}}}, \qquad q^* = (a^2 + b^2 + v - q^2)^{\frac{1}{2}}$$

$$F_n(t) = \frac{1}{\sqrt{n\pi}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \int_{-\infty}^{t} \frac{dw}{\left(1+\frac{w^2}{2}\right)^{\frac{n+1}{2}}} .$$

A number of properties of the ILS estimator are the same as when (v_1, v_2) are normally distributed. In particular, the key parameters of the density are the same as is given in (2.3) and the estimator possesses no moments of finite order (see, for example, Press (1969) Knight (1986)). However, using the results of Press (1969), the asymptotic distribution is not normal but Student-t so that,

$$\sqrt{\frac{\delta^2}{1+\alpha^2}}\,(\hat{\alpha}-\alpha) \rightarrow MT_1(0,1), \quad \text{as} \quad \delta^2 \rightarrow \infty ,$$

with asymptotic variance $\frac{v}{v-2}$ for $v > 2$. Table 2.2 illustrates the approach of the standardized distributions to their limit distributions for various v, δ^2 and $\alpha = 0.5$ and 5.33. These values were obtained by calculating a number of points of the pdf using (2.10) and then numerically integrating to obtain the DF. As in the case of normally distributed (v_1, v_2), the limiting distribution provides a good approximation to the finite-sample distribution,

Table 2.2: Points of the Distribution Function of $\hat{\alpha}$ in the Exactly-Identified Limited-Information SEM with Reduced-form Disturbances Distributed as in (2.6), $\alpha = 0.5$ and 5.33 and Various δ^2.

	$\alpha = 0.5$			$\alpha = 5.33$			
v=1							
	δ^2=25	δ^2=100	δ^2=1000	δ^2=25	δ^2=100	δ^2=1000	Cauchy
60%	0.24041	0.28651	0.31778	0.13693	0.23600	0.30057	0.31831
70%	0.63526	0.68883	0.71979	0.52169	0.63168	0.70358	0.71520
80%	1.2812	1.3423	1.3830	1.1575	1.2805	1.3525	1.3602
90%	2.9283	3.0222	3.0777	2.8108	2.9838	3.0265	3.0662
95%	5.9985	6.1196	6.2236	6.0695	6.1050	6.0939	6.3985
99%	31.241	31.576	31.145	30.034	30.287	29.532	32.197

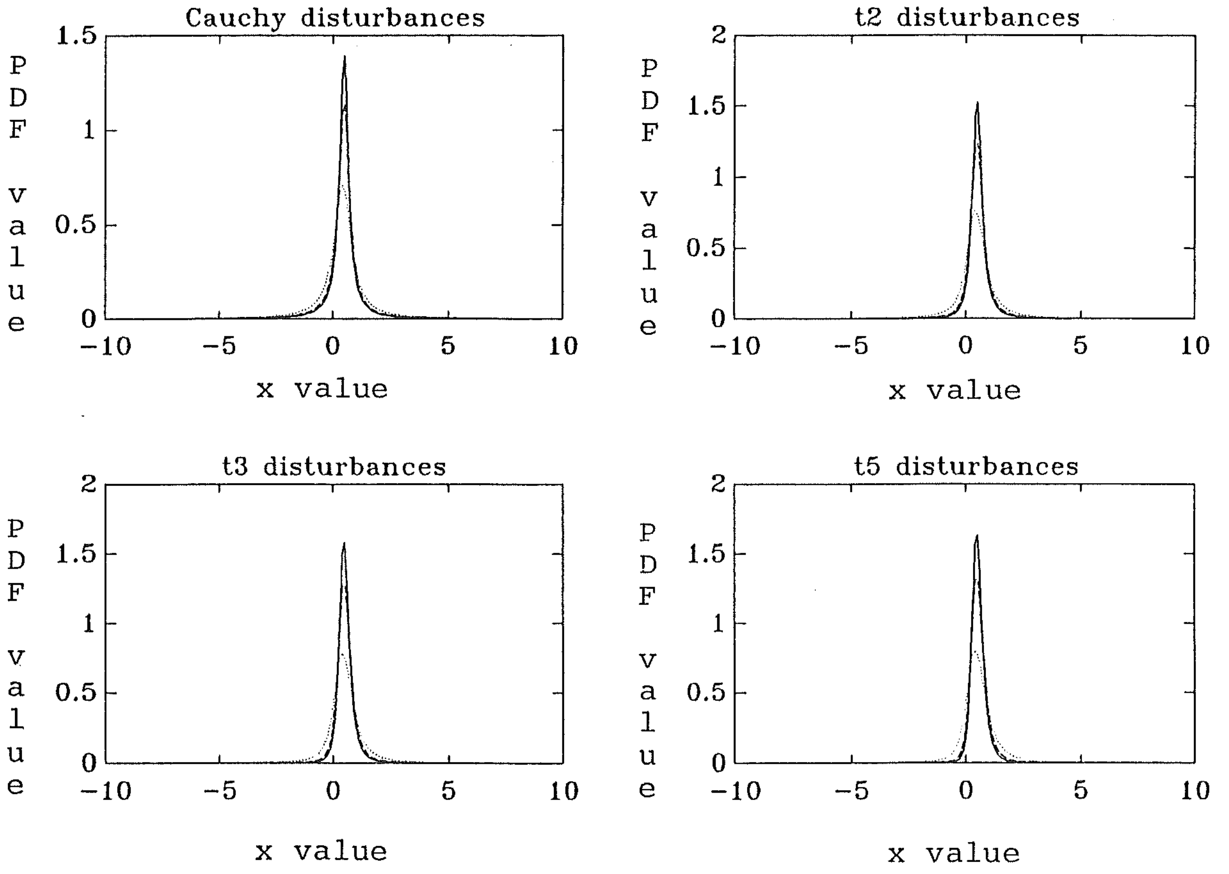
	$\alpha = 0.5$			$\alpha = 5.33$			
v=2							
	δ^2=25	δ^2=100	δ^2=1000	δ^2=25	δ^2=100	δ^2=1000	Cauchy
60%	0.27260	0.29383	0.29863	0.24485	0.28731	0.29502	0.28404
70%	0.62077	0.63302	0.63103	0.61038	0.63785	0.62702	0.61725
80%	1.1224	1.1102	1.0825	1.1763	1.1538	1.0951	1.0639
90%	2.1786	2.0548	1.9540	2.4641	2.2408	2.0090	1.8893
95%	3.8042	3.3758	3.0874	4.5815	3.8077	3.1957	2.9351
99%	15.048	10.680	7.8185	19.786	13.494	8.8381	6.9584

Table 2.2 continued

	$\alpha = 0.5$			$\alpha = 5.33$			
v=3							
	δ^2=25	δ^2=100	δ^2=1000	δ^2=25	δ^2=100	δ^2=1000	Cauchy
60%	0.27952	0.28781	0.28691	0.26960	0.28250	0.28020	0.27355
70%	0.60416	0.60384	0.59406	0.62805	0.61572	0.59477	0.58387
80%	1.0609	1.0275	0.99644	1.1599	1.0763	1.0098	0.98337
90%	1.9352	1.7819	1.6841	2.2626	1.9256	1.7150	1.6426
95%	3.1230	2.6900	2.4534	3.8935	3.0223	2.5373	2.3498
99%	9.6094	6.2814	5.0255	13.414	7.7844	5.3318	4.5500

	$\alpha = 0.5$			$\alpha = 5.33$			
v=5							
	δ^2=25	δ^2=100	δ^2=1000	δ^2=25	δ^2=100	δ^2=1000	Cauchy
60%	0.26697	0.26173	0.26564	0.27528	0.27383	0.26919	0.26381
70%	0.58174	0.56849	0.55634	0.62259	0.59355	0.57063	0.55853
80%	1.0017	0.96015	0.92868	1.1052	1.0063	0.94319	0.92556
90%	1.7480	1.5911	1.5062	2.0220	1.7071	1.5322	1.4741
95%	2.6245	2.2519	2.0756	2.2534	2.5098	2.1471	2.0146
99%	6.0475	4.1986	3.5601	8.7804	5.0229	3.7640	3.3705

FIGURE 2.3 Distributions of Maximum Likelihood Estimator in Exactly-Identified SEM with Student-t Distributed Reduced-form Disturbances given by (2.6) and Corresponding to $\alpha = 0.5$

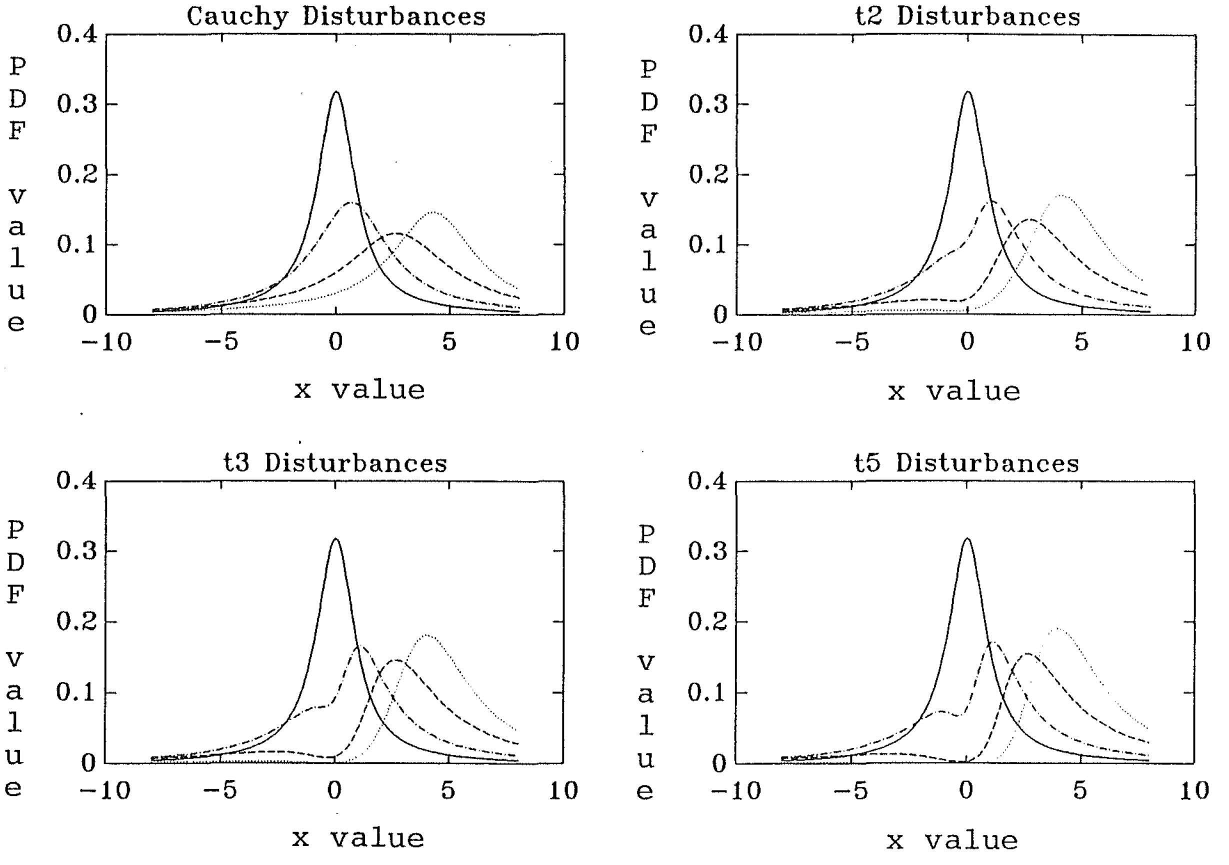


KEY : $\delta^2 = 5.0$

-- $\delta^2 = 15.0$

\- $\delta^2 = 23.0$

FIGURE 2.4 Distributions of Maximum Likelihood Estimator in Exactly-Identified SEM with Student-t Distributed Reduced-form Disturbances given by (2.6) and Corresponding to $\alpha = 5.33$



KEY

- – $\delta^2 = 0.0001$
- –· $\delta^2 = 0.1410$
- –– $\delta^2 = 1.0$
- ·· $\delta^2 = 4.0$

even for small v and δ^2, and hence the usual methods with asymptotic standard deviation give reasonable inference.

From a numerical analysis, Press (1969, p.245) concludes the finite- sample distribution of z^* is similar to that of the ratio of normal variates. Figures 2.3 and 2.4 illustrate a number of distributions for the ILS estimator for different v and $\alpha = 0.5$ and 5.33 respectively. These Figures are calculated using (2.10). Comparing these illustrations to those of Figures 2.1 and 2.2, the conclusion reached by Press (1969) seems valid, although bimodality does not tend to be as pronounced for small v.

Therefore, most of the properties obtained for the ILS estimator with normally-distributed disturbances remain valid when these disturbances have a joint multivariate Student-t pdf. The only major difference between the two error assumptions is that with multivariate Student-t errors the limiting distribution of the ILS estimator is Student-t.

(iii) Independent Student-t Disturbances

For the structural equation (5.3.1) with corresponding reduced-form (5.3.2), if (v_{1n},v_{2n}) is assumed to be independently distributed bivariate Student-t for all n, we have,

$$\text{pdf}(v_{1n},v_{2n}) = \frac{\Gamma\left(\frac{v+2}{2}\right)}{\Gamma\left(\frac{v}{2}\right)v\pi}\left[1 + \frac{1}{v}(v_{1n}^2+v_{2n}^2)\right]^{-\left(\frac{v+2}{2}\right)} , \qquad (2.11)$$

so that the joint-distribution of (v_1,v_2) is given by,

$\text{pdf}(v_{11}\ldots v_{1N}v_{21}\ldots v_{2N}) =$

$$\left[\frac{\Gamma\left(\frac{v+2}{2}\right)}{\Gamma\left(\frac{v}{2}\right)v\pi}\right]^N \prod_{i=1}^{N}\left[1 + \frac{1}{v}(v_{1n}^2+v_{2n}^2)\right]^{-\left(\frac{v+2}{2}\right)} . \qquad (2.12)$$

The LIMLK estimator of α can be obtained by first maximizing the log- likelihood equation for the reduced-form parameters π_{21} and π_{22}, that is, maximizing the expression,

$$-\left(\frac{v+2}{2}\right)\prod_{i=1}^{N} \text{Log}\left[1 + \frac{1}{v}(\tilde{v}_{1n}^2 + \tilde{v}_{2n}^2)\right] \qquad (2.13)$$

where,

$$\tilde{v}_{1n} = y_{1n}-X_{1n}\pi_{11}-X_{2n}\pi_{21} \ , \ \tilde{v}_{2n} = y_{2n}-X_{1n}\pi_{12}-X_{2n}\pi_{22} \ ,$$

and then, secondly, by using the relationship between the structural and reduced-form parameters in the exactly-identified SEM, to obtain,

$$\tilde{\alpha} = \tilde{\pi}_{21}/\tilde{\pi}_{22} \ . \qquad (2.14)$$

The log-likelihood equation (2.13) needs to be solved numerically. Recently, Koenker and Portnoy (1988) have considered classes of robust estimators for this type of model.[2] In particular, they show that the usual LAE, although asymptotically inefficient, is asymptotically normally distributed, which suggests that it provides a useful starting value. Furthermore, since to date no analytical expression exists for the probability density function of $\tilde{\alpha}$, preliminary numerical analysis was required to determine the key parameters of the density of $\tilde{\alpha}$. However, this analysis indicated that the key parameters of the density are the same as is given in (2.3).

Kelejian and Prucha (1984) show that the asymptotic distribution of $\tilde{\alpha}$ for $v \geq 5$ is,

$$\mathrm{Sqrt}\left(\frac{v+2}{v+4}\,\frac{\delta^2}{1+\alpha^2}\right)(\tilde{\alpha}-\alpha) \rightarrow N(0,1) \text{ as } \delta^2 \rightarrow \infty . \qquad (2.15)$$

Table 2.3 contains a number of points of the distribution function of (2.14) for $v = 5$, $\alpha = 0.5$ and 5.33 and $\delta^2 = 25$, 100, 1000. These points are obtained via Monte-Carlo methods. In particular the empirical DF is estimated, (see (4.2.1)), with appropriate bivariate Student - t random numbers generated using the relationship, (see e.g. (2.3.4)),

$$X_i = Z_i\left(\frac{\chi^2}{v}\right)^{-\frac{1}{2}} \qquad i = 1, \ldots K \qquad (2.16)$$

---

[2] In fact, Koenker and Portnoy (1988) consider classes of robust estimators for Seemingly Unrelated Regression Models. However, the reduced- form of a SEM is just a special class of these (see, for example, Srivastava and Giles (1987, p.6)).

where $Z_1, \ldots Z_K$ are K, (K in this case equals 2), independent standard normal variables and χ^2 is an independent chi-square variable with v degrees of freedom. From Table 2.3 it is seen that the finite-sample distributions are well approximated by the asymptotic distribution N(0,1). Similar points are also given for v = 1, 2 and 3, although these are values of v that are not covered in the proof of Kelejian and Prucha (1984). However, comparing these points with the appropriate values from N(0,1) it is conjectured that (2.15) is, in fact, the asymptotic distribution for all v. Furthermore, (2.15) provides a good approximation to the finite-sample distributions for these v, except for small δ^2 and v = 1, which tends in these cases to have very "fat tails".

Graphs of the distribution of $\tilde{\alpha}$ corresponding to $\alpha = 0.5, 5.33$ for various v and δ^2 are illustrated in Figures (2.5) and (2.6) respectively. The densities illustrated in these Figures are estimated via the integration of the kernel density estimator with the naive Monte-Carlo method. The kernel estimate at point X is equal to,

$$\hat{pdf}(X) = \frac{1}{N^*h(N^*)} \sum_j k\left[\frac{X - X_j}{h(N^*)}\right] \qquad (2.17)$$

where k[.] is the standard N(0,1) density. The window width h(N*) is chosen using the interactive approach of Tapia and Thompson (1978). In all cases this approach led to the use of a window width between 0.02 and 0.09. N* is simply the number of replications in the simulation experiment, and is chosen using the bound of estimation. For example, the results of Parzen (1962) and Cacoullos (1966) imply,

$$\left(N{*}h^{m}(N{*})\right)^{\frac{1}{2}}\left[\hat{pdf}(x) - E\left(\hat{pdf}(x)\right)\right] \sim N\left(0, pdf(x)\int K^{2}\right) \qquad (2.18)$$

holds. The result given in (2.18) can be achieved if $\left(N{*}\ h^{m}(N{*})\right)^{\frac{1}{2}}{*}$ $Bias\left[\hat{pdf}(x)\right]$ tends to zero asymptotically since,

$$\left(N{*}h^{m}(N{*})\right)^{\frac{1}{2}}\left[\hat{pdf}(x) - pdf(x)\right] = \left(N{*}h^{m}(N{*})\right)\left[\hat{pdf}(x) - E\left(\hat{pdf}(x)\right)\right]$$

$$+ \left(N{*}h^{m}(N{*})\right)^{\frac{1}{2}}Bias\left[\hat{pdf}(x)\right]$$

Ullah (1988, p.642) shows that $Bias\left[\hat{pdf}(x)\right]$ is proportional to $h^{2}(N{*})$. This implies that if $N{*}h^{(4+m)/2}(N{*})$ tends to zero asymptotically then (2.18) holds. Therefore, for the normal kernel $\frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}y^{2})$, the 99% asymptotic confidence interval for $\hat{pdf}(X)$ is given by,

$$\hat{pdf}(X) \pm 2.58\left[\frac{\hat{pdf}(X)}{2N{*}h\sqrt{\pi}}\right]^{\frac{1}{2}},$$

so that B is given by,

$$B = 2.58\left[\frac{\hat{pdf}(X)}{2N{*}h(N{*})\pi}\right]^{\frac{1}{2}}.$$

Table 2.3: Points of the Distribution Function of $\tilde{\alpha}$ in the Exactly-Identified Limited-Information SEM with Reduced-form Disturbances Distributed as (2.12), α = 0.5 and 5.33 and Various δ^2.

	$\alpha = 0.5$			$\alpha = 5.33$			
v=1							
	δ^2=25	δ^2=100	δ^2=1000	δ^2=25	δ^2=100	δ^2=1000	Normal
60%	0.26088	0.26366	0.26549	0.27049	0.27005	0.26369	0.25547
70%	0.57163	0.55658	0.55000	0.61162	0.57664	0.54971	0.53010
80%	0.99046	0.93255	0.90903	1.1022	0.97887	0.90342	0.84690
90%	1.7387	1.5319	1.4572	2.0866	1.6785	1.4644	1.28263
95%	2.6404	2.1384	1.9623	3.3645	2.4145	1.9952	1.6450
99%	6.3461	3.9129	3.1764	9.3176	4.7121	3.3461	2.3260

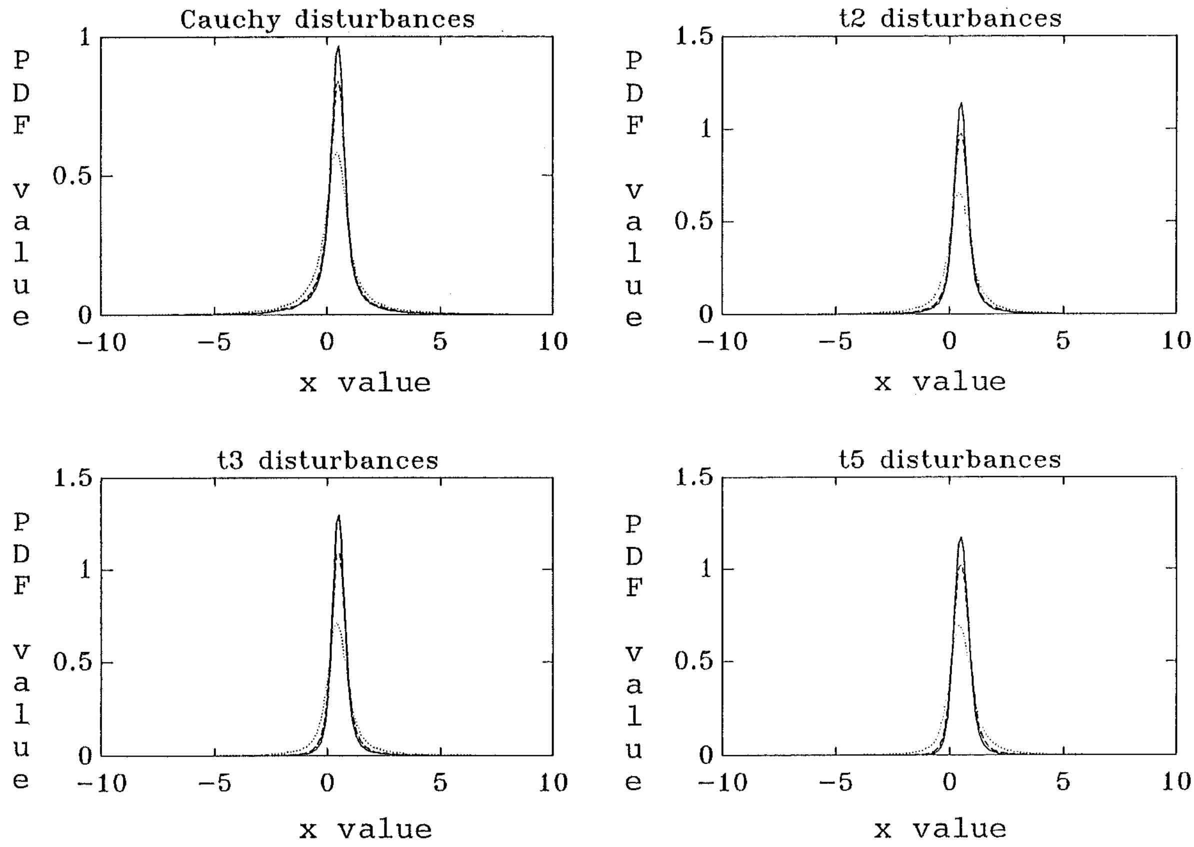
	$\alpha = 0.5$			$\alpha = 5.33$			
v=2							
	δ^2=25	δ^2=100	δ^2=1000	δ^2=25	δ^2=100	δ^2=1000	Normal
60%	0.26767	0.26338	0.25969	0.27274	0.26465	0.25934	0.25547
70%	0.57062	0.54520	0.53491	0.60480	0.56615	0.54184	0.53010
80%	0.96609	0.89471	0.86492	1.0657	0.94926	0.88200	0.84690
90%	1.6448	1.4320	1.3429	1.8785	1.5411	1.3738	1.28263
95%	2.3577	1.9067	1.7585	2.8370	2.1302	1.8187	1.6450
99%	4.7057	3.0127	2.6072	5.9174	3.5108	2.7387	2.3260

Table 2.3 continued

	$\alpha = 0.5$			$\alpha = 5.33$			
v=3							
	δ^2=25	δ^2=100	δ^2=1000	δ^2=25	δ^2=100	δ^2=1000	Normal
60%	0.26635	0.26338	0.25969	0.27274	0.26465	0.25934	0.25547
70%	0.56557	0.54520	0.53491	0.60480	0.56615	0.54184	0.53010
80%	0.95315	0.89471	0.86492	1.0657	0.94926	0.88200	0.84690
90%	1.6044	1.4320	1.3429	1.8785	1.5411	1.3738	1.28263
95%	2.2618	1.9067	1.7585	2.8370	2.1302	1.8187	1.6450
99%	4.2261	3.0127	2.6072	5.9174	3.5108	2.7387	2.3260

	$\alpha = 0.5$			$\alpha = 5.33$			
v=5							
	δ^2=25	δ^2=100	δ^2=1000	δ^2=25	δ^2=100	δ^2=1000	Normal
60%	0.25368	0.25128	0.24975	0.26435	0.25682	0.25146	0.25547
70%	0.55655	0.53874	0.52672	0.58863	0.55318	0.53865	0.53010
80%	0.94102	0.88659	0.85668	1.0335	0.92662	0.86537	0.84690
90%	1.5465	1.3982	1.3209	1.8151	1.5076	1.3517	1.28263
95%	2.1585	1.8556	1.7226	2.6626	2.0533	1.7724	1.6450
99%	3.8551	2.8602	2.5256	5.2774	3.3075	2.6357	2.3260

FIGURE 2.5 Distributions of Maximum Likelihood Estimator in Exactly-Identified SEM with Student-t Distributed Reduced-form Disturbances given by (2.12) and Corresponding to $\alpha = 0.5$

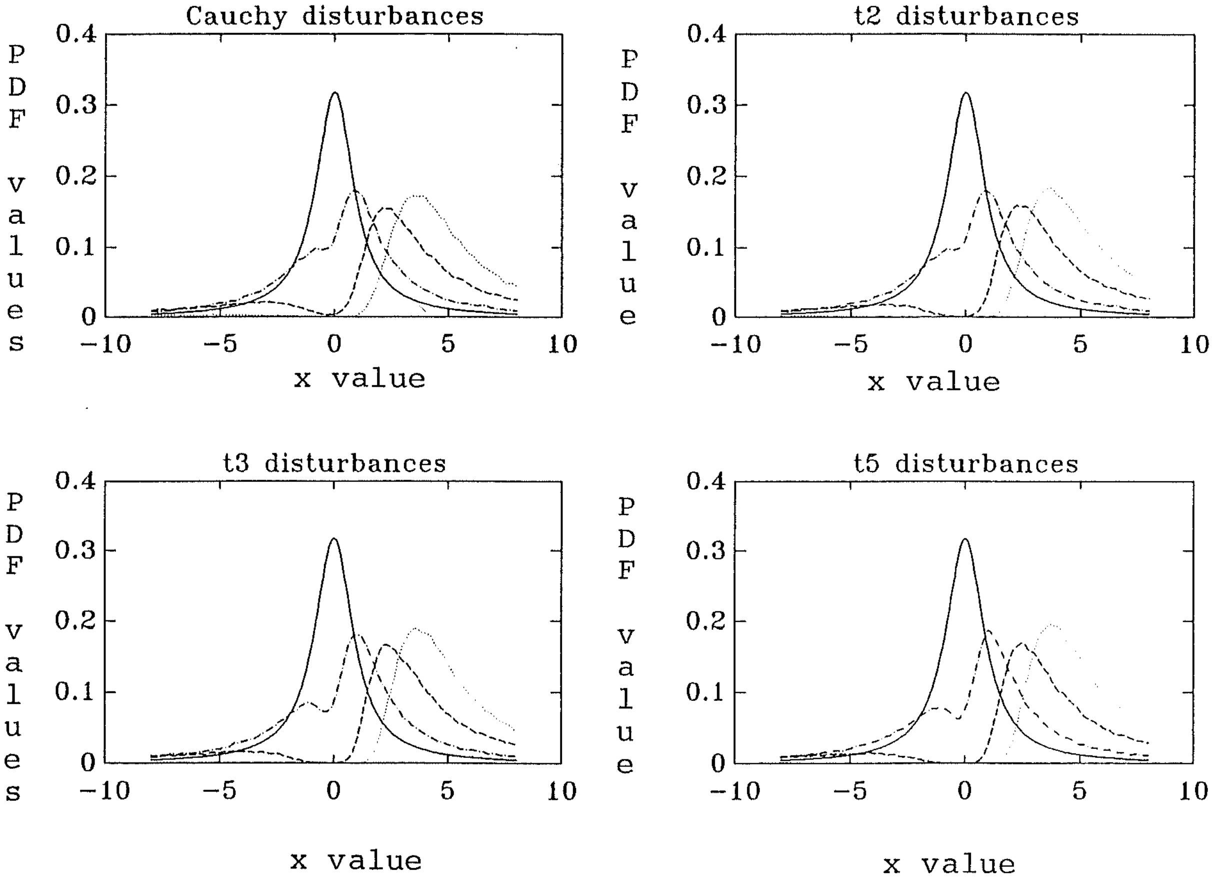


KEY : $\delta^2 = 5.0$

-- $\delta^2 = 15.0$

\- $\delta^2 = 23.0$

FIGURE 2.6 Distributions of Maximum Likelihood Estimator in Exactly-Identified SEM with Student-t Distributed Reduced-form Disturbances given by (2.12) and Corresponding to $\alpha = 5.33$



KEY
- $-$ $\delta^2 = 0.0001$
- $-\cdot$ $\delta^2 = 0.1410$
- $--$ $\delta^2 = 1.0$
- $\cdot\cdot$ $\delta^2 = 4.0$

N* is varied until B is less than 0.01 for all points at which the density is estimated. In all experiments, N* varies between 60,000 and 90,000 replications[3]. The input of X_j in (2.17) involves numerically maximizing the likelihood function (2.16). Two algorithms from the Harwell Subroutine library are used, these being algorithms VAI3AD and VF04AD, which both use the BFGS formula, (Broydon (1970), Fletcher (1970), Goldfard (1970) and Shanno (1970)). All computations are performed in double precision to 7 decimal places of accuracy. The final results, however, are not dependent upon which algorithm is used in this step. Furthermore, the solutions of each of the algorithms used were compared with those in the standard Econometric packages TSP and SHAZAM, and were found to give similar results. Appropriate bivariate Student-t variates, are generated as described above. Further details of the Monte Carlo methodology are given in Chapter 4.

Generally, these figures illustrate that the maximum likelihood estimator behaves similarly in finite-samples to the maximum likelihood estimator associated with normally-distributed reduced-form disturbances. In particular, we again see that the maximum likelihood estimator is bimodal over part of the parameter space.

---

[3]Empirical densities were also computed using the Epanechnikov (1969) kernel. However, given the number of replications used, the results proved not to depend on which kernel is used. This situation is similar to the comparison of different kernels for the Cauchy distribution using a "large sample", as is illustrated in Figure 5.1 in Chapter 3.

Overall Comments

The finite-sample distribution of the LIMLK estimator with normally- distributed reduced-form disturbances has a number of interesting properties. In particular, the LIMLK estimator reduces to ILS and the computations presented in this section indicate that the limiting distribution is a good approximation to the finite-sample distribution. Further, the numerical computations of Marsaglia (1965) illustrate that the distribution is bimodal over part of the parameter space.

When the distribution of the reduced-form disturbances are widened to include the Student-t family, there are two different error assumptions to consider. These are given by equations (2.6) and (2.12), and they lead to quite different estimation techniques with different properties. In particular, when (2.6) is assumed, the LIMLK estimator is ILS whereas when (2.12) is assumed, the LIMLK estimator needs to be numerically computed. Further, each of the estimators converges to different limiting distributions. However, the computations of the finite-sample distributions of each of these estimators indicates that they both have distributions with similar properties to the LIMLK estimator when the reduced-form disturbances are normally distributed. That is, in each case, the limiting distribution is a good approximation to the finite-sample distribution and the distribution is bimodal over part of the parameter space.

However, because there are differences between the two assumptions it is important to consider the consequences of misspecifying the type of Student- assumption. This is the topic of the next section.

10.3 CONSEQUENCES OF MISSPECIFICATION

In this section, the statistical consequences of misspecifying the jointness versus independence assumption of Student-t distributed reduced- form disturbances in the exactly-identified limited-information SEM is considered. One implication of this analysis is to determine how important it is to make this distinction by applying appropriate "powerful" specification tests. In particular, the consequences of misspecification on the following three measures are considered.

(i) Median and Interquartile Range (IQR) of the finite-sample distribution, which is used to determine the consequences of misspecification in finite- sample distributions.

(ii) Asymptotic Variance, which is considered because even though the finite-sample variance does not necessarily exist, the asymptotic variance is often reported as an approximate measure of dispersion.

(iii) Limiting Distribution, which is considered because this is often used as an approximation to the finite-sample distribution for the purposes of inference.

Throughout this section the superscripts D and I will be used to distinguish between the assumptions given by (2.6) and (2.12) respectively. In particular, the following notation will be used:

$IML^{(I)}$ = Maximum Likelihood Estimator associated with (2.12) when the pdf of (v_1,v_2) is given by (2.12).

$IML^{(D)}$ = Maximum Likelihood Estimator associated with (2.12) when the pdf of (v_1,v_2) is given by (2.6).

$DML^{(I)}$ = Maximum Likelihood Estimator associated with (2.6) when the pdf of (v_1,v_2) is given by (2.12).

$DML^{(D)}$ = Maximum Likelihood Estimator associated with (2.6) when the pdf of (v_1,v_2) is given by (2.6).

(i) Median and IQR of finite-sample distribution

Table 3.1 compares values of the median and IQR for the estimators $DML^{(I)}$ and $IML^{(I)}$, corresponding to α = 5.33, 0.5 and various δ^2. These are estimated on the basis of a simple Monte - Carlo experiment using at least 40,000 replications. Appropriate random numbers are generated using (2.16).

In this case the reduced-form disturbances are assumed to be distributed as in (2.6) but actually have the distribution given by (2.12). Consequently, the appropriate maximum likelihood estimator to use is IML but due to this misspecification, DML is used instead. From Table 3.1 we can see that this results in the use of an estimator similarly dispersed as $IML^{(I)}$, but one which locates around the true parameter value much more slowly. These features are also illustrated by comparing Figure 3.1, which shows various

Table 3.1: Comparison of Median and IQR for Estimators $IML^{(I)}$ and $DML^{(I)}$

$\alpha = 5.33$

$IML^{(I)}$

	v=1		v=2		v=3		v=5	
δ^2	Median	IQR	Median	IQR	Median	IQR	Median	IQR
0.141	0.66162	4.0574	0.7555	4.2283	0.81975	4.3676	0.91612	4.5412
1.00	2.9093	4.807	3.0817	4.7230	3.2148	4.7045	3.3607	4.7124
4.00	4.7298	4.3867	4.8843	4.3118	4.9304	4.2705	5.0155	4.1438

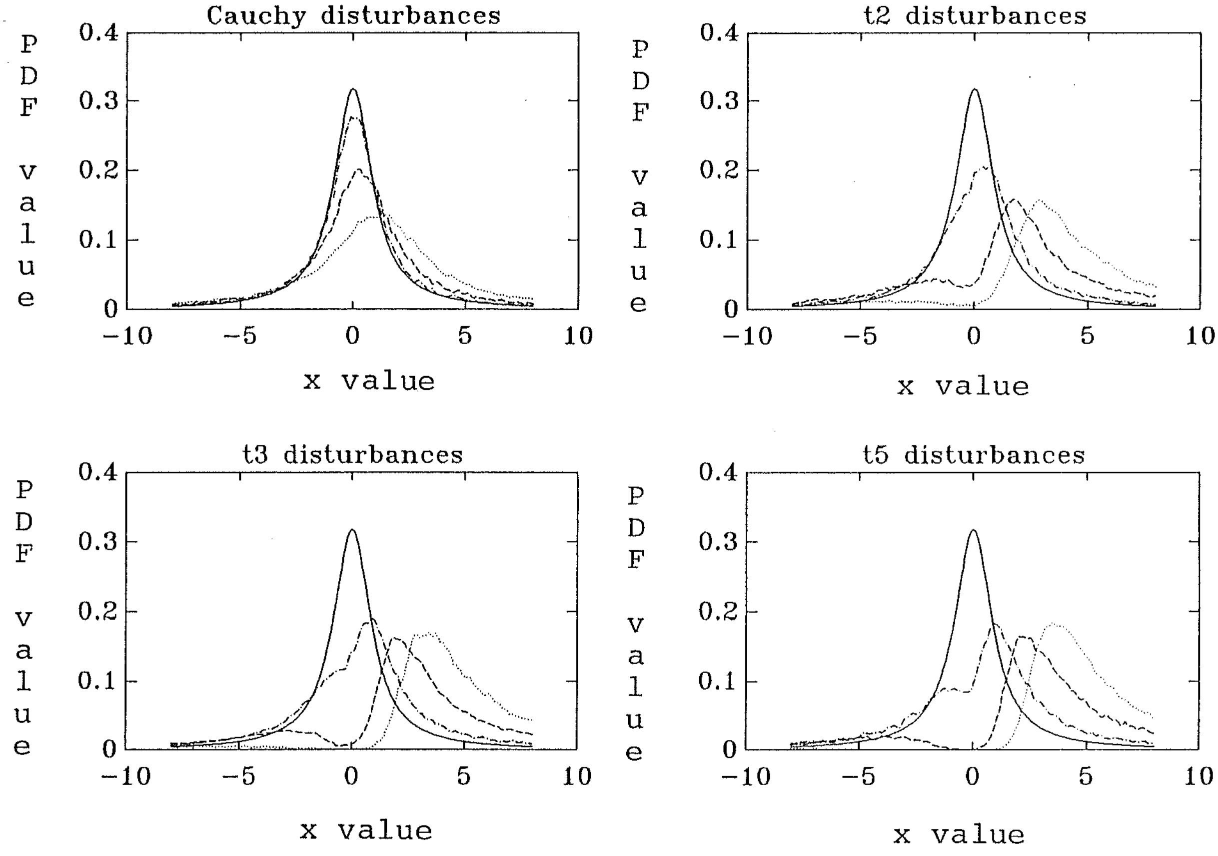
$DML^{(I)}$

	v=1		v=2		v=3		v=5	
δ^2	Median	IQR	Median	IQR	Median	IQR	Median	IQR
0.141	0.04132	9.2247	0.26942	3.0298	0.49154	3.6245	0.58650	3.840318
1.00	0.31977	3.2247	1.7682	5.8761	2.5061	6.1140	2.70228	5.2526
4.00	1.11304	4.6842	3.6970	4.8161	0.1875	4.28779	4.5308	4.4199

$\alpha = 0.05$

	$IML^{(I)}$				$DML^{(I)}$			
	v=1		v=5		v=1		v=5	
δ^2	Median	IQR	Median	IQR	Median	IQR	Median	IQR
5	0.44461	0.91239	0.47269	0.79927	0.11716	1.79085	0.45683	0.874497
23	0.49865	0.42616	0.49971	0.36419	0.29884	1.35172	0.49963	0.40088

FIGURE 3.1 Graphs of $DML^{(I)}$



KEY
- – $\delta^2 = 0.0001$
- −· $\delta^2 = 0.1410$
- −− $\delta^2 = 1.0$
- ·· $\delta^2 = 4.0$

Table 3.2: Comparison of Median and IQR for Estimators $IML^{(D)}$ and $DML^{(D)}$

$\alpha = 5.33$

$IML^{(D)}$

δ^2	v=1 Median	v=1 IQR	v=2 Median	v=2 IQR	v=3 Median	v=3 IQR	v=5 Median	v=5 IQR
0.141	0.49067	3.6467	0.71379	4.0723	0.83722	4.344	0.95121	4.5641
1.00	2.1830	5.6321	3.0526	5.2346	3.1560	4.9883	3.3943	4.7916
4.00	3.6319	4.7006	4.4795	4.0724	5.0522	5.5123	4.9172	3.8686

$DML^{(D)}$

δ^2	v=1 Median	v=1 IQR	v=2 Median	v=2 IQR	v=3 Median	v=3 IQR	v=5 Median	v=5 IQR
0.141	0.6416	4.0185	0.8384	4.3621	0.9385	4.58123	1.01731	4.6965
1.00	2.6615	5.5760	3.1725	5.0887	3.3724	4.8793	3.5055	4.7981
4.00	4.2718	4.3901	4.7105	3.8734	5.2526	5.4833	4.9931	3.8163

$\alpha = 0.5$

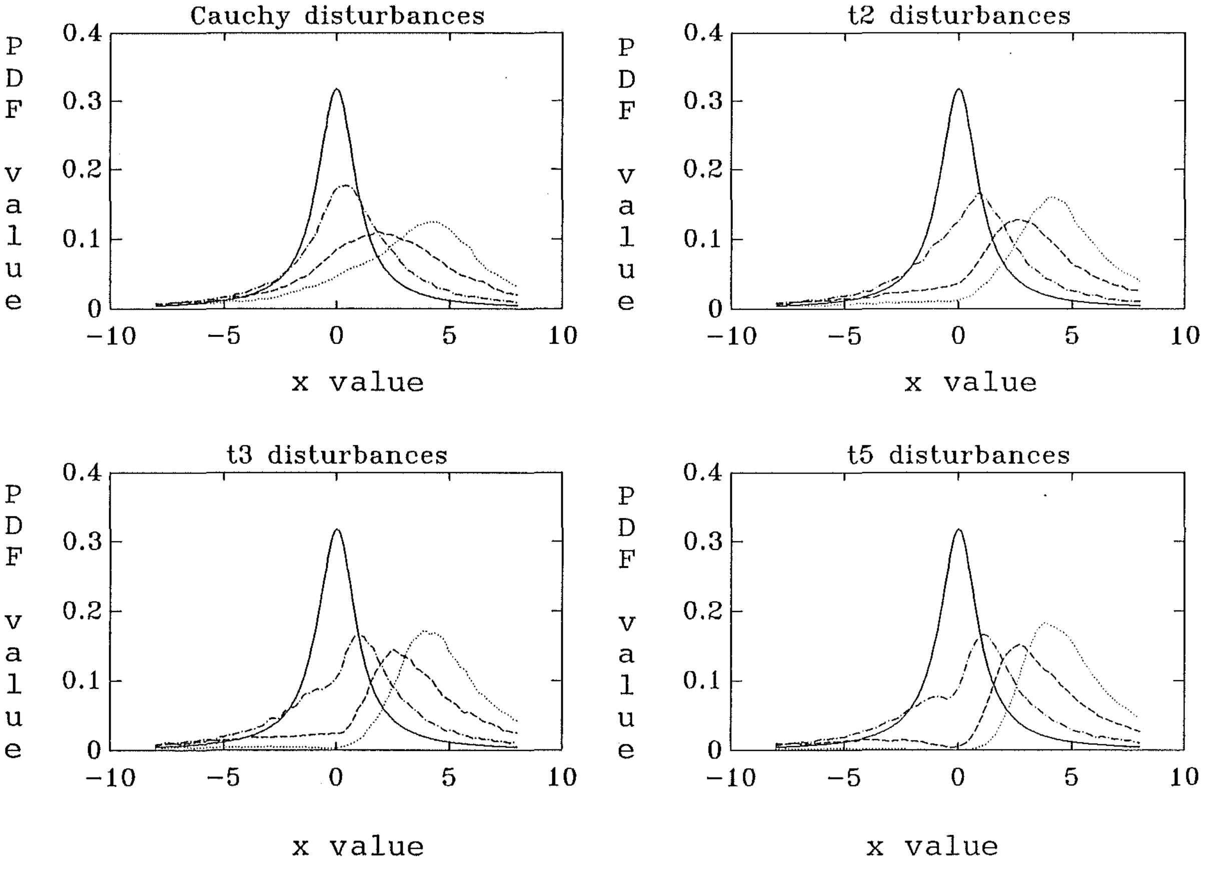
δ^2	$IML^{(D)}$ v=1 Median	$IML^{(D)}$ v=1 IQR	$IML^{(D)}$ v=5 Median	$IML^{(D)}$ v=5 IQR	$DML^{(D)}$ v=1 Median	$DML^{(D)}$ v=1 IQR	$DML^{(D)}$ v=5 Median	$DML^{(D)}$ v=5 IQR
5	0.3921	1.0195	0.4617	0.7679	0.4014	0.8387	0.4679	0.7385
23	0.4679	0.5479	0.4963	0.3583	0.4813	0.4569	0.4979	0.3430

graphs of $DML^{(I)}$, and Figure 2.6 which shows the corresponding graphs of $IML^{(I)}$. The graphs of $DML^{(I)}$ are estimated via an integration of the kernel density estimator with the Monte-Carlo method, as described in the previous Section, except with inputs into (2.17) obtained by calculating (2.9) with bivariate Student-t variates generated using (2.16). We consider different v, $\alpha = 5.33$ and numerous δ^2, Further, we also see from Figure 3.1 that the distribution of $DML^{(I)}$ tends to be bimodal over the same parameter space as $IML^{(I)}$ except for $v = 1$.

Alternatively, Table 3.2 compares values of the median and IQR for the estimators $DML^{(D)}$ and $IML^{(D)}$, corresponding to α = 5.33, 0.5 and various δ^2. For $DML^{(D)}$ these values are calculated exactly via the numerical integration of points of the pdf calculated from (2.10). For $IML^{(D)}$, they are estimated using a simple Monte-Carlo experiment using at least 40,000 replications and N, (= sample size, arbitrarily chosen since sample size is not a key parameter), jointly distributed random variates generated using the relationship given in (2.16).

In this case the reduced-form disturbances are assumed to be distributed as in (2.12) but actually have the distribution given by (2.6). However, the resulting estimator that is used, $IML^{(D)}$, has a distribution that is similar both in location and dispersion, to the correct maximum likelihood estimator $DML^{(D)}$. This feature is also illustrated by comparing Figure 3.2, which shows various graphs of $IML^{(D)}$, (estimated via the integration of the kernel density estimator with the Monte-Carlo method as described in the

FIGURE 3.2 Graphs of $IML^{(D)}$



KEY

- $-$ $\delta^2 = 0.0001$
- $-\cdot$ $\delta^2 = 0.1410$
- $--$ $\delta^2 = 1.0$
- $\cdot\cdot$ $\delta^2 = 4.0$

previous Section, with N jointly distributed random Student-t variates generated using (2.16)), corresponding to various v, α = 5.33 and numerous δ^2, and Figure 2.4, which shows the corresponding graphs of $DML^{(D)}$. In particular, the two figures can be seen to be essentially identical.

Consequently, on the basis of this measure, we would conclude that $IML^{(D)}$ is a more robust estimator in comparison to $DML^{(I)}$.

(ii) Asymptotic Variance

The implications of misspecification on the asymptotic variances are similar to those discussed by Kelejian and Prucha (1985) for the linear regression model. In particular, some examples of the standardized asymptotic variances of $\hat{\alpha}$ and $\tilde{\alpha}$ (i.e. $\frac{1+\alpha^2}{\delta^2}(\hat{\alpha}-\alpha)$, $\frac{1+\alpha^2}{\delta^2}(\tilde{\alpha}-\alpha)$), associated with reduced-form disturbances distributed as in (2.6) and (2.12) are given in Table 3.3. These variances are calculated using the known results of Kelejian and Prucha (1985); for $IML^{(I)}$ for $v \geq 5$ and calculated using at 40,000 replications in a simple Monte - Carlo experiment for $v < 5$; from Theil (1971, p.505) for $DML^{(I)}$ and $DML^{(D)}$ for $v > 2$; and are calculated on the basis of a simulation experiment using at least 40,000 replications for $IML^{(D)}$. From the values given in Table 3.3 the following general comments can be seen to hold for $v \geq 3$:

(1) If the reduced-form disturbances are jointly-distributed as in (2.6) but are assumed to be iid-distributed as in (2.12), then IML will be taken as the appropriate maximum likelihood estimator

Table 3.3: Asymptotic Variances for $\frac{1+\alpha^2}{\delta^2}$(DML-$\alpha$) and $\frac{1+\alpha^2}{\delta^2}$(IML-$\alpha$) when reduced-form disturbances are distributed as (2.6) and (2.12).

	v		
	3	4	5
Var($DML^{I=D}$)	3.0	2.0	1.6667
Var(IML^{D})	3.4649	2.2366	1.7999
Var(IML^{I})	1.40	1.3333	1.2857

to use. In this case the asymptotic variances of $IML^{(D)}$ are similar to those of $DML^{(D)}$, which again emphasizes the robustness of this estimator. However, the asymptotic variances reported for IML will be those associated with $IML^{(I)}$, which substantially underestimate those for $IML^{(D)}$. Consequently, on the basis of asymptotic variance, under this type of misspecification $IML^{(D)}$ is robust, but incorrect asymptotic variances will be reported.

(2) On the other hand, if the reduced-form disturbances are iid-distributed as in (2.12) but are assumed to be jointly-distributed as in (2.6), then DML will be used, with associated asymptotic variances given by $DML^{(I)}$. These variances can be seen to be substantially greater than those corresponding to the correct maximum likelihood estimator $IML^{(D)}$. Consequently, on the basis of asymptotic variance, this type of misspecification is associated with an inefficient estimator.

For the infinite-variance distributions, that is $v = 1, 2$, the asymptotic variances for $DML^{(I)}$ and $DML^{(D)}$ do not exist, so in this case the consequences of misspecifying the type of Student-t distribution are even more serious.

(ii) Limiting Distribution

If it is thought that the reduced - form disturbances are independent Student - t distributed, that is with joint distribution given by (2.12), then it will be assumed that the associated maximum likelihood estimator, $\tilde{\alpha}$, has the asymptotic distribution,

$$\sqrt{\frac{\delta^2}{1+\alpha^2}}\ (\tilde{\alpha}-\alpha) \rightarrow N\left(0,\frac{v+4}{v+2}\right) \text{ as } \delta^2 \rightarrow \infty . \qquad (3.1)$$

This result is based on the results of Kelejian and Prucha (1984) and simulation results presented in this chapter. However, if the reduced-form disturbances are actually dependent Student - t distributed as in (2.6), this limiting distribution will be incorrect. It is important to consider the implications of the use of the wrong limiting distribution since it is this distribution that is often used as an approximation to the finite - sample distribution for purposes of inference.

Various points of the distribution of $\sqrt{\frac{\delta^2}{1+\alpha^2}}\ (\tilde{\alpha}-\alpha)$, assuming dependent Student - t distributed reduced - form disturbances (2.6), are given in Table 3.4, corresponding to α = 0.5, 5.33 and δ^2 = 25, 1000 for various v. These points are obtained via the estimation of the empirical DF (see (4.2.1)), with appropriate N, jointly distributed Student- t random numbers generated using the relationship (2.16). These points are compared with the corresponding points of the incorrect limiting distribution given by (3.1). In particular, we can see that this misspecification will result in the use of an asymptotic approximation that has tails that are much thinner than the actual finite-sample distribution. This suggests that conventional hypothesis testing about a structural coefficient based on the incorrect limiting distribution is very likely to seriously overestimate the actual

Table 3.4 Effect Of Using The Wrong Limiting Distribution For Standardized $\tilde{\alpha}$ When Errors are Jointly - Distributed but are Thought to be Independently-Distributed

$v = 1$

$\alpha = 0.5$		$\alpha = 5.33$			
$\delta^2=25$	$\delta^2=1000$	$\delta^2=25$	$\delta^2=1000$	Normal with variance=$\frac{5}{3}$	
-0.05615	-0.015	-0.14882	-0.0314	0.0	50%
0.18537	0.2518	0.09788	0.22974	0.33588	60%
0.49478	0.5794	0.40091	0.55844	0.68278	70%
0.98739	1.1449	0.88458	1.0984	1.095589	80%
2.562	2.8233	2.1704	2.7086	1.65957	90%
4.7073	6.6137	4.5275	6.1978	2.124746	95%
23.681	37.092	21.736	32.9330	2.99854	99%

$v = 2$

$\alpha = 0.5$		$\alpha = 5.33$			
$\delta^2=25$	$\delta^2=1000$	$\delta^2=25$	$\delta^2=1000$	Normal with variance=$\frac{3}{2}$	
-0.02517	-0.0639	-0.0018	-0.045	0.0	50%
0.20929	0.23842	0.19702	0.23873	0.318641	60%
0.49204	0.52747	0.48951	0.52533	0.64774	70%
0.90398	0.92969	0.93528	0.93676	1.03937	80%
1.7620	1.7470	1.9842	1.7781	1.57441	90%
3.0839	2.9367	3.7235	3.0806	2.01571	95%
11.138	9.3376	14.706	10.7560	2.84466	99%

Table 3.4 continued

v = 3

$\alpha = 0.5$		$\alpha = 5.33$			
δ^2=25	δ^2=1000	δ^2=25	δ^2=1000	Normal with variance=$\frac{7}{5}$	
-0.01329	-0.006	-0.02055	0.007	0.0	50%
0.22135	0.2402	0.22078	0.23508	0.307836	60%
0.50103	0.5169	0.51246	0.51531	0.625775	70%
0.89663	0.8808	0.95179	0.88275	1.00412	80%
1.6140	1.5245	1.8937	1.5687	1.52102	90%
2.6488	2.3036	3.2677	2.4067	1.94736	95%
7.9429	5.3103	10.608	5.8209	2.74820	99%

v = 5

$\alpha = 0.5$		$\alpha = 5.33$			
δ^2=25	δ^2=1000	δ^2=25	δ^2=1000	Normal with variance=$\frac{9}{7}$	
-0.006	0.0	-0.0110	-0.007	0.0	50%
0.23491	0.295004	0.23621	0.23839	0.295004	60%
0.51026	0.599689	0.54425	0.51720	0.599689	70%
0.88533	0.962267	0.97429	0.86824	0.9622767	80%
1.5478	1.45762	1.8132	1.4365	1.45762	90%
2.2832	1.86619	2.8796	2.0349	1.86619	95%
5.3451	2.63364	8.0263	3.7498	2.63364	99%

significance.

Similar comments can be made when the reduced-form disturbances are assumed to be distributed as (2.6) but are actually distributed as (2.12). In this case if disturbances are thought to be distributed as (2.6) then DML will be taken as the appropriate maximum likelihood estimator to use with corresponding limiting distribution,

$$\sqrt{\frac{\delta^2}{1+\alpha^2}}\,(\hat{\alpha}-\alpha) \rightarrow MT_1(0,1,v) \quad \text{as} \quad \delta^2 \rightarrow \infty \,. \qquad (3.2)$$

However, if the reduced - form disturbances are actually distributed as (2.12) then (3.2) will be wrong. We again examine the consequences of the use of this wrong limiting distribution by comparing the finite - sample distribution of $\hat{\alpha}$, assuming (2.12) with the limit distribution given in (3.2).

In Table 3.5 various points of the distribution function of $\sqrt{\frac{\delta^2}{1+\alpha^2}}\,(\hat{\alpha}-\alpha)$, assuming reduced - form disturbances distributed as in (2.12) and corresponding to $\alpha = 0.5$, 5.33 and $\delta^2 = 25$, 1000 for various v are given. These are obtained via the estimation of the empirical DF, (see (4.2.1)), with bivariate Student - t random numbers generated using the relationship (2.16) with $K=2$. A comparison of these points with the corresponding points of the incorrect limiting distribution $MT_1(0,1,v)$, when the reduced-form disturbances are actually distributed as (2.12) illustrate that the use of the wrong limit distribution results in the use of an approximation to the finite-sample distribution that has much thinner tails. This suggests that conventional hypothesis testing

Table 3.5: Effect Of Using The Wrong Limiting Distribution for Standardized $\hat{\alpha}$ when Errors are Independently-Distributed but are Thought to be Jointly-Distributed

v = 1

$\alpha = 0.5$, $\delta = 25$	$\alpha = 0.5$, $\delta = 1000$	$\alpha = 5.33$, $\delta = 25$	$\alpha = 5.33$, $\delta = 1000$	Standardized Cauchy	
-0.79209	-0.21434	-1.8473	-0.48310	0.0	50%
0.17617	1.0627	-1.0483	0.75251	0.31831	60%
1.3876	2.5909	-0.07521	2.3163	0.71520	70%
3.3597	5.1025	1.5267	4.8041	1.3602	80%
8.3335	11.838	5.6863	11.399	3.0662	90%
18.109	24.506	13.466	23.042	6.3985	95%
93.673	104.21	76.542	112.58	33.197	99%

v = 2

$\alpha = 0.5$, $\delta^2 = 25$	$\alpha = 0.5$, $\delta^2 = 1000$	$\alpha = 5.33$, $\delta^2 = 25$	$\alpha = 5.33$, $\delta^2 = 1000$	Standardized t_2	
-0.076	0.0056	-0.11832	0.01946	0.0	50%
0.43516	0.51300	0.39167	0.52170	0.28404	60%
1.0621	1.0754	1.0909	1.1028	0.61725	70%
1.9535	1.7917	2.1923	1.8479	1.0639	80%
3.8423	2.9663	4.6354	3.0860	1.8839	90%
6.4801	4.2071	8.8531	4.5621	2.9351	95%
27.866	8.6984	41.896	9.8816	6.9584	99%

Table 3.5 continued

$v = 3$

$\alpha = 0.5$		$\alpha = 5.33$			
δ^2=25	δ^2=1000	δ^2=25	δ^2=1000	Standardized t_3	
0.00	0.01	-0.01005	0.00848	0.0	50%
0.38074	0.38478	0.40673	0.40234	0.7335	60%
0.8632	0.81032	0.91962	0.82241	0.58387	70%
1.4920	1.3192	1.6809	1.3485	0.98337	80%
2.6055	2.0940	3.1914	2.1459	1.6426	90%
3.9398	2.8111	5.2853	2.8976	2.3498	95%
10.219	4.4269	16.101	4.9267	4.5500	99%

$v = 5$

$\alpha = 0.5$		$\alpha = 5.33$			
δ^2=25	δ^2=1000	δ^2=25	δ^2=1000	Standardized t_5	
0.00	0.0046	0.01676	0.01841	0.0	50%
0.3215	0.31422	0.35601	0.33849	0.25546	60%
0.69624	0.65384	0.77380	0.68782	0.55802	70%
1.1789	1.0684	1.3359	1.0964	0.91295	80%
2.0125	1.6709	2.4058	1.7259	1.44792	90%
2.9204	2.1872	3.6145	2.2755	1.96776	95%
5.5950	3.2260	8.5127	3.5476	3.33237	99%

about a structural coefficient is very likely to seriously overestimate the actual significance.

Consequently, misspecifying the type of error distribution results in the use of the wrong limiting distribution, which in each case has much thinner tails than the actual finite-sample distribution, and this will have adverse implications for inference.

(iv) Overall Comments

The purpose of this section has been to illustrate the importance of distinguishing between reduced-form distributed disturbances given by (2.6) and (2.12). In particular, we see that when reduced-form disturbances are assumed to be distributed as (2.6) but actually have distribution (2.12), the effects on the resulting maximum likelihood estimator used are two-fold. This estimator is slow to locate around the true parameter value and an incorrect asymptotic distribution is used to approximate the finite-sample distribution, resulting in an approximation that has much thinner tails than the actual distribution, which will have implications for inference. On the other hand, when the reduced-form disturbances are assumed to be distributed as (2.12) but actually have distribution (2.6) the resulting maximum likelihood estimator used is robust in the sense that its finite-sample distribution is essentially identical to the correct maximum likelihood estimator. However, once again an incorrect asymptotic distribution is used to approximate the finite-sample distribution of the maximum likelihood estimator used, which results in an approximation that is much thinner tailed than the actual distribution, and this will have implications for inference. Consequently, the results suggest that it is worthwhile to have appropriate specification tests to distinguish between (2.6) and (2.12). This is one of the topics of the next chapter.

Before closing this Chapter however, note that the results presented could have been extended to include the more general SEM, by using for example the Godfrey and Wickens (1982) approach of treating LIML as a special case of FIML. However, it was decided to restrict attention simply to the exactly-identified SEM because of the interesting bimodality feature of the resulting density in this case, and also because it was considered that the results obtained would illustrate the general features of misspecification. Furthermore, the use of the LIMLK estimator, by assuming a known covariance matrix, simplified numerical computations considerably. More generally, what is required assuming iid nonnormal errors is a comphrehensive theory, including computational aspects, of robust estimators in the multivariate case. The maximum likelihood estimator for Student - t errors can be considered to be an example of a robust estimator, see e.g. Koenker and Prucha (1984). This problem has been set aside for future work, and it seems more appropriate to consider the more general SEM in this context.

CHAPTER 11

TESTING THE ASSUMPTION OF JOINTLY-DISTRIBUTED VERSUS INDEPENDENTLY-DISTRIBUTED NONNORMAL DISTURBANCES

11.1 INTRODUCTION

A widely used assumption in econometrics is that regression disturbances are normally distributed, and in this case there is no need to distinguish between independence and uncorrelatedness. Recently, however, as is illustrated in Chapter 7, there has been much interest in nonnormally distributed disturbances, and in this case a distinction needs to be made between assuming independently-distributed nonnormal disturbances and jointly-distributed nonnormal disturbances. In particular, if the appropriate moments exist, then this is a distinction between independence and uncorrelatedness. Chapters 8, 9 and 10 illustrate the importance of making this distinction in two models, these being the linear regression model and the exactly-identified limited-information SEM. In the linear regression model and in the exactly-identified linear-information SEM, the consequences of misspecifying the jointness/independence distinction are such that it is important to construct appropriate specification tests that make this distinction. In this chapter, such specification tests are presented to make this distinction in the elliptically-symmetric family of distributions, by adopting the use of existing tests for normality.

As the specification tests for jointness versus independence presented here adopt the use of existing tests for normality, we begin this chapter by first reviewing tests for normality. This review begins in Section 2 with tests for univariate normality and multivariate normality. In particular, attention is given to Shapiro and Wilk's (1965) test used to test for univariate normality, and a modification of this test used to test for multivariate normality. Section 3 discusses the application of these tests to the types of models considered in Chapters 9 and 10, while Section 4 considers the use of these tests for testing the jointness versus independence assumption. In Section 5 a Monte Carlo experiment is presented which illustrates the power of these tests for testing the jointness versus independence assumption assuming that the disturbances are Student-t distributed, and Section 6 concludes with some final comments.

11.2 TESTS OF NORMALITY

(i) Univariate Normality

Research into tests of normality of observations has a long history, with attention being given to one-directional tests such as skewness and kurtosis tests and tests that are sensitive to any form of departure from normality such as omnibus tests. Recent contributions to the literature are the skewness, kurtosis and omnibus tests proposed by D'Agostino and Pearson (1973), Bowman and Shenton (1975), Pearson, D'Agostino and Bowman (1977), Shapiro and Wilk (1965) and Shapiro and Francia (1972) and the use of the score test on a general family of distributions by Jarque and Bera (1987).

Shapiro et al. (1968) launched the first major power study into the behaviour of a number of tests for normality. They concluded that the Shapiro and Wilk (1965), W, test provides a general omnibus measure of nonnormality. Similar conclusions were obtained in studies by Dyer (1974), Stephens (1974) and Pearson et al. (1977).

Given these conclusions then, this chapter will focus on the W-test. This test has further appeal since Royston (1982) has provided a simple algorithm which enables it to be computed for sample sizes up to 2000 as well as providing appropriate significance levels. It is defined as follows. Let $m' = (m_1, \ldots, m_N)$ denote the vector of expected values of standard normal order statistics, and let $V = (v_{ij})$ be the corresponding N*N covariance matrix; that is

$$E(x_i) = m_i (i=1,\ldots,N) \text{ and } \mathrm{cov}(x_i,x_j) = v_{ij}(i,j=1,\ldots,N)$$

where $x_1 < x_2 < \ldots < x_N$ is an ordered random sample from a standard normal distribution N(0,1). Suppose $y' = (y_1,\ldots,y_N)$ is a random sample on which the W test of normality is to be carried out, ordered $y_{(1)} < y_{(2)} < \ldots < y_{(N)}$. Then

$$W = \left[\sum_i a_i y_i\right]^2 / \sum_i (y_i - \bar{y})^2 \qquad (2.1)$$

where

$$\underline{a}' = (a_1,\ldots,a_N) = m'V^{-1}\left[(m'V^{-1})(V^{-1}m)\right]^{-1/2} .$$

The coefficients (a_i) are the normalized "best linear unbiased" coefficients tabulated for $N < 20$ by Sarhan and Greenberg (1956). The covariance matrix V which features in $\underline{a}$ may be obtained using the algorithm of Davis and Stephens (1978). However, V is not required explicitly, and Shapiro and Wilk (1965) offer a satisfactory approximation for $\underline{a}$ which improves with increasing sample size N, and this approximation is usually adopted. By definition, $\underline{a}$ has the property $\underline{a}'\underline{a} = 1$. Let $a^* = m'V^{-1}$; approximations $\hat{a}^*$ for a^* are

$$\hat{a}_i = \begin{cases} 2m_i & , \quad i = 2,3 \ldots N-1 \\ \left[\dfrac{\hat{a}_i^2}{1-2\hat{a}_i^2} \displaystyle\sum_{j=2}^{N-1} \hat{a}_j^{*2}\right]^{\frac{1}{2}} & , \quad i = 1,\ i = N \end{cases}$$

where,

$$\hat{a}_1^2 = \hat{a}_N^2 = \begin{cases} g(N-1) & , \ N \le 20 \\ g(N) & , \ N > 20 \end{cases}$$

and $g(N) = \Gamma(N+1)/\sqrt{2}\Gamma(\frac{1}{2}N+1)$.

In the algorithm developed by Royston (1982) for computing W, these approximations are used throughout the range $7 < N < 2000$, while exact values are used for the (a_i) for $N < 7$. The values of m_i required in the computation are calculated using Blom (1958, pp.69-71) and are accurate to 0.0001. Values of the significance levels are also given, and are obtained by approximating the null distribution of W. That is, Royston (1982) showed that W could be transformed to an approximately standard normal variate, Z, under the hypothesis that the unordered observations come from a normal distribution with unspecified mean and variance, so that,

$$Z = \left((1-W)^{\lambda} - \mu\right)/\sigma \ , \qquad (2.2)$$

and λ, μ, σ are all functions of N, for which polynomial formulae are provided (Royston (1982, p.119)).

(ii) Multivariate Normality

Let $X = (x_1,\dots,x_m)$ be m variates each with N observations and let $\bar{X}$ and S be the sample mean vector and covariance-matrix respectively, corresponding to the population statistics μ and Σ, that is,

$$\bar{X} = \left[\bar{x}_j\right], \; j = 1,\dots,m, \text{ where } \bar{x}_j = 1/N \sum_{i=1}^{N} x_{ji} \; .$$

and

$$S = \frac{1}{N} \sum_{i=1}^{N} (x_{ji}-\bar{x})(x_{ki}-\bar{x}), \; j = k = 1,\dots,m.$$

The null hypothesis to be tested is that X is multivariate normally distributed. One simple procedure is to test the marginal normality of each of the m components by using univariate procedures. However, marginal normality does not imply multivariate normality although the presence of nonnormality is often reflected in the marginal distributions. Hence, it is usually claimed (see, for example, Mardia (1980)) that tests which exploit the multivariate structure will be more sensitive.

A number of test procedures for multivariate normality have been proposed in the literature, and reviews of these procedures are given in, for example, Mardia (1980) and Cox and Small (1978). Generally though these procedures have concentrated either on combinations of univariate tests of normality such as those of Small (1980), Malkovich and Afifi (1973), or on the geometrical properties in R^m of two or more variates taken together such as

Healy (1968) and Cox and Small (1978). However, often the suggested tests have intractable null hypothesis distributions, are difficult to calculate and further convincing power studies are rare.

Recently, a number of authors such as Royston (1983) and Srivastava and Hui (1987) have suggested extensions of W which solve some of these problems. However, the statistic proposed by Royston (1983) requires certain approximations to be made, in order for the statistic to have large-sample justification. Consequently, in this section we review only the statistic suggested by Srivastava and Hui (1987) which has large-sample justification.

Srivastava and Hui (1987) propose the test statistic M_1 for testing multivariate normality, and this statistic may be considered as a generalisation of both the univariate-W statistic, and also the statistic proposed by Shapiro and Wilk (1968) for the joint assessment of normality of several independent samples. In particular, M_1 is based on principal components. That is, let $\Gamma = (\gamma_1,\ldots,\gamma_m)$ be an orthogonal matrix such that $\Gamma'\Sigma\Gamma = D_\lambda$, where D_λ is a diagonal matrix with diagonal elements $\lambda_1,\ldots,\lambda_m$, then, $\gamma_1'X,\ldots,\gamma_m'X$ are called m principal components which are independently distributed with means $\gamma_1'\mu,\ldots,\gamma_m'\mu$ and variances $\lambda_1,\ldots,\lambda_m$ respectively, if X is normally distributed. When Σ is not known it is estimated from the sample by S and approximately independent principal components are obtained. That is, let $H = (h_1,\ldots,h_m)$, be an orthogonal matrix such that $H'SH = D_W$ where $D_W = \mathrm{diag}(W_1,\ldots,W_m)$ and let

$$y_{ij} = h_i' x_{ij}, \quad i = 1,\dots,m, \quad j = 1,\dots,N.$$

Then y_{ij} is the ith principal component for the jth sample, where Σ is estimated by S. Thus under the null hypothesis of normality, $(y_{i1},\dots,y_{iN})$, $i = 1,\dots,m$ is treated as m approximately independent samples and the procedures of Shapiro and Wilk (1968) can then be used. That is, for sample i, the univariate W is calculated, denoted as W(i), where

$$W(i) = \left[\sum_i a_i y_{(i)} \right]^2 / \sum_i \left(y_{(i)} - \bar{y} \right)^2 ,$$

as in (2.1) and,

$$M_1 = -2\sum_i \ln\left\{ \phi\left(G\left(W(i) \right) \right) \right\} \qquad (2.3)$$

where $G\left(W(i)\right)$ is the transformation of W(i) to a standard normal variate, suggested by Shapiro and Wilk (1968), and is equal to

$$G\left(W(i)\right) = \gamma + \delta \log\left\{ \frac{W(i) - \epsilon}{1 - W(i)} \right\}$$

with values for γ, δ and ϵ obtained in Table 1 of Shapiro and Wilk (1968) up to N = 50 and Royston (1983) for larger sample sizes, and,

$$\phi(x) = (2\Pi)^{-\frac{1}{2}} \int_{-\infty}^{x} \exp(-\tfrac{1}{2}t^2)\,dt .$$

Asymptotically M_1 is distributed as χ^2_{2m}.

A Monte Carlo study is reported by Srivastava and Hui (1987) in which it is concluded that the null distribution of M_1 is well approximated by χ^2_{2m} for sample sizes that are small as 10. However, no power results are given.

11.3 EXTENSION TO TESTING THE NORMALITY ASSUMPTION IN REGRESSION MODELS

(i) Using Univariate Tests in the Classical Regression Model

In the classical regression model (1.1.1), the disturbances, although assumed iid are unobservable. They can be estimated with the least squares residual vector,

$$e = \left(I - X(X'X)^{-1}X'\right)y$$

but $E(ee') = \sigma^2\left(I - X(X'X)^{-1}X'\right)$ and so the elements of e are correlated. This problem can be overcome by transforming e to a new N-k vector of uncorrelated residuals such as the BLUS or recursive residual vector (see, for example, Judge (1985, pp.172-173)), and then these can be used in conjunction with the univariate tests of normality in 11.2.

Alternatively, Mukantseva (1977), Pierce and Kopecky (1979), Loynes (1980) and White and MacDonald (1980) provide conditions under which several well-known tests for univariate normality have the same limiting null distribution when used to test the normality assumption of the regression disturbances and are calculated using least-squares residuals. These conditions are summarized in Condition (11.1):

Condition (11.1) (X_i), which denotes a vector of all of the observations on the regressors at point i, is a sequence of uniformly bounded fixed vectors such that $\lim_{N\Rightarrow\infty}\left[X'X/N\right] = Q$, a positive definite matrix.

Given this assumption then, for W we have, in probability,

$$|\hat{W} - W| \longrightarrow 0 \ ,$$

where the ^ refers to W calculated with least-squares residuals.

In finite-samples, using the same significance levels as appropriate for the univariate tests, Huang and Bolch (1974) and Ramsey (1974) report on Monte Carlo studies, including $\hat{W}$, where use of the least squares vector leads to a more powerful test than that obtained using the BLUS residual vector e. Furthermore, Monte Carlo studies carried out by Weisberg (1980) suggest that the significance level of $\hat{W}$ is near the nominal level, therefore suggesting that in finite-samples the significance levels of W can be used, although the accuracy of this approximation depends on the regressors as well as sample size. However, in all of these studies emphasis is mainly given to the nominal 10% level. The Monte Carlo study carried out in the last section of this chapter extends this analysis to all the common nominal significance values used.

(ii) Using Multivariate Tests in SEM's

If we consider the reduced-form of the SEM, such that,

$$Y = X\Pi + v \ , \tag{3.1}$$

where X is assumed to be strictly exogenous such that $\lim\frac{(X'X)}{N} = Q$, and Π is the matrix of corresponding reduced-form parameters, then the usual assumption made about the reduced-form errors is $v \sim N(0,\Omega)$, and this then also implies that the structural disturbances are multivariate-normal. If $v \sim N(0,\Omega)$ is taken as the null hypothesis, then it is easily shown that when the reduced-form errors $\hat{v}$, with estimated covariance matrix $\hat{\Omega}$, where,

$$\hat{\Omega}_j = \frac{1}{N}\hat{v}_i'\hat{v}_j ,$$

with ith and jth residuals from the ith and jth equation (2.4), are used to calculate $\hat{M}_1$, (^ indicates use of $\hat{v}$ in constructing M_1), then the limiting distribution of $\hat{M}_1$ is as in (2.3), that is, $\hat{M}_1 \sim \chi^2_{2m}$, where m is the number of exogenous variables in the reduced-form. This result holds simply because Condition (11.1) is satisfied at the first stage of forming the marginal W(i).

However, while many Monte Carlo studies have been carried out to assess the finite-sample performance of tests of normality calculated using least-squares residuals in the linear regression model, similar studies that consider the use of the reduced-form least squares residuals in the calculation of tests of multivariate-normality in SEMs are non-existent. Consequently, in the last section of this chapter a simple Monte Carlo study is performed to consider the performance of M_1 in the exactly-identified model of Chapter 10.

11.4 TESTING FOR JOINTNESS VERSUS INDEPENDENCE

King (1980b) shows that any statistic which is invariant to the scale of the disturbances of the linear regression model has the same distribution when $u \sim N(0,\sigma^2 L)$ as it does when u is assumed to follow any other elliptically symmetric distribution with characteristic matrix L. In particular, if tests of normality satisfy this invariance property then widening the null hypothesis of these tests to include the spherically symmetric family of distributions does nothing to the size (and also the power in this case), of these tests. Examples King (1980b) gives of tests of normality where this property holds are those suggested by Putter (1967), Koerts and Abrahamse (1969), Louter and Koerts (1970), Huang and Bolch (1974) and Mukantseva (1977).

In the previous section emphasis is given to the Shapiro and Wilk (1965) univariate test W and its modifications to testing normality in the classical regression model and SEM's. These tests also satisfy the above invariance property as shown by Shapiro and Wilk (1965). Consequently, it is more accurate to regard them as tests for spherical symmetry rather than tests for normality.

In the linear regression model, using this testing strategy implies the following. If the null hypothesis is accepted then there is no need to distinguish between the spherically distributed distributions. This is because all of the common test-statistics used satisfy this invariance property, such as, for example, the classical F-test of fixed linear restrictions on β, tests for serial correlation in regression disturbances proposed by Durbin and Watson (1950), tests for heteroscedastic disturbances suggested by Goldfeld and Quandt (1965), tests for regression misspecification such as those outlined by Ramsey (1969) (see also King (1980b, p.14)). However, if the alternative hypothesis is accepted then robust estimation and inference techniques are needed, at least for moderately-sized linear regression models.

A similar strategy can be adopted in SEM's using the recently suggested ways of defining structural-form residuals by Harvey and Phillips (1980), Phillips (1988), for limited- and full-information SEM's.

For example, for the limited-information SEM with two endogenous variables, the topic of Chapters 5 and 6,

$$y_1^* = y_2^*\beta + X_1\gamma_1 + u_1 \qquad (4.1)$$

we have the following relationship between the structural- and reduced-form disturbances,

$$v_1 = v_2\beta + u_1 .$$

Harvey and Phillips (1980) showed that all the usual tests, such as those mentioned above, have the same exact size as in the general linear model when based on the estimate,

$$\hat{u}_1 = (\hat{v}_1 \; \hat{v}_2)\begin{pmatrix} 1 \\ -\hat{\beta} \end{pmatrix},$$

where $\hat{v}_1$, $\hat{v}_2$ are the OLS estimates of the reduced-form disturbances corresponding to (3.1), and $\hat{\beta}$ is a consisitent estimate of β. Therefore, since these statistics are also invariant to scale, this implies there is no need to make any distinction between the elliptically-symmetric distributed disturbances when the null hypothesis of multivariate normality is accepted. If the alternative hypothesis is accepted then estimation and inference procedures can be based on maximum-likelihood methods with Student-t distributed disturbances for example, or some other robust method such as those suggested by Amemiya (1982) and Powell (1983).

11.5 MONTE CARLO EXPERIMENTS

In this section, results of Monte Carlo experiments are given to illustrate the performance, in terms of both size and power, of the $\hat{W}_1$ and $\hat{M}_1$ tests, in the linear regression model and the exactly-identified SEM. The results of a Monte- Carlo experiment to determine the size and power of the W_1 test are presented in Table 5.1 and for the $\hat{M}_1$ test in Tables 5.2 and 5.3 respectively. Every Monte - Carlo experiment in this section consists of generating 5,000 random numbers from a given distribution; computing the values of the test - statistics and seeing whether H_0 is rejected by each individual test. Assuming an underlying normal distribution gives the size of the test, and assuming the independent nonnormal Student - t alternatives yields an estimate of the power of the test. The estimates of the size and power of the tests are obtained by dividing 5,000 the number of times H_0 is rejected.

(i) Linear Regression Model

In this part, results of a Monte Carlo experiment are presented which illustrate the size and power associated with $\hat{W}$, where $\hat{W}$ is used to test the assumption,

$$H_0: \quad \epsilon \sim N(0,\sigma^2 I) \ ,$$

which, from Section 4, is equivalent to assuming

$$H_0: \epsilon \sim MT_N(0,\sigma^2 I,v) \ ,$$

and where the alternative hypothesis is,

$$H_1: \epsilon_n \sim \text{iid } MT_1(0,\sigma^2,v) \text{ for } n = 1,\ldots N,$$

and the associated significance levels are taken from those calculated by Royston (1982) for the W test. These significance levels rather than the asymptotic significance levels are used, due to the results obtained by Weisberg (1980), (see e.g. 11.3). Each regression model contains a number of nonstochastic regressors, as well as a constant term, and the total number of regressors is denoted by K. These numbers of regressors were chosen to illustrate the consequences of both size and power when, first, N is fixed and K is increased and, secondly, when K is fixed and N is increased. For N = 20, the three data sets of Weisberg (1980) are used. It is well known (see e.g. Jarque and Bera (1987, p.170)), that the matrix $V = I - X(X'X)^{-1}X'$ influences both the actual size and power of the normality test. The three data sets of Weisberg (1980) illustrate the effects of different V on the normality test. For N = 50, four nonstochastic regressors are generated from independent uniform, normal and χ^2_{10} distributions. Uniform variates are generated using the NAG subroutine G05CAF, which uses a multiplicative congruential method; normal variates are generated using the NAG subroutine G05DDF, which is based on Brent's (1974) algorithm and χ^2_{10} variates are generated using the formula, $-2\ln\left[\prod_{i=1}^{5} U_i\right]$, where U_i are uniform variates. These regressors are used as it is considered that they cover a wide range of alternatives. To obtain the estimates of the size and power of W, normal and iid Student - t variates are generated. Normal variates are generated as above, and for v = 1, the Cauchy distribution, standard Cauchy variates are generated as,

$$X = \tan\left(\pi(U-\tfrac{1}{2})\right)$$

and for v = 2, the t_2-distribution,

$$X = \sqrt{2}(U-\tfrac{1}{2})/SQRT\left(U(1-U)\right) ,$$

where U is from U(0,1).

For the rest of the Student-t family, $v \geq 3$, X is generated via a transformation of a symmetric beta variate, (see, for example, Devroye (1986, p.446)). This can be written in terms of independent uniform random numbers U_1, U_2 as,

$$X = \frac{2\sqrt{v}\ \sin\ (2\pi U_1)(1-U_2^{2/v-1})}{(1-\sin^2(2\pi U_1))(1-U_2^{2/v-1})} .$$

This formula is useful as it is valid for all members of the Student-t family with $v \geq 3$. It also does not require the generation of as many random uniform deviates as does the traditional method of generating a t-random variable via its interpretation as a ratio of a standard normal to the square root of an independent normalized chi-square variable.

The results of the simulations for three significance levels α= 0.01,0.05 and 0.10, are presented in Table 5.1, from which the following two points can be made:

- Except for small N and large K, the actual size of $\hat{W}$ is very close to the normal size.
- The power of $\hat{W}$ is large for the infinite-variance disturbances ($v \leq 2$) even for small sample sizes, and in comparison falls dramatically for the finite-variance disturbances ($v > 2$).

Consequently, for Student-t disturbances, for small samples and moderate values of K, the significance values of $\hat{W}$ are well approximated by those computed for W. Furthermore, $\hat{W}$ is very

Table 5.1: Results of Monte Carlo Experiments for Linear Regression Models using 5000 Replications

α	N	K	Normal	v = 1	v = 2	v = 3	v = 5	v = 10
Data Set 1			size	power				
0.10	20	4	0.0978	0.7782	0.4618	0.3132	0.1952	0.1304
	20	8	0.1068	0.5746	0.3274	0.1920	0.1682	0.1284
Data Set 2								
0.10	20	4	0.0994	0.7838	0.8066	0.3118	0.1964	0.1330
	20	8	0.1012	0.5386	0.5542	0.2232	0.1586	0.1316
Data Set 3								
0.10	20	4	0.1040	0.7848	0.9010	0.3110	0.1904	0.1242
	20	8	0.2440	0.7478	0.7650	0.4160	0.3320	0.2778
Data Set 3								
0.05	20	4	0.042	0.7346	0.7592	0.2416	0.1302	0.0686
	20	8	0.153	0.6840	0.7058	0.3106	0.2362	0.0930
Data Set 3								
0.01	20	4	0.0080	0.6254	0.6514	0.1340	0.0534	0.0162
	20	8	0.0486	0.4542	0.5736	0.1774	0.1032	0.0670
Data Set 4								
0.10	50	4	0.0980	0.9892	0.9928	0.568	0.3172	0.1688
Data Set 4								
0.01	50	4	0.005	0.970	0.6592	0.3804	0.1528	0.0418

powerful in distinguishing between joint Student-t and infinite-variance iid Student-t disturbances, and moderately powerful otherwise. Whether this is a feature of all the normality tests is a matter for future analysis. Another interesting question is the behaviour of $\hat{W}$ in the elliptically-symmetric family of distributions generally.

(ii) Exactly Identified SEM

In this section, Monte Carlo results are presented for the $\hat{M}_1$-test, used to test the assumption of multivariate normality in the exactly- identified SEM.[1] The reduced-form model considered is,

$$y_1^* = c_1 + X_2\pi_{21}^* + v_1$$

$$y_2^* = c_2 + X_2\pi_{22}^* + v_2 ,$$

where c_1 and c_2 are constants, and X_2 is a $N \times 1$ vector of observations on a strictly exogenous variable. The null hypothesis is taken to be,

$$(v_{1t}, v_{2t}) \sim N(0,\Omega) \quad \forall\ t$$

where

$$\Omega = \begin{pmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{pmatrix} . \qquad (5.1)$$

This can also be written as,

---

[1] For the purposes of the simulation experiment here, the SEM is considered in standard form as opposed to the canonical form presented in Chapters 5, 6 and 10. This is so we could determine the effect of ρ, see for example (5.3).

$$(v_1, v_2) \sim N(0, \Omega \otimes I) \ , \tag{5.2}$$

and using the "invariance of scale" property of $\hat{M}_1$ as discussed in Section 4, implies that $\hat{M}_1$ has the same distribution when (5.2) is assumed as it does when it is assumed,

$$(v_1, v_2) \sim MT_{2N}(0, \Omega \otimes I, v).$$

The alternative hypothesis is taken as

$$(v_{1t}, v_{2t}) \sim \text{iid } MT_2(0, \Omega, v) \quad \forall \ t.$$

As the results on the Monte Carlo experiment depend upon the values of Ω and X_2, these are varied in a number of ways. In particular, it is assumed,

$$\Omega = \begin{bmatrix} \sigma^2_{v_1} & \rho\sigma_{v_1}\sigma_{v_2} \\ \rho\sigma_{v_1}\sigma_{v_2} & \sigma^2_{v_2} \end{bmatrix} \tag{5.3}$$

where ρ represents the correlation between v_{1t} and v_{2t} and is set equal to 0.3, 0.6 and 0.9. Two different data sets for X_2 are included. In the first data set X_2 is assumed orthogonal ($X_2'X_2 = 1$), and in the second X_2 is assumed to be χ^2_{10} distributed, and these variates are generated as above. To obtain estimates of the size and power of the test it is necessary to generate bivariate normal and Student - t variates with precision matrix Ω. Normal variates are generated as described above. Student - t variates are generated using the relationship

$$X_i = Z_i\left(\frac{\chi^2}{v}\right)^{-\frac{1}{2}} \qquad i = 1,2$$

Table 5.2: Results of Monte Carlo Experiments for Exactly-Identified Limited-Information SEM using 5000 Replications and Corresponding to data set 1.

	$\rho = 0.30$		$\rho = 0.60$		$\rho = 0.90$
nominal size 10%					
actual size	0.108		0.093		0.091
power		power		power	
v = 1	0.949	v = 1	0.964	v = 1	0.972
2	0.713	2	0.741	2	0.872
3	0.518	3	0.568	3	0.674
5	0.304	5	0.330	5	0.392
10	0.190	10	0.200	10	0.204
nominal size 5%					
actual size	0.046		0.048		0.054
power		power		power	
v = 1	0.931	v = 1	0.952	v = 1	0.959
2	0.645	2	0.677	2	0.784
3	0.437	3	0.473	3	0.590
5	0.231	5	0.246	5	0.325
10	0.112	10	0.116	10	0.140
nominal size 1%					
actual size	0.01		0.012		0.006
power		power		power	
v = 1	0.891	v = 1	0.926	v = 1	0.936
2	0.508	2	0.583	2	0.687
3	0.297	3	0.335	3	0.464
5	0.129	5	0.136	5	0.209
10	0.037	10	0.041	10	0.065

Table 5.3: Results of Monte Carlo Experiments for Exactly-Identified Limited-Information SEM Using 5000 Replications and Corresponding to Data set 2.

	ρ = 0.30		ρ = 0.60		ρ = 0.90
nominal size 10%					
actual size	0.113		0.10		0.095
power		power		power	
v = 1	0.940	v = 1	0.975	v = 1	0.979
2	0.716	2	0.741	2	0.807
3	0.520	3	0.549	3	0.670
5	0.319	5	0.344	5	0.409
10	0.177	10	0.181	10	0.205
nominal size 5%					
actual size	0.056		0.051		0.048
power		power			
v = 1	0.950	v = 1	0.965	v = 1	0.970
2	0.643	2	0.702	2	0.763
3	0.434	3	0.476	3	0.598
5	0.239	5	0.250	5	0.332
10	0.111	10	0.113	10	0.113
nominal size 1%					
actual size	0.09		0.012		0.010
power		power		power	
v = 1	0.890	v = 1	0.935	v = 1	0.955
2	0.523	2	0.591	2	0.677
3	0.297	3	0.349	3	0.486
5	0.129	5	0.137	5	0.212
10	0.030	10	0.037	10	0.067

where Z_1, Z_2 are K independent standard normal variables and χ^2 is an independent chi-square variable with v degrees of freedom. The generated normal and Student - t variates are then appropriately transformed so as to have precision matrix Ω.

The results of the Monte Carlo experiment are presented in Tables 5.2 and 5.3. These results indicate that the size of the test is well approximated by the corresponding size of M_1 and furthermore, the power of the test is reasonably large even for rather high values of v.

11.6 SOME FINAL COMMENTS

The objective of this Chapter was to illustrate the use of existing normality tests to test for the distinction between jointness versus independence in the elliptically-symmetric family of distributions. In particular, results of Monte Carlo experiments suggest that the use of Shapiro and Wilk's (1965) test and various modifications to this test are useful methods of testing this assumption in moderately-sized linear regression models and in exactly-identified limited-information SEM's.

CHAPTER 12

SUMMARY AND CONCLUSIONS

12.1 OVERVIEW

The OLS estimator is the most common procedure for estimation in the classical multiple linear regression model. This estimation technique is justified on the basis of its well known finite-sample behaviour. However, in empirical work, the assumptions of this model, including nonstochastic regressors and normally distributed disturbances, are often violated, and as a result OLS often has no statistical justification. This has led to the relaxation of these assumptions and consequently to the development of a number of estimation and inference techniques which are alternatives to those based on OLS. The introduction of these techniques though, has usually been justified on the basis of their behaviour in large samples. However, generally the sample sizes used in empirical work are small, and in small samples the behaviour of these techniques may be very different. Consequently, this suggests that the choice of appropriate statistical techniques to use should be based on finite-sample behaviour.

Early investigations into the finite-sample behaviour of various statistics date back to Haavelmo (1947), Anderson and Rubin (1949), and Hurwicz (1950) and since the 1960's substantial progress has been made, particularly in the finite-sample analysis of SEM's, (see e.g. Phillips (1982)). In this thesis, finite-sample properties of estimators used in three well known econometric models have been extended and developed. Each of these models are extensions of the classical multiple linear regression model when the assumptions of either nonstochastic regressors, or normally distributed disturbances, or a combination of these assumptions are relaxed. In particular, all of the estimators considered are now included in standard and widely-used econometric packages such as SHAZAM and TSP. The three models considered are, the limited-information SEM, the nonnormal linear regression model and the nonnormal limited-information SEM.

12.2 METHODS USED

The approach taken in the development of finite-sample properties of estimators used in each of the models considered, was to calculate or approximate the exact distribution function, or density function, or various descriptions of these functions, such as moments, medians and inter- quartile ranges. Further, the 'key parameters' of these functions were identified, and varied in the computations of these functions, so as to make the results as general as possible.

In the calculation of exact results, a FORTRAN version of Davies' (1980) algorithm was used. This algorithm has been well tested (see e.g. Davies (1980)), and results obtained using this algorithm were considered to be very accurate. However, when computations of exact results were impossible, due either to analytical intractability, or infeasible numerical calculations, Monte Carlo techniques were employed. In particular, empirical distributions were estimated using order statistics. To empirically estimate density functions, the nonparametric density estimator (Rosenblatt (1956)) was integrated with a simple Monte

Carlo approach, and the empirical measures of location and dispersion were estimated on the basis of their sample definitions. In each of the experiments, the number of replications used was chosen using the Kolmogorov-Smirnov statistic for empirical distribution functions, and a technique similar to this for the empirical density functions, and empirical measures of dispersion. On the basis of these statistics, it is considered that the results obtained are accurate to at least two decimal places.

12.3 RESULTS AND CONCLUSIONS OBTAINED

This section summarizes the results and conclusions obtained in each of the models considered.

(i) The Limited-Information Simultaneous Equations Model

The Limited-Information SEM considered in this thesis is defined as the structural equation,

$$y_1 = y_2\alpha + X_1\gamma_1 + u \qquad (3.1)$$

where y_1 and y_2 are N-component vectors of observations on the endogenous variables, X_1 is a $N * G_1$ matrix of observations on exogenous variables, α is a scalar parameter, γ_1 is a G_1-component vector of parameters and u is a N-component vector of structural disturbances. Further, the reduced-form of the system of structural equations includes,

$$(y_1, y_2) = (X_1, X_2)\begin{bmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{bmatrix} + (v_1, v_2)$$

where X_2 is a $N * G_2$ matrix of observations on G_2 exogenous variables, π_{11}, π_{12}, π_{21}, π_{22} are reduced-form coefficients, and (v_1, v_2) is a $N * 2$ matrix of reduced-form disturbances. A number of assumptions were made. In particular it was assumed that, the rows of the reduced-form disturbances are independently normally distributed, each row having mean 0 and non- singular covariance matrix, I, (that is, the model was assumed to be in canonical form); the matrix $X = (X_1, X_2)$ is of rank (G_1+G_2) and finally it was assumed (3.1) was identified by exclusion type restrictions.

This model was analyzed in two stages. In the first stage, a useful method of numerically evaluating the distribution function of many of the commonly used estimators of α was presented. These estimators include those that can be written as a ratio of quadratic forms, so that existing numerical algorithms, such as those of Imhof (1961) and Davies (1980) can be used to calculate exact points of the distribution function. This method was applied to estimators in both correctly-specified and misspecified limited-information SEM's. An example of estimators included is the DK family where,

$$\hat{\alpha}_{DK} = (y_2'A_1y_2)^{-1}(y_2'A_2y_1) , \qquad (3.3)$$

for $A_j = K_j(P_X - P_{X_1}) + (1-K_j)\bar{P}_{X_1}$, $j = 1, 2$; $P_D = D(D'D)^{-1}D'$, and $\bar{P}_D = I - P_D$, for any matrix D of full column rank, and nonstochastic K_1 and K_2. This class of estimators provides considerable appeal as a summary statement of several commonly used estimators, including TSLS, which is equal to (3.3) when $K_1 = K_2 = 1$.

In the second stage, comparisons were made between the TSLS and LIML estimators, where the LIML estimator is a member of DK corresponding to stochastic parameters $K_1 = K_2 = \hat{1}$, where $\hat{1}$ is the smallest root of the determinental equation,

$$|Y_\Delta' P_{X_1} Y_\Delta - \hat{l} Y_\Delta' P_X Y_\Delta| .$$

It is well known, (see e.g. Phillips (1982)) that the distribution functions of these estimators in a correctly specified limited-information SEM depend only on a small number of 'key parameters'. For the TSLS estimator, these 'key parameters' are, K_2, α and a noncentrality parameter δ^2, which is related to the proportion of the variation in y_2 explained in its reduced form equation by the excluded exogenous variables X_2. The 'key parameters' of the LIML estimator are the same as for TSLS, plus N - G. Anderson et al. (1982) compare the distribution functions of these asymptotically equivalent estimators by covering a wide range of values for the key parameters. They conclude from this comparison that the LIML estimator is a superior estimation technique to the TSLS estimator. This is because the distribution of the LIML estimator approaches its limit distribution much faster than TSLS and furthermore, LIML is essentially median-unbiased whereas the distribution of the TSLS estimator is, in general, badly distorted.

In this thesis, this comparison was extended to misspecified limited- information SEM's. In particular, it was assumed that (3.1) is misspecified by the exclusion of relevant exogenous variables. This is an important area of analysis as typically in applied econometric studies, economic theory provides some guidance, but falls short of specifying the precise form of structural relationship. The key parameters of the misspecified distributions were identified. These are seen to be the same as those for the correctly specified case, plus a number of combinations of reduced-form parameters associated with the incorrectly excluded exogenous variables.

The distribution functions for the estimators were calculated, using the method given in the first stage for the TSLS estimator, and using Monte Carlo methods for the LIML estimator. From these computations it was concluded that under this type of misspecification, the LIML estimator is generally more robust than is the TSLS estimator, as it is better located around the true parameter value. Hence, the superiority of the LIML estimator is maintained in the presence of misspecification.

Finally, the numerical results obtained were shown to be applicable to the analysis of other types of misspecification, specifically the inclusion of irrelevant exogenous variables and a combination of inclusion and exclusion of relevant exogenous variables. In each of these cases, the LIML estimator is robust to misspecification.

(ii) The Nonnormal Linear Regression Model

Recently, models with possible nonnormally distributed disturbances have attracted more attention. This is because there is a large body of empirical literature, (e.g. Mandelbrot (1963, 1967, 1969), and Fama (1963, 1965)), which suggests that many economic time series are well represented by nonnormal disturbances.

In particular, to broaden the assumption of normality in the linear regression model,

$$y = X\beta + \epsilon \tag{3.4}$$

where $y' = (y_1,\ldots,y_N)$, X is an N * K matrix of nonstochastic regressors, $\beta' = (\beta_1,\ldots,\beta_K)$ is a vector of unknown parameters and $\epsilon' = (\epsilon_1,\ldots,\epsilon_N)$ is a vector of homoskedastic and serially-independent distributed disturbances; it has often been assumed that the error components follow a joint multivariate elliptical distribution, of the form,

$$\mathrm{pdf}(\epsilon) = C_N\, g\left[(\epsilon'\epsilon)\right] ,$$

where g is a one-dimensional real-valued function independent of N and C_N is a scalar proportionality constant. The results obtained, (e.g. Zellner (1976), Thomas (1970)), indicate that provided the resulting likelihood function is a monotonically decreasing function of $\epsilon'\epsilon$, the maximum likelihood estimator of β is the same as for normally distributed disturbances. This illustrates the robustness of the OLS estimator of β, in the presence of nonnormality.

However, the marginal distributions of the disturbance terms which are multivariate elliptically symmetric distributed, are identical to those when it is assumed the disturbances are distributed identically and independently elliptically symmetric, that is, when it is assumed

$$\mathrm{pdf}(\epsilon) = [C_1]^N \mathrm{pdf}(\epsilon_1) \ \ldots\ \mathrm{pdf}(\epsilon_N).$$

In this case it is well known (see e.g. Judge (1985)), that the OLS estimator is asymptotically inefficient, and furthermore, a class of 'robust estimation' methods have been introduced which possess superior asymptotic properties to OLS. These differences suggest it is important to distinguish between 'jointly-distributed' and 'iid distributed' disturbances, as they lead to quite different estimation techniques.

Using (3.4) the statistical consequences of distinguishing between "jointness" and "independence" was considered when it was assumed the disturbances were Student-t distributed. The Student-t distribution belongs to the elliptically symmetric family, and furthermore, this distribution is an important nonnormal distribution as it is considered that it is a reasonable way of modelling tails that are fatter than those of the normal distribution (see e.g. Jeffreys (1961)). This is relevant for many economic data series such as prices in financial and commodity markets, (see e.g. Judge et al. (1985, p.825) and Lange et al. (1989)).

There were a number of stages used to develop the statistical consequences of distinguishing between "jointness" and "independence". First, finite-sample properties of the appropriate maximum likelihood estimator under each assumption were considered, and these properties were also compared with a number of "general type robust estimators". Secondly, these properties were then used to consider the consequences of misspecifying the "jointness" and "independence" assumptions. Finally, specification tests were presented, which test for "jointness versus independence". These tests are applicable for the elliptically symmetric family, in general.

Each of these stages was examined for two separate cases of the linear regression model. In the first case, the location-scale model was assumed, which corresponds to (3.4) when $K = 1$, and in the second case, the more general model was assumed, which corresponds to $K > 1$. This distinction was made because a number of techniques could be used to examine this topic, (e.g. order statistics), in the location-scale model that are not applicable in the more general case. A further distinction was made on the basis of the variance of the disturbances. In particular, two cases were considered, one where the variance of the disturbances is finite, and the other where it is infinite. This distinction follows Judge (1985, p.823), and was made because the consequences of misspecification were generally more serious for the infinite-variance case than the finite- variance case.

Generally in both models the consequences of misspecifying the "jointness" versus "independence" assumptions can be summarized as follows. First, suppose the disturbances are assumed to be independently-distributed Student-t, but are actually jointly-distributed Student-t. Then the maximum likelihood estimator associated with independently-distributed Student-t disturbances is used in estimation, denoted $\hat{\beta}_{ML}$, where this estimator belongs to a class of "robust estimators". However, the "correct" maximum likelihood estimator to use in this case is the OLS estimator, denoted b. The use of a "robust estimator" rather than the OLS estimator results in an inefficient estimator, and the inefficiency increases as the number of regressors in the model increases. Furthermore, the variances used to estimate the actual variances are based on the use of the distribution of $\hat{\beta}_{ML}$ assuming the disturbances are independently distributed. This results in estimates of variances that seriously underestimate the actual variances, and consequently one concludes that the estimator is substantially more precise than it actually is. On the other hand, when the disturbances are assumed to be jointly-distributed

Student-t but are actually independently- distributed Student-t, the OLS estimator is used in estimation. However, in this case it is well known that b is in general asymptotically inefficient, and the use of robust procedures such as $\hat{\beta}_{ML}$ has been suggested. In particular, it is shown here that in finite samples the OLS estimator is inefficient with respect to $\hat{\beta}_{ML}$, although this inefficiency decreases as K increases, and for large K corresponding to fixed N, b is more efficient. However, the distribution of b will be assumed to be as for jointly- distributed disturbances. This assumption was shown to be incorrect, and the use of this incorrect assumption will have implications for inference.

Although the extent of these consequences depends on the particular Student-t distribution and K assumed, in general they are serious. Hence, this justifies the use of specification tests to test for "jointness versus independence", just as "serial independence versus autocorrelation" and "homoskedasticity versus heteroskedasticity" are tested for. In particular, King's (1980) invariance property of statistics for elliptically symmetric disturbances was used to adopt existing tests of normality to test for "jointness versus independence". An examination of the size and power of these tests using Student-t disturbances showed them to be useful for moderately-sized regression models.

(iii) The Nonnormal Limited-Information Simultaneous Equations Model

The nonnormal limited-information SEM provides a relatively new area of analysis as there are few published results available on the effects of nonnormal disturbances in the limited-information

SEM (see e.g. Knight (1985b, 1986), Raj (1980), Donatos (1989)).

The objective here was simply to combine the themes that were pursued separately in the limited-information SEM and nonnormal linear regression model. That is, we considered the finite-sample distribution of the LIMLK estimator of α in (3.1) when the corresponding reduced-form disturbances are assumed either to be multivariate Student-t distributed or iid Student-t distributed. The LIMLK corresponds to the LIML estimator when it is assumed the covariance matrix of the reduced-form disturbances is known. This estimator was considered rather than the LIML estimator itself, because in the cases when the LIMLK estimator is not equivalent to the LIML estimator, it is numerically easy to compute, and it is considered that the distribution functions of the two estimators will have similar features.

Further, it was also assumed (3.1) was exactly-identified by exclusion-type restrictions. An exactly-identified model was chosen simply as a way of narrowing the range of possible models to consider. More importantly, the exactly-identified model has a number of interesting properties when the reduced-form disturbances are assumed to be normally distributed. Hence, it was interesting to see how these properties changed when the assumption of normally distributed disturbances was widened to Student-t disturbances. In particular, when the disturbances are normally distributed the following properties of the LIMLK estimator were illustrated:

(1) - the LIMLK estimator reduced to ILS (which corresponds to (3.2) when $K_1 = K_2 = 1$ and $G_2 = 1$).

(2) - the computations presented, indicated that the limiting

distribution was a good approximation to the finite-sample distribution.

(3) - the computations of Marsaglia (1965) illustrated that the distribution was bimodal over part of the parameter space.

However, when the reduced-form disturbances were assumed to be Student-t distributed the following properties were obtained. When the disturbances were multivariate Student-t distributed the LIMLK estimator was ILS, whereas when the disturbances were iid Student-t distributed, the LIMLK estimator needed to be computed numerically. Further, each of the estimators converged to different limiting distributions, although the computations of the finite-sample distributions of each of these estimators indicated that they both had distributions with similar properties to the LIMLK estimator when the reduced-form disturbances were normally distributed. That is, in each case, the limiting distribution was a good approximation to the finite-sample distribution and the distribution was bimodal over part of the parameter space.

Therefore, as in the nonnormal linear regression model, the assumptions of joint and independent disturbances lead to quite different estimation methods that have different properties. Hence we also considered the consequences of misspecifying the type of Student-t assumption.

The consequences of misspecifying the type of Student-t assumption can be summarized as follows. If the reduced-form disturbances are assumed to be jointly Student-t distributed, but are actually independently Student-t distributed, the effects on the resulting maximum likelihood estimator are two-fold. This estimator is slow to locate around the true parameter value and an incorrect asymptotic distribution will be used to approximate the finite-sample distribution, which will have implications for inference. On the other hand, when the reduced-form disturbances are assumed to be independently Student-t distributed but are actually jointly Student-t distributed, the resulting maximum likelihood estimator used is robust in the sense that its finite-sample distribution is essentially identical to the correct maximum likelihood estimator. However, once again an incorrect asymptotic distribution will be used to approximate the finite-sample distribution of the maximum likelihood estimator used, which will have implications for inference.

These results suggested that it would be worthwhile to have specification tests that distinguish between jointly-distributed reduced-form disturbances and independently-distributed reduced-form disturbances. Tests of "jointness versus independence" in the exactly-identified limited- information SEM were constructed by applying King's (1980) invariance property of statistics for elliptically symmetric disturbances, to existing tests of multivariate normality. An examination of the size and power of these tests using Student-t disturbances showed that this was a useful method of testing this assumption in the exactly-identified limited- information SEM.

12.4 SOME FURTHER ISSUES

This thesis has extended and developed finite-sample properties of estimators used in three well known econometric models. However, even within these models a number of interesting aspects still remain to be considered. Some of these warrant brief mention.

In the Limited-Information SEM, results were obtained assuming that the structural equation of interest contained only two endogenous variables, and all of the predetermined variables in the system were strictly exogenous. Therefore it would be of interest to see how the results obtained change when these two assumptions are relaxed, that is, when there are three or more endogenous regressors included in the structural equation of interest, and/or dynamic predetermined variables are included in the system.

In the Nonnormal Linear Regression model, the statistical consequences for estimation, of distinguishing between jointly-distributed and independently-distributed nonnormal disturbances were considered. It would also be of interest to consider the implications of this distinction for inference. This is of particular interest for some distributions with infinite-variance, as the misspecification of the type of distribution in this case leads to the use of bimodal distributions under the null hypothesis, (see e.g. Logan et al. (1973)). Another interesting topic to pursue for the iid nonnormal linear regression model is an analysis of the implications of increasing the number of regressors on the resulting finite-sample mean-square error (provided this measure exists), as the evidence presented in this thesis suggests that the "robust estimators" become inefficient with respect to the ordinary least squares estimator.

In the Nonnormal Limited-Information SEM, a general theory for robust estimators for iid nonnormal disturbances needs to be developed. While there is a rapidly growing literature on robust estimation for univariate linear models, the multivariate case has received little attention. Therefore, a comprehensive treatment of the multivariate case is long overdue, and as well as theory, this should also include computational aspects.

REFERENCES

Amemiya, T. (1982), "Two Stage Least Absolute Deviations Estimators", *Econometrica*, 50, 689-712.

Amemiya, T. (1985), *Advanced Econometrics*, Basil Blackwell Ltd, Oxford.

Anderson, T.W. (1982), "Some recent developments on the distribution of single-equation estimators", in *Advances in Econometrics: Invited Papers for the Fourth World Congress of the Econometric Society*, ed. W. Hildenbrand, 109-122, Cambridge University Press, Cambridge.

Anderson, T.W., K. Fang and H. Hsu (1986), "Maximum likelihood estimators and likelihood-ratio criteria for multivariate elliptically contoured distributions", *The Canadian Journal of Statistics*, 14, 55-59.

Anderson, T.W., N. Kunitomo and T. Sawa (1982), "Evaluation of the distribution function of the Limited-Information Maximum Likelihood Estimator", *Econometrica*, 50, 1009-1028.

Anderson, T.W., N. Kunitomo and K. Morimume (1986), "Comparing Single Equation Estimators in a Simultaneous Equation System", *Econometric Theory*, 2, 1-32.

Anderson, T.W. and T. Sawa (1973), "Distributions of estimators of coefficients of a single equation in a simultaneous system and their asymptotic expansions", *Econometrica*, 41, 683-714.

Anderson, T.W. and T. Sawa (1977), "Two-Stage Least Squares: In which direction should the residuals be minimized?" *Journal* *of the American Statistical Association*, 72, 187-191.

Anderson, T.W. and T. Sawa (1979), "Evaluation of the distribution function of the Two-Stage Least Squares Estimate", *Econometrica*, 47, 163-18?

Anderson, T.W. and H. Rubin (1949), "Estimation of the parameters of a single equation in a complete system of stochastic equations", *Annals of Mathematical Statistics*, 20, 46-63.

Andrews, D.F., P.J. Bickel, F.R. Hampel, P.J. Huber, W.H. Rogers and J.W. Tukey (1972), *Robust Estimators of Location*, Princeton: Princeton University Press.

Andrews, D.K. and P.C.B. Phillips (1987), "Best Median-Unbiased Estimators in Linear Regression with Bounded Asymmetric Loss Functions", *Journal of the American Statistical Association*, 82, 886-893.

Antle, C.E. and L.J. Bain (1969), "A property of Maximum Likelihood estimators of location and scale parameters", *SIAM Review*, 11, 251-253.

Bai, Z.D. and X.R. Chen (1987), "Necessary and sufficient conditions for the convergence of integrated and mean-integrated p-th order error of the Kernel density estimates", Centre for Multivariate Analysis, Technical Report No. 87-06, University of Pittsburg.

Barnett, V.C. (1966), "Order statistics estimators of the location of the Cauchy distribution", *Journal of the American Statistical Association*, 61, 1205-1218, Correction 63, 383-385.

Barrodale, I. and F.D.K. Roberts (1974), "Solution of an overdetermined system of equations in the l_1-norm", *Communications of the Association of Computing Machinery*, 17, 319-320.

Bartels, R. (1977), "On the use of Limit Theorem Arguments in Economic Statistics", *American Statistician*, 31, 85-87.

Bartlett, M.S. (1935), "The effect of nonnormality on the t-distribution, *Proceedings of the Cambridge Philosophical Society*, 31, 223-231.

Basmann, R.L. (1961), "A note on the exact finite-sample frequency functions of the Generalized Classical Linear Estimators in Two Overidentified cases", *Journal of the American Statistical Association*, 56, 619-636.

Basmann, R.L. (1963), "Remarks concerning the application of exact finite- sample distribution functions of GCL estimators in Econometric Statistical Inference", *Journal of the American Statistical Association*, 58, 943-976.

Basmann, R.L. (1974), "Exact finite sample distributions for some econometric estimators and test statistics: A survey and appraisal", Chapter 4 of *Frontiers of Quantitative Economics*, Vol. II, eds. M.D. Intriligator and D.A. Kendrick, Amsterdam: North Holland, 209-285.

Benjamini, Y. (1983), "Is the t-test really conservative when the parent distribution is long-tailed?", *Journal of the American Statistical Association*, 78, 645-654.

Berger, J.O. (1975), "Minimax estimation of location vectors for a wide class of densities", *The Annals of Statistics*, 3, 1318-1328.

Bergstrom, A.R. (1962), "The exact sampling distribution of Least Squares and Maximum Likelihood estimators of the marginal propensity to consume", *Econometrica*, 30, 480-490.

Bickel, P.J. (1976), "Another look at robustness", *Scandinavian Journal of Statistics*, 3, 145-168.

Bierens, H.J. (1986), "Kernel estimators of regression functions", in *Advances in Econometrics - Fifth World Congress*, Vol. 1, ed. T.F. Bewley, Cambridge University Press, Cambridge, 99-146.

Blom, G. (1958), *Statistical Estimates and Transformed Beta-variables*, Wiley, New York.

Bondesson, L. (1976), "When is the sample mean BLUE?", *Scandinavian Journal of Statistics*, 3, 116-120.

Bondesson, L. (1983), "When is the t-statistic t-distributed?", *SANKHYA The Indian Journal of Statistics*, Series A, 45, 338-345.

Bowden, R.J. and D.A. Turkington (1984), *Instrumental Variables*, Cambridge University Press, Cambridge.

Bowman, K.O. and L.R. Shenton (1975), "Omnibus Tests contours for departures from normality based on $\sqrt{b_1}$ and b_2", *Biometrika*, 62, 243-250.

Brandwein, A.C. (1979), "Minimax estimation of the mean of spherically symmetric distributions under general quadratic loss", *Journal of Multivariate Analysis*, 9, 579-588.

Brandwein, A.C. and W.E. Strawderman (1978), "Minimax estimation of location parameters in spherically symmetric unimodal distributions under quadratic loss", *The Annals of Statistics*, 6, 377-416.

Brandwein, A.C. and W.E. Strawderman (1980), "Minimax estimation of location parameters for spherically symmetric distributions with concave loss", *The Annals of Statistics*, 8, 279-284.

Bratley, B., B.L. Fox and L.E. Schrage (1983), *A Guide to Simulation*, Springer-Verlag, New York.

Breiman, L., W. Meisel and E. Purcell (1977), "Variable kernel estimators of multivariate densities", *Technometrics*, 19, 135-144.

Brent, R.B. (1974), "A Gaussian pseudo-random number generator", *Communications of the Association of Computing Machinery*, 17, 704-706.

Brown, L.D. (1966), "On the admissibility of invariant estimators of one or more location parameters", *The Annals of Mathematical Statistics*, 37, 1087-1136.

Broyden, C.G. (1970), "The convergence of a class of double rank minimization algorithms, 2. The new algorithm". *Journal of the Institute of Mathematical Applications*, 6, 222-231.

Cacoullos, T. (1966), "Estimation of a multivariate density", *Annals of the Institute of Mathematical Statistics*, 18, 176-189.

Cambanis, S., S. Huang and G. Simons (1981), "On the theory of Elliptically Contoured Distributions", *Journal of Multivariate Analysis*, 11, 368-385.

Chmielewski, M.A. (1981), "Elliptically Symmetric Distributions", *International Statistical Review*, 49, 67-74.

Coope, I.D. (1987), "A conjugate direction implementation of the BFGS algorithm with automatic scaling", *Research Report, Department of Mathematics, University of Canterbury.*

Copas, J.B. (1975), "On the unimodality of the likelihood for the Cauchy distribution", *Biometrika*, 62, 701-704.

Cramér, H. (1946), *Mathematical Methods of Statistics*, Princeton University Press, Princeton.

Cressie, N. (1980), "Relaxing assumptions in the one-sample t-test", *Australian Journal of Statistics*, 22, 143-153.

Cribbett, P.F., J.N. Lye and A. Ullah (1989), "Evaluation of the Two-Stage Least Squares Distribution Function by Imhof's Procedure", *Journal of Quantitative Economics*, 5, 91-96.

Cox, D.R. and D.V. Hinkley (1974), *Theoretical Statistics*, Chapman and Hall, London.

Cox, D.R. and N.J.H. Small (1978), "Testing multivariate normality", *Biometrika*, 65, 263-272.

Dagpunar, J. (1988), *Principles of Random Variate Generation*, Oxford University Press, New York.

David, H.A. (1970), *Order Statistics*, John Wiley, New York.

Davies, R.B. (1973), "Numerical inversion of a characteristic function", *Biometrika*, 60, 415-417.

Davies, R.B. (1980), "Alg.AS155. The distribution of a linear combination of χ^2 random variables", *Applied Statistics*, 29, 323-333.

Davis, A.W. (1976), "Statistical distributions in univariate and multivariate Edgeworth populations", *Biometrika*, 63, 661-670.

Davis, C.S. and M.A. Stephens (1978), "Algorithm AS 128, Approximating the covariance matrix of normal order statistics", *Applied Statistics*, 27, 206-212.

Davis, K.B. (1975), "Mean square error properties of density estimates", *Annals of Statistics*, 3, 1025-1030.

Davis, K.B. (1977), "Mean integrated square error properties of density estimates", *Annals of Statistics*, 5, 530-535.

Dawid, A.P. (1977), "Spherical matrix distributions and a multivariate model", *Journal of the Royal Statistical Society, Series B*, 39, 254-261.

D'Agostino, R.B. and E.S. Pearson (1973), "Tests for departures from normality. Empirical results for the distribution of b_2 and $\sqrt{b_1}$", *Biometrika*, 60, 613-222.

De Groot, M.H. (1970), *Optimal Statistical Decisions*, McGraw-Hill, New York.

De Moivre, A. (1733), *Miscellanea Analytica*, second supplement.

Deheuvels, P. (1974), "Conditions necessoires et suffisante de convergence ponctuelle presque sure et uniforme sure des estimateurs de la densité, *Comptes Rendus Academie des Sciences de Paris*, Ser. A, 178, 1217-1220.

Devroye, L. (1986), *Non-Uniform Random Variate Generation*, Springer-Verlag.

Devroye, L. (1987), *A Course in Density Estimation*, Donnelley, Harrisonburg.

Devroye, L. and T.J. Wagner (1976), "Nonparametric discrimination and density estimation", *Technical Report 183*, University of

Texas, Austin.

Dhrymes, P. (1970), *Econometrics*, Harper and Row, New York.

Donatos, G.S. (1989), "A Monte-Carlo study of K-class estimators for small- samples with normal and nonnormal disturbances", *The Statistician*, 38, 11-20.

Duin, R.P.W. (1976), "On the choice of smoothing parameters for Parzen estimators of probability density functions", *IEEE Transactions on Computers*, C-25, 1175-1179.

Durbin, J. and G.S. Watson (1950), "Testing for serial correlation in Least Squares Regression I", *Biometrika*, 37, 409-428.

Dyer, A.R. (1974), "Comparisons of tests for normality with a cautionary note", *Biometrika*, 61, 185-189.

Eaton, M.L. (1983), *Multivariate Statistics: A Vector Space Approach*, Wiley, New York.

Efron, B. (1969), "Student's t-test under symmetry conditions", *Journal of the American Statistical Association*, 73, 536-544.

Epanechnikov, V.A. (1969), "Nonparametric estimates of a multivariate probability density", *Theory of Probability and Applications*, 14, 153-158.

Epstein, R.J. (1987), *A History of Econometrics*, North-Holland, Amsterdam.

Fama, E.F. (1963), "Mandelbrot and the stable Paretian hypothesis", *Journal of Business*, 36, 420-429.

Fama, E.F. (1965), "The behaviour of stock market prices", *Journal of Business*, 38, 34-105.

Fama, E.F. (1970), "Efficient capital markets: A review of theory and empirical work", *Journal of Finance*, 25, 383-417.

Farebrother, R.W. (1983), "A remark on Algorithm AS155", *Department of Econometrics*, University of Manchester.

Feller, W. (1966), *An Introduction to Probability Theory and its Applications*, Vol. 2, John Wiley and Sons, New York.

Feller, W. (1968), *An Introduction to Probability Theory and its Applications*, Vol. 1, 3rd edition, John Wiley and Sons, New York.

Ferguson, T.S. (1978), "Maximum Likelihood Estimators of the parameters of the Cauchy distribution for samples of size 3 and 4", *Journal of the American Statistical Society*, 73, 211-213.

Fieller, E.C. (1932), "The distribution of the index in a normal bivariate population", *Biometrica*, 24, 428-440.

Fisher, F.M. (1961), "On the cost of approximate specification in simultaneous equation estimators", *Econometrica*, 29, 139-170.

Fisher, F.M. (1966), "The relative sensitivity to specification error of different K-class estimators", *Journal of the American Statistical Association*, 61, 345-356.

Fisher, F.M. (1967), "Approximate specification and the choice of a K-class estimator", *Journal of the American Statistical Association*, 62, 1265-1276.

Fisher, R.A. (1925), *Statistical Methods for Research Workers*, Oliver and Boyd, Edinburgh.

Fletcher, R. (1970), "A new approach to variable metric algorithms", *Computer Journal*, 13, 317-322.

Forsythe, A.B. (1972), "Robust estimation of straight line regression coefficients by minimizing p-th power deviations", *Technometrics*, 14, 159-166.

Fraser, D.A.S. and K.W. Ng (1980), "Multivariate regression analysis with spherical errors", in *Multivariate Analysis 5*, ed. P.R. Krishnaiah, North-Holland, New York, 369-386.

Frisch, R. (1933), *Pitfalls in the Statistic Construction of Demand and Supply Curves*, Hans Burke, Leipzig.

Frisch, R. (1934), *Statistical Regression Analysis by Means of Complete Regression Systems*, University Economics Institute, Oslo.

Fuller, W.A. (1977), "Some properties of a modification of the Limited- Information Estimator", *Econometrica*, 45, 939-953.

Gabrielsen, G. (1982), "On the unimodality of the likelihood for the Cauchy distribution: Some comments", *Biometrika*, 69, 677-678.

Gasser, T. and H. Müller (1979), "Kernel estimation of regression functions" in *Smoothing Techniques for Curve Estimation: Lecture Notes in Mathematics 757*, ed. T. Gasser and M. Rosenblatt (Springer-Verlag, Berlin).

Gauss, C.F. (1809), *Theoria motus corporum celestium*, Hamburg, Perthes et Besser, translated 1857 as *Theory of Motion of the Heavenly Bodies Moving About the Sun in Conic Sections* , translator C.H. Davis, Boston: Little Brown.

Gayen, A. (1949), "The distribution of Student's-t in random samples of any size drawn from non-normal universes", *Biometrika*, 36, 353-369.

Geary, R.C. (1930), "The frequency distribution of the quotient of two normal variates", *Journal of the Royal Statistical Society*, 9, 442-446.

Geary, R.C. (1947), "Testing for normality", *Biometrika*, 34, 209-220.

Gerald, C.F. and P.O. Wheatley (1984), *Applied Numerical Analysis* (Addison- Wesley, Reading, 3rd edition).

Giles, J.A. (1990), *Preliminary - Test Estimation of a Mis-Specified linear Model with Spherically Symmetric disturbances*, Ph.D. Thesis, Department of Economics, University of Canterbury.

Gil - Pereaz, J. (1951), "Note on the inversion theorem", *Biometrika*, 38, 481-482.

Godfrey L.G. and M.R. Wickens, (1982), "A simple derivation of the limited information maximum likelihood estimator', *Economics Letters*, 10, 277-284.

Goldfard, D. (1970), "A family of variable metric methods derived by variational means", *Mathematics of Computation*, 24, 23-26.

Goldfeld, S.M. and R.E. Quandt (1965), "Some tests for homoskedasticity", *Journal of the American Statistical Association*, 60, 539-547.

Haas, G., L. Bain and C. Antle (1970), "Inferences for the Cauchy distribution based on maximum likelihood estimators", *Biometrika*, 403-408.

Haavelmo, T. (1943), "The statistical implications of a system of simultaneous equations", *Econometrica*, 11, 1-12.

Haavelmo, T. (1944), "The probability approach in Econometrics", *supplement to Econometrica*, 12.

Haavelmo, T. (1947), "Methods of measuring the marginal propensity to consume", *Journal of the American Statistical Association*, 42, 105-122.

Hale, C., R.S. Mariano and J.G. Ramage (1980), "Finite-sample analysis in simultaneous equation models", *Journal of the American Statistical Association*, 75, 418-427.

Hall, P. (1983), "Large sample optimality of least-squares cross-validation in density estimation", *Annals of Statistics*, 11, 1156-1174.

Harvey, A.C. (1981), *The Econometric Analysis of Time Series*, Halsted Press, New York.

Harvey, A.C. and G.D.A. Phillips (1980), "Testing for serial correlation in Simultaneous Equations Models; some further results", *Journal of Econometrics*, 17, 99-105.

Healy, M.J.R. (1968), "Multivariate normal plotting", *Applied Statistics*, 17, 157-161.

Hill, R.W. and P.W. Holland (1977), "Two robust alternatives to least squares regression", *Journal of the American Statistical Association*, 72, 828-833.

Hillier, G. (1985), "On the Joint and Marginal Densities of instrumental variable estimators in a general structural equation", *Econometric Theory*,1, 53-72.

Hillier, G. (1987), "Joint distribution theory for some statistics based on LIML and TSLS", *Cowles Foundation Discussion Paper*, No. 840.

Hillier, G. (1988), "On the interpretation of exact results for structural equation estimators", W.P. No. 88, *Department of Econometrics and Operations Research, Monash University*.

Hillier, G., T.W. Kinal and V.K. Srivastava, (1984), "On the moments of Ordinary Least Squares and Instrumental Variable Estimators in a general structural equation", *Econometrica*, 52, 185-202.

Hotelling, H. (1961), "The behaviour of some standard statistical tests under nonstandard conditions", *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, 1, 319-360.

Huang, C.J. and B.W. Bolch (1974), "On the testing of regression disturbances for normality", *Journal of the American Statistical Association*, 69, 330-335.

Huber, P.J. (1972), "Robust statistics: A review", *The Annals of Mathematical Statistics*, 43, 1041-1067.

Huber, P.J. (1973), "Robust regression: Asymptotics, Conjectures and Monte Carlo", *The Annals of Statistics*, 1, 799-821.

Huber, P.J. (1977), *Robust Statistical Procedures*, SIAM, Philadelphia.

Huber, P.J. (1981), *Robust Statistics*, Wiley, New York.

Hurwicz, L. (1950), "Least Squares Bias in Time Series", in T.C. Koopmans (ed.), *Statistical Inference in Dynamic Economic*

Models, John Wiley, New York.

Imhof, J.P. (1961), "Computing the distribution of quadratic forms in normal variables", *Biometrika*, 48, 419-426.

Jaeckel, L.A. (1972), "Estimating regression coefficients by minimizing the dispersion of the residuals", *The Annals of Mathematical Statistics*, 42, 1328-1338.

Jammalamadaka, S.R., R.C. Tiwari and S. Chib (1987), "Bayes prediction in the linear model with spherically symmetric errors", *Economics Letters*, 24, 39-44.

Jammalamadaka, S.R. S. Chib and R.C. Tiwari (1988), "Bayes prediction in regressions with elliptical errors", *Journal of Econometrics*, 38, 349-360.

James, W. and C. Stein (1961), "Estimation with quadratic loss", *Proceedings of the Fourth Berkeley Symposium*, University of California Press, Berkeley, CA, 361-379.

Jarque, C.M. and A.K. Bera (1987), "A test for Normality of observations and regression residuals", *International Statistical Review*, 55, 163-172.

Jeffreys, H. (1961), *Theory of Probability*, Oxford, Clarendon.

Jensen, D.R. and I.J. Good (1988), "Invariant distributions associated with matrix laws under structural symmetry", *Journal of the Royal Statistical Society, Series B*, 43, 327-332.

Johnston, J. (1984), *Econometric Methods*, McGraw-Hill, Auckland, 3rd edition.

Johnston, N. (1978), "Modified t-tests and confidence intervals for asymmetrical populations", *Journal of the American Statistical*

Association, 73, 536-544.

Johnston, N.I. and S. Kotz (1970), *Continuous Univariate Distributions 1*, Houghton Mifflin Company, Boston.

Johnston, N.I. and S. Kotz (1972), *Continuous Multivariate Distributions*, Houghton Mifflin Company, Boston.

Judge, G.G., W.G. Griffiths, R. Carter Hill, H. Lütkepohl and Tsoung-Chao Lee (1985), *The Theory and Practice of Econometrics*, John Wiley and Sons, 2nd edition.

Judge, G.G., S. Miyazaki and T. Yancey (1985), "Minimax estimators for the location vectors of spherically symmetric densities", *Econometric Theory*, 1, 409-417.

Judge, G.G. and T.A. Yancey (1986), *Improved Methods of Inference in Econometrics*, North-Holland, Amsterdam.

Jung, J. (1962), "Approximation of least-squares estimators of location and scale parameters", in *Contributions to Order Statistics*, eds. A.G. Sarhan and B.G. Greenberg, New York, Wiley, 28-46.

Kadane, J.B. (1971), "Comparison of K-class estimators when the disturbances are small", *Econometrica*, 39, 723-737.

Kariya, T. (1981), "Robustness of multivariate tests", *Annals of Statistics*, 9, 1267-1275.

Kelejian, H.H. and I.R. Prucha (1984), "The structure of simultaneous equation estimators: A generalization towards nonnormal disturbances, *Econometrica*, 52, 721-735.

Kelejian, H.H. and I.R. Prucha (1985), "Independent or uncorrelated disturbances in linear regression: An illustration of the difference", *Economics Letters*, 19, 35-38.

Kelejian, H.H. and N. Adams (1989), "Pseudorandom number generation on supercomputers", FEW 378, Tilburg University.

Kelker, D. (1970), "Distribution theory of spherical distributions and a location-scale parameter generalization", *Sankhyā*, A, 32 419-430.

Kendall, M.G. and A. Stuart (1969), *The Advanced Theory of Statistics*, Vol. I, 3rd edition, Charles Griffen and Co., London.

Kinal, T.W. (1980), "The existence of moments of K-class estimators", *Econometrica*, 48, 240-249.

King, M.L. (1979), *Some Aspects of Statistical Inference in the Linear Regression Model*, Ph.D. thesis, Department of Economics, University of Canterbury.

King, M.L. (1980a), "Robust tests for spherical symmetry and their application to least squares regression", *Annals of Statistics*, 8, 1265-1271.

King, M.L. (1980b), "Small sample properties of econometric estimators and tests assuming elliptically symmetric disturbances", W.P. No. 15/80, *Department of Econometrics and Operations Research, Monash University*.

Knight, J. (1982), "A note on finite sample analysis of misspecification in simultaneous equation models", *Economics Letters*, 9, 275-279.

Knight, J. (1985a), "The joint characteristic function of linear and quadratic forms of nonnormal variables", *Sankhyā*, A, 231-238.

Knight, J. (1985b), "The moments of OLS and TSLS when the disturbances are non-normal", *Journal of Econometrics*, 27, 39-60.

Knight, J. (1986), "Non-normal errors and the distribution of OLS and TSLS structural estimators", *Econometric Theory*, 2, 75-106.

Koenker, R.W. (1982), "Robust methods in econometrics", *Econometric Reviews*, 1, 213-290.

Koenker, R.W. (1987), discussion on paper by A.H. Welsh, "The trimmed mean in the linear model", *Annals of Statistics*, 15, 20-43.

Koenker, R.W. and G.W. Bassett (1978), "Regression quantiles", *Econometrica*, 46, 33-50.

Koenker, R.W. and G.W. Bassett (1982), "Tests of linear hypotheses and l_1- estimation", *Econometrica*, 50, 1577-1583.

Koenker, R.W. and V. D'Orey (1987), "Algorithm AS229: Computing regression quantiles", *Applied Statistics*, 36, 383-393.

Koenker, R.W. and S. Portnoy (1988), "M-Estimation of multivariate regressions", *mimeo*, Department of Economics and Statistics, University of Illinois.

Koerts, J. and A.P.J. Abrahamse (1971), *On the Theory and Application of the General Linear Model*, Rotterdam University Press, Rotterdam.

Koopman, B.O. (1936), "On distributions admitting sufficient statistics", *Transactions American Mathematical Society*, 39, 399-409.

Lange, K.L., R.J.A. Little and J.M.G. Taylor (1989), "Robust statistical modelling using the t-distribution", *Journal of the American Statistical Association*, 881-896.

Laplace, P.S. (1812), *Théorie analytique des probabilités*, Paris, Courcier (3rd edition, 1820 with supplements, reprinted in P.S. Laplace 1878- 1912, Oeuvres Complétes de Laplace, 14 vols, Paris, Gauthier-Villars, Vol. 7).

Lehmann, E.L. (1983), *Theory of Point Estimation*, Wiley, New York.

Lévy, P. (1925), *Calcul des Probabilités*, Paris.

Lloyd, G.H. (1952), "Least-squares estimation of location and scale parameters using order statistics, *Biometrika*, 39, 88-95.

Logan, B.F., C.C. Mallows, S.O. Rice and L.A. Shepp (1973), "Limit distributions of self normalized sums", *Annals of Probability*, 1, 788-809.

Louter, A.S. and J. Koerts (1970), "On the Kuiper Test for normality with mean and variance unknown", *Statistical Neerlandica*, 24, 83-87.

Loynes, R.M. (1980), "The empirical distribution function of residuals from generalized regression", *Annals of Statistics*, 8, 285-298.

Lukacs E. and R.G. Laha (1964), *Applications of Characteristic Functions*, Griffin, London.

Lye, J.N. (1988), "On the exact distribution of a ratio of bilinear to quadratic forms in normal variables with econometric applications", *Paper presented to the International Economics Postgraduate Research Conference*, Department of Economics, The University of Western Australia.

Malkovich, J.F. and A.A. Afifi (1973), "On tests for multivariate normality", *Journal of the American Statistical Association*, 68, 176-179.

Mandelbrot, B.B. (1963a), "The variation of certain speculative prices", *Journal of Business*, 36, 394-419.

Mandelbrot, B.B. (1963b), "New methods in statistical economics", *Journal of Political Economy*, 71, 421-440.

Mandelbrot, B.B. (1966), "Forecasts of future prices, unbiased markets and martingale models", *Journal of Business*, 39, 242-255.

Mandelbrot, B.B. (1967), "The variation of some other speculative prices", *Journal of Business*, 40, 393-413.

Mandelbrot, B.B. (1969), "Long run linearity, locally Gaussian process, H-spectra and infinite variances", *International Economic Review*, 10, 82-111.

Mardia, K.V. (1980), "Tests of univariate and multivariate normality", in *Handbook of Statistics, Vol. 1*, ed. P.R. Krishnaiah, North-Holland, Amsterdam, 279-320.

Mariano, R.S. (1972), "The existence of moments of the ordinary least squares and two stage least squares estimators", *Econometrica*, 40, 643-652.

Mariano, R.S. (1973a), "Approximations to the distribution functions of the ordinary least squares and two-stage least squares estimators in the case of two included endogenous variables", *Econometrica*, 41, 67-77.

Mariano, R.S. (1973b), "Approximations to the distribution functions of Theil's K-class estimators", *Econometrica*, 41, 715-721.

Mariano, R.S. (1977), "Finite-sample properties of instrumental variable estimators of structural coefficients", *Econometrica*, 45, 487-496.

Mariano, R.S. (1982), "Analytical small-sample distribution theory in econometrics: The simultaneous equations case", *International Economic Review*, 23, 503-533.

Mariano, R.S. and J.G. Ramage (1978), "Ordinary least squares versus other single equation estimators: A return bout under misspecification in simultaneous systems", *Discussion Paper*, No. 400, University of Pennsylvania, Department of Economics.

Mariano, R.S. and J.G. Ramage (1983), "Specification error analysis in linear simultaneous equation models", eds. R.L. Basmann and G.F. Rhodes Jr., JAI Press, Greenwich, 51-96.

Mariano, R.S. and T. Sawa (1972), "The exact finite-sample distribution of the limited-information maximum-likelihood estimator in the case of two-included endogenous variables", *Journal of the American Statistical Association*, 67, 159-163.

Marsaglia, G. (1965), "Ratios of normal variables and ratios of sums of uniform variables", *Journal of the American Statistical Association*, 60, 193-204.

Mood, M.A., F.A. Graybill and D.C. Boes (1986), *Introduction to the Theory of Statistics*, McGraw-Hill, Auckland, 3rd edition.

Mosteller, F. and J.W. Tukey (1977), *Data Analysis and Regression*, Addison- Wesley, Reading, Mass.

Muirhead, R.J. (1982), *Aspects of Multivariate Statistical Theory*, John Wiley and Sons, New York.

Mukantseva, L.A. (1977), "Testing normality in one-dimensional and multi- dimensional linear regression", *Theory of Probability* *and Its Applications*, 22, 591-602.

Nagar, A.L. (1959), "The bias and moment matrix of the general K-class estimators of the parameters in simultaneous equations", *Econometrica*, 27, 525-595.

Nagar, A.L. (1962), "Double K-class estimators of parameters in simultaneous equations and their small-sample properties", *International Economic Review*, 3, 168-188.

Norden, R.H. (1972), "A survey of maximum likelihood estimation", *International Statistical Review*, 40, 329-354.

Parzen, E. (1962), "On the estimation of probability density and mode", *Annals of Mathematical Statistics*, 33, 1065-1076.

Pearson, G.S. and N.K. Adyantaya (1929), "The distribution of frequency constants in small samples from non-normal symmetrical and skew distributions", *Biometrika*, 21, 259-286.

Pearson, G.S., R.B. D'Agostino and K.O. Bowman (1977), "Tests for departures from normality: Comparison of powers", *Biometrika*, 231-246.

Peters, T.A. (1989), "The Exact Moments of OLS in Dynamic Regression Models with nonnormal errors", *Journal of Econometrics*, 40, 279-305.

Phillips, G.D.A. (1988), "Testing for serial correlation after Three Stage Least Squares Estimation", *Bulletin of Economic Research*, 40, 145-151.

Phillips, P.C.B. (1980a), "The exact distribution of instrumental variable estimators in an equation containing n+1 endogenous variables", *Econometrica*, 48, 861-878.

Phillips, P.C.B. (1980b), "Finite-sample theory and the distribution of alternative estimators of the marginal propensity to consume", *Review of Economic Studies*, 47, 183-224.

Phillips, P.C.B. (1982), "Small-sample distribution theory in econometric models of simultaneous equations", *Cowles Foundation Discussion Paper*, No. 617, Yale University.

Phillips, P.C.B. (1983), "Marginal densities of instrumental variable estimators in the general single equation case" in *Advances in Econometrics: Exact Distribution Analysis in Linear SEM's*, Vol. 2, eds. R. Basmann and G.F. Rhodes, JAI Press Inc., 1-24.

Phillips, P.C.B. (1984a), "The exact distribution of Liml: I", *International Economic Review*, 25, 249-261.

Phillips, P.C.B. (1984b), "The exact distribution of exogenous variable coefficient estimators", *Journal of Econometrics*, 26, 387-398.

Phillips, P.C.B. (1985), "The exact distribution of Liml: II", *International Economic Review*, 26, 21-36.

Phillips, P.C.B. and V. Hajivassiliou (1987), "Bimodal t-ratios", *Cowles Foundation Discussion Paper* No. 842.

Pierce, D.A. and K.J. Kopecky (1979), "Testing goodness of fit for the distribution of errors in regression models", *Biometrika*, 66, 1-5.

Pitman, E.J.G. (1936), "Sufficient statistics and intrinsic accuracy", *Proceedings of the Cambridge Philosophical Society*, 32, 567-579.

Poincaré, H. (1912), *Calcul des Probabilités*, Paris, Gauthier-Villars, 2nd edition.

Powell, J.L. (1983), "The asymptotic normality of two-stage least absolute deviations estimators", *Econometrica*, 51, 1569-1576.

Press, S.J. (1969), "The t-ratio distribution, *Journal of the American Statistical Association*, 64, 242-252.

Putter, J. (1967), "Orthonormal bases of error spaces and their use for investigating the normality and variances of residuals", *Journal of the American Statistical Association*, 62, 1022-1036.

Raj, B. (1980), "Simultaneous equation estimators with normal and nonnormal disturbances", *Journal of the American Statistical Association*, 75, 221-229.

Ramsey, J.B. (1969), "Tests for specification errors in classical linear least-squares regression analysis", *Journal of the Royal Statistical Society* B, 31, 350-371.

Ramsey, J.B. (1974), "Classical model selection through specification error", in *Frontiers of Econometrics*, ed. P. Zarembka, Academic, New York.

Rao, C.R. (1965), *Linear Statistical Inferences and its Applications*, Wiley, New York.

Rao, C.R. (1981), "Some comments on the minimum mean square error as a criterion of estimators", in *Statistics and Related Topics*, eds. M. Csorgo, D. Dawson, J. Rao and A. Saleh, Amsterdam, North-Holland, 123-143.

Rhodes, G. and D. Westbrook (1981), "A study of estimator densities and performance under misspecification", *Journal of Econometrics*, 16, 311-337.

Rhodes, G. and D. Westbrook (1983), "Simultaneous equations estimators, identifiability test-statistics and structural form", in *Advances in Econometrics Vol 2: Exact Distribution Analysis in Linear Simultaneous Equation Models*, eds. R.L. Basmann and G.F. Rhodes, JAI Press, Greenwich, 129-196.

Richardson, D.H. (1968), "The exact distribution of a structural coefficient estimator", *Journal of the American Statistical Association*, 63, 1214-1226.

Rosenblatt, M. (1956), "Remarks on some nonparametric estimates of density function", *Annals of Mathematical Statistics*, 27, 832-837.

Royston, J.P. (1982), "An extension of Shapiro and Wilk's test for normality to large samples", *Applied Statistics*, 31, 115-124.

Royston, J.P. (1983), "Some techniques for assessing multivariate normality based on the Shapiro-Wilk W", *Applied Statistics*, 32, 121-133.

Rubinstein, R.Y. (1981), *Simulation and the Monte-Carlo Method*, John Wiley, New York.

Ruppert, D. and J. Carroll (1980), "Trimmed least squares in the linear model", *Journal of the American Statistical Association*, 75, 825-838.

Sargan, J.D. (1970), "The finite-sample distribution of Fiml estimators", *Paper presented to the World Congress of the Econometric Society*, Cambridge.

Sarhan, A.E. and B.G. Greenberg (1956), "Estimation of location and scale parameters by order statistics from singly and doubly censored samples. Part I", *Annals of Mathematical Statistics*, 27, 427-451.

Sawa, T. (1969), "The exact distribution of ordinary least squares and two-stage least squares estimators", *Journal of the American Statistical Association*, 64, 923-937.

Sawa, T. (1972), "Finite-sample properties of K-class estimators", *Econometrica*, 40, 653-680.

Sawa, T. (1978), "The exact moments of the Least Squares estimate for the autoregressive model", *Journal of Econometrics*, 8, 159 - 172.

Schmidt, P. (1976a), *Econometrics*, Marcel Dekker Inc., New York.

Schmidt, P. (1976b), "On the statistical estimation of Parametric Frontier Production Functions", *Review of Economics and Statistics*, 58, 238-239.

Schrader, R.M. and T.P. Hettmansperger (1980), "Robust analysis of variance based upon a likelihood ratio criterion", *Biometrika*, 67, 93-101.

Scott, D.W. and L.E. Factor (1981), "Monte-Carlo study of three data-based nonparametric probability density estimators", *Journal of the American Statistical Association*, 76, 9-15.

Shanno, D.F. (1970), "Conditioning of quasi-Newton methods for function minimization", *Mathematics of Computation*, 24, 647-656.

Shapiro, S.S. and R.S. Francia (1972), "An approximate analysis of variance test for normality", *Journal of the American*

Statistical Association, 67, 215-216.

Shapiro, S.S. and M.B. Wilk (1965), "An analysis of variance test for normality", *Biometrika*, 52, 591-611.

Shapiro, S.S., M.B. Wilk and H.J. Chen (1968), "A comparative study of various tests of normality", *Journal of the American Statistical Association*, 63, 1343-1372.

Silverman, B.W., (1986), *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, New York.

Singh, R.S. (1987), "A family of improved estimators in linear regression models with errors having multivariate Student-t distribution", W.P., 4, *Department of Econometrics and O.R.*, Monash University.

Singh, R.S. (1988), "Estimation of error variance in linear regression models with errors having multivariate Student-t distribution with unknown degrees of freedom", *Economics Letters*, 27, 47-54.

Singh, R.S., A. Ullah and R.A.L. Carter (1987), "Nonparametric inference in econometrics: New applications", in *Time Series and Econometric Modelling*, eds. I. MacNeill and G. Umphrey, D. Reidel, Holland.

Skeels, C. (1988), *Estimation in misspecified systems of equations: some exact results*, Ph. D. thesis, Department of Econometrics, University of Monash.

Small, N.J.H. (1980), "Marginal skewness and kurtosis in testing multivariate normality", *Applied Statistics*, 29, 85-87.

Srivastava, A.B.L. (1958), "Effect of nonnormality on the power function of the t-test", *Biometrika*, 45, 421-429.

Srivastava, M.S. and T.K. Hui (1987), "On assessing multivariate normality based on Shapiro-Wilk W statistic", *Statistics and Probability Letters*, 5, 15-18.

Srivastava, V.K. and D.E.A. Giles (1987), *Seemingly Unrelated Regression Equations Models Estimation and Inference*, Marcel Dekker, Inc., New York.

Stein, C. (1955), "Inadmissibility of the usual estimator for the mean of a multivariate normal distribution", *Proceedings of the Third Berkeley Symposium*, University of California Press, Berkeley, CA, 197-206.

Stephens, M.A. (1974), "EDF statistic for goodness of fit and some comparisons", *Journal of the American Statistical Association*, 69, 730-737.

Stigler, S.M. (1973), "The history of robust estimation", *Journal of the American Statistical Association*, 68, 872-879.

Stigler, S.M. (1986), *The History of Statistics, The Measurement of Uncertainty Before 1900*, The Belknap Press of Harvard University Press, Cambridge.

Stone, C.J. (1984), "An asymptotic optimal window selection rule for kernel density estimates", *Annals of Statistics*, 12, 1285-1297.

Strawderman, W.E. (1974), "Minimax estimation of location parameters for certain spherically symmetric distributions", *Journal of Multivariate Analysis*, 4, 255-264.

Sutradhar, B.C. and M.M. Ali (1986), "Estimation of the parameters of a regression model with a multivariate-t error variable", *Communications in Statistics A Theory and Methods*, 15, 429-450.

Sutradhar, B.C. (1988), "Testing linear hypotheses with t-error variables", *Sankhya Series B*, 50, 175-180.

Tapia, R.A. and J.R. Thompson (1976), *Nonparametric Probability Density Estimation*, Baltimore: John Hopkins Press.

Taylor, W.E. (1983), "On the relevance of finite-sample distribution theory", *Econometric Reviews*, 2, 1-139.

Theil, H. (1961), *Economic Forecasts and Policy*, 2nd edition, North-Holland Publishing Company, Amsterdam.

Theil, H. (1971), *Principles of Econometrics*, North-Holland Publishing Company, Amsterdam.

Thomas, D.H. (1970), "Some contributions to radial probability distributions, statistics and the operational calculi", Ph.D. dissertation, Wayne State University.

Tiku, M.L. (1971), "Student's t-distribution under non-normal situations", *Australian Journal of Statistics*, 13, 142-145.

Tiku, M.L. and S. Kumra (1985), "Expected values and variances and covariances of order statistics for a family of symmetric distributions (Student's-t)" in *Selected Tables in Mathematical Statistics*, Vol. 8, Providence, R.I., American Mathematical Society, 141-270.

Tinbergen, J. (1930), "Bestimmung und Deutung von Amgebotskurven: Ein Beispiel", Zeitschrift fur Nationalökonomiel, 669-679.

Ullah, A. (1985), "Unanticipated macro model estimation", *Econometric Theory*, 1,, 419-420.

Ullah, A. (1988), "Nonparametric estimation of econometric functionals", *Canadian Journal of Economics*, 21, 625-658.

Ullah, A. and P.C.B. Phillips (1986), "Distribution of the F-ratio", *Econometric Theory*, 2, 449.

Ullah, A. and R.S. Singh (1985), "The estimation of probability density functions and its application in econometrics", *Technical Report 6*, University of Western Ontario.

Ullah, A. and V. Zinde-Walsh (1984), "On the robustness of LM, LR and W tests in regression models", *Econometrica*, 52, 1055-1066.

Ullah, A. and V. Zinde-Walsh (1985), "Estimation and testing in regression models with spherically symmetric errors", *Economics Letters*, 17, 127-132.

Ullah, A. and V. Zinde-Walsh (1987), "On robustness of tests of linear restrictions in regression models with elliptical error distributions" in *Time Series and Economic Modelling*, eds. I.B. MacNeill and G.J. Umphrey, D. Reidel, Holland.

Von Neumann, J. (1951), "Various techniques used in connection with random digits", *Collected Works Vol. 5*, 768-770, Pergaman Press 1963, also in *Monte Carlo Method*, National Bureau of Standard Series, Vol. 12, 36-38.

Waldman, D.M. (1982), "A stationary point for the stochastic frontier likelihood", *Journal of Econometrics*, 18, 275-280.

Weisberg, S. (1980), "Comment on a paper by White and MacDonald", *Journal of the American Statistical Association*, 75, 28-31.

Wertz, W. (1978), *Statistical Density Estimation: A Survey*, Vandenhoeck and Ruprecht in Göttingen.

White, H. and G.M. MacDonald (1980), "Some large-sample tests for non- normality in the linear regression model", *Journal of the American Statistical Association*, 75, 16-27.

Wright, S. (1934), "The method of path coefficients", *Annals of Mathematical Statistics*, 5, 161-215.

Yohai, V.J. and R.A. Maronna (1979), "Asymptotic behaviour of M-estimators for the linear model", *Annals of Statistics*, 7, 258-268.

Yuen, K.K. and V.K. Murthy (1974), "Percentage points of the distribution of the t-statistic when the parent is Student's-t", *Technometrics*, 16, 495-497.

Zellner, A. (1976), "Baynesian and non-Baynesian analysis of the regression model with multivariate Student-t error terms", *Journal of the American Statistical Association*, 71, 400-405.

Zellner, A. (1979), "Statistical analysis of econometric models", *Journal of the American Statistical Association*, 74, 628-651.

Zellner, A. (1986), "Further results on Baynesian Minimum Expected Loss (MELO) estimates with posterior distributions for structural coefficients", in *Advances in Econometrics: Innovations in Quantitative Economics - Essays in Honor of Robert L. Basmann*, Vol. 5, eds. R. Basmann and G. Rhodes, JAI Press, 171-182.

APPENDIX A

CALCULATION OF EIGENVALUES AND NONCENTRALITY PARAMETERS

In this appendix, the method used to derive the eigenvalues and noncentrality parameters given in Chapter 5 is illustrated for the DK family.

In addition to the assumptions of Chapter 5, orthonormality of the exogenous variables is assumed. That is,

$$X'X = I_G,$$ [1]

and in particular

$$X_1 = \begin{bmatrix} I \\ 0 \\ 0 \end{bmatrix} \begin{matrix} G_1 \times G \\ G_2 \times G_1 \\ (N\times G) \times G_1 \end{matrix} \qquad X_2 = \begin{pmatrix} 0 \\ I \\ 0 \end{pmatrix} \begin{matrix} G_1 \times G_2 \\ G_2 \times G_2 \\ (N\times G) \times G_2 \end{matrix} . \qquad A(1)$$

For the Double K-Class estimator family, the eigenvalues and corresponding eigenvectors of the matrix

$$(B_1 - qB_2) = \begin{bmatrix} 0 & , & \frac{K_2}{2}(P_X - P_{X_1}) + \left(\frac{1-K_2}{2}\right)\bar{P}_{X_1} \\ \frac{K_2}{2}(P_X - P_{X_1}) + \left(\frac{1-K_2}{2}\right)\bar{P}_{X_1} & , & -K_1 q(P_X - P_{X_1}) - (1-K_1)q\bar{P}_{X_1} \end{bmatrix}$$

will be derived.

<u>Theorem A1.</u> The eigenvalues and eigenvectors of the matrix $(B_1 - qB_2)$ are as listed in Table A1.

---

[1] This assumption is the orthonormalization of the exogenous variables. It helps to reduce the parameter space to an essential set. It is discussed in Phillips (1983, p.467).

Table A1: The Eigenvalues of $(B_1 - qB_2)$ and their multiplicities

λ	Multiplicity
0	$2G_1$
$\frac{1}{2}(-q + \sqrt{1+q^2})$	G_2
$\frac{1}{2}(-q - \sqrt{1+q^2})$	G_2
$\frac{1}{2}\left\{-q(1-\bar{K}_1) + \sqrt{q^2(1-\bar{K}_1)^2 + (1-\bar{K}_2)^2}\right\}$	N-G
$\frac{1}{2}\left\{-q(1-\bar{K}_1) - \sqrt{q^2(1-\bar{K}_1)^2 + (1-\bar{K}_2)^2}\right\}$	N-G

Proof:

Let $(B_1 - qB_2) = \begin{bmatrix} 0 & A_2^* \\ A_2^* & A_1^* \end{bmatrix}$, where the definitions of A_1^* and A_2^* are obvious. To find the eigenvalues of $(B_1 - qB_2)$, the equation

$$\det.\left(\begin{bmatrix} 0 & A_2^* \\ A_2^* & A_1^* \end{bmatrix} - \begin{bmatrix} \lambda I & 0 \\ 0 & \lambda I \end{bmatrix}\right) = 0$$

or equivalently

$$\det.\left(\begin{bmatrix} -\lambda I & A_2^* \\ A_2^* & (A_1^* - \lambda I) \end{bmatrix}\right) = 0 \qquad \text{(A2)}$$

needs to be solved.

Equation (A2) may be written as

$$|-\lambda I|\left|(A_1^* - \lambda I) - A_2^*(-\lambda I)^{-1}A_2^*\right| = 0$$

or, as

$$(-\lambda)^T \begin{vmatrix} -\lambda I & 0 & 0 \\ 0 & (-q+\lambda-\frac{1}{4\lambda}I) & 0 \\ 0 & & \left(-(q-qK_1-\lambda)+\frac{1}{\lambda}\left(\frac{1-K_2}{2}\right)^2\right)I \end{vmatrix} = 0$$

and the multiplicities can be determined, since

$$\begin{bmatrix} -\lambda I & 0 & 0 \\ 0 & (-q+\lambda-\frac{1}{4\lambda})I & 0 \\ 0 & 0 & \left(-(q-qK_1+\lambda)+\frac{1}{\lambda}\left(\frac{1-K_2}{2}\right)^2\right)I \end{bmatrix}$$

is partitioned into blocks with the following orders

$$\begin{bmatrix} G_1 \times G_1 & G_1 \times G_2 & G_1 \times (N-G) \\ G_2 \times G_1 & G_2 \times G_2 & G_2 \times (N-G) \\ (N-G) \times G_1 & (N-G) \times G_2 & (N-G) \times (N-G) \end{bmatrix}$$

#

Theorem A2: The components for the eigenvectors corresponding to each eigenvalue are as listed in Table A2.

Table A2: Components of the Eigenvectors

λ	Components
0	$(1 \quad , \quad 0)$
$\frac{1}{2}\left(-q + \sqrt{1+q^2}\right)$	$(a_1 \quad , \quad a_2)$
$\frac{1}{2}\left(-q - \sqrt{1+q^2}\right)$	$(a_3 \quad , \quad a_4)$
$\frac{1}{2}\left\{-q(1-\bar{K}_1) + \sqrt{q^2(1-\bar{K}_1)^2 + (1-\bar{K}_2)^2}\right\}$	$(b_1 \quad , \quad b_2)$
$\frac{1}{2}\left\{-q(1-\bar{K}_1) - \sqrt{q^2(1-\bar{K}_1)^2 + (1-\bar{K}_2)^2}\right\}$	$(b_3 \quad , \quad b_4)$

where,

$$a_1 = q + \sqrt{1+q^2}/\sqrt{2(1+q^2+q\sqrt{1+q^2}}$$

$$a_2 = 1/\sqrt{2(1+q^2+q\sqrt{1+q^2}}$$

$$a_3 = q-\sqrt{1+q^2}/\sqrt{2(1+q^2+q\sqrt{1+q^2}}$$

$$a_4 = 1/\sqrt{2(1+q^2-q\sqrt{1+q^2})}.$$

Let $\hat{K}_1 = (1-\bar{K}_1)$ and $\hat{K}_2 = (1-\bar{K}_2)$, then

$$b_1 = (q\hat{K}_1+\sqrt{q^2\hat{K}_1^2+\hat{K}_2^2})/\sqrt{2(q^2\hat{K}_1^2+\hat{K}_2^2+q\hat{K}_1\sqrt{q^2\hat{K}_1^2+\hat{K}_2^2})}$$

$$b_2 = \hat{K}_2/\sqrt{2(q^2\hat{K}_1^2+\hat{K}_2^2+q\hat{K}_1\sqrt{q^2\hat{K}_1^2+\hat{K}_2^2})}$$

$$b_3 = (q\hat{K}_1 - \sqrt{q^2\hat{K}_1^2+\hat{K}_2^2})/\sqrt{2(q^2\hat{K}_1^2+\hat{K}_2^2-q\hat{K}_1\sqrt{q^2\hat{K}_1^2+\hat{K}_2^2})}$$

$$b_4 = \hat{K}_2/\sqrt{2(q^2\hat{K}_1^2+\hat{K}_2^2+q\hat{K}_1\sqrt{q^2\hat{K}_1^2+\hat{K}_2^2})} .$$

Proof:

Given that the matrix (B_1-qB_2) can be diagonalized orthogonally, then if an eigenvalue λ of (B_1-qB_2) has algebraic multiplicity j its geometric multiplicity is also j. Therefore, instead of solving the $2N \times 2N$ system of linear equations we need only solve 2×2 sets and insert the resulting eigenvector into the appropriate position in the P matrix. Taking, for example $\lambda = \frac{1}{2}(-q+\sqrt{1+q^2})$, we need to solve

$$\begin{bmatrix} -\lambda & \frac{1}{2} \\ \frac{1}{2} & -(q+\lambda) \end{bmatrix} \underline{x} = \underline{0}$$

or, as the equations are linearly dependent, we have

$$\frac{1}{2}x_1 = \frac{1}{2}\left(q+\sqrt{1+q^2}\right)x_2 ,$$

so that after normalization we have (a_1,a_2). A similar process can be repeated for the other eigenvalues.

Theorems A1 and A2 derive results for the implementation of the Imhof or Davies technique for one particular case corresponding to $X'X = I_G$. However, when the orthogonality of the data matrix is assumed it is well known (e.g. Phillips (1983, p.467)) that the results do not depend on the form of the X matrix. Therefore, there is no loss in generality in assuming (A1).

The quadratic form $z'(B_1-qB_2)z$ corresponding to (5.3.8) can be written as $\sum_j \lambda_j \chi^2(1,\delta_j^2)$ where λ_j are the eigenvalues from Table A1 and $\chi^2(1,\delta_j^2)$ is a noncentrality parameter which equals the square of the j'th element of the vector P'E(z) where the components of the P matrix are given in Table A.2. However, since the matrix (B_1-qB_2) has multiple roots then we can rewrite $z'(B_1-qB_2)z$ as follows. Suppose m_r is the order of multiplicity of the different roots and n_r is the number of different roots, then rewrite

$$z'(B_1-qB_2)z = \sum_{r=1}^{n_r} \chi^2_{m_r};\delta^2_r$$

where $$\chi^2_m = \sum_{r=1}^{m_r} \chi^2_1;\delta^2_i$$

and

$$\delta^2_r = \sum_{r=1}^{m_r} \delta^2_i$$

It is this form that is used in Chapter 5.

APPENDIX B

THE ASYMPTOTIC PROPERTIES OF THE SECOND ROUND ESTIMATOR IN QUASI - NEWTON METHODS

Quasi - Newton methods are based on the quadratic approximation of the maximand (or minimand if relevant):

$$Q(\theta) \cong Q(\hat{\theta}_1) + g_1'(\theta - \hat{\theta}_1) + 1/2(\theta - \hat{\theta}_1)' H_1(\theta - \hat{\theta}_1), \qquad \text{(A3)}$$

where $\hat{\theta}_1$ is an initial estimate and

$$g_1 = \partial Q/\partial\theta \,\Big|_{\hat{\theta}_1} \quad \text{and} \quad H_1 = \partial^2 Q/\partial\theta\partial\theta' \,\Big|_{\hat{\theta}_1}.$$

The second round estimator, $\hat{\theta}_2$, is obtained by maximizing the right-hand side of the approximation (A3) so that,

$$\hat{\theta}_2 = \hat{\theta}_1 - H_1^{-1} g_1, \qquad \text{(A4)}$$

where,

$$g_1 = \partial Q/\partial\theta \,\Big|_{\theta_0} + \partial^2 Q/\partial\theta\partial\theta' \,\Big|_{\theta^*} (\hat{\theta}_1 - \theta_0)$$

and θ^* lies between $\hat{\theta}_1$ and θ_0. Inserting (A4) into (A3) yields,

$$\sqrt{N}(\hat{\theta}_2 - \theta_0) = \left[I - (\partial^2 Q/\partial\theta\partial\theta' \,\Big|_{\hat{\theta}_1})^{-1} \partial^2 Q/\partial\theta\partial\theta' \,\Big|_{\theta^*} \right] \sqrt{N}(\hat{\theta}_1 - \theta_0)$$

$$- (N^{-1}\partial^2 Q/\partial\theta\partial\theta' \,\Big|_{\hat{\theta}_1})^{-1} (1/\sqrt{N})\, \partial Q/\partial\theta \,\Big|_{\theta_0}. \qquad \text{(A5)}$$

Under the following conditions:

(i) $\partial^2 Q_N/\partial\theta\partial\theta'$ exists and is continuous in an open, convex neighbourhood of θ_0

(ii) $N^{-1}(\partial^2 Q_N/\partial\theta\partial\theta')\,\big|_{\theta_0}$ converges to a finite nonsingular matrix

$\lim E\, N^{-1}(\partial^2 Q_N/\partial\theta\partial\theta')\,\big|_{\theta_0}$ in probability for any sequence θ_N^*

such that plim $\theta_N^* = \theta_0$,

where Q_N is the set of roots of the equations $\partial Q_N/\partial\theta = 0$ corresponding to the local maxima, (or minima as the case may be), then,

$$\text{plim } N^{-1}(\partial^2 Q/\partial\theta\partial\theta')\,\big|_{\hat{\theta}_1} = \text{plim } N^{-1}(\partial^2 Q/\partial\theta\partial\theta')\,\big|_{\theta_0}$$

$$= \text{plim } N^{-1}(\partial^2 Q/\partial\theta\partial\theta')\,\big|_{\theta^*} \qquad \text{(A6)}.$$

Therefore, substituting (A6) into (A5) gives,

$$\sqrt{N}(\hat{\theta}_2 - \theta_0) \overset{\mathcal{D}}{\Rightarrow} - \left[\text{plim } N^{-1}(\partial^2 Q/\partial\theta\partial\theta')\big|_{\theta_0}\right]^{-1} (1/\sqrt{N})\ \partial Q/\partial\theta\,\big|_{\theta_0},$$

so that $\hat{\theta}_2$ is asymptotically efficient.

Further details of this result are given in Amemiya (1985, p.138).