THE JACKKNIFE STATISTIC:
AN APPLICATION IN ECONOMETRICS

by

Anthony David Owen
B.A. (Leicester)
M.A. (Kent)

Thesis submitted to the University of Kent in fulfilment of
the requirements for the degree of
Doctor of Philosophy
1977

I hereby declare that the research embodied in this thesis is my own work and that it has not previously been submitted for a degree at this or any other University.

A.D. Owen

A.D. Owen

ABSTRACT

Quenouille has developed a procedure, later termed the jackknife by Tukey, for reducing the bias of a consistent estimator of an unknown parameter. A measure of the variance of the resulting estimator can be obtained and used to provide approximate confidence intervals and tests of significance. Thus the jackknife technique may be especially interesting when the estimator under consideration is biased but consistent and mathematically intractable distribution theory prevents the construction of exact confidence intervals.

Considerable research has been devoted to studying the jackknife technique, predominantly in the fields of biometrics, statistics and numerical analysis. So far the use of the jackknife method in econometrics has been negligible, although one very important class of econometric estimators, the simultaneous equation estimators, is biased in finite samples and, in general, has a mathematically intractable distribution.

In this thesis we investigate the application of the jackknife technique to the Two-Stage Least Squares (2SLS) structural parameter estimator in a simultaneous equation system. The bias reducing property was found to be present in the majority of cases considered in an investigation of the effects of jackknifing on the exact bias of the 2SLS estimator in a two equation model. Conditions are given for which it is unlikely that jackknifing will reduce the bias of the 2SLS estimator.

Since the exact variance of the jackknifed 2SLS estimator is unknown, an examination of the effect on the variance of 2SLS of applying the jackknife had to be made by a simulation experiment. Whilst the 2SLS estimator always had a smaller mean square error than

the jackknifed 2SLS estimator, a comparison of absolute errors rarely produced a significant difference between them.

Finally, it was observed that t statistics formed using the 2SLS estimator may not be distributed according to the Student t distribution. The actual distribution may be highly skewed and serious errors could result if the postulated theoretical distribution was used for statistical inference. In general, this feature was less noticeable for the J2SLS estimator which appeared to have a reasonably symmetric distribution, and consequently there is less likelihood of serious errors being made if the postulated theoretical distribution is used for the purpose of statistical inference.

ACKNOWLEDGEMENTS

Certain parts of the work contained in this thesis have been presented at

The Third World Econometric Congress (with G.D.A. Phillips)(1975)

and The Third Australian Statistical Conference (1976).

I am grateful for comments received at these two conferences, and for comments received from a Staff Seminar in Statistics at the Australian National University (1976).

Material extracted from Chapter 5 has been published in Econometrica (1976), whilst material extracted from Chapter 4 and Appendix C is forthcoming in the Journal of Statistical Computation and Simulation. My thanks are due to the editors and anonymous referees of both journals for their constructive comments.

For advice, comments and criticism during the course of the research work on which this thesis is based, I am indebted to my supervisor, Mr G.D.A. Phillips.

To my parents, to whom I dedicate this work, I owe a great debt for their encouragement throughout my University career. Finally, to my wife, Jackie, who has been a source of continual encouragement I extend my most affectionate thanks.

TABLE OF CONTENTS

LIST OF TABLES

"Grown-ups love figures. When you tell them you have made a new friend, they never ask you any questions about essential matters. They never say to you, 'What does his voice sound like? What games does he love best? Does he collect butterflies?' Instead, they demand: 'How old is he? How many brothers has he? How much does he weigh? How much money does his father make?' Only from these figures do they think they have learned anything about him."

Antoine de Saint-Exupéry, The Little Prince

CHAPTER 1

INTRODUCTION

Quenouille [45] has developed a technique, later termed the jackknife by Tukey [72], for reducing the bias which may be present in an (otherwise consistent) estimator of an unknown parameter. Quenouille's original justification for using the technique was based upon the assumption of the existence of a Taylor series expansion for the bias of an estimator whereupon, by applying the jackknife technique, the bias term to order ($^1/N$) could be removed. In addition to its bias reducing properties, the jackknife technique can also be used to provide approximate confidence intervals and tests of significance. Thus the jackknife technique is a viable proposition where the estimator under consideration is biased, but consistent, and/or where mathematically intractable distribution theory prohibits the formation of exact confidence intervals.

Considerable research has been devoted to studying the jackknife technique, predominantly in the fields of biometrics, statistics and numerical analysis. Its use in econometrics has been negligible, yet a class of consistent econometric estimators possess both bias and intractable distribution theory in finite samples, which would suggest that application of the jackknife technique may be a fruitful exercise. This class of estimators is the class of simultaneous equation estimators.

This thesis considers the effects of applying the jackknife technique to one of this class of estimators, the Two-Stage Least Squares (2SLS) estimator.

2SLS is a "limited information" estimator in the sense that it estimates the equations comprising a simultaneous economic system one at a time. In order to estimate any one equation, 2SLS only requires a specification of the equation being estimated and a list of the other predetermined variables appearing in the system. It does not therefore take account of contemporaneous correlation between the disturbances of the equations in the system. Neither does it use the information contained in the overidentifying restrictions on the other equations in the system. Consequently, if the entire system has been specified, 2SLS may not make the most effective use of all the available information and a "full-information" estimator may be preferred. Under the assumption that the hypothesized model is correctly specified, the most efficient method of estimation would be one of the full information methods. Most economists, however, would consider this assumption rather heroic and would select one of the limited information estimators in order to isolate the deleterious effects of any specification errors to the equations in which they arise.

There are two reasons for selecting the 2SLS estimator from such a wide class of estimators.

Firstly, the exact bias (and higher order moments) of the 2SLS estimator have been derived for a two equation model and this allows an exact investigation of the jackknife's bias reducing ability vis-a-vis 2SLS, albeit under rather restrictive assumptions.

Secondly, the other limited information simultaneous equation estimators of any importance are the Ordinary Least Squares (OLS) and the Limited Information Maximum Likelihood (LIML) estimators. OLS is not a candidate for jackknifing since it contravenes Quenouille's assumption of a consistent estimator, whilst the non-finite moments of the LIML estimator (see Mariano and Sawa [30]) precludes any examination of the effects of the jackknife technique on its "bias". In addition, within the class of limited information simultaneous equation estimators, on the basis of numerous Monte Carlo results (the major works are summarized in Johnston [20], Chapter 13, Section 8) 2SLS is generally preferred on the grounds of "all-round" performance and computational efficiency and simplicity.

"Full information" methods of estimation were not considered as possible candidates for jackknifing as this would seem to be the logical step forward after the limited information estimators had been considered. This point is discussed further in Chapter 8.

The general form of the simultaneous equation system which will be used throughout this thesis, together with the relevant notation and assumptions, is defined in Chapter 2. The 2SLS estimator and its asymptotic properties are derived for the parameters of any single equation in the system. Conditions and assumptions under which the exact finite sample results of the 2SLS estimator have been derived are also stated.

Chapter 2 continues with a description of the jackknife statistic, its bias reducing properties, and its use in formulating approximate confidence intervals and tests of significance. The literature on the jackknife and its applications is so extensive that only (what the author considers to be) the more relevant works are cited, although a bibliographical reference is given.

The asymptotic properties of the jackknife 2SLS (J2SLS) estimator are investigated in Chapter 3. A proof of the asymptotic equivalence of the J2SLS and 2SLS estimators is given, and a t ratio formed using the J2SLS estimator is shown to be asymptotically distributed as the standardized normal distribution.

The small sample properties of the J2SLS estimator are investigated by a series of simulation experiments in Chapters 5, 6 and 7. The computer algorithms used in the experiments are described in Chapter 4 together with results of verification where they do not already exist. A formula given in Chapter 3 reduces the computational burden involved in calculating J2SLS parameter estimates, and should reduce the probability of significant inaccuracies due to the build-up of rounding errors resulting from repeated use of the matrix inversion algorithm. Chapter 4 also contains a method for evaluating the accuracy of the asymptotic approximations to the exact moments of the 2SLS estimator.

For an equation containing just two endogenous variables the exact first and second order moments of the 2SLS estimator have been derived. It is relatively easy to adapt the exact bias of the 2SLS estimator to obtain the exact bias of the J2SLS estimator, but the exact mean square error of the J2SLS estimator has not, as yet, been derived. In Chapter 5 the exact relative biases of the 2SLS and J2SLS estimators are compared, under conditions which prevail for "exact" theory, by means of a simulation experiment. This experiment gives exact results on the ability of the jackknife to reduce the bias of the 2SLS estimator. For the general model, however, this form of analysis is not possible, and the author has only been able to derive a rather weak conditon under which jackknifing is "unlikely" to reduce the bias of the 2SLS estimator.

Chapter 6 presents the results of a Monte Carlo experiment into the properties of the two estimators. Comparisons of relative bias, mean square error and mean absolute error are made using a two equation model. The use of the jackknife statistic to form approximate confidence intervals and tests of significance using the 2SLS estimator is also investigated and the results are presented in Chapter 7. It is well known that standardized normal ratios and t ratios formed using the 2SLS estimator are only valid asymptotically, and that in small samples they could diverge significantly from their postulated theoretical distributions. A comparison of the small sample distributions of test statistics using both 2SLS and J2SLS estimators is made.

Concluding remarks are contained in Chapter 8.

CHAPTER 2

THE TWO-STAGE LEAST SQUARES ESTIMATOR AND THE BIAS REDUCING PROPERTIES OF THE JACKKNIFE STATISTIC

2.1 The General Linear Simultaneous Equations Model

2.1.1 Specification of the Model

The analysis in this thesis is concerned with a simultaneous economic system of G linear stochastic equations relating G endogenous (or jointly-dependent) variables and K exogenous variables, which can be written as

$$\bar{Y}B + X\Gamma = U \quad . \tag{2.1}$$

We are interested in the estimation of just one equation from this system, (say) the jth, which can be written as

$$\underline{y}_j = Y_j\underline{\beta}_j + X_{1j}\underline{\gamma}_{1j} + X_{2j}\underline{\gamma}_{2j} + \underline{u}_j \ , \tag{2.2}$$

and we will refer to this equation as the jth structural equation ($j = 1,2,\ldots,G$). For notational simplicity we will generally omit the j subscript.

2.1.2 Notation

$\bar{Y}$ is a matrix of N observations on the G endogenous variables in the entire system;

$\underline{y}$ is a vector of N observations on the "dependent" endogenous variable;

Y is a matrix of N observations on the other g endogenous variables included in the jth equation. In the unlikely event that all G endogenous variables appear in the jth equation then $g = G-1$ and

$[\underline{y} : Y] = \overline{Y}$, otherwise $g < G-1$;

X is a matrix of N observations on K exogenous variables partitioned as $X = [X_1 : X_2]$;

X_1 is a matrix of observations on the K_1 exogenous variables included in the jth equation;

X_2 is a matrix of observations on the K_2 exogenous variables excluded from the jth equation (i.e. $K = K_1 + K_2$);

U is a matrix of N unobservable disturbances for each of the G equations, with jth column denoted by $\underline{u}_j$;

B is a $G \times G$ matrix of unknown structural coefficients;

$\underline{\beta}$ is an unknown g component sub-vector of B with non-zero elements.

Γ is a $K \times G$ matrix of unknown structural coefficients;

$\underline{\gamma}_1$ is a K_1 component sub-vector of Γ with non-zero elements;

$\underline{\gamma}_2$ is a K_2 component sub-vector of Γ with zero elements.

2.1.3 Basic Assumptions

The following conventional assumptions are made for the system (2.1), and for the jth structural equation (2.2):

(i) B is non-singular;

(ii) the jth structural equation, (2.2), is just- or over-identified by zero restrictions on the structural coefficients, i.e. $K_2 \geq g$;

(iii) the matrix X consists of non-stochastic elements and is of full rank, K. Further, as $N \to \infty$ the matrix $N^{-1}(X'X)$ converges to a finite matrix, denoted by

$$\lim_{N \to \infty} \frac{1}{N} \cdot (X'X) = \Sigma_{XX} ,$$

where Σ_{XX} is a finite positive definite matrix ;

(iv) the sample size (N) is greater than the total number of exogenous variables (K) in the system;

(v) the N rows of U are independently and identically distributed with zero mean vector and unknown finite covariance matrix, Σ. In addition, the analysis in Chapter 3 requires that the structural disturbances have finite fourth order moment.

Postmultiplying equation (2.1) by B^{-1} we obtain the reduced form of the system, which can be written as

$$\bar{Y} = X\bar{\Pi} + \bar{V} , \qquad (2.3)$$

where $\bar{\Pi} = -\Gamma B^{-1}$,

and $\bar{V} = U B^{-1}$.

The reduced form equation for the jth "dependent" endogenous variable and the reduced form equations for the g "explanatory" endogenous variables can be written as

$$\underline{y}_j = X\underline{\pi}_j + \underline{v}_j$$

and

$$Y_j = X\Pi_j + V_j \qquad (2.4)$$

respectively. Since, for notational convenience, we are omitting the j subscripts, this explains the necessity to write equation (2.3) in the above form rather than in the more common form which would coincide with equation (2.4) when the subscripts are omitted.

2.1.4 The Two Endogenous Variables Case

The majority of results on the exact properties of the 2SLS estimator have been derived under the assumption that $g = 1$, i.e. the equation being estimated contains only two endogenous variables. In addition, it is assumed that the matrices X_1 and X_2 contain no lagged endogenous variables.

Under the above conditions, the first structural equation can be written as

$$\underline{y}_1 = \underline{y}_2\beta + X_1\underline{\gamma}_1 + X_2\underline{\gamma}_2 + \underline{u}_1 , \qquad (2.5)$$

with reduced form equations

$$\underline{y}_1 = X_1\underline{\pi}_{11} + X_2\underline{\pi}_{12} + \underline{v}_1$$

$$\underline{y}_2 = X_1\underline{\pi}_{21} + X_2\underline{\pi}_{22} + \underline{v}_2 ,$$

where $\underline{\pi}_{11}$, $\underline{\pi}_{12}$, and $\underline{\pi}_{21}$, $\underline{\pi}_{22}$ are vectors of constant coefficients. The random vector $(\underline{v}_1' : \underline{v}_2')$ is assumed to be distributed as bivariate normal with zero mean and positive definite covariance matrix $\Omega \otimes I_N$, where $\Omega = \omega_{ij}$ $(i,j = 1,2)$ is a matrix of reduced form parameters.

2.2 The Two-Stage Least Squares Estimator

It is well known that OLS is, in general, an inconsistent estimator of the parameters in the structural equation (2.2). This inconsistency is due to the correlation between the explanatory endogenous variables (Y) and the vector of structural disturbances ($\underline{u}$). Basmann [3] and Theil [70] derived, independently, an alternative estimator which "purges" Y of the stochastic component associated with the disturbance term, and then estimates the revised equation by OLS. This "alternative" estimator is called the Two-Stage Least Squares Estimator.

From equation (2.2) we write the jth structural equation as

$$\underline{y} = Y\underline{\beta} + X_1\underline{\gamma}_1 + \underline{u} .$$

If we rewrite the above equation as

$$\underline{y} = (Y - V)\underline{\beta} + X_1\underline{\gamma}_1 + \underline{u} + V\underline{\beta} ,$$

then using equation (2.4), $(Y - V) = X\Pi$ is uncorrelated with $(\underline{u} + V\underline{\beta})$ since X is non-stochastic by assumption (iii).

Since V is unobservable we must use its estimated value $\hat{V}$, where $\hat{V} = Y - X\hat{\Pi}$. Provided $\underset{N\to\infty}{\text{plim}}\, \hat{\Pi} = \Pi$, it follows that $\underset{N\to\infty}{\text{plim}}\, (Y - \hat{V}) = (Y - V)$ and hence $(Y - \hat{V})$ and $(\underline{u} + \hat{V}\underline{\beta})$ are asymptotically uncorrelated.

Thus if the least squares estimator is applied to

$$\underline{y} = (Y - \hat{V})\underline{\beta} + X_1\underline{\gamma}_1 + \underline{u} + \hat{V}\underline{\beta} ,$$

we can obtain consistent estimates of $\underline{\beta}$ and $\underline{\gamma}_1$. Since this process of estimation involves two successive applications of least squares it is known as Two-Stage Least Squares (2SLS).

In this thesis we shall work with the Instrumental Variables type formulation of the 2SLS estimator, viz:

$$\underline{\hat{\theta}} = \left[Z'X(X'X)^{-1}X'Z\right]^{-1} Z'X(X'X)^{-1} X'\underline{y} , \qquad (2.6)$$

where $Z = \left[Y : X_1\right]$ and $\underline{\hat{\theta}}' = \left[\underline{\hat{\beta}}' : \underline{\hat{\gamma}}_1'\right]$.

In order to apply tests of significance, knowledge of the distribution of the 2SLS estimator is required. The finite sample distribution of 2SLS is only known for a few specific cases, thus reliance is usually placed upon its asymptotic distribution.

Substituting for $\underline{y}$ in equation (2.6) we obtain

$$\underline{\hat{\theta}} - \underline{\theta} = \left[Z'X(X'X)^{-1}X'Z\right]^{-1} Z'X(X'X)^{-1} X'\underline{u} ,$$

and we require the limiting distribution of the sequence

$$\sqrt{N}(\underline{\hat{\theta}} - \underline{\theta}) = \left[\frac{1}{N} . Z'X\left(\frac{1}{N} . X'X\right)^{-1} \frac{1}{N} . X'Z\right]^{-1} \frac{1}{N} . Z'X\left(\frac{1}{N} . X'X\right)^{-1} \frac{1}{\sqrt{N}} . X'\underline{u} .$$

Since X is (by assumption) non-stochastic, it follows that

$$\frac{1}{N} . X'Z = \frac{1}{N} . X'[Y : X_1] = \frac{1}{N} . X'[X\Pi + V : X_1]$$

converges in probability to a finite limit, denoted by

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N} . X'Z = \Sigma_{XZ} .$$

We have already assumed the existence of a finite limit for $N^{-1} . X'X$, and thus we can denote its inverse by

$$\lim_{N\to\infty} \left(\frac{1}{N} . X'X \right)^{-1} = \Sigma_{XX}^{-1} .$$

Under assumption (V), modified application of the Lindeberg-Levy theorem (see e.g. Theil [71; pp.498-499]) using the above results will yield

$$\sqrt{N} (\underline{\hat{\theta}} - \underline{\theta}) \sim N \left[0, \sigma^2 \operatorname*{plim}_{N\to\infty} \left\{ \frac{1}{N} . Z'X \left(\frac{1}{N} . X'X \right)^{-1} \frac{1}{N} . X'Z \right\}^{-1} \right] , \qquad (2.7)$$

where σ^2 denotes the variance of the jth structural disturbance, i.e. the jjth component of Σ.

A consistent estimator of σ^2 is given by

$$\hat{\sigma}^2 = \underline{\hat{u}}'\underline{\hat{u}} / (N - K_1 - g) ,$$

where $\underline{\hat{u}} = [\underline{y} - Y\underline{\hat{\beta}} - X_1\underline{\hat{\gamma}}_1]$.

Since the asymptotic covariance matrix of the 2SLS estimator coincides with the Cramer-Rao bound (when the structural disturbances are normally distributed), 2SLS is an efficient estimator in its class of limited information simultaneous equation estimators. Its relative (small sample) efficiency however has not, in general, been ascertained.

2.3 The Jackknife Statistic

2.3.1 Definition

Let α be an unknown parameter, and let $X_1, X_2, \ldots, X_N$ be N independently and identically distributed observations from the cumulative distribution function F_α. Further, let $\hat{\alpha}$ be a biased estimator of α such that

$$E(\hat{\alpha}-\alpha) = \frac{a_1}{N} + \frac{a_2}{N^2} + \ldots. + \frac{a_r}{N^r} + \ldots. , \qquad (2.8)$$

where a_1, a_2, ..., a_r are constants and not dependent upon N. If the N observations can be divided into n groups, each of r observations (i.e. $N = nr$), then the estimator

$$J_i(\hat{\alpha}) = n\hat{\alpha} - (n-1)\hat{\alpha}_i , \qquad (i = 1,2,\ldots,n)$$

where $\hat{\alpha}_i$ denotes the estimate of α obtained with the ith group of observations omitted, removes the term in 1/N from equation (2.8).

Applying the technique to equation (2.8) gives

$$E[J_i(\hat{\alpha})] = n\alpha + \frac{a_1}{r} + \frac{a_2}{r^2n} + \frac{a_3}{r^3n^2} + \ldots\ldots$$

$$- (n-1)\alpha - \frac{a_1}{r} - \frac{a_2}{r^2(n-1)} - \frac{a_3}{r^3(n-1)^2} - \ldots.. ;$$

$$\text{i.e. } E[J_i(\hat{\alpha})] = \alpha - \frac{a_2}{r^2n(n-1)} - \frac{(2n-1)a_3}{r^3n^2(n-1)^2} - \ldots. .$$

$(i = 1,2,\ldots,n)$.

Tukey,[1] in unpublished work, has named $J_i(\hat{\alpha})$ the pseudo-jackknife estimator. He defined the jackknife estimator, $J(\hat{\alpha})$, as the average of the i pseudo-jackknife values $(i = 1,2,\ldots, n)$, i.e.

$$J(\hat{\alpha}) = \frac{1}{n}\sum_{i=1}^{n} J_i(\hat{\alpha}) = n\hat{\alpha} - \frac{(n-1)}{n}\sum_{i=1}^{n}\hat{\alpha}_i . \qquad (2.9)$$

---

1. The definition that follows is taken from Brillinger [7] who cites an unpublished paper and an abstract [72] of a conference paper by Tukey.

$J(\hat{\alpha})$ will have the same expected value as $J_i(\hat{\alpha})$, but a smaller variance. The term jackknife was coined for this procedure since it shared two characteristics with a boy scout's jackknife:

(i) wide applicability to many different problems;

(ii) inferiority to special tools for those problems for which special tools have been designed.

In most problems however the property of removing bias would not be sufficient to recommend the use of the jackknife. A comparison of the dispersion of the original estimator with that of the jackknife estimator is needed. Tukey noted that not only are the pseudo-jackknife estimates nearly unbiased, but their average sum of squares of deviations is nearly N(N-1) times the variance of their means. He proposed that in many instances the $J_i(\hat{\alpha})$ are approximately independently and identically distributed and hence an approximate estimate of the variance of $J(\hat{\alpha})$ is given by

$$\sum_{i=1}^{n} \frac{[J_i(\hat{\alpha}) - J(\hat{\alpha})]^2}{n(n-1)}, \tag{2.10}$$

whilst

$$\frac{[J(\hat{\alpha}) - \alpha]}{\sqrt{\sum_{i=1}^{n} \frac{[J_i(\hat{\alpha}) - J(\hat{\alpha})]^2}{n(n-1)}}} \tag{2.11}$$

is approximately distributed as a t variate with (n-1) degrees of freedom.

The jackknife can be re-applied in order to remove the bias term of order $1/N^2$ which remains after the initial application. Quenouille [45] and Kendall and Stuart [24] give a formula to achieve this further bias reduction, but if $a_k = 0$ for all $k > 2$ then the second application of the jackknife does not yield an exactly unbiased statistic as one would have desired. Schucany, Gray and Owen [62] give a higher order transformation which provides an algorithm for eliminating, exactly, bias terms of higher order.

This thesis considers the jackknife technique when $r = 1$ (i.e. $N = n$). Thus each of the N pseudo-jackknife estimates is calculated from the total number of observations less one, and the jackknife statistic is defined as

$$J(\hat{\alpha}) = N\hat{\alpha} - \frac{(N-1)}{N} \sum_{i=1}^{N} \hat{\alpha}_i . \tag{2.12}$$

Intuitively, choosing $r = 1$ is appealing since problems of dividing up samples and being left with awkward remainders are avoided. In addition, several studies involving applications of the jackknife have found $r = 1$ to be the "optimal" value of r (e.g. see Robson and Whitlock [56] and Rao [47]).

2.3.2 The Generalized Jackknife

Schucany et al. [62] provide a general method for bias reduction which includes the jackknife as a special case. Suppose that there are $k + 1$ biased estimators of α, viz: $\hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_{k+1}$, defined over the N(=n) observations, and further suppose that the biases of these $k + 1$ estimators can be written as

$$E(\hat{\alpha}_i) - \alpha = \sum_{j=1}^{k} f_{ij}(N) b_j(\alpha) , \qquad (i = 1,2,\ldots,k+1)$$

then the estimator

$$\tilde{\alpha}^{(k)} = \frac{\begin{vmatrix} \hat{\alpha}_1 & \hat{\alpha}_2 & \cdots\cdots & \hat{\alpha}_{k+1} \\ f_{11} & f_{12} & \cdots\cdots & f_{1,k+1} \\ \vdots & \vdots & & \vdots \\ f_{k1} & f_{k2} & & f_{k,k+1} \end{vmatrix}}{\begin{vmatrix} 1 & 1 & \cdots\cdots & 1 \\ f_{11} & f_{12} & \cdots\cdots & f_{1,k+1} \\ f_{21} & f_{22} & \cdots\cdots & f_{2,k+2} \\ \vdots & \vdots & & \vdots \\ f_{k1} & f_{k2} & \cdots\cdots & f_{k,k+1} \end{vmatrix}} \qquad (2.13)$$

reduces the order of bias to terms of order $(k+1)$ in $1/N$, i.e.

$$E\left[\tilde{\alpha}^{(k)}\right] - \alpha = O\left[N^{-(k+1)}\right] ,$$

where the argument of the f_{ij} functions has been suppressed for notational convenience, and these functions are assumed to be known. Further, it is assumed that $1 \leq k \leq N-1$ and that the denominator of equation (2.13) is non-zero.

If $k = 1$, then $\hat{\alpha}_1 = \hat{\alpha}$, $\hat{\alpha}_2 = \frac{1}{N} \sum_{i=1}^{N} \hat{\alpha}_i$,

$$f_{11}(N) = \frac{1}{N} , \quad \text{and} \quad f_{12}(N) = \frac{1}{(N-1)} ,$$

and equation (2.13) reduces to the "regular" jackknife as defined by equation (2.12).

The formula given by equation (2.13) is exact, in the sense that if the bias of the original estimator takes the form of equation (2.8) with only the first k terms non-zero, application of $\tilde{\alpha}^{(k)}$ will remove all bias.

Schucany et al. only considered the problem of bias reduction. The effect of their higher order transformation on the variance of $\tilde{\alpha}^{(k)}$ was not investigated for the general case.

2.4 Previous Applications of the Jackknife Technique in Econometrics

2.4.1 Partial Correlation Coefficient

If the estimated value of the partial correlation coefficient is used as an approximate test for serial correlation in time series, Quenouille [44] has shown that the bias of the estimator is inversely proportional to the sample size, N. He suggested using (what later became known as) the jackknife technique with $n = 2$, i.e. the sample was split in half, in order to remove the bias term of order (1/N). In a later paper, Quenouille [45] generalized this procedure by noting that the same amount of bias reduction could be achieved by splitting the sample into n groups each of size r (where $N = nr$).

2.4.2 Autoregressive Processes

Quenouille's [44] original method of jackknifing (i.e. $n = 2$) was later applied by Orcutt and Winokur [39] to the least squares estimator in an attempt to reduce the bias of $\hat{\beta}$ (the least squares estimator of β) in the autoregressive process

$$y_t = \alpha + \beta y_{t-1} + \varepsilon_t \qquad (t = 1, 2, \ldots, N)$$

(ε_t normally and independently distributed).

Using a Monte Carlo study they compared sample means and mean square errors of three estimators of β: least squares, jackknife least squares, and an estimator based upon correcting the bias of least squares using an expression derived by Marriott and Pope [31]. Whilst both modified least squares estimators reduced bias, the jackknife estimator exhibited a larger mean square error than the other two estimators and consequently was not to be preferred.

2.5 Use of the Jackknife Technique in Other Disciplines

A substantial body of literature on the application of the jackknife technique in various disciplines has evolved since Tukey's [72] initial conjecture. A survey of these applications, together with a comprehensive bibliography, has been compiled by Miller [35]. With the exception of the two papers cited in the previous section, few of the applications have any direct relevance to econometrics.

Perhaps the most successful area in which the jackknife has been used to date is that of ratio estimation. Given a bivariate sample (X_i, Y_i) $(i = 1,2,\ldots,T)$ from a population of size N $(T < N)$ with means μ and η respectively, we are interested in estimation of the ratio $R = \mu/\eta$. In many instances the classical ratio estimator $r = \overline{X}/\overline{Y}$ (i.e. the ratio of sample means) with $\overline{X}$ known, may exhibit a large bias compared to its standard error in surveys with many strata and small samples within strata. Durbin [14] suggested the jackknife with $n = 2$ as a bias reducing tool and investigated its properties under two distributional assumptions on the stochastic error term in the general linear model. Under both assumptions the jackknife not only reduced the bias of the ratio estimator, but also reduced the mean square error. Rao [47] and Rao and Webster [48] showed that the optimal choice of n under both of Durbin's [14] distributional assumptions is $n = N$.

Subsequent research investigated the performance of the jackknife in ratio estimation as compared with several other estimators. In general, the jackknife appeared to rank close behind the most efficient estimators but had the disadvantage of being more complicated to compute.

An application of the jackknife with direct relevance to econometrics is Miller's [36] proof that the jackknife OLS estimator of the vector of parameters in the general linear model is asymptotically normally distributed under conditions that do not require the vector of stochastic disturbances to be normally distributed. He conjectured that his proof extended to the case of non-linear least squares.

The jackknife has also been applied in the areas of maximum likelihood estimation, functions of a U-statistic, stochastic processes, inference on variances, and multivariate analysis. This list is far from exhaustive and the interested reader is referred to Miller's [35] bibliography for additional areas of application, and his synthesis for a review of the performance of the jackknife statistic over the many disciplines in which it has been used.

2.6 Alternative Methods of Bias Reduction Using the 2SLS Estimator

2.6.1 General Remarks

Methods designed to reduce the bias of the 2SLS estimator, without increasing the mean square error, have been devised by Nagar [37] and Sawa [60, 61]. Strictly speaking neither author "manipulates" the 2SLS estimator specifically, but since both proposed estimators converge in distribution to the 2SLS estimator as the sample size increases indefinitely, they could offer themselves as alternatives to the J2SLS estimator, at least on a bias reduction criterion.

2.6.2 Nagar's Unbiased k-Class Estimator

Nagar [37] has derived an expression for the bias to order 1/N of a distribution approximating the distribution of the k-class estimators. He noted that for $k = 1 + \nu/N$, where ν is the degree of overidentification of the equation being estimated, the bias vanishes to order $1/N$. Asymptotically, Nagar's unbiased estimator is clearly equivalent to the 2SLS estimator.

Using Klein's model I, Nagar showed that whilst this choice of k certainly exhibited a smaller "bias" than the corresponding 2SLS estimator, 2SLS dominated on a "mean square error" criterion. Sawa [59], however, has shown (for a two endogenous variables model) that if $k > 1$ and nonstochastic then no moments of the k-class estimators are finite; hence Nagar's "unbiased" k-class estimator does not possess a finite first order, or any other order, moment.

2.6.3 Sawa's Combined Estimator

On the basis of an asymptotic expansion of the exact bias of the k-class estimators in a two endogenous variables model, Sawa [60] proposed an estimator which uses a weighted combination of the 2SLS and OLS estimators in order to remove the leading term of the asymptotic expansion. The weights are such that, asymptotically, Sawa's combined estimator converges to 2SLS.

In a series of experiments, the combined estimator dominated the 2SLS estimator (on a mean square error criterion) when the number of exogenous variables excluded from the equation being estimated was very large. The reduction in bias (over 2SLS) obtained by using the combined estimator was always evident and frequently substantial.

The experiments were only conducted for an equation containing just two endogenous variables. Sawa [61] justified the extension of his combined estimator to equations containing an arbitrary number of endogenous variables by using Kadane's [23] small σ approximations. As yet, however, no Monte Carlo results have been published on the relative merits of the combined estimator vis-a-vis other limited information estimators. Clearly if the combined estimator dominates other limited information estimators on a mean square error criterion only for a large number of excluded exogenous variables, a Monte Carlo study may be impracticable, or at least very expensive.

2.7 Justification for Applying the Jackknife Technique to the Two-Stage Least Squares Estimator

The author has been unable to produce a rigorous justification for applying the jackknife technique to the 2SLS estimator, as he cannot express the bias of 2SLS as a Taylor series expansion in terms of increasing powers of 1/N. Nagar [37], however, has shown that the bias of the 2SLS estimator can be approximated by an expression involving terms of increasing powers of order $(1/N^{\frac{1}{2}})$ in probability. In addition, using Kadane's [23] approximation to the bias of the 2SLS estimator, the author has been able to derive a condition under which the jackknife is "unlikely" to reduce the bias of the 2SLS estimator. This analysis is contained in Chapter 5.

Whilst these results cannot provide a rigorous justification for using the jackknife technique as a bias reducing tool, it suggests that its application may be worth pursuing.

CHAPTER 3

ASYMPTOTIC THEORY

3.1 Derivation of the Computing Formula for the J2SLS Estimator

From equation (2.6) the 2SLS estimator can be written, in instrumental variable form, as

$$\underline{\hat{\theta}} = [Z'X(X'X)^{-1}X'Z]^{-1}Z'X(X'X)^{-1}X'\underline{y} \quad . \tag{3.1}$$

We denote the 2SLS estimator of $\underline{\theta}$ based upon (N-1) observations as

$$\underline{\hat{\theta}}_i = [Z_i'X_i(X_i'X_i)^{-1}X_i'Z_i]^{-1} Z_i'X_i(X_i'X_i)^{-1}X_i'\underline{y}_i \ , \tag{3.2}$$

where the i subscript denotes that the ith observation $(i = 1,2,\ldots,N)$ has been removed from the relevant data matrix. Using Appendix A we can show that

$$(X_i'X_i)^{-1} = (X'X - \underline{x}_i\underline{x}_i')^{-1} = (X'X)^{-1} + \frac{(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}}{1 - \underline{x}_i'(X'X)^{-1}\underline{x}_i} \ , \tag{3.3}$$

where $\underline{x}_i$ (a K dimensional column vector) denotes the ith row of X; i.e. the ith observation on X.

Using equation (3.3), we can rewrite equation (3.2) as

$$\underline{\hat{\theta}}_i = \left\{ [Z'X-\underline{z}_i\underline{x}_i'] \left[(X'X)^{-1} + \frac{(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}}{1 - \underline{x}_i'(X'X)^{-1}\underline{x}_i} \right] [X'Z-\underline{x}_i\underline{z}_i'] \right\}^{-1}$$

$$\times \left\{ [Z'X-\underline{z}_i\underline{x}_i'] \left[(X'X)^{-1} + \frac{(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}}{1 - \underline{x}_i'(X'X)^{-1}\underline{x}_i} \right] [X'\underline{y}-\underline{x}_i y_i] \right\}, \tag{3.4}$$

where $\underline{z}_i$ (a $K_1 + g$ dimensional column vector) and y_i (scalar) denote the ith observations on Z and $\underline{y}$ respectively.

Consider the term to be inverted in equation (3.4). Upon expansion we obtain

$$\left[Z'X(X'X)^{-1}X'Z + \underline{z}_i\underline{x}_i'(X'X)^{-1}\underline{x}_i\underline{z}_i' - Z'X(X'X)^{-1}\underline{x}_i\underline{z}_i' - \underline{z}_i\underline{x}_i'(X'X)^{-1}X'Z \right.$$
$$+ \frac{Z'X(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}X'Z}{1-\underline{x}_i'(X'X)^{-1}\underline{x}_i} + \frac{\underline{z}_i\underline{x}_i'(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}\underline{x}_i\underline{z}_i'}{1-\underline{x}_i'(X'X)^{-1}\underline{x}_i}$$
$$\left. - \frac{Z'X(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}\underline{x}_i\underline{z}_i'}{1-\underline{x}_i'(X'X)^{-1}\underline{x}_i} - \frac{\underline{z}_i\underline{x}_i'(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}X'Z}{1-\underline{x}_i'(X'X)^{-1}\underline{x}_i} \right]^{-1} . \qquad (3.5)$$

Let $P = Z'X(X'X)^{-1}X'Z$,

$s_i = \underline{x}_i'(X'X)^{-1}\underline{x}_i$, (scalar)

$b_i = \underline{x}_i'(X'X)^{-1}X'\underline{y} = \underline{x}_i'\hat{\underline{\pi}}$, (scalar)

and $\underline{a}_i = Z'X(X'X)^{-1}\underline{x}_i$,

then equation (3.5) can be rewritten as

$$\left[P + s_i\underline{z}_i\underline{z}_i' + \frac{\underline{a}_i\underline{a}_i'}{(1-s_i)} + \frac{s_i^2\underline{z}_i\underline{z}_i'}{(1-s_i)} - \underline{a}_i\underline{z}_i' - \underline{z}_i\underline{a}_i' - \frac{s_i\underline{a}_i\underline{z}_i'}{(1-s_i)} - \frac{s_i\underline{z}_i\underline{a}_i'}{(1-s_i)} \right]^{-1}$$

$$= \left[P - \underline{z}_i\underline{z}_i' + \frac{1}{(1-s_i)}(\underline{z}_i-\underline{a}_i)(\underline{z}_i-\underline{a}_i)' \right]^{-1} . \qquad (3.6)$$

Let $\left[P + \frac{1}{(1-s_i)}(\underline{z}_i-\underline{a}_i)(\underline{z}_i-\underline{a}_i)' \right] = C$,

then, using Appendix I, equation (3.6) can be rewritten as

$$(C - \underline{z}_i\underline{z}_i')^{-1} = C^{-1} + \frac{C^{-1}\underline{z}_i\underline{z}_i'C^{-1}}{1-\underline{z}_i'C^{-1}\underline{z}_i} , \qquad (3.7)$$

and using the same expansion, it follows that

$$C^{-1} = P^{-1} - \frac{P^{-1}(\underline{z}_i-\underline{a}_i)(\underline{z}_i-\underline{a}_i)'P^{-1}}{(1-s_i)+(\underline{z}_i-\underline{a}_i)'P^{-1}(\underline{z}_i-\underline{a}_i)} . \qquad (3.8)$$

Combining equations (3.7) and (3.8) and simplifying we obtain

$$\left[P^{-1} - \underline{z}_i\underline{z}_i' + \frac{1}{(1-s_i)}(\underline{z}_i - \underline{a}_i)(\underline{z}_i - \underline{a}_i)'\right]^{-1} =$$

$$P^{-1} - \left\{\frac{P^{-1}(\underline{z}_i-\underline{a}_i)(\underline{z}_i-\underline{a}_i)'}{(1 - s_i + d_i)} + \frac{1}{k_i}\left[P^{-1}\underline{z}_i\underline{z}_i' - \frac{P^{-1}(\underline{z}_i-\underline{a}_i)(\underline{z}_i-\underline{a}_i)'P^{-1}\underline{z}_i\underline{z}_i'}{(1 - s_i + d_i)}\right.\right.$$

$$\left.\left. - \frac{P^{-1}\underline{z}_i\underline{z}_i'P^{-1}(\underline{z}_i-\underline{a}_i)(\underline{z}_i-\underline{a}_i)'}{(1 - s_i + d_i)} + \frac{P^{-1}(\underline{z}_i-\underline{a}_i)(\underline{z}_i-\underline{a}_i)'P^{-1}\underline{z}_i\underline{z}_i'P^{-1}(\underline{z}_i-\underline{a}_i)(\underline{z}_i-\underline{a}_i)'}{(1 - s_i + d_i)^2}\right]\right\}P^{-1}, \tag{3.9}$$

where $d_i = (\underline{z}_i-\underline{a}_i)'P^{-1}(\underline{z}_i-\underline{a}_i)$, (scalar)

and $k_i = 1 - \underline{z}_i'P^{-1}\underline{z}_i + \dfrac{\underline{z}_i'P^{-1}(\underline{z}_i-\underline{a}_i)(\underline{z}_i-\underline{a}_i)'P^{-1}\underline{z}_i}{(1 - s_i + d_i)}$. (scalar)

The last term in equation (3.9) can be rewritten as

$$P^{-1}[\underline{z}_i-\underline{a}_i][\underline{z}_i-\underline{a}_i]'\left[\frac{k_i - (1-\underline{z}_i'P^{-1}\underline{z}_i)}{(1 - s_i + d_i)}\right],$$

and combining this with the first term in curly brackets gives

$$P^{-1}[\underline{z}_i-\underline{a}_i][\underline{z}_i-\underline{a}_i]'\left[-\frac{(1 - \underline{z}_i'P^{-1}\underline{z}_i)}{k_i(1 - s_i + d_i)}\right].$$

Thus equation (3.9) can be written as

$$P^{-1} + \frac{1}{k_i}\left[-\frac{(1-\underline{z}_i'P^{-1}\underline{z}_i)}{(1 - s_i + d_i)}P^{-1}(\underline{z}_i-\underline{a}_i)(\underline{z}_i-\underline{a}_i)' + P^{-1}\underline{z}_i\underline{z}_i'\right.$$

$$\left. - \frac{(\underline{z}_i-\underline{a}_i)'P^{-1}\underline{z}_i}{(1 - s_i + d_i)}P^{-1}(\underline{z}_i-\underline{a}_i)\underline{z}_i' - \frac{\underline{z}_i'P^{-1}(\underline{z}_i-\underline{a}_i)}{(1 - s_i + d_i)}P^{-1}\underline{z}_i(\underline{z}_i-\underline{a}_i)'\right]P^{-1}. \tag{3.10}$$

Now consider the second "term" in equation (3.4), viz:

$$\left\{ [Z'X - \underline{z}_i\underline{x}_i'] \left[(X'X)^{-1} + \frac{(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}}{1 - \underline{x}_i'(X'X)^{-1}\underline{x}_i} \right] [X'\underline{y} - \underline{x}_i y_i] \right\}$$

$$= \Big[Z'X(X'X)^{-1}X'\underline{y} + \underline{z}_i\underline{x}_i'(X'X)^{-1}\underline{x}_i y_i$$

$$+ \frac{1}{(1-s_i)} \cdot Z'X(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}X'\underline{y} + \frac{1}{(1-s_i)} \cdot \underline{z}_i\underline{x}_i'(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}\underline{x}_i y_i$$

$$- Z'X(X'X)^{-1}\,\underline{x}_i y_i - \underline{z}_i\underline{x}_i'(X'X)^{-1}X'\underline{y}$$

$$- \frac{1}{(1-s_i)} \cdot Z'X(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}\underline{x}_i y_i - \frac{1}{(1-s_i)} \cdot \underline{z}_i\underline{x}_i'(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}X'\underline{y} \Big] . \tag{3.11}$$

Let

$$\underline{q} = Z'X(X'X)^{-1}X'\underline{y} ,$$

then equation (3.11) can be written as

$$\underline{q} - y_i\underline{z}_i + \frac{1}{(1-s_i)} \left[(\underline{z}_i - \underline{a}_i)(y_i - \underline{x}_i'(X'X)^{-1}X'\underline{y}) \right] . \tag{3.12}$$

To obtain an expression for $\hat{\underline{\theta}}_i$ in (3.2), we must postmultiply equation (3.10) by expression (3.12). Postmultiplying equation (3.10) by $\underline{q}$ we obtain

$$\hat{\underline{\theta}} + \frac{1}{k_i} \Big[- \frac{(1 - \underline{z}_i'P^{-1}\underline{z}_i)}{(1 - s_i + d_i)} P^{-1}(\underline{z}_i - \underline{a}_i)\;(\underline{z}_i - \underline{a}_i)' + P^{-1}\underline{z}_i\underline{z}_i'$$

$$- \frac{(\underline{z}_i - \underline{a}_i)'P^{-1}\underline{z}_i}{(1 - s_i + d_i)} P^{-1}(\underline{z}_i - \underline{a}_i)\underline{z}_i' - \frac{\underline{z}_i'P^{-1}(\underline{z}_i - \underline{a}_i)}{(1 - s_i + d_i)} P^{-1}\underline{z}_i(\underline{z}_i - \underline{a}_i)' \Big] \hat{\underline{\theta}} .$$

Postmultiplying equation (3.10) by $-y_i \underline{z}_i$ and simplifying we obtain

$$-\frac{1}{k_i}\left[P^{-1}\underline{z}_i - \frac{(\underline{z}_i-\underline{a}_i)'P^{-1}\underline{z}_i}{(1-s_i+d_i)} P^{-1}(\underline{z}_i-\underline{a}_i)\right] y_i \ .$$

Postmultiplying equation (3.10) by the term in square brackets in expression (3.12) and simplifying we obtain

$$\frac{1}{k_i}\left[\frac{(1-\underline{z}_i'P^{-1}\underline{z}_i)}{(1-s_i+d_i)} P^{-1}(\underline{z}_i-\underline{a}_i) + \frac{\underline{z}_i'P^{-1}(\underline{z}_i-\underline{a}_i)}{(1-s_i+d_i)} P^{-1}\underline{z}_i\right]\left[y_i - \underline{x}_i'\hat{\underline{\pi}}\right] \ .$$

Then rearranging the above expressions we obtain

$$\hat{\underline{\theta}}_i = \hat{\underline{\theta}} + \left[\frac{(\underline{z}_i-\underline{a}_i)'P^{-1}\underline{z}_i}{k_i(1-s_i+d_i)} P^{-1}(\underline{z}_i-\underline{a}_i) - \frac{1}{k_i} P^{-1}\underline{z}_i\right]\left[y_i-\underline{z}_i'\hat{\underline{\theta}}\right]$$

$$+ \left[\frac{(1-\underline{z}_i'P^{-1}\underline{z}_i)}{k_i(1-s_i+d_i)} P^{-1}(\underline{z}_i-\underline{a}_i) + \frac{\underline{z}_i'P^{-1}(\underline{z}_i-\underline{a}_i)}{k_i(1-s_i+d_i)} P^{-1}\underline{z}_i\right] \times$$

$$\left[(y_i - \underline{x}_i'\hat{\underline{\pi}}) - (\underline{z}_i-\underline{a}_i)'\hat{\underline{\theta}}\right]. \qquad (i=1,2,\ldots,N) \qquad (3.13)$$

Note that $(y_i - \underline{x}_i'\hat{\underline{\pi}})$ is the ith component of the reduced form residual vector $\hat{\underline{v}} = (I - M_X)\underline{y}$, we denote it therefore by $\hat{v}_i$. Similarly, $\hat{u}_i = (y_i - \underline{z}_i\hat{\underline{\theta}})$ is the ith component of the structural form residual vector $\hat{\underline{u}} = (\underline{y} - Z\hat{\underline{\theta}})$, and $\hat{w}_i = (y_i - \underline{a}_i'\hat{\underline{\theta}})$ is the ith component of the "second-stage" residual vector $\hat{\underline{w}} = (\underline{y} - M_X Z\hat{\underline{\theta}})$ where $M_X = X(X'X)^{-1}X'$.

Equation (3.13) was used for computing the J2SLS estimator and its associated test statistic in the Monte Carlo study of Chapters 6 and 7.

For future analysis, it will be convenient to rewrite equation (3.13) as

$$\underline{\hat{\theta}}_i = \underline{\hat{\theta}} + P^{-1}\underline{g}_i \, , \qquad (i = 1,2,\ldots,N) \tag{3.14}$$

where $$\underline{g}_i = (\underline{h}_i + \underline{j}_i)\hat{u}_i + \underline{h}_i(\hat{v}_i - \hat{w}_i) \, , \tag{3.15}$$

$$\underline{h}_i = \frac{(1 - \underline{z}_i' P^{-1} \underline{z}_i)}{k_i(1-s_i+d_i)} (\underline{z}_i - \underline{a}_i) + \frac{\underline{z}_i' P^{-1} (\underline{z}_i - \underline{a}_i)}{k_i(1-s_i+d_i)} \underline{z}_i \, , \tag{3.16}$$

and $$\underline{j}_i = \frac{(\underline{z}_i - \underline{a}_i)' P^{-1} \underline{z}_i}{k_i(1-s_i+d_i)} (\underline{z}_i - \underline{a}_i) - \frac{1}{k_i} \underline{z}_i \, . \tag{3.17}$$

The result in (3.14) is given in Phillips [42].

3.2 An Expression for the J2SLS Estimator

To form the J2SLS estimator we are required to take the summation of equation (3.14) over all i $(i = 1,2,\ldots,N)$ omitted observations. Using equation (3.14) we can form the J2SLS estimator as

$$J(\underline{\hat{\theta}}) = N\underline{\hat{\theta}} - \frac{(N-1)}{N} \sum_{i=1}^{N} \underline{\hat{\theta}}_i$$

$$= \underline{\hat{\theta}} - \frac{(N-1)}{N} \cdot P^{-1} \sum_{i=1}^{N} \underline{g}_i \, ,$$

and using equation (3.15) we then obtain

$$J(\underline{\hat{\theta}}) = \underline{\hat{\theta}} - \frac{(N-1)}{N} \cdot P^{-1} \left[\sum_{i=1}^{N} (\underline{h}_i + \underline{j}_i)\hat{u}_i + \sum_{i=1}^{N} \underline{h}_i (\hat{v}_i - \hat{w}_i) \right] \, . \tag{3.18}$$

Substituting for $\underline{h}_i$ from (3.16) we obtain

$$\sum_{i=1}^{N} \underline{h}_i (\hat{v}_i - \hat{w}_i) = \sum_{i=1}^{N} \frac{1 - \underline{z}_i' P^{-1} \underline{z}_i}{k_i(1-s_i+d_i)} (\underline{z}_i - \underline{a}_i)(\hat{v}_i - \hat{w}_i)$$

$$+ \sum_{i=1}^{N} \frac{\underline{z}_i' P^{-1} (\underline{z}_i - \underline{a}_i)}{k_i(1-s_i+d_i)} \underline{z}_i (\hat{v}_i - \hat{w}_i) \, . \tag{3.19}$$

Since $(\underline{z}_i - \underline{a}_i)$ and $\underline{z}_i$ are the ith columns of $Z'(I-M_X)$ and Z' respectively, it follows that

$$\sum_{i=1}^{N} \frac{1-\underline{z}_i' P^{-1} \underline{z}_i}{k_i(1-s_i+d_i)} (\underline{z}_i - \underline{a}_i)(\hat{v}_i - \hat{w}_i) = Z'(I-M_X)\Lambda_3(\hat{\underline{v}} - \hat{\underline{w}}) \qquad (3.20)$$

and

$$\sum_{i=1}^{N} \frac{(\underline{z}_i - \underline{a}_i)' P^{-1} \underline{z}_i}{k_i(1-s_i+d_i)} \underline{z}_i(\hat{v}_i - \hat{w}_i) = - Z'\Lambda_2(\hat{\underline{v}} - \hat{\underline{w}}) \quad ; \qquad (3.21)$$

where Λ_3 is an $N \times N$ diagonal matrix with iith component

$$(\Lambda_3)_{ii} = \frac{1 - \underline{z}_i' P^{-1} \underline{z}_i}{k_i(1-s_i+d_i)} \quad , \qquad (i = 1,2,\ldots,N)$$

and Λ_2 is an $N \times N$ diagonal matrix with iith component

$$(\Lambda_2)_{ii} = \frac{-\underline{z}_i' P^{-1}(\underline{z}_i - \underline{a}_i)}{k_i(1-s_i+d_i)} . \qquad (i = 1,2,\ldots,N)$$

Substituting from equations (3.20) and (3.21) into equation (3.19) gives

$$\sum_{i=1}^{N} \underline{h}_i(\hat{v}_i - \hat{w}_i) = Z'(I-M_X)\Lambda_3(\hat{\underline{v}} - \hat{\underline{w}}) - Z'\Lambda_2(\hat{\underline{v}} - \hat{\underline{w}}) . \qquad (3.22)$$

Similarly it can be shown that

$$\sum_{i=1}^{N} \underline{h}_i \hat{u}_i = Z'(I-M_X)\Lambda_3\hat{\underline{u}} - Z'\Lambda_2\hat{\underline{u}} . \qquad (3.23)$$

Consider the term $\sum_{i=1}^{N} \underline{j}_i \hat{u}_i$ in equation (3.18). Substituting for $\underline{j}_i$ from equation (3.17) gives

$$\sum_{i=1}^{N} \underline{j}_i \hat{u}_i = \sum_{i=1}^{N} \frac{(\underline{z}_i - \underline{a}_i)' P^{-1} \underline{z}_i}{k_i(1-s_i+d_i)} (\underline{z}_i - \underline{a}_i)\hat{u}_i - \sum_{i=1}^{N} \frac{1}{k_i} \cdot \underline{z}_i \hat{u}_i$$

$$= - Z'(I-M_X)\Lambda_2\hat{\underline{u}} - Z'\Lambda_1\hat{\underline{u}} \quad , \qquad (3.24)$$

where Λ_1 is an $N \times N$ diagonal matrix with iith component

$$(\Lambda_1)_{ii} = \frac{1}{k_i} .$$

Substituting from equations (3.22), (3.23), and (3.24) into equation (3.18) we obtain

$$J(\underline{\hat{\theta}}) = \underline{\hat{\theta}} + \frac{(N-1)}{N}\Big[P^{-1}Z'(\Lambda_1 + 2\Lambda_2 - \Lambda_3)\underline{\hat{u}}$$

$$- P^{-1}Z'M_X(\Lambda_2 - \Lambda_3)\underline{\hat{u}}$$

$$+ P^{-1}Z'(\Lambda_2 - \Lambda_3)(\underline{\hat{v}} - \underline{\hat{w}})$$

$$+ P^{-1}Z'M_X\Lambda_3(\underline{\hat{v}} - \underline{\hat{w}})\Big] . \qquad (3.25)$$

The ensuing analysis is simplified by writing equation (3.25) in a slightly amended form. Recall that

$$\underline{\hat{u}} = [I - Z(Z'M_XZ)^{-1}Z'M_X]\underline{y} ,$$

$$\underline{\hat{v}} = [I - M_X]\underline{y} ,$$

and

$$\underline{\hat{w}} = [I - M_XZ(Z'M_XZ)^{-1}Z'M_X]\underline{y} ,$$

from which it follows that

$$Z'(\underline{\hat{v}} - \underline{\hat{w}}) = \underline{0} ,$$

$$Z'M_X\underline{\hat{u}} = \underline{0} ,$$

and $Z'M_X(\underline{\hat{v}} - \underline{\hat{w}}) = \underline{0}$.

Thus, if we define

$$\overline{\Lambda}_3 = I - \Lambda_3 ,$$

we can rewrite equation (3.25) as

$$J(\underline{\hat{\theta}}) = \underline{\hat{\theta}} + \frac{(N-1)}{N} . P^{-1}\Big[Z'(\Lambda_1 + 2\Lambda_2 - \Lambda_3)\ \underline{\hat{u}}$$

$$- Z'M_X(\Lambda_2 - \overline{\Lambda}_3)\underline{\hat{u}}$$

$$+ Z'(\Lambda_2 - \overline{\Lambda}_3)(\underline{\hat{v}} - \underline{\hat{w}})$$

$$+ Z'M_X\overline{\Lambda}_3(\underline{\hat{v}} - \underline{\hat{w}})\Big] . \qquad (3.26)$$

3.3 The Asymptotic Equivalence of J2SLS and 2SLS

3.3.1 Preliminary Results

In this section the asymptotic behaviour of the three diagonal matrices (viz: Λ_1, Λ_2, and $\bar{\Lambda}_3$) introduced earlier in this Chapter will be investigated.

Essentially we must consider the following terms:

$$k_i,\ (1 - s_i + d_i),\ \underline{z}_i' P^{-1} \underline{z}_i \text{ and } \underline{z}_i' P^{-1} (\underline{z}_i - \underline{a}_i),$$

where $$s_i = \underline{x}_i' (X'X)^{-1}\, \underline{x}_i ,$$

$$d_i = (\underline{z}_i - \underline{a}_i)' P^{-1} (\underline{z}_i - \underline{a}_i) ,$$

and $$k_i = 1 - \underline{z}_i' P^{-1} \underline{z}_i + \frac{\underline{z}_i' P^{-1} (\underline{z}_i - \underline{a}_i)(\underline{z}_i - \underline{a}_i)' P^{-1} \underline{z}_i}{(1 - s_i + d_i)} .$$

The reader is reminded of the following results which were established in Chapter 2:

(a) $$\operatorname*{plim}_{N\to\infty} \left[\frac{1}{N} . Z'X \left(\frac{1}{N} . X'X \right)^{-1} \frac{1}{N} . X'Z \right]^{-1} = \operatorname*{plim}_{N\to\infty} N.P^{-1} = \Sigma_P^{-1} ,$$

where Σ_P^{-1} is a finite positive definite matrix;

and

(b) $$\operatorname*{plim}_{N\to\infty} \frac{1}{N} . X'Z = \Sigma_{XZ} ,$$

where Σ_{XZ} is a finite matrix.

For the ensuing analysis result (b) will be expressed in a different form. Since

$$X'Z = X'[X\Pi + V : X_1] ,$$

we can rewrite result (b) as

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N} . X'Z = \left[\Sigma_{XX}\Pi : \Sigma_{XX_1} \right] .$$

It follows from assumption (iii), section 2.1.3, that

$$\lim_{N\to\infty} Ns_i = \lim_{N\to\infty} \underline{x}_i' \left(\frac{1}{N} \cdot X'X\right)^{-1} \underline{x}_i = \underline{x}_i' \Sigma_{XX}^{-1} \underline{x}_i ,$$

which is a finite constant, and consequently

$$\lim_{N\to\infty} s_i = 0 . \tag{3.27}$$

Consider the vector $\underline{a}_i'$, where

$$\underline{a}_i' = \underline{x}_i'(X'X)^{-1}X'Z = \underline{x}_i'(X'X)^{-1}X' \; [Y : X_1] ;$$

i.e. $\underline{a}_i' = [\underline{x}_i'\hat{\Pi} : \underline{x}_{1i}'] .$ (3.28)

Using result (a) it follows that

$$\operatorname*{plim}_{N\to\infty} N \cdot \underline{a}_i' P^{-1} \underline{a}_i = [\underline{x}_i'\Pi : \underline{x}_{1i}'] \Sigma_P^{-1} [\Pi'\underline{x}_i : \underline{x}_{1i}] , \tag{3.29}$$

a finite constant, where $\underline{x}_{1i}'$ is the ith row of X_1 .

This result can be shown as follows:

the matrix $\qquad (X'X)^{-1}X'X_1$

is a submatrix of $\qquad (X'X)^{-1}X'X = I_K ,$

and thus consists of K_1 columns of the $K \times K$ identity matrix. By premultiplying these columns by $\underline{x}_i'$ we obtain $\underline{x}_{1i}'$.

It will be convenient to write equation (3.29) as

$$\operatorname*{plim}_{N\to\infty} N \cdot \underline{a}_i' P^{-1} \underline{a}_i = \bar{\underline{a}}_i' \Sigma_P^{-1} \bar{\underline{a}}_i , \tag{3.30}$$

where $\operatorname*{plim}_{N\to\infty} \underline{a}_i = \bar{\underline{a}}_i$, a finite constant vector.

We can conclude, therefore, that

$$\operatorname*{plim}_{N\to\infty} \underline{a}_i' P^{-1} \underline{a}_i = 0. \tag{3.31}$$

Consider the term

$$(\underline{z}_i - \underline{a}_i)' \; P^{-1} \; (\underline{z}_i - \underline{a}_i) \;,$$

where the vector $(\underline{z}_i - \underline{a}_i)'$ is the ith observation on the matrix

$$[Z - X(X'X)^{-1}X'Z] \;.$$

Partitioning this matrix we obtain

$$[Y : X_1] - X(X'X)^{-1}X'[Y : X_1]$$

$$= [Y : X_1] - [X\hat{\Pi} : X_1]$$

$$= [\hat{V} : 0] \;,$$

which will have ith observation denoted by

$$[\hat{\underline{v}}_i' : \underline{0}'] \;.$$

It follows that

$$(\underline{z}_i - \underline{a}_i)' \; P^{-1}(\underline{z}_i - \underline{a}_i) = \hat{\underline{v}}_i' P^{-1} \hat{\underline{v}}_i \;, \qquad (3.32)$$

$$(\underline{z}_i - \underline{a}_i)' P^{-1} \underline{z}_i = [\hat{\underline{v}}_i' : \underline{0}'] \; P^{-1} [\underline{v}_i + \Pi' \underline{x}_i : \underline{x}_{1i}] \;, \qquad (3.33)$$

and

$$\underline{a}_i' P^{-1} (\underline{z}_i - \underline{a}_i) = [\underline{x}_i' \hat{\Pi} : \underline{x}_{1i}'] \; P^{-1} [\hat{\underline{v}}_i : \underline{0}] \;. \qquad (3.34)$$

Since each element of the OLS reduced form residuals matrix converges in distribution to the corresponding element of the disturbance matrix, from equation (3.32) and using result (a) we can write

$$\underset{N \to \infty}{\text{plim}} \; N \,.\, \hat{\underline{v}}_i' \; P^{-1} \hat{\underline{v}}_i = \underline{v}_i' \; \Sigma_P^{-1} \; \underline{v}_i \;. \qquad (3.35)$$

Since Σ_P^{-1} is a finite positive definite matrix, and since the $\underline{v}_i$ $(i = 1,2,\ldots,N)$ are independently and identically distributed with mean zero and finite covariance matrix (this fact follows from assumption (v), section 2.1.3, since the reduced form disturbances are just linear combinations of the structural form disturbances), it follows that

$$\underline{v}_i' \, \Sigma_P^{-1} \, \underline{v}_i$$

is a random variable with finite mean and variance. Hence

$$\underset{N \to \infty}{\text{plim}} \; \hat{\underline{v}}_i' \, P^{-1} \, \hat{\underline{v}}_i = 0 \;, \qquad (3.36)$$

since $\frac{1}{N} \cdot v_i' \, \Sigma_P^{-1} \, v_i$ converges in probability to zero.

Combining equations (3.32), (3.33), and (3.34) we can write

$$(\underline{z}_i - \underline{a}_i)'P^{-1}\underline{z}_i = (\underline{z}_i - \underline{a}_i)'P^{-1}(\underline{z}_i - \underline{a}_i) + (\underline{z}_i - \underline{a}_i)'P^{-1}\underline{a}_i \;. \qquad (3.37)$$

The probability limit of the last term in equation (3.37) can be written as

$$\underset{N \to \infty}{\text{plim}} \; N.(\underline{z}_i - \underline{a}_i)'P^{-1}\underline{a}_i = \underline{v}_i' \, \Sigma_P^{-1} \, \bar{\underline{a}}_i \;. \qquad (3.38)$$

Since $\Sigma_P^{-1} \, \bar{\underline{a}}_i$ is a finite vector, and since the $\underline{v}_i$ $(i = 1,2,\ldots,N)$ are independently and identically distributed with mean zero and finite covariance matrix, it follows that

$$\underline{v}_i' \, \Sigma_P^{-1} \, \bar{\underline{a}}_i$$

is a random variable with mean zero and finite variance. Thus

$$\underset{N \to \infty}{\text{plim}} \; (\underline{z}_i - \underline{a}_i)' \, P^{-1} \, \underline{a}_i = 0 \;,$$

since $\frac{1}{N} \cdot \underline{v}_i' \, \Sigma_P^{-1} \, \bar{\underline{a}}_i$ converges in probability to zero.

Combining the above result with that given by equation (3.36), and substituting into equation (3.37), we have shown that

$$\underset{N \to \infty}{\text{plim}} \; (\underline{z}_i - \underline{a}_i)' \, P^{-1}\underline{z}_i = 0 \;. \qquad (3.39)$$

We now consider the scalar $\underline{z}_i' \, P^{-1}\underline{z}_i$ which can be written as

$$\begin{aligned} \underline{z}_i'P^{-1}\underline{z}_i &= [(\underline{z}_i - \underline{a}_i) + \underline{a}_i]' \, P^{-1} \, [(\underline{z}_i - \underline{a}_i) + \underline{a}_i] \\ &= (\underline{z}_i - \underline{a}_i)'P^{-1}(\underline{z}_i - \underline{a}_i) + \underline{a}_i'P^{-1}\underline{a}_i \\ &\quad + 2\underline{a}_i'P^{-1}(\underline{z}_i - \underline{a}_i) \;. \end{aligned}$$

Using equations (3.29), (3.32), (3.36), and (3.39) it follows that

$$\underset{N\to\infty}{\text{plim}} \ \underline{z}_i' P^{-1} \underline{z}_i = 0 . \quad (3.40)$$

From equations (3.27), (3.32), and (3.36) we have shown that

$$\underset{N\to\infty}{\text{plim}} \ (1 - s_i + d_i) = 1 , \quad (3.41)$$

and from equations (3.39), (3.40), and (3.41) we have shown that

$$\underset{N\to\infty}{\text{plim}} \ k_i = 1 . \quad (3.42)$$

3.3.2 Proof of Asymptotic Equivalence

To prove that the 2SLS and J2SLS estimators are asymptotically equivalent, we are required to show that

$$\underset{N\to\infty}{\text{plim}} \ \sqrt{N}[J(\underline{\hat{\theta}}) - \underline{\hat{\theta}}] = \underline{0} .$$

From equation (3.26) we can write this requirement as

$$\underset{N\to\infty}{\text{plim}} \sqrt{N} \ [J(\underline{\hat{\theta}}) - \underline{\hat{\theta}}] = \underset{N\to\infty}{\text{plim}} \left(\frac{1}{N} . P\right)^{-1} \left\{ \underset{N\to\infty}{\text{plim}} \frac{1}{\sqrt{N}} . Z'(\Lambda_1 + 2\Lambda_2 - \Lambda_3)\underline{\hat{u}} \right.$$

$$- \underset{N\to\infty}{\text{plim}} \frac{1}{\sqrt{N}} . Z'M_X(\Lambda_2 - \bar{\Lambda}_3)\underline{\hat{u}} + \underset{N\to\infty}{\text{plim}} \frac{1}{\sqrt{N}} . Z'(\Lambda_2 - \bar{\Lambda}_3)(\underline{\hat{v}} - \underline{\hat{w}})$$

$$\left. + \underset{N\to\infty}{\text{plim}} \frac{1}{\sqrt{N}} . Z'M_X\bar{\Lambda}_3(\underline{\hat{v}} - \underline{\hat{w}}) \right\} = \underline{0} . \quad (3.43)$$

A term by term evaluation of equation (3.43) now follows.

Consider the first term in curly brackets in equation (3.43), viz:

$$Z'(\Lambda_1 + 2\Lambda_2 - \Lambda_3)\underline{\hat{u}} . \quad (3.44)$$

We know that:

the iith component of Λ_1 is $1/k_i$;

the iith component of Λ_2 is $\dfrac{-(\underline{z}_i - \underline{a}_i)'P^{-1}\underline{z}_i}{k_i(1 - s_i + d_i)}$;

and the iith component of Λ_3 is $\dfrac{(1 - \underline{z}_i' P^{-1}\underline{z}_i)}{k_i(1 - s_i + d_i)}$;

thus, after some algebraic manipulation, the iith component of the bracketed term in expression (3.44) can be written as

$$(\Lambda_1 + 2\Lambda_2 - \Lambda_3)_{ii} = -\frac{s_i}{k_i(1 - s_i + d_i)} + \frac{\underline{a}_i' \, P^{-1}\underline{a}_i}{k_i(1 - s_i + d_i)} .$$

Let $\underline{a}_s' P^{-1}\underline{a}_s$ be the largest of the $\underline{a}_i' \, P^{-1}\underline{a}_i$ $(i = 1,2,\ldots,N)$, then it follows from equation (3.30) that

$$\operatorname*{plim}_{N\to\infty} N.\underline{a}_s' P^{-1}\underline{a}_s = \bar{\underline{a}}_s' \, \Sigma_P^{-1} \, \bar{\underline{a}}_s ,$$

a finite positive definite quadratic form. It follows, therefore, that

$$\operatorname*{plim}_{N\to\infty} \sqrt{N} \max_{1\leq i\leq N} \underline{a}_i' \, P^{-1}\underline{a}_i = 0 . \qquad (3.45)$$

Using a similar argument it can be shown that

$$\lim_{N\to\infty} \sqrt{N} \max_{1\leq i\leq N} s_i = 0 . \qquad (3.46)$$

Combining equations (3.41), (3.42), (3.45), and (3.46) we obtain

$$\operatorname*{plim}_{N\to\infty} \sqrt{N} \max_{1\leq i\leq N} (\Lambda_1 + 2\Lambda_2 - \Lambda_3)_{ii} = 0 . \qquad (3.47)$$

The jth component of the random vector (3.44) can be written as

$$\sum_{i=1}^{N} m_{ij}(\Lambda_1 + 2\Lambda_2 - \Lambda_3)_{ii} \, \hat{u}_i$$

$$+ \sum_{i=1}^{N} (\Lambda_1 + 2\Lambda_2 - \Lambda_3)_{ii}\hat{u}_i v_{ij} \qquad (3.48)$$

where the m_{ij} $(i = 1,2,\ldots,N;\ j = 1,2,\ldots,K_1 + g)$ represent the nonstochastic part of the z_{ij} (the ijth element of Z), and the $v_{ij}(i = 1,2,\ldots,N;\ \ j = 1,2,\ldots\ K_1 + g)$ represent the reduced form disturbance part of the z_{ij} (where appropriate). Without loss of generality we can assume that the observations on the (g) explanatory endogenous variables occur in the first g columns of Z; thus $v_{ij} = 0$ for all $j > g$ (for all i).

Consider the first term in expression (3.48), viz:

$$\sum_{i=1}^{N} m_{ij}(\Lambda_1 + 2\Lambda_2 - \Lambda_3)_{ii}\ \hat{u}_i\ .$$

Since the m_{ij} are nonstochastic, it follows from equation (3.47) that

$$\operatorname*{plim}_{N\to\infty} \sqrt{N} \max_{1\leq i\leq N} \left| m_{ij}(\Lambda_1 + 2\Lambda_2 - \Lambda_3)_{ii} \right| = 0\ . \qquad (3.49)$$

We now require the following theorem which is taken from Malinvaud [27; pp. 322-323] and is cited without proof.

THEOREM I.

Let x_{tT} $(t = 1,2,\ldots,T;\ \ T = 1,2,\ldots.)$ be random variables. If

$$\operatorname*{plim}_{T\to\infty} \max_{1\leq t\leq T} \left| x_{tT} \right| = 0\ ,$$

and if the u_t are mutually independent random variables identically distributed with zero mean, then:

$$\operatorname*{plim}_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} x_{tT} = 0 \quad \text{and} \quad \operatorname*{plim}_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} u_t x_{tT} = 0\ .$$

Since each component of the 2SLS residual vector converges in distribution to the corresponding element of the disturbance vector, using Theorem I together with equation (3.49) if follows that

$$\underset{N\to\infty}{\text{plim}}\ \frac{1}{N}\sum_{i=1}^{N} m_{ij}\ \sqrt{N}\ (\Lambda_1 + 2\Lambda_2 - \Lambda_3)_{ii}\hat{u}_i = 0\ .$$

Now consider the second term in expression (3.48), viz:

$$\sum_{i=1}^{N} (\Lambda_1 + 2\Lambda_2 - \Lambda_3)_{ii}\ \hat{u}_i\ v_{ij}\ . \tag{3.50}$$

The reduced form disturbances associated with the g explanatory variables can be decomposed into a term ($\underline{u}\ \underline{\Psi}'$) which is proportional to the disturbance term in the ith structural equation, and a term (E) which is uncorrelated with $\underline{u}$ (e.g. see Nagar [37; p.577]), viz:

$$V = \underline{u}\ \underline{\Psi}' + E\ . \tag{3.51}$$

The ith row of v can be written as

$$\underline{v}_i' = u_i\ \underline{\Psi}' + \underline{e}_i'\ , \qquad (i = 1,2,\ldots,N)$$

whereupon by substituting for v_{ij} in (3.50) we obtain

$$\sum_{i=1}^{N} (\Lambda_1 + 2\Lambda_2 - \Lambda_3)_{ii}\ \psi_j\hat{u}_i u_i$$

$$+ \sum_{i=1}^{N} (\Lambda_1 + 2\Lambda_2 - \Lambda_3)_{ii}\ \hat{u}_i e_{ij}\ ,$$

where ψ_j denotes the jth element of $\underline{\Psi}'$.

Let $E(u_i^2) = \sigma^2$ then, since $\hat{u}_i \to u_i$ as $N \to \infty$, it follows that the $(u_i^2 - \sigma^2)$ are (asymptotically) independently and identically distributed random variables with mean zero.

Since the ψ_j are nonstochastic it follows, using equation (3.47), that

$$\underset{N\to\infty}{\text{plim}}\ \sqrt{N}\ \max_{1\leq i\leq N} \left|(\Lambda_1 + 2\Lambda_2 - \Lambda_3)_{ii}\ \psi_j\ \right| = 0\ .$$

Combining this result with Theorem I it follows that

$$\underset{N\to\infty}{\text{plim}} \frac{1}{N} \sum_{i=1}^{N} \sqrt{N}\ (\Lambda_1 + 2\Lambda_2 - \Lambda_3)_{ii}\ \psi_j (\hat{u}_i u_i - \sigma^2) = 0 \ .$$

This result implies that

$$\underset{N\to\infty}{\text{plim}} \frac{1}{N} \sum_{i=1}^{N} \sqrt{N}(\Lambda_1 + 2\Lambda_2 - \Lambda_3)_{ii}\ \psi_j \hat{u}_i u_i$$

$$= \underset{N\to\infty}{\text{plim}} \frac{1}{N} \sum_{i=1}^{N} \sqrt{N}(\Lambda_1 + 2\Lambda_2 - \Lambda_3)_{ii}\ \psi_j\ \sigma^2$$

$$= \sigma^2\ \underset{N\to\infty}{\text{plim}} \frac{1}{N} \sum_{i=1}^{N} \sqrt{N}(\Lambda_1 + 2\Lambda_2 - \Lambda_3)_{ii}\ \psi_j$$

$= 0$ from Theorem I .

Since $\hat{u}_i e_{ij} \to u_i e_{ij}$ (as $N \to \infty$) which are mutually independent random variables (i.e. $u_{i+1} e_{i+1,j}$ is independent of $u_i e_{ij}$) it follows from Theorem I that

$$\underset{N\to\infty}{\text{plim}} \frac{1}{N} \sum_{i=1}^{N} \sqrt{N}(\Lambda_1 + 2\Lambda_2 - \Lambda_3)_{ii} \hat{u}_i e_{ij} = 0.$$

This concludes the analysis on the first term in curly brackets in equation (3.43). To summarize, we have shown that

$$\underset{N\to\infty}{\text{plim}} \frac{1}{\sqrt{N}} \cdot Z'(\Lambda_1 + 2\Lambda_2 - \Lambda_3)\underline{\hat{u}} = \underline{0}\ .$$

Consider the second term in curly brackets in equation (3.43), viz:

$$- Z'M_X(\Lambda_2 - \overline{\Lambda}_3)\underline{\hat{u}}\ , \tag{3.52}$$

where the iith component of the term in brackets can be written as

$$(\Lambda_2 - \overline{\Lambda}_3)_{ii} = - \frac{(\underline{z}_i - \underline{a}_i)' P^{-1} \underline{z}_i}{k_i (1 - s_i + d_i)} - 1 + \frac{(1 - \underline{z}_i' P^{-1} \underline{z}_i)}{k_i (1 - s_i + d_i)}\ . \tag{3.53}$$

Expressing equation (3.53) in terms of a common denominator, the numerator can be written as

$$- (\underline{z}_i - \underline{a}_i)'P^{-1}\underline{z}_i + s_i - s_i\underline{z}_i'\ P^{-1}\underline{z}_i - d_i + d_i\underline{z}_i'\ P^{-1}\underline{z}_i$$
$$- \underline{z}_i'\ P^{-1}(\underline{z}_i - \underline{a}_i)(\underline{z}_i - \underline{a}_i)'P^{-1}\underline{z}_i\ .$$

The following probability limits can now be established:

$$\operatorname*{plim}_{N\to\infty} N(s_i\underline{z}_i'P^{-1}\underline{z}_i) = \operatorname*{plim}_{N\to\infty} \frac{1}{N} \cdot \underline{x}_i'\left(\frac{1}{N} \cdot X'X\right)^{-1} \underline{x}_i \cdot \underline{z}_i'\left(\frac{1}{N} \cdot P\right)^{-1} \underline{z}_i = 0\ ,$$

using equation (3.40) and the knowledge that $\underline{x}_i'\ \Sigma_{XX}^{-1}\ \underline{x}_i$ is a finite constant;

$$\operatorname*{plim}_{N\to\infty}\left(\underline{z}_i'\ P^{-1}(\underline{z}_i - \underline{a}_i)(\underline{z}_i - \underline{a}_i)'P^{-1}\underline{z}_i\right) = \operatorname*{plim}_{N\to\infty}\left[\underline{z}_i'P^{-1}(\underline{z}_i - \underline{a}_i)\right]^2 = 0,$$

using equation (3.39);

and

$$\operatorname*{plim}_{N\to\infty} N(d_i\underline{z}_i'P^{-1}\underline{z}_i) = \operatorname*{plim}_{N\to\infty} \frac{1}{N}\left\{\hat{\underline{v}}_i'\left(\frac{1}{N} \cdot P\right)^{-1}\hat{\underline{v}}_i \cdot \underline{z}_i'\left(\frac{1}{N} \cdot P\right)^{-1}\underline{z}_i\right\} = 0\ ,$$

using equations (3.35) and (3.40).

Combining the above three results with equations (3.41) and (3.42) we have shown that

$$\operatorname*{plim}_{N\to\infty} N(\Lambda_2 - \overline{\Lambda}_3)_{ii} = \underline{x}_i'\ \Sigma_{XX}^{-1}\ \underline{x}_i - 2\underline{v}_i'\ \Sigma_P^{-1}\ \underline{v}_i\ , \tag{3.54}$$

and, by the same proof, that

$$\operatorname*{plim}_{N\to\infty} N(\overline{\Lambda}_3)_{ii} = -\ \underline{x}_i'\ \Sigma_{XX}^{-1}\ \underline{x}_i + \underline{v}_i'\ \Sigma_P^{-1}\ \underline{v}_i\ . \tag{3.55}$$

Expression (3.52) can be written as

$$- \frac{1}{N} \cdot Z'X\left(\frac{1}{N} \cdot X'X\right)^{-1} \cdot X'(\Lambda_2 - \overline{\Lambda}_3)\hat{\underline{u}}\ . \tag{3.56}$$

The rth element of

$$X'(\Lambda_2 - \bar{\Lambda}_3)\hat{\underline{u}}$$

can therefore be written as

$$\frac{1}{N} \sum_{i=1}^{N} x_{ir} \underline{x}_i' \left(\frac{1}{N} \cdot X'X\right)^{-1} \underline{x}_i \hat{u}_i$$

$$- \frac{2}{N} \sum_{i=1}^{N} x_{ir} \hat{\underline{v}}_i' \left(\frac{1}{N} \cdot P\right)^{-1} \hat{\underline{v}}_i \hat{u}_i , \qquad (3.57)$$

where x_{ir} $(r = 1,2,\ldots,K)$ is the rith element of X'.

Rearranging the first term in expression (3.57) and taking its probability limit in the context of expression (3.56) gives

$$\Sigma_{ZX} \Sigma_{XX}^{-1} \sum_{i=1}^{N} \lim_{N\to\infty} \underline{x}_i' \left(\frac{1}{N} \cdot X'X\right)^{-1} \underline{x}_i \cdot \underset{N\to\infty}{\text{plim}} \frac{1}{N} \sum_{i=1}^{N} x_{ir} \hat{u}_i = \underline{0} .$$

This result is obtained by noting that the limit term is a finite constant, whilst the Law of Large Numbers (e.g. see Malinvaud [27; Proposition 12, p.322]) ensures that the probability limit term is zero.

Since each element of the 2SLS residual vector converges in distribution to the corresponding element of the disturbance vector, and since each element of the OLS reduced form residuals matrix converges in distribution to the corresponding element of the disturbance matrix, it follows that

$$\hat{\underline{v}}_i' \left(\frac{1}{N} \cdot P\right)^{-1} \hat{\underline{v}}_i \hat{u}_i$$

converges in probability to

$$\underline{v}_i' \Sigma_P^{-1} \underline{v}_i u_i . \qquad (3.58)$$

Using equation (3.51), expression (3.58) can be written as

$$(u_i \underline{\Psi}' + \underline{e}_i') \Sigma_P^{-1} (\underline{\Psi} u_i + \underline{e}_i) u_i ,$$

which upon expansion gives

$$\underline{\Psi}'\Sigma_P^{-1}\,\underline{\Psi}\,u_i^3 + 2\underline{e}_i'\,\Sigma_P^{-1}\,\underline{\Psi}\,u_i^2 + \underline{e}_i'\,\Sigma_P^{-1}\,\underline{e}_i u_i \;. \tag{3.59}$$

In the context of the second term in expression (3.57), the first term in expression (3.59) can be written as

$$-\,\underline{\Psi}'\,\Sigma_P^{-1}\,\underline{\Psi}\,\operatorname*{plim}_{N\to\infty}\frac{2}{N}\sum_{i=1}^{N}x_{ir}[u_i^3 - E(u_i^3)] = 0 \;.$$

Noting that the quadratic form in the above equation is a constant, the Law of Large Numbers ensures the result.

In the context of expression (3.56) this result implies that

$$-\;\Sigma_{ZX}\,\Sigma_{XX}^{-1}\,\underline{\Psi}'\,\Sigma_P^{-1}\,\underline{\Psi}\,\operatorname*{plim}_{N\to\infty}\frac{2}{N}\sum_{i=1}^{N}x_{ir}u_i^3$$

in a finite matrix, provided $E(u_i^3)$ is finite.

Since u_i and $\underline{e}_i$ are uncorrelated random variables (by assumption) it follows, using Theorem I, that the second and third terms in expression (3.59), in the context of expression (3.57), converge in probability to zero.

Collecting results, we have shown that

$$\operatorname*{plim}_{N\to\infty} Z'M_X(\Lambda_2 - \overline{\Lambda}_3)\hat{\underline{u}}$$

is a finite constant, and hence

$$\operatorname*{plim}_{N\to\infty}\frac{1}{\sqrt{N}}\cdot Z'M_X(\Lambda_2 - \overline{\Lambda}_3)\hat{\underline{u}} = \underline{0} \;.$$

The third term in curly brackets in equation (3.43) is

$$Z'(\Lambda_2 - \overline{\Lambda}_3)(\hat{\underline{v}} - \hat{\underline{w}}) \; . \tag{3.60}$$

We know that

$$\hat{\underline{u}} = (y - Z\hat{\underline{\theta}}) \; ,$$

$$\hat{\underline{v}} = (I - M_X)\underline{y} \; ,$$

and

$$\hat{\underline{w}} = (\underline{y} - M_X Z\hat{\underline{\theta}}) \; ,$$

from which it follows that

$$(\hat{\underline{v}} - \hat{\underline{w}}) = - M_X\hat{\underline{u}} \; . \tag{3.61}$$

Substituting from equation (3.61), expression (3.60) can be written as

$$Z'(\Lambda_2 - \overline{\Lambda}_3)X\left(\frac{1}{N} . X'X\right)^{-1}\left(\frac{1}{N} . X'\hat{\underline{u}}\right) . \tag{3.62}$$

Consider the term

$$Z'(\Lambda_2 - \overline{\Lambda}_3)X$$

which has jsth element given by

$$\frac{1}{N}\sum_{i=1}^{N} z_{ij}\underline{x}_i'\left(\frac{1}{N} . X'X\right)^{-1}\underline{x}_i \; x_{is}$$

$$- \frac{2}{N}\sum_{i=1}^{N} z_{ij}\hat{\underline{v}}_i'\left(\frac{1}{N} . P\right)^{-1}\hat{\underline{v}}_i \; x_{is} \quad , \tag{3.63}$$

where x_{is} (s = 1,2,...,K) is the isth element of X.

The first term in expression (3.63) can be expanded as

$$\frac{1}{N}\sum_{i=1}^{N} m_{ij}\underline{x}_i'\left(\frac{1}{N} . X'X\right)^{-1}\underline{x}_i x_{is}$$

$$+ \frac{1}{N}\sum_{i=1}^{N} \underline{x}_i'\left(\frac{1}{N} . X'X\right)^{-1}\underline{x}_i x_{is} v_{ij} \; ,$$

whence the first term of the above expression converges to a finite constant and the second term converges in probability to zero by the

Law of Large Numbers.

The second term in expression (3.63) can be expanded as

$$- \frac{2}{N} \sum_{i=1}^{N} m_{ij} \; x_{is} \; \hat{\underline{v}}_i' \left(\frac{1}{N} \cdot P\right)^{-1} \hat{\underline{v}}_i$$

$$- \frac{2}{N} \sum_{i=1}^{N} x_{is} \; \hat{\underline{v}}_i' \left(\frac{1}{N} \cdot P\right)^{-1} \hat{\underline{v}}_i \; v_{ij} \; . \qquad (3.64)$$

Using the argument preceding expression (3.58), and substituting for $\underline{v}_i$ from equation (3.51), expression (3.64) can be rewritten as

$$- \frac{2}{N} \sum_{i=1}^{N} m_{ij} x_{is} \left[\underline{\Psi}' \Sigma_P^{-1} \; \underline{\Psi} u_i^2 + 2\underline{e}_i' \; \Sigma_P^{-1} \; \underline{\Psi} u_i + \underline{e}_i' \; \Sigma_P^{-1} \underline{e}_i \right]$$

$$- \frac{2}{N} \sum_{i=1}^{N} x_{is} \left[\underline{\Psi}' \Sigma_p^{-1} \underline{\Psi} u_i^2 + 2\underline{e}_i' \; \Sigma_p^{-1} \; \underline{\Psi} u_i + \underline{e}_i' \; \Sigma_p^{-1} \; \underline{e}_i \right] \; (\psi_j u_i + e_{ij}) \; .$$

From our analysis to date, it follows that both terms in the above expression converge in probability to finite constants.

Since, from our initial assumptions, the term to be inverted in expression (3.62) converges to a finite matrix and the term on its right converges in probability to a null vector, we have shown that

$$\operatorname*{plim}_{N \to \infty} \frac{1}{\sqrt{N}} \cdot Z'(\Lambda_2 - \bar{\Lambda}_3)(\hat{\underline{v}} - \hat{\underline{w}}) = \underline{0} \; .$$

The fourth term in curly brackets in equation (3.43) is

$$Z' M_X \; \bar{\Lambda}_3 (\hat{\underline{v}} - \hat{\underline{w}})$$

which, using equation (3.61), can be rewritten as

$$\left(\frac{1}{N} \cdot Z'X\right)\left(\frac{1}{N} \cdot X'X\right)^{-1} X' \; \bar{\Lambda}_3 X \left(\frac{1}{N} \cdot X'X\right)^{-1}\left(\frac{1}{N} \cdot X'\hat{\underline{u}}\right) .$$

From our initial assumptions, the first three terms in round brackets converge in probability to finite matrices, whilst the fourth converges in probability to a null vector.

Using equation (3.55) the rsth element of the "middle" term in the above expression can be written as

$$- \frac{1}{N} \sum_{i=1}^{N} x_{ir} \, \underline{x}_i' \, \Sigma_{XX}^{-1} \, \underline{x}_i \, x_{is}$$

$$+ \frac{1}{N} \sum_{i=1}^{N} x_{ir} \, \underline{v}_i' \, \Sigma_P^{-1} \, \underline{v}_i \, x_{is} \, ,$$

whence the first term converges to a finite limit whilst the second term converges in probability to a finite limit. Thus we have shown that

$$\underset{N \to \infty}{\text{plim}} \; \frac{1}{\sqrt{N}} \; . \; Z' M_X \, \bar{\Lambda}_3 \, (\hat{\underline{v}} - \hat{\underline{w}}) = \underline{0} \; .$$

Using the above results, we have shown that

$$\underset{N \to \infty}{\text{plim}} \; \sqrt{N} \; [J(\hat{\underline{\theta}}) - \hat{\underline{\theta}}] = \underline{0} \; ;$$

i.e. the J2SLS and 2SLS estimators are asymptotically equivalent.

3.4 Asymptotic Normality of J2SLS t-Ratios

From equation (2.10) the variance of the J2SLS estimator of θ can be written as

$$V[J(\hat{\underline{\theta}})] = \frac{1}{N(N-1)} \sum_{i=1}^{N} \left[J_i(\hat{\underline{\theta}}) - J(\hat{\underline{\theta}}) \right] \left[J_i(\hat{\underline{\theta}}) - J(\hat{\underline{\theta}}) \right]' . \qquad (3.65)$$

Using the definition of the jackknife it can be shown that

$$J_i(\underline{\hat{\theta}}) - J(\underline{\hat{\theta}}) = \left[N.\underline{\hat{\theta}} - (N-1)\underline{\hat{\theta}}_{i}\right] - \left[N.\underline{\hat{\theta}} - \frac{(N-1)}{N} \sum_{i=1}^{N} \underline{\hat{\theta}}_{i}\right] , \qquad (3.66)$$

whereupon, if we let $\bar{\underline{\hat{\theta}}}_{i} = \frac{1}{N} \sum_{i=1}^{N} \underline{\hat{\theta}}_{i}$,

equation (3.65) can be rewritten as ,

$$V[J(\underline{\hat{\theta}})] = \frac{1}{N(N-1)} (N-1)^2 \sum_{i=1}^{N} (\bar{\underline{\hat{\theta}}}_{i} - \underline{\hat{\theta}}_{i})(\bar{\underline{\hat{\theta}}}_{i} - \underline{\hat{\theta}}_{i})' . \qquad (3.67)$$

Using equations (3.25) and (3.66) we can write

$$\begin{aligned} \bar{\underline{\hat{\theta}}}_{i} = \underline{\hat{\theta}} - N^{-1} . P^{-1}[&Z'(\Lambda_1 + 2\Lambda_2 - \Lambda_3)\underline{\hat{u}} \\ &- Z'M_X(\Lambda_2 - \Lambda_3)\underline{\hat{u}} \\ &+ Z'(\Lambda_2 - \Lambda_3)(\underline{\hat{v}} - \underline{\hat{w}}) \\ &+ Z'M_X\Lambda_3(\underline{\hat{v}} - \underline{\hat{w}})] . \end{aligned} \qquad (3.68)$$

Let $\bar{\underline{g}} = \frac{1}{N} \sum_{i=1}^{N} \underline{g}_i$

represent the terms within the square brackets in equation (3.68), then equation (3.68) can be written as

$$\bar{\underline{\hat{\theta}}}_{i} = \underline{\hat{\theta}} - N^{-1} . P^{-1} \bar{\underline{g}} ,$$

and thus

$$\sum_{i=1}^{N} \left(\bar{\underline{\hat{\theta}}}_{i} - \underline{\hat{\theta}}_{i}\right)\left(\bar{\underline{\hat{\theta}}}_{i} - \underline{\hat{\theta}}_{i}\right)' = \sum_{i=1}^{N} \left(\underline{\hat{\theta}} - \underline{\hat{\theta}}_{i} + N^{-1}.P^{-1}\bar{\underline{g}}\right)\left(\underline{\hat{\theta}} - \underline{\hat{\theta}}_{i} + N^{-1}.P^{-1}\bar{\underline{g}}\right)' .$$

Expanding the right hand side of the above expression gives

$$\sum_{i=1}^{N}\left(\hat{\underline{\theta}}_i - \hat{\underline{\theta}}\right)\left(\hat{\underline{\theta}}_i - \hat{\underline{\theta}}\right)' - \frac{1}{N}\sum_{i=1}^{N}\left(\hat{\underline{\theta}}_i - \hat{\underline{\theta}}\right)\bar{\underline{g}}' P^{-1}$$

$$- \frac{1}{N}\sum_{i=1}^{N} P^{-1}\bar{\underline{g}}(\hat{\underline{\theta}}_i - \hat{\underline{\theta}}) + \frac{1}{N^2}\sum_{i=1}^{N} P^{-1}\bar{\underline{g}}\,\bar{\underline{g}}' P^{-1} \quad .$$

Since

$$\sum_{i=1}^{N}(\hat{\underline{\theta}}_i - \hat{\underline{\theta}}) = \sum_{i=1}^{N}\hat{\underline{\theta}}_i - N\hat{\underline{\theta}} = -P^{-1}\bar{\underline{g}} \quad ,$$

we can write

$$\sum_{i=1}^{N}(\hat{\underline{\theta}}_i - \hat{\underline{\theta}})\bar{\underline{g}}'P^{-1} = -\sum_{i=1}^{N} P^{-1}\bar{\underline{g}}\,\bar{\underline{g}}' P^{-1} \quad .$$

From equation (3.14) it follows that

$$\sum_{i=1}^{N}(\hat{\underline{\theta}} - \hat{\underline{\theta}}_i)(\hat{\underline{\theta}} - \hat{\underline{\theta}}_i)' = \sum_{i=1}^{N} P^{-1}\underline{g}_i\,\underline{g}_i' P^{-1} \quad ,$$

and using the definition of $\underline{g}_i$ as given by equations (3.15), (3.16), and (3.17) we can write

$$\sum_{i=1}^{N}\underline{g}_i\underline{g}_i = \sum_{i=1}^{N}\left[(\hat{v}_i - \hat{w}_i + \hat{u}_i)\underline{h}_i + \hat{u}_i\underline{j}_i\right]\left[(\hat{v}_i - \hat{w}_i + \hat{u}_i)\underline{h}_i + \hat{u}_i\underline{j}_i\right]'$$

$$= \sum_{i=1}^{N}\Big[(\hat{v}_i - \hat{w}_i + \hat{u}_i)^2\underline{h}_i\underline{h}_i' + (\hat{v}_i - \hat{w}_i + \hat{u}_i)\hat{u}_i\underline{h}_i\underline{j}_i'$$

$$+ \hat{u}_i(\hat{v}_i - \hat{w}_i + \hat{u}_i)\underline{j}_i\underline{h}_i' + \hat{u}_i^2\,\underline{j}_i\,\underline{j}_i'\Big] \quad .$$

Letting $\hat{e}_i = (\hat{v}_i - \hat{w}_i + \hat{u}_i)$ and expanding the above terms individually we obtain the following four expressions:

$$\sum_{i=1}^{N} (\hat{v}_i - \hat{w}_i + \hat{u}_i)^2 \underline{h}_i\underline{h}_i' = \sum_{i=1}^{N} \hat{e}_i^2 (\Lambda_3)^2_{ii} (\underline{z}_i - \underline{a}_i)(\underline{z}_i - \underline{a}_i)'$$

$$+ \sum_{i=1}^{N} \hat{e}_i^2 (\Lambda_2)^2_{ii} \underline{z}_i\underline{z}_i'$$

$$+ \sum_{i=1}^{N} \hat{e}_i^2 (\Lambda_3)_{ii} (\Lambda_2)_{ii} (\underline{z}_i - \underline{a}_i)\underline{z}_i'$$

$$+ \sum_{i=1}^{N} \hat{e}_i^2 (\Lambda_2)_{ii} (\Lambda_3)_{ii} \underline{z}_i(\underline{z}_i - \underline{a}_i)' \ ;$$

$$\sum_{i=1}^{N} (\hat{v}_i - \hat{w}_i + \hat{u}_i)\hat{u}_i\underline{h}_i\underline{j}_i' = \sum_{i=1}^{N} \hat{e}_i\hat{u}_i (\Lambda_3)_{ii} (\Lambda_2)_{ii} (\underline{z}_i - \underline{a}_i)(\underline{z}_i - \underline{a}_i)'$$

$$- \sum_{i=1}^{N} \hat{e}_i\hat{u}_i (\Lambda_3)_{ii} (\Lambda_1)_{ii} (\underline{z}_i - \underline{a}_i)\ \underline{z}_i'$$

$$+ \sum_{i=1}^{N} \hat{e}_i\hat{u}_i (\Lambda_2)^2_{ii} \underline{z}_i(\underline{z}_i - \underline{a}_i)'$$

$$- \sum_{i=1}^{N} \hat{e}_i\hat{u}_i (\Lambda_2)_{ii} (\Lambda_1)_{ii} \underline{z}_i\underline{z}_i' \ ;$$

$$\sum_{i=1}^{N} u_i (\hat{v}_i - \hat{w}_i + \hat{u}_i)\underline{j}_i\underline{h}_i' = \sum_{i=1}^{N} \hat{u}_i\hat{e}_i (\Lambda_3)_{ii} (\Lambda_2)_{ii} (\underline{z}_i - \underline{a}_i)(\underline{z}_i - \underline{a}_i)'$$

$$- \sum_{i=1}^{N} \hat{u}_i\hat{e}_i (\Lambda_3)_{ii} (\Lambda_1)_{ii} \underline{z}_i(\underline{z}_i - \underline{a}_i)'$$

$$+ \sum_{i=1}^{N} \hat{u}_i\hat{e}_i (\Lambda_2)^2_{ii} (\underline{z}_i - \underline{a}_i)\underline{z}_i'$$

$$- \sum_{i=1}^{N} \hat{u}_i\hat{e}_i (\Lambda_2)_{ii} (\Lambda_1)_{ii} \underline{z}_i\underline{z}_i' \ ;$$

$$\sum_{i=1}^{N} \hat{u}_i^2 \, \underline{j}_i \underline{j}_i' = \sum_{i=1}^{N} \hat{u}_i^2 (\Lambda_2)_{ii}^2 (\underline{z}_i - \underline{a}_i)(\underline{z}_i - \underline{a}_i)'$$

$$- \sum_{i=1}^{N} \hat{u}_i^2 (\Lambda_2)_{ii} (\Lambda_1)_{ii} (\underline{z}_i - \underline{a}_i) \underline{z}_i'$$

$$- \sum_{i=1}^{N} \hat{u}_i^2 (\Lambda_2)_{ii} (\Lambda_1)_{ii} \, \underline{z}_i (\underline{z}_i - \underline{a}_i)'$$

$$+ \sum_{i=1}^{N} \hat{u}_i^2 (\Lambda_1)_{ii}^2 \, \underline{z}_i \underline{z}_i' \; ;$$

where, as before, the ii subscript on a matrix indicates the iith component of that matrix.

Gathering terms, we can write

$$\sum_{i=1}^{N} \underline{g}_i \underline{g}_i' = \sum_{i=1}^{N} \left[\hat{e}_i^2 (\Lambda_3)_{ii}^2 + 2\hat{e}_i \hat{u}_i (\Lambda_3)_{ii} (\Lambda_2)_{ii} + \hat{u}_i^2 \, (\Lambda_2)_{ii}^2 \right] (\underline{z}_i - \underline{a}_i)(\underline{z}_i - \underline{a}_i)'$$

$$+ \sum_{i=1}^{N} \left[\hat{e}_i^2 (\Lambda_2)_{ii}^2 - 2\hat{e}_i \hat{u}_i (\Lambda_2)_{ii} (\Lambda_1)_{ii} + \hat{u}_i^2 (\Lambda_1)_{ii}^2 \right] \underline{z}_i \underline{z}_i'$$

$$+ \sum_{i=1}^{N} \Big[\hat{e}_i^2 (\Lambda_3)_{ii} (\Lambda_2)_{ii} - \hat{e}_i \hat{u}_i (\Lambda_3)_{ii} (\Lambda_1)_{ii} + \hat{u}_i \hat{e}_i (\Lambda_2)_{ii}^2$$

$$- \hat{u}_i^2 (\Lambda_2)_{ii} (\Lambda_1)_{ii} \Big] \; (\underline{z}_i - \underline{a}_i) \underline{z}_i'$$

$$+ \sum_{i=1}^{N} \Big[\hat{e}_i^2 (\Lambda_2)_{ii} (\Lambda_3)_{ii} + \hat{e}_i \hat{u}_i (\Lambda_2)_{ii}^2 - \hat{e}_i \hat{u}_i (\Lambda_3)_{ii} (\Lambda_1)_{ii}$$

$$- \hat{u}_i^2 (\Lambda_2)_{ii} (\Lambda_1)_{ii} \Big] \; \underline{z}_i (\underline{z}_i - \underline{a}_i)' \; . \tag{3.69}$$

We define the following matrices

$$\left.\begin{aligned} R_1 &= (\underline{y} - Z\hat{\underline{\theta}})(\underline{y} - Z\hat{\underline{\theta}})' = \hat{\underline{u}} \, \hat{\underline{u}}' \; , \\ R_2 &= \left[(I - M_X)(\underline{y} - Z\hat{\underline{\theta}}) \right] \left[(I - M_X)(\underline{y} - Z\hat{\underline{\theta}}) \right]' = (I - M_X) \hat{\underline{u}} \, \hat{\underline{u}}' (I - M_X) \; , \\ \text{and} \quad R_3 &= (y - Z\hat{\underline{\theta}}) \left[(I - M_X)(y - Z\hat{\underline{\theta}}) \right]' = \hat{\underline{u}} \, \hat{\underline{u}}' (I - M_X) \; , \end{aligned}\right\} \tag{3.70}$$

which allows us to rewrite equation (3.69) as

$$\sum_{i=1}^{N} \underline{g}_i \underline{g}_i' = Z'(I - M_X)[\Lambda_2(\text{diag } R_1)\Lambda_2 + \Lambda_3 (\text{diag } R_2)\Lambda_3$$

$$+ 2\Lambda_2(\text{diag } R_3)\Lambda_3] (I - M_X)Z$$

$$+ Z'[\Lambda_2(\text{diag } R_2)\Lambda_2 - 2\Lambda_2(\text{diag } R_3)\Lambda_1 + \Lambda_1(\text{diag } R_1)\Lambda_1]Z$$

$$+ Z'(I-M_X)[\Lambda_3(\text{diag } R_2)\Lambda_2 - \Lambda_3(\text{diag } R_3)\Lambda_1$$

$$+ \Lambda_2(\text{diag } R_3)\Lambda_2 - \Lambda_2(\text{diag } R_1)\Lambda_1]Z$$

$$+ Z'[\Lambda_2(\text{diag } R_2)\Lambda_3 + \Lambda_2(\text{diag } R_3)\Lambda_2$$

$$- \Lambda_3(\text{diag } R_3)\Lambda_1 - \Lambda_2(\text{diag } R_1)\Lambda_1](I - M_X)Z \ , \qquad (3.71)$$

where (diag) denotes that the relevant matrix has all off-diagonal components equal to zero.

If we also define $S_1 = Z'\Lambda_1 + Z'(I - M_X)\Lambda_2$

and $S_2 = - Z'\Lambda_2 + Z'(I - M_X)\Lambda_3$,

then equation (3.71) can be rewritten as

$$\sum_{i=1}^{N} \underline{g}_i \underline{g}_i' = S_1(\text{diag } R_1)S_1' + S_2(\text{diag } R_2)S_2'$$

$$- S_1(\text{diag } R_3)S_2' - S_2(\text{diag } R_3)S_1' \ .$$

We also require $\bar{\underline{g}}$ which can be written, using equation (3.68), as

$$\bar{\underline{g}} = Z'(\Lambda_1 + 2\Lambda_2 - \Lambda_3)\hat{\underline{u}} - Z'M_X(\Lambda_2 - \Lambda_3)\hat{\underline{u}}$$

$$+ Z'(\Lambda_2 - \Lambda_3)(\hat{\underline{v}} - \hat{\underline{w}}) + Z'M_X\Lambda_3(\hat{\underline{v}} - \hat{\underline{w}}) \ ;$$

i.e. $\bar{\underline{g}} = S_1\hat{\underline{u}} - S_2(\hat{\underline{v}} - \hat{\underline{w}} + \hat{\underline{u}})$,

and hence

$$\bar{\underline{g}}\,\bar{\underline{g}}' = S_1R_1S_1' + S_2R_2S_2' - S_1R_3S_2' - S_2R_3S_1' \ .$$

Upon substituting the above results into equation (3.68), and then into equation (3.67), we obtain

$$V[J(\underline{\hat{\theta}})] = \frac{(N-1)}{N} \cdot P^{-1} \left[S_1 \left(\text{diag } R_1 + \frac{1}{N} \cdot R_1 \right) S_1' + S_2 \left(\text{diag } R_2 + \frac{1}{N} \cdot R_2 \right) S_2' - S_1 \left(\text{diag } R_3 + \frac{1}{N} \cdot R_3 \right) S_2' - S_2 \left(\text{diag } R_3 + \frac{1}{N} \cdot R_3 \right) S_1' \right] P^{-1} . \quad (3.72)$$

It is shown in Appendix B that the expression in square brackets in equation (3.72) converges to

$$\sigma^2 \Sigma_P$$

in probability as $N \to \infty$.

It follows from equation (3.72) that

$$\underset{N \to \infty}{\text{plim}} \; V[J(\underline{\hat{\theta}})] = \sigma^2 \Sigma_P^{-1} ,$$

since $(N-1)/N \to 1$ as $N \to \infty$.

Since $J(\underline{\hat{\theta}})$ has been shown to be asymptotically equivalent to $\underline{\hat{\theta}}$ it follows that, asymptotically,

$$\frac{(J(\hat{\theta}_j) - \theta_j)}{\sqrt{V[J(\hat{\theta}_j)]}} \sim N(0,1) . \qquad (j = 1,2,\ldots, K_1 + g) .$$

CHAPTER 4

COMPUTATIONAL ASPECTS

4.1 Computer Algorithms and their Certification

From equation (2.6) the 2SLS estimator of $\underline{\theta}$ can be written as

$$\underline{\hat{\theta}} = [Z'X(X'X)^{-1}X'Z]^{-1}Z'X(X'X)^{-1}X'\underline{y} \qquad (4.1)$$

In all but the simplest cases, equation (4.1) must be evaluated using a computer. Matrix manipulations can be performed using either standard algorithms designed for a specific computer and usually incorporated in the software library, or machine independent algorithms published in computer programming journals. Alternatively one could write one's own algorithms although this might be inadvisable for the more complicated operations such as matrix inversion.

In all computational work in this thesis, matrix manipulations were performed with algorithms written by the author, except for the matrix inversion algorithm. To perform inversions an algorithm written by Devine [11], which inverts a symmetric positive definite matrix by the Choleski decomposition method was selected. All programs were written in Algol 60.

Certification of Devine's algorithm was carried out by the author. This was performed by multiplying the original matrix by its calculated inverse and then obtaining the maximum absolute deviation of elements from the unit matrix. These maximum absolute deviations are given in Table 4.1 for the eight different data matrices which are inverted during the Monte Carlo study in Chapters 6 and 7. The column headed K represents the dimensions of the matrix (i.e. the number of exogenous variables in the model), whilst the column headed λ

denotes the theoretical pairwise correlation between the K variables. The sample correlation matrices are given in Table 6.2.

Table 4.1: Maximum Absolute Deviations (M.A.D.) of $(X'X)^{-1}(X'X)$ from the Unit Matrix

K	λ	M.A.D.
5	0.00	5.46×10^{-12}
5	0.45	8.19×10^{-12}
8	0.00	2.32×10^{-11}
8	0.45	2.46×10^{-11}
8	0.90	1.36×10^{-11}
11	0.00	2.91×10^{-11}
11	0.45	2.18×10^{-11}
11	0.90	2.18×10^{-11}

The accuracy of the matrix inversion, as reflected by the maximum absolute deviations given in Table 4.1, is certainly satisfactory for our purposes.

For $K = 5$ and $\lambda = 0.90$, whilst the moment matrix of predetermined variables was inverted satisfactorily, a further inversion incorporating stochastic matrices which is required at each replication in the Monte Carlo experiment exhibited substantial "inversion errors" and consequently "inconsistent" results were obtained. This problem is discussed in Chapter 6.

A machine independent pseudo-random number generator devised by Pike and Hill [43] was used for generating uniformly distributed pseudo-random numbers for the experiments in Chapters 5, 6 and 7. Favourable evidence of randomness for this algorithm is given by serial and poker tests conducted by Pike and Hill, and by frequency tests in the certification by Sullins [65].

The Box and Muller [6] transformation for generating normally distributed pseudo-random variates is given by

$$\left.\begin{aligned} x_1 &= (-2 \log_e r_1)^{\frac{1}{2}} \sin 2\pi r_2 \\ x_2 &= (-2 \log_e r_1)^{\frac{1}{2}} \cos 2\pi r_2 \end{aligned}\right) \quad (4.2)$$

where x_1 and x_2 are two uncorrelated pseudo-random standardized normal variates, and r_1 and r_2 are uniformly distributed pseudo-random variates defined on the [0,1] interval. This transformation produces exact results conditional upon the accuracy of evaluation of the sin and cos functions and the correct distribution of r_1 and r_2. When used in conjunction with a multiplicative congruential pseudo-random number generator however, Neave [38] has shown how the transformation may break down. Amendments to equation (4.2), as suggested by Chay, Fardo and Mazumbar [9], were used in this research, therefore, to avoid Neave's objections. With these amendments the transformation becomes

$$x_1 = (-2 \log_e r_2)^{\frac{1}{2}} \sin 2\pi r_1 ,$$

where it should be noted that only the sin transformation is used and the uniformly distributed variates have been interchanged.

The Monte Carlo study reported in Chapters 6 and 7 necessitated the generation of 4,400 pseudo-random standardized normal variates (this figure excludes the additional normally distributed variates required to calculate the power functions in Chapter 7). The Kolmogorov-Smirnov test was conducted to test for any significant divergence between the theoretical (standardized normal) and empirical distributions of the pseudo-random variates. The maximum absolute value of D (the difference between the two distributions) was 0.01306. At the 5% level of significance the hypothesis of equality cannot be rejected.

The pseudo-random normal variates were subsequently transformed into pseudo-random bivariate normal variates by using the transformation

$$Z_1 = \omega_{11}^{\frac{1}{2}} x_1$$

$$Z_2 = \omega_{22}^{\frac{1}{2}} (\delta x_1 + \sqrt{1 - \delta}\, x_2) ,$$

where Z_1 and Z_2 are correlated normal variates with coefficient of correlation equal to δ. ω_{11} and ω_{22} are the specified population variances of Z_1 and Z_2 respectively, and the covariance of Z_1 and Z_2 is given by $\delta\omega_{12}$.

4.2 Computing J2SLS Parameter Estimates

In order to apply the jackknife to the 2SLS estimator we must have some method by which the ith observation can be extracted from equation (4.1). Clearly one could calculate equation (4.1) N times using a 2SLS program and omitting a different observation on each occasion, but this would be a tedious and computationally expensive procedure especially for "large" N and/or K as it would require inverting both matrices in square brackets in equation (4.1) (minus one observation) at each iteration. In addition, rounding errors from the inversion algorithm may lead to a build-up of inaccuracies.

In Chapter 3 we derived equation (3.13) for calculating the 2SLS estimator with the ith observation removed which obviates the need to perform matrix inversions additional to those required for 2SLS with all N observations included. This formula was checked by calculating the J2SLS estimator both ways with a test program and noting that the parameter estimates were identical to at least the sixth decimal place.

4.3 Computing Exact Results

Calculation of the exact moments of the 2SLS estimator, and exact bias in the case of J2SLS, requires evaluation of the confluent hypergeometric function

$$_1F_1\ (\alpha;\gamma\ ;\ x)\ . \tag{4.3}$$

Although tables are available (e.g. see Slater [64]), relatively few values of α, γ and x have been tabulated. In general, therefore, the function must be calculated by direct summation of an infinite series or via an asymptotic approximation.

An algorithm for calculating the confluent hypergeometric function with complex parameters via the method of direct summation has been written by Relph [49]. Thacher [69] in his certification of this algorithm mentioned its inefficiency for real arguments.

A problem frequently encountered in this thesis was that of relatively small α and γ, but relatively large x, whence evaluation of equation (4.3) is characterized by slow convergence. When this problem arose it was resolved by using an asymptotic approximation to the confluent hypergeometric function, which for integer α and $\gamma = \alpha + 1$ contains a finite number of terms. A check on the error involved in using the approximation can be made if α is an integer and, if necessary, a correction made.

For a model containing just two endogenous variables, Richardson and Wu [55] have derived the bias of the 2SLS estimator ($\hat{\beta}$) of β in equation (2.5) as

$$E(\hat{\beta}-\beta) = -\frac{\omega_{22}\beta-\omega_{12}}{\omega_{22}}\ e^{-\mu^2/2}\ {}_1F_1\left(\frac{K_2}{2}-1\ ;\ \frac{K_2}{2}\ ;\ \frac{\mu^2}{2}\right), \tag{4.4}$$

where $\mu^2 = \omega_{22}^{-1} \underline{\pi}_{22}' X_2' [I_N - X_1(X_1' X_1)^{-1} X_1'] X_2 \underline{\pi}_{22}$ is the concentration parameter so named because for every $\varepsilon > 0$

$$\lim_{\mu^2 \to \infty} \Pr(|\hat{\beta} - \beta| > \varepsilon) = 0 .$$

All other notation was explained in Chapter 2.

Clearly $\alpha = (K_2/2 - 1)$ is an integer if K_2 is even.

From Appendix C (equation (C.1)) the asymptotic (in μ^2) expansion of the confluent hypergeometric function (for $\gamma = \alpha + 1$) can be written as

$${}_1F_1\ (\alpha; \alpha + 1; x) \sim \frac{\alpha}{x} e^x \sum_{r=0}^{\infty} (1 - \alpha)_r \left(\frac{1}{x}\right)^r , \tag{4.5}$$

and thus the asymptotic approximation to the bias (4.4) is

$$E(\hat{\beta} - \beta) \sim - \frac{\omega_{22}\beta - \omega_{12}}{\omega_{22}} \frac{(K_2 - 2)}{\mu^2} \sum_{r=0}^{\infty} \left(2 - \frac{K_2}{2}\right)_r \left(\frac{2}{\mu^2}\right)^r . \tag{4.6}$$

The error incurred by applying this approximation for finite μ^2 and integer α is given (from Appendix C (equation C.8)) by

$$\frac{\omega_{22}\beta - \omega_{12}}{\omega_{22}} e^{-\mu^2/2} \, \Gamma\left(\frac{K_2}{2}\right) \left(- \frac{2}{\mu^2}\right)^k , \tag{4.7}$$

where $k = (K_2 - 2)/2$.

It is interesting to note from equation (4.6) that for "large" μ^2 and $K_2 = 2$ the 2SLS estimator is unbiased.

Thus provided the asymptotic approximation of the confluent hypergeometric function terminates after a finite number of terms, equations (4.5) and (4.7) will ensure exact evaluation of this function. The gain in computational efficiency will be particularly marked when the summation of the infinite series required for direct evaluation of the confluent hypergeometric function is slow to converge.

For α non-integer, equation (4.5) is an infinite series, although it can be truncated after (say) n terms. If this is done the error involved in truncating the infinite series after the nth term will not exceed the (n + 1)th term, and will be of the same sign as the (n + 1)th term (Luke [25; p.127]).

In this thesis, when α is not an integer the confluent hypergeometric function had to be truncated in such a way as to ensure that all values of bias and mean square error were correct to at least the number of decimal places given in the text. For integer α, all results are "exact".

CHAPTER 5

THE EXACT BIASES OF THE TWO-STAGE LEAST SQUARES AND JACKKNIFE TWO-STAGE LEAST SQUARES ESTIMATORS

5.1 Résumé of "Exact" Studies

In his pioneering work on the exact finite sample distribution function of the 2SLS estimator, Basmann [4] demonstrated analytically that for a two equation simultaneous equations model, under certain conditions, the moments may not exist (i.e. they may not be finite).

Prior to Basmann's [4] paper, Monte Carlo studies of the relative properties of simultaneous equations estimators had frequently used as their objective function the mean square error in order to compare the relative properties of the estimators. Basmann remarked that an objective function which involved moments of the estimators would have little significance if the moments of the estimators did not exist. In addition, non-finite moments could give rise to "outliers" when this form of objective function is used in Monte Carlo studies, and thus uncritical rejection of these outliers is not a valid procedure.

On the basis of his early work, Basmann [4] conjectured that the moments of the 2SLS estimator exist up to the order of over-identification of the equation being estimated. Basmann's proof was only valid for a two-equation model with $K_1 = K_2 = 2$ and $K_1 = 1$, $K_2 = 3$, although in a later paper (Basmann [5]) he extended it to a three equation model with $g = 2$, $K_1 = 1$ and $K_2 = 3$.

Kabe [21, 22] greatly simplified Basmann's derivations, and this was followed by analytical proofs of Basmann's conjecture for $g = 1$, $K_2 \geq 2$, by Richardson [52] and Sawa [58].

For the general case (i.e. g and K_2 both arbitrary) Mariano [28] has provided a proof of Basmann's conjecture for the even-ordered moments of the 2SLS estimator, whilst Hatanaka [17] has shown that the same conjecture provides a sufficient condition for the existence of the odd-ordered moments.

Sawa [58] and Richardson and Wu [55] derived, independently, the distribution function of the OLS estimator, and then showed how the distribution function of the 2SLS estimator could be derived as a corollary to the derivation of the OLS estimator. For g = 1 the exact moments of the coefficient (β) of the right-hand side endogenous variable in equation (2.5) have been calculated by Sawa [58], Takeuchi [67], and Richardson and Wu [55] for both estimators. From Richardson and Wu [55], the first order moment of the 2SLS estimator can be written as

$$E(\hat{\beta} - \beta) = - \frac{\omega_{22}\beta - \omega_{12}}{\omega_{22}} e^{-\mu^2/2} {}_1F_1\left(\frac{K_2}{2} - 1 ; \frac{K_2}{2} ; \frac{\mu^2}{2}\right) \qquad (5.1)$$

Second and higher order moments take a more complicated form and the interested reader is referred to the literature previously cited.

The fundamental parameter in all "exact" studies is the concentration parameter μ^2, and not the sample size which does not enter equation (5.1) explicitly, although it is implicit in μ^2.

As μ^2 increases indefinitely, the 2SLS estimator of β converges to its true parameter value (i.e. it is a consistent estimator). A sufficient, but not a necessary, condition for μ^2 to increase indefinitely is for the sample size to increase indefinitely.

In general, the concentration parameter for the jth equation is defined by

$$\mu^2_j = \text{trace } (M_j \textstyle\sum_*^{-1}) \quad ,$$

where $M_j = \Pi_{22}' X_{2j}' [I - X_{1j}(X_{1j}'X_{1j})^{-1} X_{1j}'] X_{2j}\Pi_{22}$

and $\sum_*^{-1}$ is the covariance matrix of non-normalized endogenous variables included in the structural equation.

Essentially, therefore, the moments of the 2SLS estimator are derived in terms of "nuisance" parameters. Sawa [58] assigned "reasonable" values to these nuisance parameters in order to ascertain the relative importance of N, $\rho = \frac{\omega_{12}}{\omega_{22}}$ and K_2. He observed that the bias of 2SLS is an increasing function of $|\rho|$ and that frequently it "is not negligible". In addition, he found that the distribution of the 2SLS estimator was often considerably asymmetric.

Mariano and Ramage [29] considered the effects on 2SLS of excluding relevant exogenous variables and including extraneous exogenous variables in the equation to be estimated. Mathematical complexity precludes useful analysis of the former specification error, but under the latter type of misspecification both the concentration parameter and the degrees of freedom are smaller than for a correctly specified model. The decrease in the concentration parameter increases the bias and mean square error of both estimators, whilst the effect of the decrease in the degrees of freedom is indefinite and depends on other unknown parameters in the model.

5.2 The Concentration Parameter and a Change in Sample Size

Let μ_N^2 and μ_{N-1}^2 denote the concentration parameter based upon N and (N-1) observations respectively, then

$$\mu_N^2 = \omega_{22}^{-1} \underline{\pi}_{22}' X_2' \left[I - X_1(X_1' X_1)^{-1} X_1' \right] X_2 \underline{\pi}_{22} \tag{5.2}$$

and

$$\mu^2_{N-1} = \omega_{22}^{-1} \underline{\pi}_{22}' X_2^{*'} \left[I - X_1^*(X_1^{*'} X_1^*)^{-1} X_1^{*'} \right] X_2^* \underline{\pi}_{22} \quad , \tag{5.3}$$

where the asterisk superscript refers to the relevant data matrix with one observation removed. Without loss of generality we assume that the Nth observation has been removed, i.e.

$$X_1 = \begin{bmatrix} X_1^* \\ \underline{x}_1' \end{bmatrix} \quad \text{and} \quad X_2 = \begin{bmatrix} X_2^* \\ \underline{x}_2' \end{bmatrix}$$

where $\underline{x}_1$ and $\underline{x}_2$ are K_1 and K_2 dimensional column vectors representing the omitted observation from X_1 and X_2 respectively.

Noting that

$$(X_1^{*'} X_1^*) = (X_1' X_1 - \underline{x}_1 \underline{x}_1') ,$$

$$(X_2^{*'} X_2^*) = (X_2' X_2 - \underline{x}_2 \underline{x}_2') ,$$

and $$(X_1^{*'} X_2^*) = (X_1' X_2 - \underline{x}_1 \underline{x}_2') ,$$

equation (5.3) can be written as

$$\mu^2_{N-1} = \omega_{22}^{-1} \underline{\pi}_{22}' \Big\{ (X_2' X_2 - \underline{x}_2 \underline{x}_2') - (X_2' X_1 - \underline{x}_2 \underline{x}_1')(X_1' X_1 - \underline{x}_1 \underline{x}_1')^{-1} (X_1' X_2 - \underline{x}_1 \underline{x}_2') \Big\} \underline{\pi}_{22}$$

It can be shown (see Appendix A) that

$$(X_1' X_1 - \underline{x}_1 \underline{x}_1')^{-1} = (X_1' X_1)^{-1} + \frac{(X_1' X_1)^{-1} \underline{x}_1 \underline{x}_1' (X_1' X_1)^{-1}}{1 - \underline{x}_1' (X_1' X_1)^{-1} \underline{x}_1} .$$

Using this result, and after considerable algebraic manipulation, equation (5.3) can be written as

$$\mu^2_{N-1} = \mu^2_N - \frac{1}{(1-c)} \omega_{22}^{-1} \underline{\pi}_{22}' (\underline{x}_2 - \underline{d})(\underline{x}_2 - \underline{d})' \underline{\pi}_{22} , \qquad (5.4)$$

where $c = \underline{x}_1' (X_1' X_1)^{-1} \underline{x}_1$, $\qquad 0 < c < 1$

and $d = (X_2' X_1)(X_1' X_1)^{-1} \underline{x}_1$.

Since $\underline{\pi}_{22}'(\underline{x}_2 - \underline{d})(\underline{x}_2 - \underline{d})'\underline{\pi}_{22}$ is a positive semi-definite quadratic form, and as ω_{22} and $(1 - c)$ are both greater than zero, it follows that

$$\mu^2_N \geq \mu^2_{N-1} \quad ,$$

i.e. the concentration parameter is a monotonically non-decreasing function of sample size.

5.3 The Exact Bias of the Jackknife Two-Stage Least Squares Estimator

5.3.1 Introduction

Since only μ^2 is dependent upon changes in N, the bias of the 2SLS estimator of β with the ith observation omitted ($\hat{\beta}_i$) can be written, using equation (5.1), as

$$E(\hat{\beta}_i - \beta) = -\frac{\omega_{22}\beta - \omega_{12}}{\omega_{22}} \exp\left(-\frac{\mu^2_{N-1}}{2}\right) {}_1F_1\left(\frac{K_2}{2} - 1 ; \frac{K_2}{2} ; \frac{\mu^2_{N-1}}{2}\right) . \quad (5.5)$$

Thus, when the exact bias of the 2SLS estimator can be calculated, it is relatively easy to calculate the exact bias of the J2SLS estimator.

Differentiating the absolute bias with respect to $\mu^2_N/2$, and utilizing the contiguity relations of the confluent hypergeometric function (e.g. see Slater [64; p.19] gives

$$\frac{d|E(\hat{\beta} - \beta)|}{d\, \mu^2_N/2} = -\frac{\omega_{22}\beta - \omega_{12}}{\omega_{22}} \cdot \frac{2}{K_2} \cdot e^{-\mu^2_N/2}\, {}_1F_1\left(\frac{K_2}{2} - 1; \frac{K_2}{2} + 1; \frac{\mu^2_N}{2}\right) . \quad (5.6)$$

From equation (5.6) is is apparent that the absolute value of the bias is a monotonically decreasing function of the concentration parameter μ^2_N, provided $\beta > \omega_{12}/\omega_{22}$. If $\beta = \omega_{12}/\omega_{22}$ no bias exists, whilst if $\beta < \omega_{12}/\omega_{22}$ it follows that the actual bias is a monotonically decreasing function of μ^2. Similarly, the mean square error of the

2SLS estimator can be shown to be a monotonically decreasing function of the concentration parameter (see Owen [40]).

Earlier in this Chapter is was shown that the concentration parameter is a monotonically non-decreasing function of sample size. Thus, combining these two results, it has been shown that the bias (and the mean square error) of the 2SLS estimator are monotonically non-increasing functions of the sample size; conditional, of course, on the exogenous variables.

We have already seen that the bias of the J2SLS estimator can be written as

$$E(\underline{\hat{\theta}} - \underline{\theta}) + (N-1)\left[E(\underline{\hat{\theta}} - \underline{\theta}) - \frac{1}{N}E\left\{\sum_{i=1}^{N}(\underline{\hat{\theta}}_i - \underline{\theta})\right\}\right]. \qquad (5.7)$$

It follows from the above result that the term in square brackets in equation (5.7) will be either zero or opposite in sign to $E(\underline{\hat{\theta}} - \underline{\theta})$. Consequently, application of the jackknife will have one of three possible effects on the bias of the 2SLS estimator:

1. The absolute bias decreases but its sign remains unchanged;
2. The absolute bias decreases and its sign changes;
3. The absolute bias increases and its sign changes.

If the bias decreases slowly or approximately linearly with sample size, then it seems reasonable to expect possibilities 1. or 2. to occur. When the bias is decreasing rapidly with sample size however, there could be a tendency for the jackknife to "over-correct" for bias and possibility 3. could occur.

Since the above eventualities are somewhat vague, we turn from heuristic analysis to consider an analytical investigation of the conditions under which jackknifing is unlikely to decrease the bias of the 2SLS estimator. First we consider the exact bias of the 2SLS estimator of β as given by equation (5.1) for the special case of $K_2 = 2$, then we consider a more general approach using Kadane's [23] approximation to the bias of the 2SLS estimator.

5.3.2 Effect of Jackknifing on the Exact Bias of 2SLS when $K_2 = 2$

From equation (5.1), if $K_2 = 2$ the exact bias of the 2SLS estimator of β degenerates to

$$E(\hat{\beta} - \beta) = - \frac{(\omega_{22}\beta - \omega_{12})}{\omega_{22}} e^{-\mu^2/2} , \qquad (5.8)$$

since $_1F_1(0,1,\mu^2/2) = 1$.

Expanding the exponential term in equation (5.8) and setting $\rho = \frac{\omega_{12}}{\omega_{22}}$ gives

$$E(\hat{\beta} - \beta) = -(\beta-\rho) \left[1 + \left(-\frac{\mu^2}{2}\right) + \left(-\frac{\mu^2}{2}\right)^2 . \frac{1}{2!} + \ldots . + \left(-\frac{\mu^2}{2}\right)^r . \frac{1}{r!} + \ldots \right] \qquad (5.9)$$

Since μ^2 is of order N, when $K_2 = 2$ the bias of the 2SLS estimator is clearly a function of terms (with alternating signs) of increasing powers of order N. Whilst alternating signs will not weaken the jackknife's bias reducing properties, equation (5.9) clearly contravenes Quenouille's basic assumption regarding the application of the jackknife, viz: that the bias can be expressed as an expansion in terms of increasing powers of order $(\frac{1}{N})$. This suggests that application of the jackknife technique is unlikely to be successful if $K_2 = 2$.

When $K_2 > 2$

$$_1F_1 \left(\frac{K_2}{2} - 1; \frac{K_2}{2} ; \frac{\mu^2}{2} \right)$$

takes the form of an infinite series and the bias cannot be expanded into an expression such as equation (5.9). We can however fall back on equation (4.6) which gives a "large" μ^2 expansion of the bias, equation (5.1), in terms of increasing powers of order $\frac{1}{N}$, provided $K_2 > 2$ (although if K_2 is an integer this expansion will terminate after a finite number of terms). This suggests that for "large" μ^2 and $K_2 > 2$, application of the jackknife technique could reduce the bias of the 2SLS estimator.

Both of the above observations will be investigated by means of a simulation experiment in Section 5.5.

5.4 Jackknifing the Approximate Bias of the 2SLS Estimator

Kadane [23] has derived the leading terms of the first two moments of a distribution approximating the exact distribution of the 2SLS estimator, although it should be emphasized that the moments of approximate distributions are not necessarily identical to approximations to the moments of the exact distribution.

Nagar's [37] work in this field carries a similar interpretation.

Kadane's approximations are based on a "small" σ asymptotic expansion of the moments of the k-class estimators (N.B. in our notation $\sigma = \omega_{11} - \omega_{12}\rho + \omega_{22}(\beta - \rho)^2$ and is not to be confused with the σ used elsewhere in this thesis. The definition of σ given here is restricted solely to this Section). For N fixed, $\mu^2 \to \infty$ if $\sigma \to 0$ and it can be shown (see Sawa [59; Appendix C]) that Kadane's (and Nagar's) expansion coincides with "large" μ^2 expansions of the exact moments, provided the latter exist.

Kadane [23] has approximated the bias of the 2SLS estimator by

$$E(\hat{\underline{\theta}} - \underline{\theta}) = \sigma^2(L-1)Q\underline{q} + O(\sigma^3), \tag{5.10}$$

where $L = K_2 - g$, i.e. the degree of overidentification of the equation being estimated ,

$$W = [X\Pi : X_1],$$

$$Q = (W'W)^{-1}$$

$$\underline{q} = \frac{1}{N} E[Y : X_1]'\underline{u},$$

and $\underline{\theta} = \begin{bmatrix} \beta \\ .. \\ \underline{\gamma}_1 \end{bmatrix}$

Let $\hat{\underline{\theta}}_i$ denote the 2SLS estimator of $\underline{\theta}$ with the ith observation removed then,

$$E(\hat{\underline{\theta}}_i - \underline{\theta}) = \sigma^2(L-1)Q_i\underline{q} + O(\sigma^3) , \tag{5.11}$$

where $Q_i = (W'W - \underline{w}_i\underline{w}_i')^{-1}$ and $\underline{w}_i$ is a $K_1 + g$ dimensional column vector representing the omitted observation from W.

From Appendix A, it can be shown that

$$Q_i = (W'W)^{-1} + \frac{(W'W)^{-1}\underline{w}_i\underline{w}_i'(W'W)^{-1}}{1 - \underline{w}_i'(W'W)^{-1}\underline{w}_i} = Q + \frac{Q\underline{w}_i\underline{w}_i'Q}{1 - \underline{w}_i'Q\underline{w}_i} ,$$

and hence

$$E(\hat{\underline{\theta}}_i - \underline{\theta}) = \sigma^2(L-1)Q\underline{q} + \sigma^2(L-1)\frac{Q\underline{w}_i\,\underline{w}_i'Q}{1 - \underline{w}_i'Q\underline{w}_i}\underline{q}$$

$$= E(\hat{\underline{\theta}} - \underline{\theta}) + \sigma^2(L-1)\frac{Q\underline{w}_i\,\underline{w}_i'Q}{1 - \underline{w}_i'Q\underline{w}_i}\underline{q} , \tag{5.12}$$

where terms of higher order in σ have been neglected.

From the definition of the jackknife, and using equation (5.12), we obtain

$$E[J(\hat{\underline{\theta}}) - \underline{\theta}] = NE(\hat{\underline{\theta}} - \underline{\theta}) - \frac{(N-1)}{N}\sum_{i=1}^{N} E(\hat{\underline{\theta}}_i - \underline{\theta})$$

$$= NE(\underline{\hat{\theta}} - \underline{\theta}) - (N-1)E(\underline{\hat{\theta}} - \underline{\theta}) - \sigma^2 \frac{(N-1)}{N} \sum_{i=1}^{N} (L-1) \frac{Q\underline{w}_i \underline{w}_i' Q}{1-\underline{w}_i' Q\underline{w}_i} \underline{q}$$

$$= E(\underline{\hat{\theta}} - \underline{\theta}) - \sigma^2 (L-1) \frac{(N-1)}{N} \sum_{i=1}^{N} \frac{Q\underline{w}_i \underline{w}_i' Q}{1-\underline{w}_i' Q\underline{w}_i} \underline{q} \,. \tag{5.13}$$

For jackknifing not to increase the absolute value of the bias of the 2SLS estimator over all parameters being estimated, we require

$$E[J(\underline{\hat{\theta}}) - \underline{\theta}]\, E[J(\underline{\hat{\theta}}) - \underline{\theta}]' - E[\underline{\hat{\theta}} - \underline{\theta}]\, E[\underline{\hat{\theta}} - \underline{\theta}]' \tag{5.14}$$

to have all main diagonal components ≤ 0.

Consider the last term in equation (5.13) which can be rewritten as

$$\sigma^2 (L-1) \frac{(N-1)}{N} \sum_{i=1}^{N} \frac{Q\underline{w}_i \underline{w}_i' Q}{1-\underline{w}_i' Q\underline{w}_i} \underline{q} = \sigma^2 (L-1) \frac{(N-1)}{N} Q \sum_{i=1}^{N} \left[\frac{\underline{w}_i \underline{w}_i'}{1-\underline{w}_i' Q\underline{w}_i} \right] Q\underline{q} \,.$$

Let Λ be an $N \times N$ diagonal matrix with iith component equal to $\underline{w}_i' Q \underline{w}_i$, then

$$[I - \Lambda]^{-1}$$

is an $N \times N$ diagonal matrix with iith component equal to

$$\frac{1}{1 - \underline{w}_i' Q\underline{w}_i} \,,$$

and hence

$$\sum_{i=1}^{N} \frac{\underline{w}_i \underline{w}_i'}{1-\underline{w}_i' Q\underline{w}_i} = W'[I - \Lambda]^{-1} W \,. \tag{5.15}$$

Thus equation (5.13) can be rewritten as

$$E[J(\underline{\hat{\theta}}) - \underline{\theta}] = E(\underline{\hat{\theta}} - \underline{\theta}) - \sigma^2 (L-1) \frac{(N-1)}{N} QW'[I - \Lambda]^{-1} WQ\underline{q} \,, \tag{5.16}$$

and upon substituting into equation (5.14) we obtain

$$E[J(\underline{\hat{\theta}}) - \underline{\theta}]E[J(\underline{\hat{\theta}}) - \underline{\theta}]' = E[\underline{\hat{\theta}} - \underline{\theta}]E[\underline{\hat{\theta}} - \underline{\theta}]'$$

$$- \sigma^2(L-1)\frac{(N-1)}{N} QW' \; [I - \Lambda]^{-1} \; WQ\underline{q} \; E[\underline{\hat{\theta}} - \underline{\theta}]'$$

$$- \sigma^2(L-1)\frac{(N-1)}{N} E[\underline{\hat{\theta}} - \underline{\theta}]\underline{q}' \; QW' \; [I - \Lambda]^{-1}WQ$$

$$+ \sigma^4(L-1)^2\frac{(N-1)^2}{N^2} QW'[I - \Lambda]^{-1}WQ\underline{q}\,\underline{q}' QW'[I - \Lambda]^{-1}WQ \quad ,$$

which can be rewritten as

$$E[J(\underline{\hat{\theta}})-\underline{\theta}]E[J(\underline{\hat{\theta}})-\underline{\theta}]' = \sigma^4(L-1)^2 QW'\left[I-\frac{(N-1)}{N}(I-\Lambda)^{-1}\right] WQ\underline{q}\,\underline{q}' QW' \cdot\left[I - \frac{(N-1)}{N} (I - \Lambda)^{-1}\right]WQ \quad ,$$

where $\sigma^2(L-1)Q\underline{q}$ has been substituted for $E[\underline{\hat{\theta}} - \underline{\theta}]$.

Thus, for the jackknife not to increase the bias of the 2SLS estimator, we are required to show that

$$\sigma^4(L-1)^2 QW'\left[I - \frac{(N-1)}{N} (I-\Lambda)^{-1}\right] WQ\underline{q}\,\underline{q}' QW' \left[I - \frac{(N-1)}{N} (I - \Lambda)^{-1}\right] WQ - \sigma^4(L-1)^2 Q\underline{q}\,\underline{q}' Q \tag{5.17}$$

has all main diagonal components ≤ 0.

If we denote the iith component of Λ by λ_i, then Teekens [68; pp.103-106] has shown that, in general,

$$\frac{1}{N} \leq \lambda_i \leq 1, \qquad (i = 1, 2, \ldots, N)$$

and it follows that

$$\left[I - \frac{(N-1)}{N} (I - \Lambda)^{-1}\right]_{ii} < 0 \; , \; (i=1, 2, \ldots, N) \tag{5.18}$$

where the ii subscript refers to the iith component of the matrix formed by those terms in the square brackets.

Thus, when equation (5.18) holds,

$$- QW' \left[I - \frac{(N-1)}{N} (I - \Lambda)^{-1} \right] WQ$$

must be positive definite.

We now require the following theorem from Rao [46; p.37]:

THEOREM

Let A and B be real m x m symmetric matrices of which B is positive definite. Then there exists a matrix R such that

$$A = R'^{-1} \Delta R^{-1} \quad \text{and} \quad B = R'^{-1} R^{-1}$$

where Δ is a diagonal matrix.

Using this theorem, there exists a matrix R such that

$$-R'QW' \left[I - \frac{(N-1)}{N} (I - \Lambda)^{-1} \right] WQR = \Delta$$

and $\quad R'QR = I$,

where Δ is a diagonal matrix whose main diagonal components are positive and equal to the roots of the equation

$$\left| - QW' \left[I - \frac{(N-1)}{N} (I - \Lambda)^{-1} \right] WQ - \lambda Q \right| = 0 ,$$

or $\quad \left| - Q^{\frac{1}{2}}W' \left[I - \frac{(N-1)}{N} (I - \Lambda)^{-1} \right] WQ^{\frac{1}{2}} - \lambda I \right| = 0 .$

Thus, from equation (5.17), for the jackknife not to increase bias we require

$$\sigma^4 (L-1)^2 (R')^{-1} R'QRR^{-1}\underline{q}\underline{q}'(R')^{-1}R'QRR^{-1}$$

$$- \sigma^4 (L-1)^2 (R')^{-1}R'QW \left[I - \frac{(N-1)}{N} (I - \Lambda)^{-1} \right] W\,QRR^{-1}\underline{q}\underline{q}'R'^{-1}$$

$$.R'QW' \left[I - \frac{(N-1)}{N} (I - \Lambda)^{-1} \right] WQRR^{-1}$$

$$= \sigma^4 (L-1)^2 \left[(R')^{-1} R^{-1}\underline{q}\underline{q}'R'^{-1} (R^{-1}) - (R')^{-1} \Delta R^{-1}\underline{q}\underline{q}' R'^{-1} \Delta R^{-1} \right]$$

to have non-negative main diagonal components. This cannot be shown but the sum of squared biases will be reduced in the general case and, in the case of two included endogenous variables, the squared bias of the endogenous coefficient estimator will be reduced if the roots λ_i, of the matrix

$$-Q^{\frac{1}{2}}W'\left[I - \frac{(N-1)}{N}(I-\Lambda)^{-1}\right]WQ^{\frac{1}{2}}$$

are such that

$$0 \leq \lambda_i \leq 1 \quad . \qquad (i = 1, 2, \ldots, K_1 + g) \tag{5.20}$$

Since this condition is dependent upon W it is not possible to give a general statement concerning its existence. However, a sufficient condition for equation (5.20) to hold is that

$$\left[I - \frac{(N-1)}{N}(I-\Lambda)^{-1}\right]_{ii} \geq -1 \; , \qquad (i = 1, 2, \ldots, N)$$

i.e. $\frac{(N-1)}{N} \frac{1}{1 - \underline{w}_i' Q \underline{w}_i} - 1 \leq 1$

or $\quad \underline{w}_i' Q \underline{w}_i \leq \frac{N+1}{2N} \quad . \qquad (i = 1, 2, \ldots, N)$

It is known that

$$\sum_{i=1}^{N} \underline{w}_i' Q \underline{w}_i = \text{trace } W(W'W)^{-1}W' = K_1 + g \; ;$$

and so the "average" value of $\underline{w}_i' Q \underline{w}_i$ is $(K_1 + g)/N$.

But for

$$\frac{K_1 + g}{N} \geq \frac{N+1}{2N} \; , \quad \text{or identically} \quad K_1 + g \geq \frac{N+1}{2} \; ,$$

the sufficient condition cannot hold.

This suggests that when the number of observations is not at least twice the number of included variables, the jackknife should not be used.

5.5 A Comparison of the Exact Bias of the 2SLS and J2SLS Estimators

The analytical results derived in this Chapter can be summarized as follows:

(i) for a structural equation containing just two endogenous variables, if $K_2 = 2$ jackknifing is unlikely to be successful;

(ii) in general, even when $K_2 > 2$ and μ^2 is "reasonably large", jackknifing is unlikely to be successful unless the number of observations is at least twice the number of variables included in the equation being estimated.

It is apparent from these results that analytical guidelines on criteria for applying the jackknife to the 2SLS estimator are rather vague. A series of experiments was conducted therefore to observe circumstances in which the jackknife is successful in reducing the bias of the 2SLS estimator.

The experiments compare the exact biases of 2SLS and J2SLS as given by equations (5.1) and (5.7) (using equation (5.5)) respectively, but take no account of any resulting change in variance.

The exogenous variables were generated as pseudo- random numbers from the uniform distribution in the range 0 to 100. A specified level of theoretical multicollinearity (λ) was applied such that the theoretical pairwise correlation between exogenous variables was the same for each experiment. λ took values from 0.0 to 0.8 in steps of 0.2.

The relative biases of the 2SLS and J2SLS estimators were calculated exactly for specified values of N, K_1, K_2, ω_{12}, ω_{22}, and the sub-vector of reduced form coefficients, π_{22}.

The values of ω_{12} and ω_{22} were set at 0.0 and 1000.0 respectively for all experiments. From equations (5.1) and (5.5) it can be seen that ω_{12} and ω_{22} enter the expressions for bias only through ρ. Consequently a change in either or both of these parameters only has a simple multiplicative effect on the biases and can be ignored without loss of generality.

K_1 was fixed at 2 for the majority of the experiments, whilst K_2 took on values of 2, 4 and 6. N took values of 10, 20 and 30.

Tables 5.1 - 5.7 give the results of the experiments. The relative bias of both estimators is given, together with the corresponding value of the concentration parameter, μ_N^2.

Table 5.7 gives the results of experiments designed to test the conclusion derived in Section 5.4, viz: if the number of observations is not at least twice the number of included variables the jackknife should not be used. For the purpose of these experiments N and K_2 were fixed at 20 and 4 respectively, whilst K_1 took values of 4, 6 and 8.

An asterisk indicates experiments where the jackknife did not reduce the bias of the 2SLS estimator.

It was suggested in Section 5.3.2 that if $K_2 = 2$ jackknifing may not be successful in reducing bias. From Tables 5.1 and 5.4 it is apparent that jackknifing is indeed generally unsuccessful. In addition, in Section 4.3 it was shown that for "large" μ^2 and $K_2 = 2$ the 2SLS estimator is "nearly" unbiased. The results in Table 5.4 indicate the deleterious effects of using the jackknife under such conditions, even though μ_N^2 is not very "large".

For $K_2 > 2$ application of the jackknife, in general, produces a fairly substantial reduction in the bias of the 2SLS estimator. Note that for fixed N, J2SLS does not exhibit a consistent pattern of bias as λ increases, whereas the bias of 2SLS always increases with increasing λ.

In general, except for very small values of μ_N^2, jackknifing changes the sign of the 2SLS bias.

The results in Table 5.7 indicate that it would be unwise to apply the jackknife to the 2SLS estimator when the number of observations is not at least twice the number of included variables. For "small" μ_N^2 the jackknife produces a substantial reduction in bias, but the ensuing Monte Carlo study will indicate that there is likely to be a substantial increase in the variance of the J2SLS estimator when μ_N^2 is "small". However, since μ_N^2 is never known in practice, it would be unwise to use the jackknife when this condition prevails.

These exact results suggest that the jackknife can be most useful in reducing bias when the equation being estimated is "well" over-identified. It would certainly be unwise to use the jackknife when $K_2 = 2$ or when the number of observations is not at least twice the number of included variables.

Table 5.1: Exact Relative Biases of the 2SLS and J2SLS Estimators

$K_2 = 2$ $\pi_{22} = (0.5, -0.5)$

	N = 10			N = 20			N = 30		
λ	μ_N^2	Relative Bias		μ_N^2	Relative Bias		μ_N^2	Relative Bias	
		2SLS	J2SLS		2SLS	J2SLS		2SLS	J2SLS
0.0	5.8775	-0.0529	+0.1873*	8.9645	-0.0113	+0.0592*	12.0464	-0.0024	+0.0179*
0.2	4.7927	-0.0910	+0.2156*	7.4405	-0.0242	+0.0924*	10.3583	-0.0056	+0.0317*
0.4	3.5482	-0.1696	+0.2259*	5.6726	-0.0586	+0.1418*	8.3496	-0.0154	+0.0611*
0.6	2.2820	-0.3195	+0.1421	3.7775	-0.1513	+0.1732*	6.0012	-0.0498	+0.1189*
0.8	1.0954	-0.5783	-0.1826	1.8411	-0.3983	-0.0028	3.2606	-0.1959	+0.1490

$K_1 = 2$ $\omega_{12} = 0.0$ $\omega_{22} = 1000.0$

Table 5.2: Exact Relative Biases of the 2SLS and J2SLS Estimators

$K_2 = 4$ $\pi_{22} = (0.5, -0.5, 0.5, -0.5)$

	N = 10			N = 20			N = 30		
λ	μ_N^2	Relative Bias		μ_N^2	Relative Bias		μ_N^2	Relative Bias	
		2SLS	J2SLS		2SLS	J2SLS		2SLS	J2SLS
0.0	4.8967	-0.3731	-0.0161	15.5754	-0.1284	+0.0194	29.7400	-0.0672	+0.0071
0.2	3.6992	-0.4556	-0.0906	12.3167	-0.1620	+0.0241	21.4978	-0.0930	+0.0101
0.4	2.8151	-0.5366	-0.1842	9.4623	-0.2095	+0.0242	15.0562	-0.1328	+0.0143
0.6	2.0354	-0.6275	-0.3146	6.6289	-0.2907	+0.0011	9.6338	-0.2059	+0.0131
0.8	1.1894	-0.7538	-0.5275	3.5345	-0.4692	-0.1395	4.6532	-0.3879	-0.0745

$K_1 = 2$ $\omega_{12} = 0.0$ $\omega_{22} = 1000.0$

Table 5.3: Exact Relative Biases of the 2SLS and J2SLS Estimators

$K_2 = 6 \qquad \pi_{22} = (0.5, -0.5, 0.5, -0.5, 0.5, -0.5)$

	N = 20			N = 30		
		Relative Bias			Relative Bias	
λ	μ_N^2	2SLS	J2SLS	μ_N^2	2SLS	J2SLS
0.0	32.8269	-0.1144	+0.0101	52.3695	-0.0735	+0.0046
0.2	23.0495	-0.1585	+0.0070	36.8341	-0.1027	+0.0039
0.4	15.3959	-0.2261	-0.0054	24.5283	-0.1498	+0.0001
0.6	9.2195	-0.3407	-0.0544	14.6172	-0.2362	-0.0191
0.8	4.2646	-0.5502	-0.2455	6.6967	-0.4252	-0.1300

$K_1 = 2 \qquad \omega_{12} = 0.0 \qquad \omega_{22} = 1000.0$

Table 5.4: Exact Relative Biases of the 2SLS and J2SLS Estimators

$K_2 = 2$ $\Pi_{22} = (1.0, -1.0)$

	N = 10			N = 20			N = 30		
λ	μ_N^2	Relative Bias		μ_N^2	Relative Bias		μ_N^2	Relative Bias	
		2SLS	J2SLS		2SLS	J2SLS		2SLS	J2SLS
0.0	23.5099	0.0000	+0.0008*	35.8579	0.0000	0.0000*	48.1854	0.0000	-0.0005*
0.2	19.1710	-0.0001	+0.0027*	29.7621	0.0000	0.0000*	41.4331	0.0000	0.0000*
0.4	14.1927	-0.0008	+0.0141*	22.6905	0.0000	+0.0003*	33.3983	0.0000	0.0000*
0.6	9.1281	-0.0104	+0.0777*	15.1101	-0.0005	+0.0058*	24.0050	0.0000	+0.0001*
0.8	4.3816	-0.1118	+0.2585	7.3645	-0.0252	+0.0918*	13.0424	-0.0015	+0.0107*

$K_1 = 2$ $\omega_{12} = 0.0$ $\omega_{22} = 1000.0$

Table 5.5: Exact Relative Biases of the 2SLS and J2SLS Estimators

$K_2 = 4$ $\pi_{22} = (1.0, -1.0, 1.0, -1.0)$

	N = 10			N = 20			N = 30		
λ	μ_N^2	Relative Bias		μ_N^2	Relative Bias		μ_N^2	Relative Bias	
		2SLS	J2SLS		2SLS	J2SLS		2SLS	J2SLS
0.0	19.5866	-0.1021	+0.0409	62.3016	-0.0321	+0.0051	118.9602	-0.0168	+0.0018
0.2	14.7967	-0.1351	+0.0578	49.2668	-0.0406	+0.0070	85.9910	-0.0233	+0.0026
0.4	11.2603	-0.1770	+0.0778	37.8492	-0.0528	+0.0098	60.2250	-0.0332	+0.0036
0.6	8.1416	-0.2415	+0.0849	26.5155	-0.0754	+0.0145	38.5351	-0.0519	+0.0061
0.8	4.7575	-0.3814	+0.0017	14.1380	-0.1413	+0.0248	18.6126	-0.1074	+0.0127

$K_1 = 2$ $\omega_{12} = 0.0$ $\omega_{22} = 1000.0$

Table 5.6: Exact Relative Biases of the 2SLS and J2SLS Estimators

$K_2 = 6$ $\pi_{12} = (1.0, -1.0, 1.0, -1.0, 1.0, -1.0)$

	N = 20			N = 30		
λ	μ_N^2	Relative Biases		μ_N^2	Relative Bias	
		2SLS	J2SLS		2SLS	J2SLS
0.0	131.3077	-0.0300	+0.0046	209.4781	-0.0189	+0.0019
0.2	92.1980	-0.0424	+0.0058	147.3363	-0.0268	+0.0024
0.4	61.5836	-0.0628	+0.0081	98.1132	-0.0399	+0.0031
0.6	36.8780	-0.1026	+0.0102	58.4687	-0.0661	+0.0040
0.8	17.0585	-0.2070	+0.0024	26.7869	-0.1382	+0.0016

$K_1 = 2$ $\omega_{12} = 0.0$ $\omega_{22} = 1000.0$

Table 5.7: Exact Relative Biases of the 2SLS and J2SLS Estimators ($K_1 + g$ "large" relative to N)

$K_2 = 4$ π_{22} = (0.5, -0.5, 0.5, -0.5)

	$K_1 = 8$ N = 20			$K_1 = 10$ N = 20			$K_1 = 12$ N = 30		
λ	μ_N^2	Relative Bias 2SLS	Relative Bias J2SLS	μ_N^2	Relative Bias 2SLS	Relative Bias J2SLS	μ_N^2	Relative Bias 2SLS	Relative Bias J2SLS
0.0	10.5786	-0.1881	+0.1529	11.7998	-0.1690	+0.2040*	4.4360	-0.4018	+0.5649*
0.2	8.6514	-0.2281	+0.1626	8.8120	-0.2242	+0.2356*	2.9796	-0.5199	+0.3831
0.4	6.7927	-0.2846	+0.1608	6.2274	-0.3069	+0.2373	1.8676	-0.6500	+0.0091
0.6	4.8248	-0.3774	+0.1174	3.8990	-0.4399	+0.1422	1.0282	-0.7819	-0.2801
0.8	2.5090	-0.5698	-0.1007	1.8195	-0.6566	-0.1904	0.4652	-0.8922	-0.6428

$\omega_{12} = 0.0$ $\omega_{22} = 1000.0$

CHAPTER 6

MONTE CARLO STUDY

6.1 Design of Experiments

An evaluation of the effects of applying the jackknife technique to the 2SLS estimator necessitates the use of Monte Carlo methods. Although the exact finite sample distribution and exact moments (where they exist) have been derived for several simultaneous equation estimators in the context of the model used in the ensuing study (e.g. see the bibliographical paper compiled by Owen and Knight [41]), neither the exact finite sample distribution nor exact second and higher order moments of the J2SLS estimator have been derived. Consequently, a Monte Carlo analysis is our only method of evaluating the effects of applying the jackknife technique to the 2SLS estimator.

The model used for one-third of the experiments was

$$\underline{y}_1 = \beta_{12}\underline{y}_2 + \gamma_{10} + \gamma_{11}\underline{X}_1 + \underline{u}_1 \tag{6.1}$$

$$\underline{y}_2 = \beta_{21}\underline{y}_1 + \gamma_{20} + \gamma_{22}\underline{X}_2 + \gamma_{23}\underline{X}_3 + \gamma_{24}\underline{X}_4 + \underline{u}_2 , \tag{6.2}$$

whilst for the remaining experiments equation (6.2) was augmented by an additional three or six exogenous variables.

The reduced form of this two-equation model is given by

$$\underline{y}_1 = \underline{\pi}_{10} + \sum_{i=1}^{4} \underline{X}_i \pi_{1i} + \underline{v}_1 \tag{6.3}$$

$$\underline{y}_2 = \underline{\pi}_{20} + \sum_{i=1}^{4} \underline{X}_i \pi_{2i} + \underline{v}_2 , \tag{6.4}$$

where both equations should be augmented by the relevant additional terms when $K_2 = 6$ and $K_2 = 9$.

The set of parameter values used in the experiments is given in Table 6.1.

Table 6.1: Parameter Values Used in Monte Carlo Experiments

β_{12}	0.8	β_{21}	-0.7	γ_{25}	-1.0	γ_{28}	1.2
γ_{10}	50.0	γ_{20}	50.0	γ_{26}	1.9	γ_{29}	-1.5
γ_{11}	1.2	γ_{22}	1.3	γ_{27}	-1.1	γ_{210}	0.9
		γ_{23}	1.6				
		γ_{24}	-2.0				

The exogenous variables were generated as rectangularly and independently distributed pseudo-random variables in the range 0 to 100, but were then transformed in order to obtain a specified theoretical pairwise correlation (λ) between them. The sample correlations are given in Table 6.2. Values for experiments using less than the full set of exogenous variables (i.e. less than 10, excluding the constant) should be read-off from the upper left corner of the table.

All experiments were based on a sample size of 20.

The reduced form disturbances, the v_{it} $(i = 1,2)$, were generated as bivariate normal variates with zero mean and covariance matrix

$$\Omega = \begin{bmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{bmatrix} = \begin{bmatrix} 1600 & 1520\delta \\ 1520\delta & 1444 \end{bmatrix},$$

where the coefficient δ was given the value 0.19 in half of the experiments, and 0.76 in the other half.

Each estimate of the parameters in equation (6.1) (i.e. the first equation only) was calculated as the mean of 100 replications of the

Table 6.2: Matrix of Sample Correlations Between the Exogenous Variables

		X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
X_1	0.00	+1.0000									
	0.45	+1.0000									
	0.90	+1.0000									
X_2	0.00	-0.2977	+1.0000								
	0.45	+0.1317	+1.0000								
	0.90	+0.5001	+1.0000								
X_3	0.00	+0.0628	+0.0971	+1.0000							
	0.45	+0.1084	+0.5244	+1.0000							
	0.90	+0.3874	+0.7815	+1.0000							
X_4	0.00	-0.0818	-0.1558	-0.0404	+1.0000						
	0.45	-0.0215	+0.0791	+0.3923	+1.0000						
	0.90	+0.2987	+0.5786	+0.7979	+1.0000						
X_5	0.00	+0.1821	-0.3951	-0.0084	-0.0513	+1.0000					
	0.45	+0.1606	-0.2720	+0.0516	+0.3636	+1.0000					
	0.90	+0.3811	+0.3987	+0.6353	+0.8106	+1.0000					
X_6	0.00	-0.3281	+0.0148	+0.1738	-0.0418	-0.2519	+1.0000				
	0.45	-0.2641	-0.2447	+0.1183	+0.1618	+0.1851	+1.0000				
	0.90	+0.1171	+0.1936	+0.5301	+0.6734	+0.7453	+1.0000				
X_7	0.00	+0.2261	-0.1912	-0.2425	+0.1720	+0.1684	-0.3381	+1.0000			
	0.45	+0.1211	-0.2045	-0.2060	+0.1037	+0.2572	+0.1454	+1.0000			
	0.90	+0.2562	+0.1639	+0.3414	+0.5461	+0.6784	+0.7272	+1.0000			
X_8	0.00	+0.1431	+0.0475	+0.3820	-0.2319	+0.0195	+0.0832	-0.0321	+1.0000		
	0.45	+0.1829	+0.0039	+0.2346	+0.0241	+0.1246	+0.1440	+0.4693	+1.0000		
	0.90	+0.2746	+0.2069	+0.4450	+0.5172	+0.6199	+0.6784	+0.8658	+1.0000		
X_9	0.00	+0.0028	+0.3188	+0.1081	-0.0222	-0.3941	+0.1093	-0.1402	-0.4943	+1.0000	
	0.45	+0.0979	+0.3869	+0.3884	+0.1222	-0.3117	+0.0447	+0.0636	-0.0715	+1.0000	
	0.90	+0.2714	+0.3866	+0.6075	+0.6528	+0.5936	+0.6996	+0.8171	+0.7982	+1.0000	
X_{10}	0.00	+0.1320	+0.1307	+0.2076	-0.1996	+0.4577	-0.0002	-0.0124	-0.0414	+0.3131	+1.0000
	0.45	+0.1426	+0.3024	+0.3545	+0.0024	+0.2026	+0.1517	+0.0708	-0.0348	+0.6672	+1.0000
	0.90	+0.2598	+0.3838	+0.5766	+0.5465	+0.6131	+0.6751	+0.7360	+0.6869	+0.9206	+1.0000

relevant estimator. All experiments were devised to ensure that at least the first two integer moments of the 2SLS estimator were finite. This would not be so with the second equation, (6.2).

6.2 Exact Results

Although exact values for the bias and mean square error (MSE) of the 2SLS estimator, and for the bias of the J2SLS estimator, in equation (6.1) are known (and for β_{12} are given in Table 6.3), for compatibility reasons comparison of variance and MSE must necessarily be based upon a Monte Carlo study.

The values in Table 6.3 can serve as a guide to the accuracy of the experiments which follow.

It should be noted that when $\delta = 0.76$, the 2SLS and J2SLS estimators of β_{12} are both unbiased.

From equation (5.1) it can be seen that the 2SLS estimator of β_{12} is unbiased if $\beta_{12} = \rho$ $(= \omega_{12}/\omega_{22})$. In the experiments conducted here, $\beta_{12} = 0.8$ whilst

$$\rho = \frac{1520\delta}{1444} = 0.8 \text{ (if } \delta = 0.76) \ .$$

It follows from equation (5.5) and the definition of the jackknife that the J2SLS estimator of β_{12} will also be unbiased under the same conditions.

Richardson and Wu [55, pp.977-978] have shown that if the 2SLS estimator of β_{12} is unbiased, then the 2SLS estimator of the coefficients of the exogenous variables must also be unbiased.

If $\beta_{12} = \rho$, then it follows that $\underline{y}_1$ is independent of $\underline{u}_1$, and hence estimation of equation (6.1) becomes a mixed stochastic regression problem. In these circumstances ordinary least squares would be an unbiased estimator and would be the appropriate method of estimation.

Table 6.3: Exact Values of Relative Bias and M.S.E. (β_{12} only)

K_2	λ	δ	μ^2	Relative Bias	M.S.E.	Relative Bias
				2SLS		J2SLS
3	0.00	0.19	41.2725	-0.01865	0.03552	+0.00631
	0.45	0.19	29.1234	-0.02675	0.05094	+0.00968
	0.00	0.76	41.2725	0.0	0.01163	0.0
	0.45	0.76	29.1234	0.0	0.01668	0.0
6	0.00	0.19	95.7945	-0.03066	0.01483	+0.01093
	0.45	0.19	56.3108	-0.05138	0.02514	+0.01633
	0.90	0.19	8.4440	-0.27237	0.01576	-0.04342
	0.00	0.76	95.7945	0.0	0.00478	0.0
	0.45	0.76	56.3108	0.0	0.00801	0.0
	0.90	0.76	8.4440	0.0	0.04246	0.0
9	0.00	0.19	118.3348	-0.04252	0.01156	+0.01082
	0.45	0.19	61.3857	-0.07889	0.02405	+0.01527
	0.90	0.19	9.1349	-0.35015	0.15709	-0.12009
	0.00	0.76	118.3349	0.0	0.00379	0.0
	0.45	0.76	61.3857	0.0	0.00701	0.0
	0.90	0.76	9.1349	0.0	0.03113	0.0

For an equation containing an arbitrary number (g) of explanatory endogenous variables, Revankar and Hartley [51] have generalized the above result. An F test was derived by Revankar and Hartley for testing the hypothesis of equality of β_{12} and ρ.

The selection of δ to be 0.76 for half of the experiments allowed a comparison of test statistics to be made (see Chapter 7) without the added complications of bias and skewness entering the comparisons.

6.3 Computational Considerations

6.3.1 The Problem of "Outliers"

The satisfactory inversion of all moment matrices for all sets of exogenous variables was commented upon in Chapter 4. At each replication of the experiments however, it was necessary to invert the matrix

$$Z'X(X'X)^{-1}X'Z$$

and to check against singularity (or near-singularity) caused by the build-up of rounding errors. If singularity was found to be present, the relevant sample values were disgarded and an additional replication performed.

For experiments involving $K_2 = 3$ and $\lambda = 0.9$, although no replication was rejected, the 2SLS and J2SLS parameter estimates were grossly in error as compared with their exact values for β_{12}. Rather than design an ad hoc procedure to allow rejection of "unrepresentative" sample values, or outliers, in order to achieve "reasonable" parameter estimates, it was decided to reject this particular experiment completely.

It is difficult to justify the rejection of "outliers" since any cut-off point obviously suffers from a great degree of arbitrariness. Indeed, one could very well be rejecting "true" sample values as well as "rounding error" sample values by applying such a procedure.

6.3.2 Antithetic Variates

The technique of antithetic variates was used in an attempt to reduce (to an unknown degree) the sampling error of the Monte Carlo study when estimating the biases of both estimators (see Hammersley and Handscomb [16] for a description of the technique).

Whilst the antithetic method produced estimates of β_{12} which were marginally closer (than direct simulation) to their exact values for the majority of experiments, there was little to choose between the two methods for estimating the MSE of the 2SLS estimator of β_{12}. This latter feature was noticed by Mikhail [33] in a similar experiment, although he managed to achieve a substantial reduction in sampling error when estimating the bias of the 2SLS estimator.

The additional computer time and storage required to calculate parameter estimates using antithetic variates is minimal, as it merely requires a sign change at an advanced stage in the calculations. However, there is a considerable increase in computer time and storage involved in constructing, storing and sorting twice as many test statistics as were generated by direct simulation. Since this study was already facing computer time and storage constraints using direct simulation, the author did not feel that the small decrease in sampling error justified the increased computer time and storage.

6.4 Results of Monte Carlo Study

Tables 6.4, 6.5 and 6.6 (which are situated at the end of this Chapter) summarize the Monte Carlo results on relative bias, variance, MSE and mean absolute error (MAE) for the three parameters of interest; viz β_{12}, γ_{10} and γ_{11}. Values of the standardized normal statistic for the Wilcoxon Matched-Pairs Signed-Ranks test (e.g. see Siegel [63;

pp.47-52]) under the hypothesis of equality of absolute errors of the two estimators are given in the final column.

Each of these three tables is subdivided into two parts, (a) and (b). Results for $\delta = 0.19$ are given in part (a) of each table, whilst part (b) contains the results for the situation where both estimators are unbiased, i.e. $\delta = 0.76$.

We now consider, in turn, four criteria for discriminating between the two estimators.

6.4.1 Bias

The "large" relative bias of 2SLS which was evident in the exact study (Chapter 5) for high levels of multicollinearity was also apparent in the Monte Carlo study when $\delta = 0.19$. For these experiments the jackknife never failed to reduce the bias of the 2SLS estimator, although this reduction was more marked for β_{12} than for the coefficients of the (2) exogenous variables, γ_{10} and γ_{11}.

All estimates of relative bias had the correct sign. From Table 6.3 it can be seen that the exact relative bias of β_{12} for both 2SLS and J2SLS were very close to the simulation results when $K_2 = 6$. For $K_2 = 3$ and $K_2 = 9$, however, the degree of agreement between the simulated and exact results was not as good.

For $\delta = 0.76$ (i.e. both estimators unbiased) the "relative bias" figures obtained from the experiments must be due to sampling and rounding errors. These errors are particularly noticeable when the level of multicollinearity (λ) is high.

We can be reasonably pleased with the degree of agreement between the exact and experimental results on bias. It is interesting to note that in Summer's [66] experiments 1A - 4A and 1B - 4B, with a model which only differed from the one used in this study by the inclusion of fewer exogenous variables, the mean of the 2SLS estimator of β_{12} over 50 replications had an incorrect bias sign on four (of the eight) occasions.

6.4.2 Variance

In general, 2SLS exhibited a smaller variance than J2SLS for all three parameter estimates, and this was particularly noticeable as the degree of multicollinearity increased. Where the jackknife produced a smaller variance, its superiority was never significant. As K_2 increased, the discrepancy between the 2SLS variance and the larger J2SLS variance widened for all parameter estimates.

6.4.3 Mean Square Error

In general, the reduction in bias due to the application of the jackknife was not of sufficient size to offset the smaller variance of 2SLS. In most cases (for both estimators) the square of the bias was small and had little additional effect when added to the variance. Consequently, in common with the variance, 2SLS was generally superior (for all parameters) on a MSE criterion.

It should be noted, however, that this superiority was particularly marked for "small" values of μ^2 (e.g. when $\mu^2 = 8.440$ and $\mu^2 = 9.1349$). For "larger" values of μ^2, the MSEs of the two estimators did not differ greatly. Frequently, the Wilcoxon test picks up this substantial difference between the two estimators for "small" μ^2, but this statistic is based on testing absolute errors.

With only one exception, the MSE of the 2SLS estimator of β_{12} obtained from the experiments underestimated the exact MSE. Despite this, the exact and experimental values were very close for all values of K_2 and λ.

6.4.4 Mean Absolute Error

In general, 2SLS was superior on a MAE criterion, although its superiority was not as marked as for the MSE criterion. Again, "small" values of μ^2 lead to a great discrepancy between the MAEs of the 2SLS and J2SLS estimators.

6.5 Difference of Absolute Errors

At each replication the absolute error of both estimators was calculated. Let $\hat{b}_i$ and $\tilde{b}_i$ be the absolute errors at the ith replication of the 2SLS and J2SLS estimators of β_{12} respectively, then the difference score is defined as

$$d_i = \hat{b}_i - \tilde{b}_i . \qquad (i = 1,2,\ldots,R)$$

We wish to test the hypothesis of equality of $\hat{b}_i$ and $\tilde{b}_i$ over all R replications.

The usual parametric technique for handling such a problem is Student's t distribution, but this requires the assumption that the difference scores (the d_i) are normally and independently distributed in the population from which the sample was drawn. Since this assumption has no theoretical justification for the case being considered here, the Wilcoxon Matched-Pairs Signed-Ranks test (e.g. see Siegel [63; pp.75-83]) was used to test the hypothesis of equality of absolute errors. If the assumptions of the parametric t test are in fact met, the asymptotic efficiency near the null hypothesis of the Wilcoxon test compared with the t test is 95.5%.

Under the stated hypothesis, the Wilcoxon test was conducted for all three parameters being estimated, and the resulting Z statistics are given in the last column of Tables IV, V, and VI. Negative values favour 2SLS.

At a 5% level of significance, the hypothesis of equality of absolute errors is rejected only twice over all parameters when $\delta = 0.19$. Both rejections are in favour of the 2SLS estimator, and both occur for $K_2 = 9$ and $\lambda = 0.90$ (i.e. when μ^2 is "small").

When $\delta = 0.76$, however, the hypothesis is rejected on four occasions for β_{12} alone, all four rejections in favour of the 2SLS estimator. Surprisingly, this result did not carry over to the 2SLS estimates of γ_{10} and γ_{11}.

6.6 Conclusion

The results of the Monte Carlo study are not encouraging for proponents of the jackknife technique. Whilst 2SLS was clearly superior when there existed a high degree of multicollinearity, application of the jackknife technique, in general, could not produce superior results using either a MSE or MAE criterion. In view of the increased complexity and computation time involved in applying the jackknife, its use cannot be recommended on the basis of the above results alone.

On the basis of the above results, the following statements can be made:

(i) for a relatively high degree of overidentification (i.e. $K_2 = 6$ or $K_2 = 9$ in these experiments), application of the jackknife technique produces a substantial reduction in the bias of the 2SLS estimator;

(ii) over all experiments 2SLS is superior on a MSE criterion, this superiority being particularly marked when μ^2 is "small";

(iii) when $\delta = 0.19$, there appears to be little significant difference between the two estimators over all parameters, using the absolute error criterion, on the basis of the Wilcoxon Matched-Pairs Signed-Ranks test;

(iv) over all experiments, differences between 2SLS and J2SLS estimates of β_{12} using MAE, MSE, and variance criteria are far less marked than the same differences for γ_{10} and γ_{11};

(v) when the 2SLS estimator is unbiased (i.e. $\delta = 0.76$), application of the jackknife is clearly unwarranted and its application in error is likely to have a detrimental effect on the parameter estimates. Clearly, to avoid this possibility, Revankar and Hartley's [51] test should be used prior to estimation.

Table 6.4(a): Results of Monte Carlo Experiments

Parameter = β_{12}

K_2	λ	μ^2	RELATIVE BIAS		VARIANCE		M.S.E.		M.A.E.		WILCOXON TEST
			2SLS	J2SLS	2SLS	J2SLS	2SLS	J2SLS	2SLS	J2SLS	
3	0.00	41.2725	-0.02739	-0.00151	0.03014	0.03044	0.03062	0.03044	0.14138	0.14030	+0.7324
	0.45	29.1234	-0.04351	-0.00052	0.04022	0.04433	0.04143	0.04433	0.16927	0.17156	+0.1977
6	0.00	95.7945	-0.02911	+0.01092	0.01329	0.01473	0.01382	0.01480	0.09423	0.09717	-0.4487
	0.45	56.3108	-0.05176	+0.01428	0.01963	0.02315	0.02134	0.02327	0.11690	0.12211	-0.6223
	0.90	8.4440	-0.26531	-0.03903	0.09687	0.32332	0.14192	0.32429	0.30989	0.38605	-0.4986
9	0.00	118.3349	-0.03588	+0.02243	0.01123	0.01327	0.01205	0.01359	0.08642	0.09118	-0.4590
	0.45	61.3857	-0.06731	+0.03637	0.01977	0.02815	0.02267	0.02900	0.12234	0.13239	-0.4590
	0.90	9.1349	-0.30998	-0.03315	0.08634	0.37555	0.14784	0.37626	0.32287	0.44477	-2.3088

Sample size = 20 δ = 0.19

Table 6.4(b): Results of Monte Carlo Experiments

Parameter = β_{12}

			RELATIVE BIAS		VARIANCE		M.S.E.		M.A.E.		WILCOXON
K_2	λ	μ^2	2SLS	J2SLS	2SLS	J2SLS	2SLS	J2SLS	2SLS	J2SLS	TEST
3	0.00	41.2725	-0.00629	-0.00215	0.00939	0.00946	0.00942	0.00946	0.08139	0.08130	+0.4109
	0.45	29.1234	-0.00814	-0.00136	0.01326	0.01431	0.01330	0.01431	0.09392	0.09762	-1.2705
6	0.00	95.7945	-0.00524	-0.00328	0.00438	0.00490	0.00439	0.00490	0.05223	0.05582	-1.9444
	0.45	56.3108	-0.00668	-0.00305	0.00655	0.00799	0.00657	0.00800	0.06450	0.07098	-2.5616
	0.90	8.4440	-0.00844	+0.01385	0.03462	0.10562	0.03466	0.10573	0.14633	0.22304	-5.1540
9	0.00	118.3349	-0.00235	+0.00143	0.00371	0.00445	0.00371	0.00445	0.04988	0.05334	-1.5576
	0.45	61.3857	-0.00191	+0.00588	0.00624	0.00880	0.00624	0.00883	0.06383	0.07461	-2.7232
	0.90	9.1349	-0.00188	+0.03911	0.02810	0.09701	0.02811	0.09799	0.12742	0.22593	-6.4812

Sample size = 20 δ = 0.76

Table 6.5(a): Results of Monte Carlo Experiments

Parameter = γ_{10}

			RELATIVE BIAS		VARIANCE		M.S.E.		M.A.E.		WILCOXON
K_2	λ	μ^2	2SLS	J2SLS	2SLS	J2SLS	2SLS	J2SLS	2SLS	J2SLS	TEST
3	0.00	41.2725	+0.05783	+0.03567	494.50	509.77	502.86	512.95	17.80	18.16	-0.8218
	0.45	29.1234	+0.05958	+0.03691	532.11	552.15	540.98	555.55	18.52	18.69	-0.3335
6	0.00	95.7945	+0.07142	+0.01475	521.61	552.85	534.36	553.40	18.30	18.89	-0.8011
	0.45	56.3108	+0.07951	+0.01679	554.61	590.48	570.42	591.19	18.87	19.14	-0.1169
	0.90	8.4440	+0.21776	+0.07183	1712.25	2487.70	1830.80	2500.60	34.24	37.75	-1.7811
9	0.00	118.3349	+0.08609	-0.00401	524.09	560.12	542.62	560.16	18.60	18.63	+0.1994
	0.45	61.3857	+0.10527	-0.00937	568.82	636.35	596.53	636.57	19.52	19.55	+0.2571
	0.90	9.1349	+0.26538	+0.02130	1717.59	3005.34	1893.66	3006.47	35.04	39.50	-1.3169

Sample size = 20 δ = 0.19

Table 6.5(b): Results of Monte Carlo Experiments

Parameter = γ_{10}

			RELATIVE BIAS		VARIANCE		M.S.E.		M.A.E.		WILCOXON
K_2	λ	μ^2	2SLS	J2SLS	2SLS	J2SLS	2SLS	J2SLS	2SLS	J2SLS	TEST
3	0.00	41.2725	+0.02198	+0.01713	159.76	161.79	160.97	162.52	10.10	10.19	-0.2372
	0.45	29.1234	+0.02128	+0.01546	175.59	178.15	176.73	178.75	10.66	10.74	-0.6292
6	0.00	95.7945	+0.02206	+0.01941	174.43	177.16	175.65	178.10	10.58	10.69	-0.0688
	0.45	56.3108	+0.02064	+0.01770	188.26	190.72	189.32	191.50	10.95	10.89	+0.6464
	0.90	8.4440	+0.01336	+0.00231	664.16	789.39	664.61	789.40	20.49	21.88	-0.4573
9	0.00	118.3349	+0.01833	+0.01135	176.50	182.49	177.34	182.81	10.68	10.86	-0.5123
	0.45	61.3857	+0.01673	+0.00717	194.44	205.75	195.14	205.88	11.10	11.35	-0.7427
	0.90	9.1349	+0.01557	-0.01026	642.99	798.66	643.59	798.92	20.24	22.14	-1.2103

Sample size = 20 δ = 0.76

Table 6.6(a): Results of Monte Carlo Experiments

Parameter = γ_{11}

			RELATIVE BIAS		VARIANCE		M.S.E.		M.A.E.		WILCOXON
K_2	λ	μ^2	2SLS	J2SLS	2SLS	J2SLS	2SLS	J2SLS	2SLS	J2SLS	TEST
3	0.00	41.2725	-0.01578	-0.00552	0.12200	0.12560	0.12236	0.12564	0.27814	0.27567	+1.1175
	0.45	29.1234	-0.01522	-0.00789	0.13579	0.14115	0.13612	0.14124	0.29161	0.29209	+0.0499
6	0.00	95.7945	-0.02943	+0.00569	0.13652	0.13946	0.13777	0.13951	0.29307	0.29189	+0.5794
	0.45	56.3108	-0.03245	+0.00222	0.14783	0.15137	0.14934	0.15138	0.30267	0.29797	+1.1346
	0.90	8.4440	-0.11355	-0.03921	0.56781	0.79627	0.58637	0.79849	0.61435	0.67577	-1.7501
9	0.00	118.3349	-0.03078	+0.01239	0.12644	0.13457	0.12780	0.13479	0.27999	0.28116	+0.2201
	0.45	61.3857	-0.03260	+0.01124	0.14321	0.15645	0.14474	0.15664	0.29747	0.30058	-0.1221
	0.90	9.1349	-0.06151	+0.02542	0.55360	0.79839	0.55905	0.79932	0.59897	0.67042	-2.1954

Sample size = 20 δ = 0.19

Table 6.6(b): Results of Monte Carlo Experiments

Parameter = γ_{11}

			RELATIVE BIAS		VARIANCE		M.S.E.		M.A.E.		WILCOXON
K_2	λ	μ^2	2SLS	J2SLS	2SLS	J2SLS	2SLS	J2SLS	2SLS	J2SLS	TEST
3	0.00	41.2725	-0.00167	+0.00144	0.04010	0.04035	0.04010	0.04036	0.15781	0.15610	+0.4968
	0.45	29.1234	-0.00082	+0.00282	0.04498	0.04577	0.04498	0.04578	0.16653	0.16748	-0.6533
6	0.00	95.7945	-0.00307	-0.00161	0.04666	0.04552	0.04667	0.04552	0.16974	0.16665	+1.0762
	0.45	56.3108	-0.00165	+0.00019	0.05082	0.04981	0.05083	0.04981	0.17583	0.17240	+1.2808
	0.90	8.4440	+0.00305	+0.00767	0.21889	0.26339	0.21890	0.26348	0.36742	0.40131	-1.5163
9	0.00	118.3349	+0.00022	+0.00334	0.04273	0.04391	0.04273	0.04393	0.16162	0.16309	-0.1083
	0.45	61.3857	+0.00155	+0.00551	0.04836	0.05005	0.04836	0.05010	0.17096	0.17022	+0.6017
	0.90	9.1349	+0.00250	+0.00808	0.19668	0.22882	0.19670	0.22892	0.34628	0.36884	-0.7375

Sample size = 20 δ = 0.76

CHAPTER 7

INFERENCE

7.1 Tests of Significance

7.1.1 Conventional Tests of Significance

So far we have only considered point estimation of the parameters in a simultaneous equation system. In applied economics however it is usual to test for significance of the parameter estimates, or (identically) to formulate interval estimates.

From equation (2.6), the 2SLS estimator of $\underline{\theta}$ is written as

$$\hat{\underline{\theta}} = \left[Z'X(X'X)^{-1}X'Z\right]^{-1} Z'X(X'X)^{-1} X'\underline{y} ,$$

and from equation (2.7) the limiting distribution of the sequence $\sqrt{N}\,(\hat{\underline{\theta}} - \underline{\theta})$ is given by

$$\sqrt{N}\,(\hat{\underline{\theta}} - \underline{\theta}) \sim N\left[0,\ \sigma^2 \operatorname*{plim}_{N\to\infty}\left[\frac{1}{N}\,.\,Z'X\left(\frac{1}{N}\,.\,X'X\right)^{-1}\frac{1}{N}\,.\,X'Z\right]^{-1}\right], \tag{7.1}$$

provided $\lim\limits_{N\to\infty}\left(\frac{1}{N}\,.\,X'X\right)^{-1}$ exists .

The correct asymptotic test of significance therefore is the standardized normal test statistic, and a consistent estimator of σ^2 is given by

$$\tilde{\sigma}^2 = \hat{\underline{u}}'\hat{\underline{u}}\big/ N , \tag{7.2}$$

where $\hat{\underline{u}} = \underline{y} - Y\hat{\underline{\beta}} - X_1\,\hat{\underline{\gamma}}_1$.

It has become common practice however to adjust the estimator of σ^2 for loss of degrees of freedom and use the t statistic, rather than the standardized normal, when dealing with finite samples (e.g. see

Johnston [20; p.384]. Thus in finite samples a consistent estimator of σ^2 is given by

$$\hat{\sigma}^2 = \hat{\underline{u}}'\hat{\underline{u}} \Big/ (N - K_1 - g) \; .$$

From equations (7.1) and (7.2) it follows that, asymptotically,

$$\frac{\sqrt{N}\,(\hat{\theta}_k - \theta_k)}{\tilde{\sigma}\,\sqrt{\bar{S}_k}} \sim N(0,1) \; , \tag{7.3}$$

where $\hat{\theta}_k$ and θ_k are the kth components of $\hat{\underline{\theta}}$ and $\underline{\theta}$ respectively $(k = 1,2,\ldots,K_1 + g)$, and $\bar{S}_k$ is the kkth component of

$$\left[\frac{1}{N}\,.\,Z'X\left(\frac{1}{N}\,.\,X'X\right)\frac{1}{N}\,.\,X'Z\right]^{-1} .$$

Let S_k denote the kkth component of $[Z'X(X'X)^{-1}\,X'Z]^{-1}$, then $\bar{S}_k = NS_k$ and expression (7.3) can be rewritten as

$$\frac{(\hat{\theta}_k - \theta_k)}{\tilde{\sigma}\,\sqrt{S_k}} \sim N(0,1) \; . \tag{7.4}$$

The conventional finite sample counterpart of expression (7.4) is the statistic

$$\frac{(\hat{\theta}_k - \theta_k)}{\hat{\sigma}\,\sqrt{S_k}} \; , \tag{7.5}$$

which is tested as though it is distributed as Student t with $N - K_1 - g$ degrees of freedom.

7.1.2 Dhrymes' Alternative Test of Significance

An alternative asymptotic test of significance based on Student's t distribution has been proposed by Dhrymes [12]. Use of the t statistic is customary for testing the significance of 2SLS parameter estimates yet, until Dhrymes showed the asymptotic validity of his test, no theory existed to justify the practice. On the basis of the asymptotic distribution of the 2SLS estimator, the relevant test of significance should have been based on the standardized normal distribution as described by expression (7.4).

Rewrite the equation being estimated as

$$\underline{y} = Z\underline{\theta} + \underline{u} \tag{7.6}$$

where $Z = [Y : X_1]$ and $\underline{\theta}' = [\underline{\beta}' : \underline{\gamma}_1']$, then define a square, non-singular matrix R of order K such that $RR' = X'X$. Further define $P = R^{-1}X'$, then premultiplying equation (7.6) by P gives

$$\underline{w} = Q\underline{\theta} + \underline{e} ,$$

where $\underline{w} = P\underline{y}$, $Q = PZ$ and $\underline{e} = P\underline{u}$. Dhrymes showed that the 2SLS estimator of $\underline{\theta}$ in equation (7.6) is the OLS estimator of $\underline{\theta}$ in this transformed system. Further, by analogy with least squares, Dhrymes showed that, asymptotically,

$$\frac{(\hat{\theta}_k - \theta_k)}{\bar{\sigma}\sqrt{S_k}} \sim t_{K_2-g} , \tag{7.7}$$

where an asymptotically unbiased, but inconsistent, estimator of σ^2 is given by

$$\bar{\sigma}^2 = \hat{\underline{e}}'\hat{\underline{e}}/(K_2 - g) = \hat{\underline{u}}'X(X'X)^{-1}X'\hat{\underline{u}}/(K_2 - g).$$

Thus the test is only valid if the structural equation in question is over-identified.

Revankar [50], however, has shown that information is lost when a dimension reducing transformation is used as a basis for testing, thus Dhrymes' test could be expected to be inefficient compared to the conventional test based on the standardized normal distribution.

In a Monte Carlo study, Maddala [26] observed that the Dhrymes test had low power compared with the conventional tests in a two equation model. Richardson and Rohr [54] came to the same conclusion on the basis of a Monte Carlo study using a three equation model.

7.2 The Exact Distribution of a t Statistic

The exact finite sample distribution functions of several t statistics for hypothesis testing and the construction of confidence intervals on 2SLS parameter estimates have been studied by Richardson and Rohr [53] and Rohr [57]. As with many other finite sample studies into the properties of the 2SLS estimator, the results were derived for a model with just two jointly-dependent variables.

Richardson and Rohr [53] considered the finite sample distribution of Dhrymes' t statistic, expression (7.7), which Dhrymes had already shown to be asymptotically distributed as Student t with $K_2 - g$ degrees of freedom. However, since the sample size does not appear explicitly in their finite sample derivations, convergence of the t statistic to Student's t distribution was analysed for μ^2 (the concentration parameter) increasing indefinitely.

The moments of the exact distribution were found not to exist to order $K_2 - g$ and higher, but where they did exist they converged to the moments of Student's t distribution with $K_2 - g$ degrees of freedom as $\mu^2 \to \infty$. On the basis of their results Richardson and Rohr conjectured that, for large μ^2, the exact distribution function of the t statistic can be adequately approximated by Student's t distribution with $K_2 - g$ degrees of freedom.

Richardson and Rohr investigated their conjecture for one degree of freedom and for several values of μ and β. On the basis

of their computations they concluded that the actual probability of Type I error (for a significance level of 5%) will be less than 5% if β is positive, and greater than 5% if β is negative. If $\mu > 3$ the exact t statistic was found to be a good approximation to the Student t, but for small β and $\mu \leq 3$ differences between the two could lead to serious errors.

Richardson and Rohr also tabulated the exact value of the second moment and the exact absolute values for the first and third moments of the t statistic for various values of degrees of freedom, β, and μ^2, from which they concluded that the density function is highly skewed and that often the moments differ considerably from those of Student's t distribution with $K_2 - g$ degrees of freedom.

Rohr [57] has derived the exact distribution of two "more conventional" test statistics, only one of which is used in this study, viz:

$$\frac{(\hat{\theta}_k - \theta_k)}{\hat{\sigma} \sqrt{S_k}} ,$$

which is identical to expression (7.5).

Rohr showed that asymptotically (in μ^2) expression (7.5) converges to Student's t distribution with $N - K_1 - g$ degrees of freedom, but that in finite samples the moments of the statistic (7.5) exist only up to order $N - K_1 - g + 1$.

It should be noted, however, that mathematical complexity in the derivation of the moments of expression (7.5) forced Rohr to consider only the special case where $\beta = \sigma_{12}/\sigma_{22}$; i.e. 2SLS unbiased. Under this restriction, expression (7.5) has all odd moments (where they exist) equal to zero, and 2SLS and OLS are equivalent.

Rohr also showed that the variance of expression (7.5) is always less than or equal to the variance of its limiting distribution.

7.3 Student's t Distribution and its use with the Two-Stage Least Squares Estimator

The ratio

$$\frac{w}{\sqrt{v/r}} = \frac{\hat{\theta}_k - E(\hat{\theta}_k)}{S.E.\ (\hat{\theta}_k)} \qquad (k = 1,2,\ldots,K_1 + g) \qquad (7.8)$$

is distributed as Student t if w is normally distributed with zero mean and unit variance and if v has a χ^2 distribution with r degrees of freedom, provided that v and w are stochastically independent.

For 2SLS, in general, $E(\hat{\theta}_k) \neq \theta_k$ and $\hat{\sigma}^2$ is a consistent, but not unbiased, estimator of σ^2. Consequently the denominator of expression (7.8) only approximates a χ^2 distribution. In addition, $\hat{\theta}_k - E(\hat{\theta}_k)$ is not stochastically independent of its standard error (S.E.) in finite samples. It should be noted that $E(\hat{\theta}_k)$ may not even be finite, although in the ensuing Monte Carlo analysis the experiments were designed in such a way as to ensure that the first two moments of the 2SLS estimator were always finite.

7.4 An Approximate t Statistic constructed using the Jackknife Technique

Tukey [72] has suggested that the N pseudo-jackknife estimates could be treated as approximately independent, identically distributed observations from which an approximate t statistic could be constructed as

$$\frac{\sqrt{N}\,[J(\hat{\theta}_k) - \theta_k]}{\left\{(N-1)^{-1}\sum_{i=1}^{N}\left[J_i(\hat{\theta}_k) - J(\hat{\theta}_k)\right]^2\right\}^{\frac{1}{2}}} \qquad (7.9)$$

We have already shown (in Chapter 3) that expression (7.9) is asymptotically distributed as the standardized normal distribution in the context of the J2SLS estimator.

Although in general $E[J(\hat{\theta}_k)] \neq \theta_k$, in many instances it will exhibit a smaller deviation from θ_k than 2SLS, as was observed in Chapter 4. In common with 2SLS, the numerator and the denominator of expression (7.9) will not be stochastically independent in finite samples.

Miller [34] gives several counterexamples to Tukey's conjecture, but Arvesen [1] gives a wide class of situations where this suggestion is valid, i.e. when $J_i(\hat{\alpha}_k)$ and $J(\hat{\alpha}_k)$ are U statistics (see Hoeffding [19]) or functions of U statistics.

Recently, Miller [36] provided an asymptotic justification of Tukey's conjecture for a function of the regression parameters in a general linear model.

7.5 Independence of the Pseudo-Jackknife Estimates

Walsh [73] has demonstrated the deleterious effects of using correlated samples for the construction of certain significance tests. If the N pseudo-jackknife estimates could be considered as a single observation of a normal multivariate population, for which the N variables have common mean μ and variance σ^2, the effect on the t statistic of a common level of pairwise correlation between the pseudo-jackknife estimates would be to raise or lower the true confidence coefficient depending on whether the correlation was positive or negative. Thus if the pairwise correlation (r) was positive, a test result which would be significant for a random sample need no longer be so. To correct the t statistic the multiplying factor

$$\sqrt{\frac{(1-r)}{1+(N-1)r}}$$

is required.

Walsh illustrated the error incurred in assuming $r = 0$ by tabulating the true value of the confidence coefficient for varying values of N and r. Even for small r the deleterious effect of correlation was very marked; e.g. for $N = 8$ and $r = 0.1$ the true value of the 95% confidence coefficient is 86.5%, and for $N = 32$ and $r = 0.1$ the true value falls to 68%. Thus the dangers of ignoring the possibility of $|r|>0$ are evident.

Miller [34], using different initial assumptions, has also shown the deleterious effect on the t statistic of correlation among the pseudo-jackknife estimates.

Three statistics were selected, therefore, to test for the "approximate" independence of the pseudo-jackknife 2SLS estimates, and for this purpose the pseudo-jackknife estimates were expressed as deviations from their mean, viz:

$$d_{ik} = J_i(\hat{\theta}_k) - J(\hat{\theta}_k), \quad (i = 1,2, \ldots, N)$$

for all k $(k = 1,2, \ldots, K_1 + g)$. The three tests used for this purpose are well known tests for departures from randomness, and a detailed explanation of all three (the Swed-Eisenhart One Sample Runs Test, the Fisher Exact Probability Test, and Spearman's Rank Correlation Coefficient) is given in Siegel [63].

The Swed-Eisenhart test (denoted by SE in Table 7.1) was used to ascertain whether the sequence of signs of the d_{ik} was random. The Fisher test (denoted by FI) was also based on sign sequences. A 2 x 2 contingency table was set-up for each value of k and scores allotted according to the sequence of the signs of successive d_{ik} over the i observations. Spearman's Rank Correlation Coefficient (denoted by SR) was used to test for association between the natural ordering of the d_{ik} and their ranked ordering. All three tests were repeated over all replications.

The problem with using these aforementioned tests is that no general statement can be made about the efficiency of any of them. In the context in which they are used in this study, each of these three tests will produce a different "measure" of randomness. All three reject a certain amount of relevant information and therefore, at best, the test results can only be used as an approximate guide to departures from randomness of the pseudo-jackknife 2SLS estimates.

The number of times the hypothesis of randomness was rejected for each test over the 100 replications is given in Table 7.1. A visual appraisal of the results indicates that the hypothesis of randomness is upheld "approximately" 95% of the time. These results appear to offer some support to Tukey's conjecture for this particular application.

7.6 Validity of Test Statistics

It is essential to examine the validity of the standard tests of significance to ensure that the test statistics do not diverge significantly from their postulated theoretical distribution. To this end, the Kolmogorov-Smirnov One-Sample Test (see e.g. Siegel [63; pp.47-52]) was employed to test five hypotheses:

$$\frac{\hat{\theta}_k - \theta_k}{\hat{\sigma}\sqrt{S_k}} \sim N(0,1) \quad , \qquad (7.10a)$$

$$\frac{\hat{\theta}_k - \theta_k}{\hat{\sigma}\sqrt{S_k}} \sim t_{N-K_1-g} \quad , \qquad (7.10b)$$

$$\frac{\hat{\theta}_k - \theta_k}{\bar{\sigma}\sqrt{S_k}} \sim t_{K_2-g} \quad , \qquad (7.10c)$$

$$\frac{\sqrt{N}\left[J(\hat{\theta}_k)-\theta_k\right]}{\left[(N-1)^{-1}\sum_{i=1}^{N}\left(J_i(\hat{\theta}_k) - J(\hat{\theta}_k)\right)^2\right]^{\frac{1}{2}}} \sim N(0,1) \quad , \qquad (7.10d)$$

and

$$\frac{\sqrt{N}\left[J(\hat{\theta}_k)-\theta_k\right]}{\left[(N-1)^{-1}\sum_{i=1}^{N}\left(J_i(\hat{\theta}_k) - J(\hat{\theta}_k)\right)^2\right]^{\frac{1}{2}}} \sim t_{N-1} \qquad (7.10e)$$

$(k = 1,2,\ldots,K_1 + g)$.

Tables 7.2(a) and 7.2(b) set out the values of the maximum deviation, D, between the relevant empirical and theoretical distributions for each of these five hypotheses. The distributional assumptions are rejected at the 5% level for $D > 0.13403$.

Over all experiments 48 "sets" of values for D were obtained, i.e. 24 sets for each value of δ. The lowest D value in each set was designated "1st", the second lowest "2nd", and so on. Table 7.3 summarizes the number of firsts, seconds, etc., for each test statistic over all parameters and all values of K_2, for $\delta = 0.19$ and for $\delta = 0.76$. The following abbreviations are used:

CT1 - "Conventional Test No. 1", formula (7.10a);

CT2 - "Conventional Test No. 2", formula (7.10b);

DT - "Dhrymes Test", formula (7.10c);

JT1 - "Jackknife Test No. 1", formula (7.10d);

JT2 - "Jackknife Test No. 2", formula (7.10e).

Care must be taken in interpreting these figures, as the postulated theoretical distribution differs across each set.

When 2SLS was biased (i.e. $\delta = 0.19$) the jackknife-based test statistics always dominated the others for β_{12}, and γ_{10}, and for six out of the eight sets of values for γ_{11}. The t statistic based upon the Dhrymes derivation (DT) consistently produced the poorest fit.

Table 7.1: Tests of Independence of Pseudo-Jackknife Estimates
(Number of rejections at 5% level of significance)

$\delta = 0.19$

β_{12}

K_2	λ	FI	SR	SE
3	0.00	4	1	2
	0.45	2	3	2
6	0.00	7	5	5
	0.45	6	3	4
	0.90	1	3	0
9	0.00	5	6	4
	0.45	5	4	2
	0.90	4	5	1

γ_{10}

K_2	λ	FI	SR	SE
3	0.00	3	8	2
	0.45	2	6	1
6	0.00	6	6	7
	0.45	6	5	5
	0.90	6	7	6
9	0.00	5	7	4
	0.45	4	9	2
	0.90	2	6	2

γ_{11}

K_2	λ	FI	SR	SE
3	0.00	6	4	3
	0.45	4	4	0
6	0.00	7	3	4
	0.45	9	4	2
	0.90	5	5	3
9	0.00	2	2	2
	0.45	1	4	1
	0.90	5	4	4

$\delta = 0.76$

β_{12}

K_2	λ	FI	SR	SE
3	0.00	9	6	7
	0.45	5	3	5
6	0.00	4	7	4
	0.45	3	4	3
	0.90	3	7	3
9	0.00	4	10	4
	0.45	4	4	3
	0.90	7	8	2

γ_{10}

K_2	λ	FI	SR	SE
3	0.00	2	7	3
	0.45	2	5	1
6	0.00	1	7	2
	0.45	1	7	4
	0.90	3	7	2
9	0.00	3	9	1
	0.45	5	8	3
	0.90	2	5	2

γ_{11}

K_2	λ	FI	SR	SE
3	0.00	1	4	0
	0.45	5	3	3
6	0.00	4	3	2
	0.45	4	5	3
	0.90	3	4	2
9	0.00	6	2	4
	0.45	3	5	4
	0.90	2	3	2

Table 7.2(a): Kolmogorov-Smirnov D Statistic

	K_2	λ	2SLS Normal	2SLS t	Dhrymes t	J2SLS Normal	J2SLS t
β_{12}	3	0.00	0.1336	0.1329	0.1342	0.0929	0.0941
		0.45	0.1478	0.1471	0.1406	0.0878	0.0871
	6	0.00	0.1454	0.1429	0.1290	0.0912	0.0944
		0.45	0.1888	0.1853	0.1743	0.1014	0.1055
		0.90	0.3608	0.3578	0.3410	0.1381	0.1385
	9	0.00	0.1406	0.1384	0.1546	0.0900	0.0925
		0.45	0.1996	0.1987	0.2039	0.1009	0.1032
		0.90	0.3965	0.3896	0.4000	0.1589	0.1530
γ_{10}	3	0.00	0.1062	0.1068	0.1166	0.0907	0.0918
		0.45	0.1115	0.1083	0.1047	0.0952	0.0965
	6	0.00	0.1085	0.1051	0.1206	0.0439	0.0455
		0.45	0.1222	0.1229	0.1199	0.0601	0.0603
		0.90	0.1560	0.1561	0.1562	0.1052	0.1065
	9	0.00	0.1224	0.1190	0.1300	0.0532	0.0551
		0.45	0.1340	0.1343	0.1345	0.0540	0.0543
		0.90	0.1968	0.1947	0.1892	0.1040	0.1039
γ_{11}	3	0.00	0.0996	0.0987	0.1066	0.0680	0.0682
		0.45	0.0778	0.0771	0.0969	0.0783	0.0790
	6	0.00	0.0893	0.0881	0.0901	0.0851	0.0893
		0.45	0.1225	0.1235	0.1121	0.0832	0.0851
		0.90	0.1087	0.1082	0.1183	0.1200	0.1200
	9	0.00	0.0957	0.0948	0.1156	0.0790	0.0840
		0.45	0.1060	0.1061	0.1086	0.0843	0.0881
		0.90	0.0761	0.0764	0.0768	0.0780	0.0779

Sample size = 20 $\delta = 0.19$

Table 7.2(b): Kolmogorov-Smirnov D Statistic

	K_2	λ	2SLS Normal	2SLS t	Dhrymes t	J2SLS Normal	J2SLS t
β_{12}	3	0.00	0.1206	0.1184	0.0792	0.0790	0.0863
		0.45	0.0710	0.0696	0.0623	0.0967	0.1023
	6	0.00	0.0424	0.0391	0.0573	0.0477	0.0525
		0.45	0.0576	0.0617	0.0474	0.0489	0.0532
		0.90	0.0761	0.0818	0.0754	0.0594	0.0614
	9	0.00	0.0843	0.0820	0.0787	0.0650	0.0620
		0.45	0.0654	0.0629	0.0551	0.0778	0.0724
		0.90	0.0558	0.0607	0.0530	0.1252	0.1204
γ_{10}	3	0.00	0.0683	0.0707	0.0971	0.0501	0.0513
		0.45	0.0798	0.0792	0.0747	0.0594	0.0614
	6	0.00	0.0764	0.0748	0.0809	0.0497	0.0513
		0.45	0.0683	0.0702	0.0709	0.0829	0.0841
		0.90	0.0602	0.0608	0.0583	0.0494	0.0549
	9	0.00	0.0693	0.0681	0.0747	0.0571	0.0559
		0.45	0.0613	0.0628	0.0611	0.0383	0.0400
		0.90	0.0500	0.0515	0.0500	0.0511	0.0570
γ_{11}	3	0.00	0.0536	0.0596	0.0633	0.0777	0.0811
		0.45	0.0799	0.0796	0.0706	0.0876	0.0913
	6	0.00	0.0523	0.0598	0.0545	0.0905	0.0957
		0.45	0.0629	0.0642	0.0584	0.0843	0.0877
		0.90	0.0646	0.0681	0.0527	0.0751	0.0793
	9	0.00	0.0546	0.0620	0.0508	0.0827	0.0871
		0.45	0.0564	0.0592	0.0493	0.0746	0.0786
		0.90	0.0536	0.0588	0.0505	0.0705	0.0757

Sample size = 20 $\delta = 0.76$

Table 7.3: Ranking of D Statistic over the Five Tests of Significance

$\delta = 0.19$

RANK	1st*	2nd*	3rd*	4th*	5th
CT1	1	1	5	11	6
CT2	1	2	9	10	2
DT	0	0	9	0	15
JT1	19	3	1	0	1
JT2	4	17	1	2	0

$\delta = 0.76$

RANK	1st*	2nd*	3rd	4th	5th
CT1	4	7	4	6	3
CT2	1	5	9	5	4
DT	11	2	7	0	4
JT1	7	3	2	10	2
JT2	2	6	2	3	11

Sample size = 20

* Denotes that column total does not sum to 24 because of ties (to 4 decimal places).

For 2SLS unbiased, however, the superiority of the jackknife-based test statistics was less marked. This was particularly noticeable for γ_{11} where the two jackknife-based test statistics always produced the poorest fit.

The number of rejections, at the 5% level of significance, of the hypothesis that each sample was drawn from the specified theoretical distribution is given in Table 7.4 In any one cell the total possible number of rejections is 24; percentages of rejections are given next to the absolute figures.

Table 7.4: Number of Rejections of the Null Hypothesis

δ =	0.19		0.76	
CT1	9	37.5%	0	-
CT2	10	41.7%	0	-
DT	10	41.7%	0	-
JT1	2	8.3%	0	-
JT2	2	8.3%	0	-

Sample size = 20

Clearly when $\delta = 0.19$ the distribution of the t statistic formed using the 2SLS estimator gives a poor approximation to both Student's t distribution and the standardized normal distribution. Thus if the bias of the 2SLS estimator is significantly different from zero, the distribution of 2SLS-based test statistics may be a poor approximation to their postulated theoretical distributions.

7.7 Inference

7.7.1 Tests of Significance

In the preceding section it was shown that the distributions of expressions (7.10a), (7.10b) and (7.10c) show a substantial divergence from their postulated theoretical distributions when $\delta = 0.19$, even for relatively large values of μ^2. It is important to ascertain the effect of this divergence on statistical inference.

In this section we consider the degree of accuracy afforded by using the relevant theoretical distributions as approximations for making statistical inference.

The hypotheses that the biases of the 2SLS and J2SLS estimators were not significantly different from zero were tested at both the 5% and 10% levels of significance. The proportion of samples falling in the .05 and .95 percentiles of the relevant theoretical distributions are given in Tables 7.5, 7.6 and 7.7. These tables are further divided into parts (a) and (b), the former for results when $\delta = 0.19$, the latter for $\delta = 0.76$.

In these tables each cell contains three values. The number of "rejections" are tabulated according to whether they were rejected in the lower or upper tail of the relevant distribution, and are given by the figures in parentheses on the left and right respectively at the top of each cell. The total number of "rejections" is given below these two figures.

For the parameter β_{12}, both JT1 and JT2 show a number of "rejections" nearer the nominal level of significance than CT1 and CT2 in just over half of the experiments for $\delta = 0.19$. There is little to choose between these two jackknife-based test statistics, although JT2 (i.e. the t statistic given by formula (7.10e)) was marginally closer to the nominal level of significance for $K_2 = 6$ and 9 and $\lambda = 0.45$ and 0.90. CT2 is to be preferred to CT1 as the number of "rejections" were, in general, nearer the nominal level of significance. Using the same criterion, CT2 is to be preferred to JT1 but not to JT2.

Dhrymes' test statistic (DT) gave a similar pattern of "rejections" to the other t statistics, but it should be noted that approximate confidence intervals using DT will be much wider than those using either CT2 or JT2.

The striking feature about these results however is the distribution of "rejections" between the tails of the relevant distributions. The downward bias of the 2SLS estimator of β_{12} ensured that virtually all rejections for CT1, CT2 and DT fell in the lower tail, this being most noticeable when K_2 was relatively large.

The constant term, γ_{10}, gave a fairly even spread of "rejections" between the tails for all tests, whereas γ_{11} showed a similar, but less marked, pattern to that for β_{12}.

For all three parameters, the three t statistics (CT2, DT and JT2) are to be preferred to those tests based on the normal distribution, although this preference is most marked for β_{12}.

The skewness of the foregoing statistics, which is particularly noticeable for the 2SLS-based statistics, can have important consequences when the postulated distributions are used as a basis for constructing approximate critical regions for one-sided tests of hypotheses. From Tables 7.5(a), 7.6(a) and 7.7(a), it can be seen that if the lower tail of the CT1, CT2 and DT distributions is used to construct an approximate test for β_{12}, the estimate of the level of significance is generally considerably higher than the postulated level of either 2.5% or 5%, i.e. the level of significance is underestimated. Conversely, if the upper tail is used then the level of significance will be overestimated. Moreover, in general, the degree of error is larger the higher the level of multicollinearity and the greater the degree of overidentification.

By comparison, test statistics for β_{12} based on the jackknife statistics JT1 and JT2 give a more even spread of rejections and consequently a smaller error of over- or under-estimation when performing one-sided tests of hypothesis. Even here, however, for large λ and $K_2 = 6$ or $K_2 = 9$ the lower tail was markedly larger than its nominal level, although generally very much less than for CT1, CT2 or DT.

For $\delta = 0.76$ all five tests generally only differ to a small degree over the three parameters, although for β_{12}, JT1, JT2 and DT tend to overestimate the total nominal level of significance in both tails by approximately the same margin as CT1 and CT2 tend to underestimate it. For $K_2 = 9$ and $\lambda = 0.9$ the jackknife-based tests produced a "wayward" result.

When $\delta = 0.76$ the 2SLS estimator is not only unbiased, but the odd order moments (those that exist) of both CT2 and DT are zero (see Section 7.2) in the model under consideration. Thus selecting $\delta = 0.76$ has not only removed the location problem but also the problem of skewness of the distribution of CT2 and DT, provided that the first three moments exist (which they do for $K_2 = 6$ and $K_2 = 9$). It is hardly surprising, therefore, that the jackknife-based test statistics cannot show superiority over CT1, CT2 and DT under such circumstances.

From the preceding results it can be concluded that the t statistic based on the J2SLS estimator (JT2) will, in general, produce confidence intervals which are at least as accurate as those produced using test statistics based on the 2SLS estimator.

7.7.2 Power of the Tests

Finally we consider the power of the alternative tests under the hypothesis that $\beta_{12} = \beta_{12}^*$, where β_{12}^* was specified to be 0.8.

Tables 7.8 (a-c) present power functions for the five tests when $\delta = 0.19$. The computational expense involved in computing power functions prohibited further calculations. The significance level for all tests was 5%.

Table 7.5(a): Distribution of Rejections of Hypothesis that the Bias of the Relevant Estimator is not Significantly Different from Zero.

PARAMETER: β_{12}

K_2	λ	CT1	CT2	DT	JT1	JT2
3	0.00	(3) (1) 4	(2) (1) 3	(3) (2) 5	(2) (2) 4	(2) (2) 4
	0.45	(4) (0) 4	(3) (0) 3	(1) (3) 4	(2) (2) 4	(2) (1) 3
6	0.00	(6) (2) 8	(6) (2) 8	(5) (3) 8	(5) (2) 7	(5) (2) 7
	0.45	(6) (2) 8	(4) (1) 5	(5) (3) 8	(6) (2) 8	(3) (2) 5
	0.90	(15) (0) 15	(12) (0) 12	(7) (3) 10	(8) (5) 13	(6) (3) 9
9	0.00	(10) (2) 12	(7) (1) 8	(6) (2) 8	(5) (4) 9	(5) (3) 8
	0.45	(7) (2) 9	(7) (1) 8	(6) (2) 8	(4) (3) 7	(4) (3) 7
	0.90	(19) (0) 19	(17) (0) 17	(16) (0) 16	(13) (7) 20	(11) (7) 18

$\alpha = 5\%$ $\delta = 0.19$

PARAMETER: β_{12}

K_2	λ	CT1	CT2	DT	JT1	JT2
3	0.00	(8) (4) 12	(5) (4) 9	(5) (3) 8	(2) (2) 4	(2) (2) 4
	0.45	(8) (3) 11	(7) (1) 8	(4) (5) 9	(4) (3) 7	(2) (2) 4
6	0.00	(8) (3) 11	(6) (3) 9	(7) (4) 11	(6) (2) 8	(6) (2) 8
	0.45	(11) (4) 15	(9) (3) 12	(5) (4) 9	(6) (5) 11	(6) (4) 10
	0.90	(21) (1) 22	(2) (0) 20	(14) (3) 17	(11) (7) 18	(11) (6) 17
9	0.00	(10) (2) 12	(10) (1) 11	(10) (2) 12	(6) (6) 12	(6) (5) 11
	0.45	(10) (2) 12	(9) (2) 11	(10) (3) 13	(6) (8) 14	(6) (6) 12
	0.90	(26) (0) 26	(23) (0) 23	(23) (0) 23	(17) (10) 27	(16) (8) 24

$\alpha = 10\%$ $\delta = 0.19$

Sample size = 20

Table 7.5(b) : Distribution of Rejections of Hypothesis that the Bias of the Relevant Estimator is not Significantly Different from Zero.

PARAMETER: β_{12}

K_2	λ	CT1	CT2	DT	JT1	JT2
3	0.00	(1) (3) 4	(1) (1) 2	(3) (3) 6	(1) (2) 3	(1) (2) 3
	0.45	(2) (1) 3	(1) (0) 1	(2) (3) 5	(2) (2) 4	(1) (1) 2
6	0.00	(4) (2) 6	(2) (2) 4	(4) (4) 8	(4) (1) 5	(3) (1) 4
	0.45	(4) (2) 6	(2) (2) 4	(3) (2) 5	(4) (3) 7	(3) (1) 4
	0.90	(3) (1) 4	(2) (0) 2	(2) (4) 6	(4) (4) 8	(4) (4) 8
9	0.00	(5) (3) 8	(4) (3) 7	(2) (2) 4	(3) (6) 9	(3) (4) 7
	0.45	(4) (4) 8	(2) (2) 4	(1) (3) 4	(3) (5) 8	(3) (5) 8
	0.90	(3) (3) 6	(3) (2) 5	(4) (4) 8	(5) (8) 13	(3) (6) 9

α = 5% δ = 0.76

PARAMETER: β_{12}

K_2	λ	CT1	CT2	DT	JT1	JT2
3	0.00	(3) (6) 9	(3) (5) 8	(7) (6) 13	(3) (4) 7	(1) (2) 3
	0.45	(4) (3) 7	(4) (2) 6	(4) (6) 10	(3) (4) 7	(2) (3) 5
6	0.00	(7) (4) 11	(6) (3) 9	(6) (5) 11	(5) (5) 10	(5) (4) 9
	0.45	(6) (4) 10	(5) (3) 8	(5) (5) 10	(6) (6) 12	(5) (5) 10
	0.90	(5) (2) 7	(5) (2) 7	(4) (4) 8	(5) (7) 12	(5) (7) 12
9	0.00	(5) (4) 9	(5) (4) 9	(7) (4) 11	(6) (7) 13	(6) (7) 13
	0.45	(5) (4) 9	(5) (4) 9	(6) (6) 12	(6) (5) 11	(5) (5) 10
	0.90	(4) (5) 9	(4) (5) 9	(5) (6) 11	(9) (12) 21	(7) (9) 16

α = 10% δ = 0.76

Sample size = 20

Table 7.6(a) : Distribution of Rejections of Hypothesis that the Bias of the Relevant Estimator is not Significantly Different from Zero.

PARAMETER: γ_{10}

K_2	λ	CT1	CT2	DT	JT1	JT2
3	0.00	(3) (4) 7	(2) (4) 6	(4) (3) 7	(1) (4) 5	(1) (4) 5
	0.45	(3) (4) 7	(2) (4) 6	(3) (3) 6	(1) (4) 5	(1) (4) 5
6	0.00	(2) (5) 7	(1) (4) 5	(2) (4) 6	(1) (6) 7	(1) (5) 6
	0.45	(3) (4) 7	(1) (4) 5	(2) (3) 5	(1) (7) 8	(1) (4) 5
	0.90	(1) (5) 6	(1) (5) 6	(1) (3) 4	(2) (6) 8	(2) (4) 6
9	0.00	(3) (5) 8	(3) (4) 7	(1) (4) 5	(2) (4) 6	(1) (4) 5
	0.45	(3) (4) 7	(2) (4) 6	(1) (5) 6	(2) (4) 6	(0) (4) 4
	0.90	(1) (5) 6	(1) (5) 6	(1) (3) 4	(6) (6) 12	(5) (4) 9

α = 5% δ = 0.19

PARAMETER: γ_{10}

K_2	λ	CT1	CT2	DT	JT1	JT2
3	0.00	(5) (8) 13	(4) (4) 8	(4) (5) 9	(2) (7) 9	(2) (6) 8
	0.45	(5) (7) 12	(4) (7) 11	(3) (7) 10	(2) (8) 10	(2) (8) 10
6	0.00	(5) (8) 13	(4) (6) 10	(2) (8) 10	(3) (7) 10	(3) (7) 10
	0.45	(5) (9) 14	(4) (7) 11	(4) (8) 12	(3) (7) 10	(3) (7) 10
	0.90	(4) (10) 14	(3) (8) 11	(4) (8) 12	(4) (11) 15	(3) (10) 13
9	0.00	(5) (5) 10	(4) (5) 9	(4) (8) 12	(4) (8) 12	(3) (7) 10
	0.45	(6) (6) 12	(5) (5) 10	(4) (7) 11	(5) (7) 12	(4) (6) 10
	0.90	(5) (5) 10	(3) (4) 7	(6) (5) 11	(6) (6) 12	(6) (5) 11

α = 10% δ = 0.19

Sample size = 20

Table 7.6(b) : Distribution of Rejections of Hypothesis that the Bias of the Relevant Estimator is not Significantly Different from Zero.

PARAMETER : γ_{10}

K_2	λ	CT1	CT2	DT	JT1	JT2
3	0.00	(1) (4) 5	(1) (3) 4	(4) (4) 8	(2) (4) 6	(0) (3) 3
	0.45	(2) (4) 6	(1) (3) 4	(3) (5) 8	(3) (4) 7	(1) (4) 5
6	0.00	(1) (4) 5	(1) (4) 5	(2) (4) 6	(1) (5) 6	(0) (4) 4
	0.45	(1) (4) 5	(1) (4) 5	(2) (4) 6	(2) (7) 9	(0) (4) 4
	0.90	(2) (4) 6	(1) (4) 5	(2) (3) 5	(2) (5) 7	(1) (3) 4
9	0.00	(3) (4) 7	(2) (3) 5	(1) (2) 3	(1) (5) 6	(1) (3) 4
	0.45	(2) (4) 6	(1) (3) 4	(1) (2) 3	(2) (5) 7	(1) (4) 5
	0.90	(3) (3) 6	(1) (3) 4	(2) (2) 4	(4) (5) 9	(3) (4) 7

α = 5% δ = 0.76

PARAMETER: γ_{10}

K_2	λ	CT1	CT2	DT	JT1	JT2
3	0.00	(5) (4) 9	(4) (4) 8	(6) (7) 13	(3) (5) 8	(3) (4) 7
	0.45	(4) (4) 8	(4) (4) 8	(5) (9) 14	(3) (5) 8	(3) (4) 7
6	0.00	(5) (6) 11	(3) (5) 8	(4) (7) 11	(3) (8) 11	(3) (8) 11
	0.45	(6) (6) 12	(3) (5) 8	(4) (7) 11	(3) (7) 10	(2) (7) 9
	0.90	(3) (5) 8	(3) (4) 7	(3) (7) 10	(5) (6) 11	(4) (5) 9
9	0.00	(5) (5) 10	(4) (5) 9	(4) (8) 12	(4) (8) 12	(3) (7) 10
	0.45	(6) (6) 12	(5) (5) 10	(4) (7) 11	(5) (7) 12	(4) (6) 10
	0.90	(5) (5) 10	(3) (4) 7	(6) (5) 11	(6) (6) 12	(6) (5) 11

α = 10% δ = 0.76

Sample size = 20

Table 7.7(a) : Distribution of Rejections of Hypothesis that the Bias of the Relevant Estimator is not Significantly Different from Zero.

PARAMETER : γ_{11}

K_2	λ	CT1	CT2	DT	JT1	JT2
3	0.00	(3) (1) 4	(2) (1) 3	(4) (2) 6	(5) (1) 6	(3) (1) 4
	0.45	(2) (1) 3	(2) (1) 3	(4) (1) 5	(3) (2) 5	(3) (1) 4
6	0.00	(5) (2) 7	(4) (2) 6	(1) (2) 3	(4) (2) 6	(3) (2) 5
	0.45	(3) (2) 5	(3) (1) 4	(1) (3) 4	(3) (2) 5	(3) (1) 4
	0.90	(4) (1) 5	(3) (1) 4	(2) (2) 4	(5) (2) 7	(4) (2) 6
9	0.00	(4) (2) 6	(3) (2) 5	(2) (1) 3	(4) (3) 7	(3) (2) 5
	0.45	(4) (2) 6	(2) (2) 4	(4) (1) 5	(4) (3) 7	(4) (1) 5
	0.90	(3) (1) 4	(3) (1) 4	(3) (1) 4	(4) (6) 10	(3) (4) 7

α = 5% δ = 0.19

PARAMETER: γ_{11}

K_2	λ	CT1	CT2	DT	JT1	JT2
3	0.00	(5) (4) 9	(5) (4) 9	(6) (6) 12	(5) (2) 7	(5) (2) 7
	0.45	(4) (5) 9	(4) (2) 6	(5) (5) 10	(4) (4) 8	(4) (2) 6
6	0.00	(7) (3) 10	(5) (2) 7	(8) (4) 12	(7) (5) 12	(5) (4) 9
	0.45	(6) (3) 9	(6) (2) 8	(6) (4) 10	(5) (5) 10	(4) (5) 9
	0.90	(7) (2) 9	(7) (1) 8	(6) (3) 9	(8) (5) 13	(6) (4) 10
9	0.00	(5) (5) 10	(4) (3) 7	(6) (6) 12	(5) (4) 9	(5) (4) 9
	0.45	(5) (4) 9	(4) (3) 7	(5) (3) 8	(5) (6) 11	(5) (4) 9
	0.90	(8) (4) 12	(4) (2) 6	(8) (3) 11	(5) (8) 13	(5) (7) 12

α = 10% δ = 0.19

Sample size = 20

Table 7.7(b) : Distribution of Rejections of Hypothesis that the Bias of the Relevant Estimator is not Significantly Different from Zero.

PARAMETER : γ_{11}

K_2	λ	CT1	CT2	DT	JT1	JT2
3	0.00	(3) (2) 5	(2) (1) 3	(4) (4) 8	(3) (2) 5	(3) (1) 4
	0.45	(2) (1) 3	(2) (1) 3	(4) (3) 7	(3) (1) 4	(3) (1) 4
6	0.00	(3) (3) 6	(2) (2) 4	(0) (2) 2	(4) (4) 8	(3) (3) 6
	0.45	(3) (2) 5	(2) (2) 4	(2) (3) 5	(3) (3) 6	(3) (3) 6
	0.90	(2) (1) 3	(2) (1) 3	(3) (2) 5	(5) (2) 7	(3) (2) 5
9	0.00	(3) (3) 6	(3) (3) 6	(2) (3) 5	(4) (4) 8	(4) (3) 7
	0.45	(4) (3) 7	(2) (3) 5	(3) (2) 5	(4) (3) 7	(4) (3) 7
	0.90	(2) (2) 4	(2) (1) 3	(2) (4) 6	(3) (3) 6	(2) (3) 5

α = 5% δ = 0.76

PARAMETER : γ_{11}

K_2	λ	CT1	CT2	DT	JT1	JT2
3	0.00	(5) (5) 10	(5) (5) 10	(5) (6) 11	(5) (5) 10	(5) (5) 10
	0.45	(4) (4) 8	(4) (3) 7	(7) (5) 12	(4) (4) 8	(4) (4) 8
6	0.00	(6) (5) 11	(6) (3) 9	(4) (4) 8	(7) (5) 12	(6) (4) 10
	0.45	(6) (5) 11	(6) (4) 10	(4) (4) 8	(5) (6) 11	(5) (4) 9
	0.90	(6) (3) 9	(6) (3) 9	(5) (3) 8	(6) (5) 11	(5) (3) 8
9	0.00	(4) (6) 10	(4) (4) 8	(4) (5) 9	(4) (4) 8	(4) (4) 8
	0.45	(4) (6) 10	(4) (4) 8	(5) (7) 12	(4) (5) 9	(4) (4) 8
	0.90	(4) (5) 9	(4) (4) 8	(4) (7) 11	(7) (5) 12	(7) (4) 11

α = 10% δ = 0.76

Sample size = 20

For each value of β_{12}, ranging from 0.0 to 1.6 in steps of 0.2, a new set of 100 replications was generated and the power of each test was evaluated.

When $\beta_{12} = 0.8$, the power reported in the tables is of course equivalent to the probability of a Type I error.

Strictly speaking the term "power" is not appropriate as one cannot compare the powers of a number of tests when the probability of Type I errors are clearly not equal. Perhaps "probabilities of rejection" would be a more appropriate term.

Tests based on the standardized normal distribution (CT1 and JT1) showed greater "power" than their counterparts based on Student's t distribution, although we have already noted that the former produce higher Type I errors. Of the two tests based on the standardized normal distribution, CT1 generally had higher "power" than JT1 except for "large" K_2(6 or 9) and high levels of multicollinearity ($\lambda = 0.9$). A similar pattern was evident for comparisons of "power" between CT2 and JT2. On the other hand, the Type I errors associated with CT1 and CT2 were often greater than those of JT1 and JT2 (which themselves were generally greater than the nominal level of significance).

Dhrymes' test (DT) consistently exhibits the lowest "power" of the five tests, a result also noted by Maddala [26], although the Type I errors associated with this test are frequently nearer the nominal level than those associated with the other tests.

As one would expect, high levels of multicollinearity reduce the "power" of all five tests.

In conclusion, CT1 and CT2 dominate JT1 and JT2 respectively (i.e. they have higher "probabilities of rejection") although rarely over the entire range of values of β_{12}. This superiority however will be offset by the lower Type I errors which JT1 and JT2 frequently exhibit. No definitive statement can be made, therefore, concerning the relative powers of these four tests. The substantially lower "power" which is generally exhibited by DT suggests that this test is not a practical proposition, despite its accuracy for estimating the level of significance.

Table 7.8(a) Power of the Test Statistics

λ	β_{12}=	0.0	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6
0.00	CT1	1.00	0.92	0.62	0.11	0.04	0.30	0.70	0.87	0.96
	CT2	0.99	0.91	0.52	0.05	0.03	0.24	0.68	0.86	0.95
	DT	0.52	0.27	0.15	0.07	0.05	0.14	0.38	0.43	0.71
	JT1	0.96	0.82	0.48	0.11	0.04	0.19	0.62	0.77	0.92
	JT2	0.95	0.79	0.44	0.09	0.04	0.18	0.60	0.77	0.90

0.45	CT1	0.96	0.82	0.43	0.06	0.04	0.24	0.65	0.77	0.88
	CT2	0.93	0.75	0.36	0.03	0.03	0.21	0.62	0.74	0.85
	DT	0.40	0.29	0.13	0.05	0.04	0.12	0.33	0.38	0.62
	JT1	0.93	0.68	0.37	0.10	0.04	0.18	0.55	0.64	0.85
	JT2	0.89	0.61	0.30	0.07	0.03	0.15	0.52	0.59	0.82

$K_2 = 3$ $N = 20$ $\delta = 0.19$

Table 7.8(b): Power of the Test Statistics

λ	β_{12}=	0.0	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6
0.00	CT1	1.00	1.00	0.90	0.35	0.08	0.50	0.94	1.00	1.00
	CT2	1.00	0.99	0.88	0.29	0.08	0.47	0.93	1.00	1.00
	DT	1.00	0.98	0.80	0.23	0.08	0.37	0.90	0.97	1.00
	JT1	1.00	0.99	0.83	0.32	0.07	0.25	0.77	0.97	0.99
	JT2	1.00	0.98	0.78	0.31	0.07	0.24	0.74	0.95	0.99
0.45	CT1	1.00	0.96	0.70	0.14	0.08	0.45	0.82	1.00	1.00
	CT2	1.00	0.95	0.65	0.12	0.05	0.41	0.79	0.98	0.98
	DT	0.99	0.87	0.54	0.11	0.08	0.31	0.69	0.91	0.99
	JT1	1.00	0.88	0.70	0.20	0.08	0.28	0.65	0.85	0.94
	JT2	1.00	0.85	0.64	0.15	0.05	0.28	0.60	0.85	0.94
0.90	CT1	0.48	0.19	0.08	0.04	0.15	0.30	0.58	0.76	0.85
	CT2	0.41	0.15	0.04	0.03	0.12	0.24	0.50	0.71	0.84
	DT	0.35	0.14	0.08	0.06	0.10	0.24	0.37	0.56	0.74
	JT1	0.45	0.28	0.16	0.12	0.13	0.19	0.27	0.47	0.64
	JT2	0.45	0.26	0.13	0.09	0.09	0.16	0.24	0.38	0.60

$K_2 = 6$ $N = 20$ $\delta = 0.19$

Table 7.8(c): Power of the Test Statistics

λ	$\beta_{12}=$	0.0	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6
0.00	CT1	1.00	1.00	0.95	0.26	0.12	0.73	0.97	1.00	1.00
	CT2	1.00	1.00	0.94	0.21	0.08	0.63	0.95	1.00	1.00
	DT	1.00	1.00	0.92	0.15	0.08	0.58	0.96	1.00	1.00
	JT1	1.00	1.00	0.92	0.29	0.09	0.35	0.87	1.00	1.00
	JT2	1.00	1.00	0.91	0.25	0.08	0.32	0.86	0.99	1.00
0.45	CT1	1.00	0.98	0.76	0.11	0.09	0.59	0.88	0.99	0.99
	CT2	1.00	0.98	0.72	0.08	0.08	0.55	0.84	0.99	0.99
	DT	1.00	0.93	0.62	0.07	0.08	0.44	0.82	0.97	0.99
	JT1	1.00	0.91	0.73	0.16	0.07	0.28	0.75	0.94	0.96
	JT2	1.00	0.89	0.72	0.15	0.07	0.27	0.74	0.90	0.96
0.90	CT1	0.36	0.17	0.07	0.10	0.19	0.48	0.75	0.88	0.96
	CT2	0.27	0.15	0.05	0.08	0.17	0.44	0.74	0.87	0.96
	DT	0.20	0.17	0.05	0.07	0.16	0.36	0.68	0.81	0.96
	JT1	0.41	0.34	0.13	0.11	0.20	0.29	0.56	0.60	0.75
	JT2	0.36	0.31	0.12	0.09	0.18	0.27	0.53	0.57	0.72

$K_2 = 9$ $N = 20$ $\delta = 0.19$

CHAPTER 8

CONCLUSION

8.1 General Remarks

In Chapter 3 it was shown that, asymptotically, the 2SLS and J2SLS estimators are equivalent. Thus, one would expect the MSEs of the two estimators not to be significantly different from each other for "large" values of μ^2. If this were indeed so, the superiority of the jackknife technique for constructing confidence intervals and performing tests of significance would justify its use in applied economics.

From the preceding Monte Carlo study it is evident that the jackknife technique, whilst reducing the bias of the 2SLS estimator is not to be recommended for "small" μ^2 if the criterion for selection of an estimator is either minimum MSE or MAE. For "large" values of μ^2 there was little difference between the MSEs and MAEs of the 2SLS and J2SLS estimators, whilst the Wilcoxon Matched Pairs Signed Ranks test indicated significant differences between the two estimators only for small μ^2.

It was then observed (Chapter 7) that t(and z) statistics formed using the 2SLS estimator were not distributed according to the Student t or standardized normal distributions when $\delta = 0.19$. The actual distributions are highly skewed and serious errors could result if these postulated distributions were used for statistical inference. In general, this feature was less noticeable for the J2SLS estimator which, on the basis of Kolmogorov Smirnov tests, appears to have a reasonably symmetric distribution, and consequently there is less likelihood of serious errors being made if the postulated theoretical distributions are used for the purpose of statistical inference.

Even under "ideal" conditions (i.e. $\delta = 0.76$), test statistics based on the 2SLS estimator cannot show superior (to J2SLS) fit to their postulated theoretical distributions for the parameter β_{12}.

Finally, the "power" functions of the alternative tests were calculated over a range of values for β_{12}. The problem involved in comparing the "power" of two or more statistics when the Type I errors are not equal was recognized, but even making allowance for this problem Dhrymes' t statistic showed considerably lower "power" than the other statistics considered. This latter result confirms Maddala's [26] conclusions.

Clearly, therefore, a decision on circumstances under which application of the jackknife would be fruitful, hinges on one's definition of "large" in the context of the concentration parameter, μ^2.

8.2 When is the Concentration Parameter "Large"?

Whilst selection by "informed guesswork" of a value of μ^2 which could be taken as "large" is a somewhat haphazard procedure, two other problems of greater magnitude present themselves:

(i) can a value of the concentration parameter which is designated as "large" for an equation containing just two endogenous variables also be designated as "large" for an equation containing three (or more) endogenous variables?

(ii) how can the value of the concentration parameter be calculated?

To date, most of our knowledge concerning μ^2 is in the context of an equation containing just two endogenous variables, but preliminary work by Richardson and Rohr [54] appears to indicate that a value of μ^2 which is considered "large" in the context of an equation containing two endogenous variables <u>may</u> be "small" in the context of an equation containing three endogenous variables.

With regard to the second problem Rohr [57] has proposed that μ^2 be estimated from the sample and that this value be used to indicate whether μ^2 was "large" or "small" (he was interested in determining if μ^2 was large enough to enable the limiting distribution function (Student's t distribution) to be used as an approximation to the conventional t statistic without involving appreciable error).

Unfortunately, in the absence of knowledge of the sampling distribution of μ^2, when σ_{22} and $\underline{\pi}_{22}$ are replaced by their estimated values it would not be possible to obtain any measure of the reliability (i.e. the sampling variance) of our estimate. It should also be noted that there would be a conflict regarding the optimal method for estimating σ_{22} and $\underline{\pi}_{22}$. The Unrestricted Least Squares estimator would, intuitively, seem to be inefficient relative to the 2SLS induced Restricted Reduced Form estimator (although Dhrymes [13] has shown that, asymptotically, this may not be so), but the latter estimator may not possess moments of any order (see McCarthy [32]).

Clearly, therefore, considerably more knowledge concerning both the distribution of μ^2 and the properties of reduced form estimators is required before Rohr's [57] proposal can be properly evaluated.

8.3 <u>Extension of the Results</u>

The Monte Carlo experiments did not investigate the effects of an increase in sample size on the two estimators, although a proof that both the bias and the MSE of the 2SLS estimator are monotonically non-increasing functions of the sample size was given in Chapter 5. As the sample size increases, other variables being constant, the concentration parameter will, in general, increase in size and hence one would expect the MSE of the J2SLS estimator to tend towards (perhaps not monotonically) that of the 2SLS estimator. Conversely, a decrease in sample size might be expected to have the opposite effect on the J2SLS estimator.

The estimation of "large" (e.g. economy-wide) models may present a problem if use of the J2SLS estimator is contemplated. In such circumstances, the computing time and storage requirements will increase more rapidly for J2SLS than for 2SLS as the size of the model increases.

It is unlikely however that 2SLS (and hence J2SLS) would be a feasible proposition anyway in large models, since it is probable that K would exceed N and consequently 2SLS would degenerate to OLS (see Fisher and Wadycki [15]). The jackknife could however be applied to an Instrumental Variables estimator which only considered a sub-set of the excluded predetermined variables when estimating any one structural equation, thus ensuring that $K < N$. Although such a procedure may yield inconsistent (perhaps of a minor nature) parameter estimates and would thus contravene Quenouille's original assumption that a consistent estimator is necessary for the jackknife to be successfully applied, Brundy and Jorgenson [8] cite conditions under which Instrumental Variables estimators based on sub-sets of the predetermined variables retain the property of consistency.

8.4 Extension to Three-Stage Least Squares

The foregoing analysis suggests that an extension of the jackknife technique to the Three-Stage Least Squares (3SLS) estimator may be an extremely tedious procedure. Having obtained 2SLS estimates of all structural coefficients in the system, the 3SLS estimator can be calculated by applying Generalized Least Squares to the entire system (where the equations are written in stacked form) to obtain

$$\left(Z'\left[\Omega^{-1} \otimes X(X'X)^{-1}X'\right]^{-1} Z\right)^{-1} Z'\left[\Omega^{-1} \otimes X(X'X)^{-1}X'\right]y, \qquad (8.1)$$

where $\otimes$ denotes the Kronecker product.

In general, Ω will be unknown and must be replaced by $\hat{\Omega}$, the matrix of mean squares and products of the 2SLS residuals. With Ω replaced by $\hat{\Omega}$ we obtained the 3SLS estimator.

If the jackknife were applied to the 3SLS estimator, Ω would have to be replaced by the matrix of mean squares and products of the J2SLS residuals, and (8.1) would have to be estimated N times with the ith observation omitted at each (of the N) replications.

It is the author's contention that this would not be a very fruitful exercise, especially as no exact results on the moments of the 3SLS estimator are available to provide an exact analysis of the jackknife's bias reducing potential. In addition, it is unlikely that the "simplifying" formula developed for J2SLS could be extended to J3SLS without considerable difficulty and, even then, the additional (to 3SLS) computer run-time involved would probably be substantial.

8.5 The Final Word

In this thesis we have demonstrated the value of the jackknife statistic for forming "accurate" confidence intervals and tests of significance when μ^2 is "large". The bias reducing property of the jackknife is generally present in the context of the 2SLS estimator, although it would certainly be unwise to jackknife the 2SLS estimator if the sample size is less than twice the number of variables included in the equation being estimated.

In applied economics, if the above condition is met <u>and</u> provided the degree of multicollinearity is not excessive, it is the author's contention that the true (unknown) value of the concentration parameter would, in general, be large enough to enable the jackknife technique to be used on the 2SLS estimator.

APPENDIX A

LEMMA

Proof of the following lemma is due to Bartlett [2].

LEMMA:

If A is a k x k non-singular matrix, and $\underline{c}$ and $\underline{d}$ are two k dimensional column vectors, then

$$(A + \underline{c}\,\underline{d}')^{-1} = A^{-1} - \frac{A^{-1}\underline{c}\,\underline{d}'A^{-1}}{1 + \underline{d}'A^{-1}\underline{c}} .$$

PROOF:

$$\begin{aligned}(A + \underline{c}\,\underline{d}')^{-1} &= A^{-1}\left(I + \underline{c}\,\underline{d}'A^{-1}\right)^{-1} \\ &= A^{-1}(I - \underline{c}\,\underline{d}'A^{-1} + \underline{c}\,\underline{d}'A^{-1}\,\underline{c}\,\underline{d}'A^{-1} - \ldots\ldots) \\ &= A^{-1} - A^{-1}\underline{c}\,\underline{d}'A^{-1}\,\{I - \underline{d}'A^{-1}\underline{c} + (\underline{d}'A^{-1}\underline{c})^2 - \ldots\ldots\} \\ &= A^{-1} - \frac{A^{-1}\underline{c}\,\underline{d}'A^{-1}}{1 + \underline{d}'A^{-1}\underline{c}} .\end{aligned}$$

APPENDIX B

DERIVATION OF A RESULT ON ASYMPTOTIC NORMALITY

Consider the first term in square brackets in equation (3.72), viz:

$$S_1(\text{diag } R_1 + N^{-1}. R_1)S_1'$$

$$= Z'(\Lambda_1 + \Lambda_2)(\text{diag } R_1 + N^{-1}. R_1)(\Lambda_1 + \Lambda_2)Z$$

$$+ Z'M_X\Lambda_2(\text{diag } R_1 + N^{-1}. R_1)\Lambda_2 M_X Z$$

$$- Z'(\Lambda_1 + \Lambda_2)(\text{diag } R_1 + N^{-1}. R_1)\Lambda_2 M_X Z$$

$$- Z'M_X\Lambda_2(\text{diag } R_1 + N^{-1}. R_1)(\Lambda_1 + \Lambda_2)Z \ , \qquad \text{(B.1)}$$

where $R_1 = \underline{\hat{u}}\,\underline{\hat{u}}' = [\underline{u} - Z(\underline{\hat{\theta}} - \underline{\theta})][\underline{u} - Z(\underline{\hat{\theta}} - \underline{\theta})]'$.

The ith element of $\underline{\hat{u}}$ can be written as

$$\hat{u}_i = u_i - \underline{z}_i'(\underline{\hat{\theta}} - \underline{\theta}) \ ,$$

and consequently the jkth element of the first term in equation (B.1) can be written as

$$\sum_{i=1}^{N} z_{ij}z_{ik}(\Lambda_1 + \Lambda_2)^2_{ii}\ [u_i - \underline{z}_i'(\underline{\hat{\theta}} - \underline{\theta})]^2 \ , \qquad \text{(B.2)}$$

ignoring, for the present, the term incorporating $N^{-1}. R_1$.

Upon expansion, equation (B.2) can be written as

$$\sum_{i=1}^{N} z_{ij}z_{ik}(\Lambda_1 + \Lambda_2)^2_{ii}\ u_i^2 - 2\sum_{i=1}^{N} z_{ij}z_{ik}(\Lambda_1 + \Lambda_2)^2_{ii}\ \underline{z}_i'(\underline{\hat{\theta}} - \underline{\theta})u_i$$

$$+ \sum_{i=1}^{N} z_{ij}z_{ik}(\Lambda_1 + \Lambda_2)^2_{ii}\ [\underline{z}_i'(\underline{\hat{\theta}} - \underline{\theta})]^2 \ . \qquad \text{(B.3)}$$

In the forthcoming analysis we will assume, without loss of generality, that the observations on the (g) explanatory endogenous variables occur in the first g elements of $\underline{z}_i'$.

Expanding the second term in equation (B.3) we obtain

$$- 2\sum_{i=1}^{N} z_{ij} z_{ik} (\Lambda_1 + \Lambda_2)^2_{ii} \Big[z_{i1}(\hat{\theta}_1 - \theta_1) + z_{i2}(\hat{\theta}_2 - \theta_2) + \ldots\ldots$$

$$\ldots\ldots + z_{i,K_1+g}(\hat{\theta}_{K_1+g} - \theta_{K_1+g}) \Big] u_i \, . \quad \text{(B.4)}$$

Consider the first (of the K_1+g) term of the above expansion:

$$- 2\sum_{i=1}^{N} z_{ij} z_{ik} (\Lambda_1 + \Lambda_2)^2_{ii} \; z_{i1}(\hat{\theta}_1 - \theta_1) u_i \, . \quad \text{(B.5)}$$

We can partition z_{ij} and z_{ik} as

$z_{ij} = m_{ij} + v_{ij}$ and $z_{ik} = m_{ik} + v_{ik}$, (for $j,k \leq g$) where m_{ij} and m_{ik} represent the nonstochastic part of z_{ij} and z_{ik} respectively ($j,k = 1,2,\ldots,K_1 + g$), and v_{ij} and v_{ik} represent the reduced form disturbance part of z_{ij} and z_{ik} respectively (for $j,k > g$ this will of course be zero).

Expression (B.5) can therefore be written as

$$- 2(\hat{\theta}_1 - \theta_1) \Big[\sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{ij} m_{ik} m_{i1} u_i + \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{ij} m_{ik} v_{i1} u_i$$

$$+ \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{ij} v_{ik} v_{i1} u_i + \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{ij} m_{i1} v_{ik} u_i$$

$$+ \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} v_{ij} v_{ik} m_{i1} u_i + \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} v_{ij} m_{ik} v_{i1} u_i$$

$$+ \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} v_{ij} m_{ik} m_{i1} u_i + \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} v_{ij} v_{ik} v_{i1} u_i \Big] \, .$$

(B.6)

Recall the decomposition introduced as equation (3.51), viz:

$V = \underline{u}\, \Psi' + E$,

the ijth element of which can be written as

$$v_{ij} = u_i \psi_j + e_{ij} . \tag{B.7}$$

Substituting for v_{ij}, v_{ik} and v_{il} in expression (B.6) we obtain

$$- 2(\hat{\theta}_1 - \theta_1)\Bigg[\sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{ij} m_{ik} m_{il} u_i$$

$$+ \left(\sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{ij} m_{ik} \psi_l u_i^2 + \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{ij} m_{ik} e_{il} u_i\right)$$

$$+ \left(\sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{ij} \psi_k \psi_l u_i^3 + \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{ij} e_{ik} e_{il} u_i\right)$$

$$+ \left(\sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{ij} m_{il} \psi_k u_i^2 + \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{ij} m_{il} e_{ik} u_i\right)$$

$$+ \left(\sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{il} \psi_j \psi_k u_i^3 + \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{il} e_{ij} e_{ik} u_i\right)$$

$$+ \left(\sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{ik} \psi_j \psi_l u_i^3 + \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{ik} e_{ij} e_{il} u_i\right)$$

$$+ \left(\sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{ik} m_{il} \psi_j u_i^2 + \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{ik} m_{il} e_{ij} u_i\right)$$

$$+ \left(\sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} \psi_j \psi_k \psi_l u_i^4 + \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} e_{ij} e_{ik} e_{il} u_i\right)\Bigg] . \tag{B.8}$$

Consider the first term in expression (B.8) and note that

$$\max_{1 \le i \le N} \left| (\hat{\theta}_1 - \theta_1)(\Lambda_1 + \Lambda_2)^2_{ii} m_{ij} m_{ik} m_{il} \right|$$

$$\le \max_{1 \le i \le N} \left\{ \left| (\hat{\theta}_1 - \theta_1) \right| \left| (\Lambda_1 + \Lambda_2)^2_{ii} m_{ij} m_{ik} m_{il} \right| \right\}$$

$$= \left| \hat{\theta}_1 - \theta_1 \right| \max_{1 \le i \le N} \left| (\Lambda_1 + \Lambda_2)^2_{ii} m_{ij} m_{ik} m_{il} \right| .$$

Since $$\operatorname*{plim}_{N\to\infty} (\hat{\theta}_j - \theta_j) = 0 , \qquad (j = 1,2,\ldots,K_1 + g) \qquad (B.9)$$

it follows that

$$\operatorname*{plim}_{N\to\infty} (|\hat{\theta}_1 - \theta_1|) \operatorname*{plim}_{N\to\infty} \max_{1\leq i\leq N} \left| (\Lambda_1+\Lambda_2)^2_{ii} m_{ij} m_{ik} m_{il} \right| = 0 . \qquad (B.10)$$

Then using Theorem I (from Chapter 3) it follows that

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N} \left[- 2(\hat{\theta}_1 - \theta_1) \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{ij} m_{ik} m_{il} u_i \right] = 0 .$$

Consider the second term in expression (B.8). Using the above logic it follows that

$$\operatorname*{plim}_{N\to\infty} (|\hat{\theta}_1 - \theta_1|) \operatorname*{plim}_{N\to\infty} \max_{1\leq i\leq N} \left| (\Lambda_1 + \Lambda_2)^2_{ii} m_{ij} m_{ik} \psi_1 \right| = 0 .$$

Then using Theorem I it follows that

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N} \left[- 2(\hat{\theta}_1 - \theta_1) \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{ij} m_{ik} \psi_1 (u_i^2 - \sigma^2) \right] = 0 ,$$

where $E(u_i^2) = \sigma^2$(finite) .

This result implies that

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N} \left[- 2(\hat{\theta}_1 - \theta_1) \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{ij} m_{ik} \psi_1 u_i^2 \right]$$

$$= \operatorname*{plim}_{N\to\infty} \frac{1}{N} \left[- 2(\hat{\theta}_1 - \theta_1) \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{ij} m_{ik} \psi_1 \sigma^2 \right]$$

$$= \sigma^2 \operatorname*{plim}_{N\to\infty} \frac{1}{N} \left[- 2(\hat{\theta}_1 - \theta_1) \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{ij} m_{ik} \psi_1 \right]$$

$= 0$, from Theorem I. (B.11)

With minor variations, this analysis can be used to show that, in the probability limit, the fourth, sixth, eighth, tenth, twelfth and fourteenth terms in expression (B.8) are all zero. Since u_i and e_{ij} are, by assumption, uncorrelated random variables with mean zero, it follows using Theorem I that the remaining terms in expression (B.8) all converge in probability to zero.

Expression (B.5) was analysed under the assumption that $j, k \leq g$. If both j and k are greater than g then no partitioning of z_{ij} and z_{ik} is necessary as they only contain nonstochastic (corresponding to X_1) elements. Under such circumstances, the resulting expansion of expression (B.5) is limited to the first two terms of expression (B.6) and thus the first three terms of expression (B.8). We have already argued that, in the probability limit, these three terms are zero.

If either j or k is less than g then one partitioning of z_{ij} (or z_{ik}) is necessary. The subsequent expansion of expression (B.5) will be limited to just four terms of expression (B.6) and we have already argued that the corresponding terms in expression (B.8) converge in probability to zero.

This concludes the analysis of the first term in expression (B.4). The remaining $K_1 + g - 1$ terms can be dealt with in an analogous manner noting, once more, that the last K_1 values of z_{ij} $(j = 1,2,\ldots.,$ $K_1 + g)$ contain no stochastic component.

Returning to expression (B.3) we have shown that the second term converges in probability to zero.

Consider the third term in expression (B.3), viz:

$$\sum_{i=1}^{N} z_{ij} z_{ik} (\Lambda_1 + \Lambda_2)^2_{ii} [\underline{z}_i' \, (\underline{\hat{\theta}} - \underline{\theta})]^2$$

$$= \sum_{i=1}^{N} z_{ij} z_{ik} (\Lambda_1 + \Lambda_2)^2_{ii} \Big[z_{i1}(\hat{\theta}_1 - \theta_1) + z_{i2}(\hat{\theta}_2 - \theta_2) + \ldots\ldots$$

$$\ldots\ldots + z_{i,K_1+g}(\hat{\theta}_{K_1+g} - \theta_{K_1+g}) \Big]^2 .$$

Upon squaring the term in square brackets we obtain

$$\sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} z_{ij} z_{ik} \Big[z^2_{i1}(\hat{\theta}_1 - \theta_1)^2 + z^2_{i2}(\hat{\theta}_2 - \theta_2)^2 + \ldots\ldots + z^2_{i,K_1+g}(\hat{\theta}_{K_1+g} - \theta_{K_1+g})^2$$

$$+ z_{i1} z_{i2}(\hat{\theta}_1 - \theta_1)(\hat{\theta}_2 - \theta_2) + z_{i1} z_{i3}(\hat{\theta}_1 - \theta_1)(\hat{\theta}_3 - \theta_3) + \ldots\ldots$$

$$+ z_{i1} z_{i,K_1+g}(\hat{\theta}_1 - \theta_1)\,(\hat{\theta}_{K_1+g} - \theta_{K_1+g})$$

$$+ z_{i2} z_{i1}(\hat{\theta}_2 - \theta_2)(\hat{\theta}_1 - \theta_1) + z_{i2} z_{i3}(\hat{\theta}_2 - \theta_2)(\hat{\theta}_3 - \theta_3) + \ldots\ldots$$

$$+ z_{i2} z_{i,K_1+g}(\hat{\theta}_2 - \theta_2)(\hat{\theta}_{K_1+g} - \theta_{K_1+g})$$

$$+ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$+ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$+ z_{i,K_1+g} z_{i1}(\hat{\theta}_{K_1+g} - \theta_{K_1+g})(\hat{\theta}_1 - \theta_1) + z_{i,K_1+g} z_{i2}(\hat{\theta}_{K_1+g} - \theta_{K_1+g})(\hat{\theta}_2 - \theta_2) + \ldots\ldots$$

$$+ z_{i,K_1+g} z_{i,K_1+g-1}(\hat{\theta}_{K_1+g} - \theta_{K_1+g})(\hat{\theta}_{K_1+g-1} - \theta_{K_1+g-1}) \Big] .$$

(B.12)

Consider the first term of the above expression, viz:

$$\sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} z_{ij} z_{ik} z^2_{i1} (\hat{\theta}_1 - \theta_1)^2 , \tag{B.13}$$

where the z_{ij} can again be partitioned only now

$$z^2_{i1} = (m_{i1} + v_{i1})^2 = m^2_{i1} + v^2_{i1} + 2m_{i1} v_{i1} . \quad (j,k \leq g)$$

Upon expansion, expression (B.13) can be written as

$$(\hat{\theta}_1 - \theta_1)^2 \left[\sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{ij} m_{ik} m^2_{i1} + \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{ij} m_{ik} v^2_{i1} \right.$$

$$+ 2 \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{ij} m_{ik} m_{i1} v_{i1} + \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{ij} v_{ik} m^2_{i1}$$

$$+ \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{ij} v_{ik} v^2_{i1} + 2 \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{ij} v_{ik} m_{i1} v_{i1}$$

$$+ \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} v_{ij} m_{ik} m^2_{i1} + \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} v_{ij} m_{ik} v^2_{i1}$$

$$+ 2 \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} v_{ij} m_{ik} m_{i1} v_{i1} + \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} v_{ij} v_{ik} m^2_{i1}$$

$$\left. + \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} v_{ij} v_{ik} v^2_{i1} + 2 \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} v_{ij} v_{ik} m_{i1} v_{i1} \right] . \tag{B.14}$$

Using the decomposition given by equation (B.7), expression (B.14) can be evaluated in a similar term by term manner to the analysis used for evaluating expression (B.6).

Consider the first term in expression (B.14). Since

$$\underset{N \to \infty}{\text{plim}} (\hat{\theta}_j - \theta_j)^2 = \underset{N \to \infty}{\text{plim}} (\hat{\theta}_j - \theta_j) \underset{N \to \infty}{\text{plim}} (\hat{\theta}_j - \theta_j) = 0, \text{ for all } j ,$$

and since m^2_{i1} is a constant, it follows by an analogous proof to that used in deriving equation (B.10) that

$$\underset{N \to \infty}{\text{plim}} \max_{1 \le i \le N} \left| (\hat{\theta}_1 - \theta_1)^2 (\Lambda_1 + \Lambda_2)^2_{ii} m_{ij} m_{ik} m_{i1} \right| = 0 .$$

Using equation (B.7) the second term in expression (B.14) can be written as

$$(\hat{\theta}_1 - \theta_1)^2 \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{ij} m_{ik} \psi^2_1 u^2_i$$

$$+ (\hat{\theta}_1 - \theta_1)^2 \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{ij} m_{ik} e^2_{i1}$$

$$+ (\hat{\theta}_1 - \theta_1)^2 \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{ij} m_{ik} \psi_1 u_i e_{i1} \quad . \qquad \text{(B.15)}$$

Since m_{ij}, m_{ik} and ψ_1 are constants the first two terms in expression (B.15), multiplied by $^1/N$, converge in probability to zero by the same proof used to derive the result given by equation (B.11), assuming $E(e^2_{i1})$ is finite. Further, since u_i and e_{i1} are uncorrelated random variables with mean zero, it follows from Theorem I that the third term in expression (B.15) (multiplied by $^1/N$) converges in probability to zero.

Evaluation of the remaining terms in expression (B.14) follows a similar pattern, all converging in probability to zero.

Returning to expression (B.13) if either j or k (or both) are greater than g then the above analysis involves fewer terms in expression (B.14), as was shown when dealing with the second term in expression (B.3). The analysis, however, is identical.

Returning to expression (B.12), a similar analysis can be used to show that the remaining terms in the first line of this expression all converge in probability to zero. The same result holds for the terms in the remaining lines of expression (B.12), although the analysis is more tedious due to the introduction of another (the fourth) term in z.

To summarize, we have shown that the second and third terms in expression (B.3) converge in probability to zero. Thus we have shown that

$$\underset{N \to \infty}{\text{plim}} \frac{1}{N} \cdot Z'(\Lambda_1 + \Lambda_2)(\text{diag } R_1)(\Lambda_1 + \Lambda_2)Z$$

$$= \underset{N \to \infty}{\text{plim}} \frac{1}{N} \cdot Z'(\Lambda_1 + \Lambda_2)(\text{diag } \underline{u}\,\underline{u}')(\Lambda_1 + \Lambda_2)Z \ .$$

Noting that

$$\operatorname*{plim}_{N\to\infty} (\Lambda_2)_{ii} = - \operatorname*{plim}_{N\to\infty} \frac{\underline{z}_i' \, P^{-1}(\underline{z}_i - \underline{a}_i)}{k_i(1 - s_i + d_i)} = 0 \, , \tag{B.16}$$

using equations (3.39), (3.41) and (3.42), the remaining terms in (diag R_1) in equation (B.1) can be analysed in an analogous manner to the first term. We have shown therefore that

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N} \, . \, S_1(\text{diag } R_1)S_1' = \operatorname*{plim}_{N\to\infty} \frac{1}{N} \, . \, S_1(\text{diag } \underline{u}\,\underline{u}')S_1' \, .$$

We now consider the terms in $N^{-1} . R_1$ in expression (B.1). The first term can be expanded as follows:

$$\begin{aligned}
& N^{-1} \, . \, Z'(\Lambda_1 + \Lambda_2)R_1(\Lambda_1 + \Lambda_2)Z \\
= \; & N^{-1} \, . \, Z'(\Lambda_1 + \Lambda_2)\hat{\underline{u}}\,\hat{\underline{u}}'(\Lambda_1 + \Lambda_2)Z \\
= \; & N^{-1} . \, Z'(\Lambda_1 + \Lambda_2) \, \underline{u}\,\underline{u}'(\Lambda_1 + \Lambda_2)Z \\
+ \; & N^{-1} \, . \, Z'(\Lambda_1 + \Lambda_2)Z(\hat{\underline{\theta}} - \underline{\theta})(\hat{\underline{\theta}} - \underline{\theta})'Z'(\Lambda_1 + \Lambda_2)Z \\
- \; & N^{-1} \, . \, Z'(\Lambda_1 + \Lambda_2)Z(\hat{\underline{\theta}} - \underline{\theta})\underline{u}'(\Lambda_1 + \Lambda_2)Z \\
- \; & N^{-1} \, . \, Z'(\Lambda_1 + \Lambda_2)\underline{u}(\hat{\underline{\theta}} - \underline{\theta})'Z'(\Lambda_1 + \Lambda_2)Z \, .
\end{aligned}$$

Since 2SLS is a consistent estimator we know that

$$\operatorname*{plim}_{N\to\infty} (\hat{\underline{\theta}} - \underline{\theta}) = \underline{0} \, ,$$

and from the preceeding analysis it is easy to show that

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N} \, . \, Z'(\Lambda_1 + \Lambda_2)Z$$

is a finite matrix. Now consider the term

$$Z'(\Lambda_1 + \Lambda_2) \, \underline{u} \, . \tag{B.17}$$

The jth component of this random vector can be written as

$$\sum_{i=1}^{N} z_{ij}(\Lambda_1 + \Lambda_2)_{ii} u_i .$$

Partitioning z_{ij} into its stochastic and nonstochastic components, and using the decomposition of v_{ij} given by equation (B.7) we obtain

$$\sum_{i=1}^{N} m_{ij}(\Lambda_1 + \Lambda_2)_{ii} u_i + \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)_{ii} \psi_j u_i^2$$

$$+ \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)_{ii} u_i e_{ij} . \qquad \text{(B.18)}$$

Since the m_{ij} are constants, and using the result that

$$\underset{N\to\infty}{\text{plim}} (\Lambda_1 + \Lambda_2)_{ii} = 1 ,$$

it follows that

$$\underset{N\to\infty}{\text{plim}} \frac{1}{N} \sum_{i=1}^{N} m_{ij}(\Lambda_1 + \Lambda_2)_{ii} u_i = 0$$

by the Law of Large Numbers. The same Law ensures that the second term in equation (B.18) (multiplied by $^1/N$) converges in probability to a finite constant, provided $E(u_i^2)$ is finite, and that, since u_i and e_{ij} are uncorrelated random variables, the third term (multiplied by $^1/N$) converges in probability to zero.

Combining the above results, we have shown that

$$\underset{N\to\infty}{\text{plim}} \frac{1}{N} \left[\frac{1}{N} . Z'(\Lambda_1 + \Lambda_2)R_1(\Lambda_1 + \Lambda_2)Z\right]$$

$$= \underset{N\to\infty}{\text{plim}} \frac{1}{N} \left[\frac{1}{N} . Z'(\Lambda_1 + \Lambda_2)\underline{u}\, \underline{u}'(\Lambda_1 + \Lambda_2)Z\right] .$$

Using the result given by equation (B.16), in addition to the above results, it can be shown that the remaining terms containing $N^{-1}.R_1$ in equation (B.1) can be analysed in an analogous manner. All three remaining terms converge in probability to zero.

Thus we have shown that the first term in square brackets in equation (3.72) can be written as

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N} . S_1(\text{diag } R_1 + N^{-1}.R_1)S_1' = \operatorname*{plim}_{N\to\infty} \frac{1}{N} . S_1(\text{diag } \underline{u}\,\underline{u}' + N^{-1}.\underline{u}\,\underline{u}')S_1' .$$

Consider the second term in square brackets in equation (3.72), <u>viz</u>: $S_2(\text{diag } R_2 + N^{-1}.R_2)S_2'$

$$= Z'(\Lambda_2 - \Lambda_3)(\text{diag } R_2 + N^{-1}.R_2)(\Lambda_2 - \Lambda_3)Z$$

$$+ Z'M_X\Lambda_3(\text{diag } R_2 + N^{-1}.R_2)\Lambda_3 M_X Z$$

$$+ Z'(\Lambda_2 - \Lambda_3)(\text{diag } R_2 + N^{-1}.R_2)\Lambda_3 M_X Z$$

$$+ Z'M_X\Lambda_3(\text{diag } R_2 + N^{-1}.R_2)(\Lambda_2 - \Lambda_3)Z , \quad \text{(B.19)}$$

where $R_2 = (I - M_X)\hat{\underline{u}}\,\hat{\underline{u}}'(I - M_X)$

$$= (I - M_X)[\underline{u} - Z(\hat{\underline{\theta}} - \underline{\theta})][\underline{u} - Z(\hat{\underline{\theta}} - \underline{\theta})]'(I - M_X) .$$

The following results can be easily derived:

$$\operatorname*{plim}_{N\to\infty} (\Lambda_2 - \Lambda_3)_{ii} = -1 , \quad \text{(B.20)}$$

$$\operatorname*{plim}_{N\to\infty} (\Lambda_2 - \Lambda_3)^2_{ii} = 1 , \quad \text{(B.21)}$$

$$\operatorname*{plim}_{N\to\infty} (\Lambda_3)^2_{ii} = 1 , \quad \text{(B.22)}$$

and $$\operatorname*{plim}_{N\to\infty} (\Lambda_2 - \Lambda_3)_{ii}(\Lambda_3)_{ii} = 1 . \quad \text{(B.23)}$$

The first term in equation (B.19) can be written as

$$Z'(\Lambda_2 - \Lambda_3)(\text{diag } R_2)(\Lambda_2 - \Lambda_3)Z$$

$$+ Z'(\Lambda_2 - \Lambda_3)(N^{-1}_{.}R_2)(\Lambda_2 - \Lambda_3)Z \text{ ,} \qquad \text{(B.24)}$$

whereupon, using the definition of R_2, the first term in equation (B.24) can be written as

$$Z'(\Lambda_2 - \Lambda_3)(\text{diag } \underline{\hat{u}}\,\underline{\hat{u}}')(\Lambda_2 - \Lambda_3)Z$$

$$- Z'(\Lambda_2 - \Lambda_3)M_X(\text{diag } \underline{\hat{u}}\,\underline{\hat{u}}')(\Lambda_2 - \Lambda_3)Z$$

$$- Z'(\Lambda_2 - \Lambda_3)(\text{diag } \underline{\hat{u}}\,\underline{\hat{u}}')M_X(\Lambda_2 - \Lambda_3)Z$$

$$+ Z'(\Lambda_2 - \Lambda_3)M_X(\text{diag } \underline{\hat{u}}\,\underline{\hat{u}}')M_X(\Lambda_2 - \Lambda_3)Z \text{ .} \qquad \text{(B.25)}$$

The jkth element of the first term in equation (B.25) can be written as

$$\sum_{i=1}^{N} z_{ij} z_{ik} (\Lambda_2 - \Lambda_3)^2_{ii} \, [u_i - \underline{z}_i'(\underline{\hat{\theta}} - \underline{\theta})]^2 \text{ .}$$

Since

$$\underset{N\to\infty}{\text{plim}} \max_{1\leq i\leq N} (\Lambda_1 + \Lambda_2)^2_{ii} = \underset{N\to\infty}{\text{plim}} \max_{1\leq i\leq N} (\Lambda_2 - \Lambda_3)^2_{ii} = 1 \text{ ,}$$

this expression does not differ, asymptotically, from expression (B.2) and can therefore be analysed in an analogous manner. It follows that

$$\underset{N\to\infty}{\text{plim}} \frac{1}{N} \,.\, Z'(\Lambda_2 - \Lambda_3)(\text{diag } \underline{\hat{u}}\,\underline{\hat{u}}')(\Lambda_2 - \Lambda_3)Z$$

$$= \underset{N\to\infty}{\text{plim}} \frac{1}{N} \,.\, Z'(\Lambda_2 - \Lambda_3)(\text{diag } \underline{u}\,\underline{u}')(\Lambda_2 - \Lambda_3)Z \text{ .}$$

Consider the remaining terms in expression (B.25). We have already shown (equation (3.27)) that

$$\lim_{N\to\infty} (M_X)_{ii} = 0 \text{ ,} \qquad (i = 1,2,\ldots.,N) \qquad \text{(B.26)}$$

and thus the remaining terms must all converge in probability to zero by an analogous proof to that employed in analysing the first term.

Further, equations (B.22), (B.23) and (B.26) allow the remaining terms containing (diag R_2) in equation (B.19) to be analysed in an identical manner to their corresponding terms containing (diag R_1) in equation (B.1). It follows that

$$\underset{N\to\infty}{\text{plim}} \frac{1}{N} . S_2(\text{diag } R_2)S_2' = \underset{N\to\infty}{\text{plim}} \frac{1}{N} . S_2(\text{diag } \underline{u}\,\underline{u}')S_2' .$$

The terms containing $N^{-1}.R_2$ in equation (B.19) can also be analysed in a similar manner to their counterparts in equation (B.1). Consider the first term containing $N^{-1}R_2$ in equation (B.19), viz

$$N^{-1}. Z' (\Lambda_2 - \Lambda_3)R_2(\Lambda_2 - \Lambda_3)Z$$

$$= N^{-1}.Z'(\Lambda_2 - \Lambda_3)(I - M_X)\underline{\hat{u}}\,\underline{\hat{u}}'(I - M_X)(\Lambda_2 - \Lambda_3)Z$$

$$= N^{-1}.Z'(\Lambda_2 - \Lambda_3)(I - M_X)\underline{u}\,\underline{u}'(I - M_X)(\Lambda_2 - \Lambda_3)Z$$

$$+ N^{-1}.Z'(\Lambda_2 - \Lambda_3)(I - M_X)Z(\underline{\hat{\theta}} - \underline{\theta})(\underline{\hat{\theta}} - \underline{\theta})'Z'(I - M_X)(\Lambda_2 - \Lambda_3)Z$$

$$- N^{-1}.Z'(\Lambda_2 - \Lambda_3)(I - M_X)\underline{u}(\underline{\hat{\theta}} - \underline{\theta})'Z'(I - M_X)(\Lambda_2 - \Lambda_3)Z$$

$$- N^{-1}.Z'(\Lambda_2 - \Lambda_3)(I - M_X)Z(\underline{\hat{\theta}} - \underline{\theta})\underline{u}'(I - M_X)(\Lambda_2 - \Lambda_3)Z . \qquad \text{(B.27)}$$

The first term in equation (B.27) can be written as

$$N^{-1}.Z'(\Lambda_2 - \Lambda_3)\underline{u}\,\underline{u}'(\Lambda_2 - \Lambda_3)Z - N^{-1}.Z'(\Lambda_2 - \Lambda_3)M_X\underline{u}\,\underline{u}'(\Lambda_2 - \Lambda_3)Z$$

$$- N^{-1}.Z'(\Lambda_2 - \Lambda_3)\underline{u}\,\underline{u}'M_X(\Lambda_2 - \Lambda_3)Z + N^{-1}.Z'(\Lambda_2 - \Lambda_3)M_X\,\underline{u}\,\underline{u}'\,M_X(\Lambda_2 - \Lambda_3)Z .$$

From our initial assumptions (specifically, Assumption (iii) in Section 2.1.3) it follows that

$$\underset{N\to\infty}{\text{plim}} \frac{1}{N} . Z'(\Lambda_2 - \Lambda_3)M_X\underline{u} = \underline{0} ,$$

and hence, asymptotically, the first term in equation (B.27) can be written as

$$\underset{N\to\infty}{\text{plim}}\ \frac{1}{N}\left[\frac{1}{N}\,.\,Z'(\Lambda_2-\Lambda_3)\underline{u}\,\underline{u}'(\Lambda_2-\Lambda_3)Z\right].$$

Since $(I - M_X)$ is a nonstochastic matrix, using equation (B.20) it follows that the last three terms in equation (B.27) (multiplied by $\frac{1}{N}$) all converge in probability to zero. Thus we have shown that

$$\underset{N\to\infty}{\text{plim}}\ \frac{1}{N}\left[\frac{1}{N}\,.\,Z'(\Lambda_2-\Lambda_3)R_2(\Lambda_2-\Lambda_3)Z\right]$$

$$=\ \underset{N\to\infty}{\text{plim}}\ \frac{1}{N}\left[\frac{1}{N}\,.\,Z'(\Lambda_2-\Lambda_3)\underline{u}\,\underline{u}'(\Lambda_2-\Lambda_3)Z\right].$$

Clearly the third and fourth terms in equation (3.72) can be analysed in an identical manner since the relevant results have already been derived.

To conclude, we have shown that

$$\underset{N\to\infty}{\text{plim}}\ \frac{1}{N}\,.\,S_1(\text{diag}\ R_1+N^{-1}.R_1)S_1' = \underset{N\to\infty}{\text{plim}}\ \frac{1}{N}\,.\,S_1(\text{diag}\ \underline{u}\,\underline{u}'+N^{-1}.\,\underline{u}\,\underline{u}')S_1',$$

$$\underset{N\to\infty}{\text{plim}}\ \frac{1}{N}\,.\,S_2(\text{diag}\ R_2+N^{-1}.R_2)S_2' = \underset{N\to\infty}{\text{plim}}\ \frac{1}{N}\,.\,S_2(\text{diag}\ \underline{u}\,\underline{u}'+N^{-1}.\,\underline{u}\,\underline{u}')S_2',$$

and

$$\underset{N\to\infty}{\text{plim}}\ \frac{1}{N}\,.\,S_1(\text{diag}\ R_3+N^{-1}.R_3)S_2' = \underset{N\to\infty}{\text{plim}}\ \frac{1}{N}\,.\,S_1(\text{diag}\ \underline{u}\,\underline{u}'+N^{-1}.\,\underline{u}\,\underline{u}')S_2'.$$

Consider the following summation

$$S_1(\text{diag}\ \underline{u}\,\underline{u}'+N^{-1}.\ \underline{u}\,\underline{u}')S_1' + S_2(\text{diag}\ \underline{u}\,\underline{u}'+N^{-1}.\ \underline{u}\,\underline{u}')S_2'$$

$$-\ S_1(\text{diag}\ \underline{u}\,\underline{u}'+N^{-1}.\ \underline{u}\,\underline{u}')S_2' + S_2(\text{diag}\ \underline{u}\,\underline{u}'+N^{-1}.\ \underline{u}\,\underline{u}')S_1'\ ,$$

which can be simplified to

$$(S_1 - S_2)(\text{diag}\ \underline{u}\ \underline{u}' + N^{-1}\underline{u}\ \underline{u}')(S_1 - S_2)'$$

$$= Z'(\Lambda_1 + 2\Lambda_2 - \Lambda_3)(\text{diag}\ \underline{u}\ \underline{u}' + N^{-1}\underline{u}\ \underline{u}')(\Lambda_1 + 2\Lambda_2 - \Lambda_3)Z$$

$$- Z'M_X(\Lambda_2 - \Lambda_3)(\text{diag}\ \underline{u}\ \underline{u}' + N^{-1}\underline{u}\ \underline{u}')(\Lambda_1 + 2\Lambda_2 - \Lambda_3)Z$$

$$- Z'(\Lambda_1 + 2\Lambda_2 - \Lambda_3)(\text{diag}\ \underline{u}\ \underline{u}' + N^{-1}\underline{u}\ \underline{u}')(\Lambda_2 - \Lambda_3)M_XZ$$

$$+ Z'M_X(\Lambda_2 - \Lambda_3)(\text{diag}\ \underline{u}\ \underline{u}' + N^{-1}\underline{u}\ \underline{u}')(\Lambda_2 - \Lambda_3)M_XZ\ . \qquad (B.28)$$

The jkth element of the first term containing (diag $\underline{u}\ \underline{u}'$) in equation (B.28) can be written as

$$\sum_{i=1}^{N} m_{ij}m_{ik}(\Lambda_1 + 2\Lambda_2 - \Lambda_3)^2_{ii}u_i^2 + \sum_{i=1}^{N} v_{ij}m_{ik}(\Lambda_1 + 2\Lambda_2 - \Lambda_3)^2_{ii}u_i^2$$

$$+ \sum_{i=1}^{N} m_{ij}v_{ik}(\Lambda_1 + 2\Lambda_2 - \Lambda_3)^2_{ii}u_i^2 + \sum_{i=1}^{N} v_{ij}v_{ik}(\Lambda_1 + 2\Lambda_2 - \Lambda_3)^2_{ii}u_i^2\ . \qquad (B.29)$$

We have already shown (equation (3.47)) that

$$\underset{N\to\infty}{\text{plim}}\ \sqrt{N} \max_{1\leq i\leq N} (\Lambda_1 + 2\Lambda_2 - \Lambda_3)_{ii} = 0, \qquad (B.30)$$

from which it follows that

$$\underset{N\to\infty}{\text{plim}}\ \sqrt{N} \max_{1\leq i\leq N} \left|m_{ij}m_{ik}(\Lambda_1 + 2\Lambda_2 - \Lambda_3)_{ii}\right| = 0\ .$$

Using Theorem I it follows that

$$\underset{N\to\infty}{\text{plim}}\ \frac{1}{N}\sum_{i=1}^{N} m_{ij}m_{ik}(\Lambda_1 + 2\Lambda_2 - \Lambda_3)^2_{ii}u_i^2 = 0\ ,$$

provided $E(u_i{}^2)$ is finite.

The decomposition given by equation (B.7) is required in order to evaluate the three remaining terms in equation (B.29). The second term can be written as

$$\sum_{i=1}^{N} (\Lambda_1 + 2\Lambda_2 - \Lambda_3)^2_{ii}m_{ik}\psi_j u_i^3 + \sum_{i=1}^{N} (\Lambda_1 + 2\Lambda_2 - \Lambda_3)^2_{ii}m_{ik}e_{ij}u_i^2\ .$$

Equation (B.29) and Theorem I ensure that the first term in this expression (multiplied by 1/N) converges in probability to zero; and since the e_{ij} and u_i are uncorrelated random variables with mean zero, Theorem I ensures that the second term in this expression (multiplied by 1/N) also converges in probability to zero. Similarly the third term in equation (B.29) (multiplied by 1/N) converges in probability to zero.

The fourth term in expression (B.29) can be written as

$$\sum_{i=1}^{N} (\Lambda_1 + 2\Lambda_2 - \Lambda_3)^2_{ii}\psi_j\psi_k u_i^4 + \sum_{i=1}^{N} (\Lambda_1 + 2\Lambda_2 - \Lambda_3)^2_{ii}\psi_j e_{ik} u_i^3$$

$$+ \sum_{i=1}^{N} (\Lambda_1 + 2\Lambda_2 - \Lambda_3)^2_{ii}\psi_k e_{ij} u_i + \sum_{i=1}^{N} (\Lambda_1 + 2\Lambda_2 - \Lambda_3)^2_{ii} e_{ij} e_{ik} u_i^2 .$$

The first term of this expression (multiplied by 1/N) converges in probability to zero by virtue of equation (B.30) and Theorem I. The three remaining terms (multiplied by 1/N) also converge in probability to zero, using Theorem I and the assumption that e_{ij} and u_i are uncorrelated random variables with mean zero.

The first term in $N^{-1}.\ \underline{u}\ \underline{u}'$ in equation (B.28) is

$$N^{-1}.Z'(\Lambda_1 + 2\Lambda_2 - \Lambda_3)\underline{u}\ \underline{u}'(\Lambda_1 + 2\Lambda_2 - \Lambda_3)Z.$$

We have already shown in Chapter 3 that

$$\underset{N\to\infty}{\text{plim}}\ \frac{1}{\sqrt{N}} . Z'(\Lambda_1 + 2\Lambda_2 - \Lambda_3)\hat{\underline{u}} = \underline{0} .$$

Since each element of the 2SLS residual vector converges in distribution to the corresponding element of the disturbance vector, this implies that

$$\underset{N\to\infty}{\text{plim}}\ \frac{1}{\sqrt{N}} . Z'(\Lambda_1 + 2\Lambda_2 - \Lambda_3)\underline{u} = \underline{0} . \qquad \text{(B.31)}$$

It follows that

$$\underset{N\to\infty}{\text{plim}}\ \frac{1}{N} [\frac{1}{N} . Z'(\Lambda_1 + 2\Lambda_2 - \Lambda_3)\underline{u}\ \underline{u}'(\Lambda_1 + 2\Lambda_2 - \Lambda_3)Z] = 0 .$$

Thus we have shown that the first term in equation (B.28) converges in probability to zero.

Consider the second term in equation (B.28), viz:

$$- Z'M_X(\Lambda_2 - \Lambda_3)(\text{diag } \underline{u}\,\underline{u}')(\Lambda_1 + 2\Lambda_2 - \Lambda_3)Z$$

$$- Z'M_X(\Lambda_2 - \Lambda_3)(N^{-1}.\, \underline{u}\,\underline{u}')(\Lambda_1 + 2\Lambda_2 - \Lambda_3)Z \,, \qquad (B.32)$$

and we will analyse the term in (diag $\underline{u}\,\underline{u}'$) first.

$$\text{Since } \operatorname*{plim}_{N\to\infty} \left(\frac{1}{N} . X'X\right)^{-1} \left(\frac{1}{N} . X'Z\right) = \Sigma_{XX}^{-1}\,\Sigma_{XZ} \,, \qquad (B.33)$$

we need only consider the limiting form of

$$X'(\Lambda_2 - \Lambda_3)(\text{diag } \underline{u}\,\underline{u}')(\Lambda_1 + 2\Lambda_2 - \Lambda_3)Z \,.$$

The rjth element of this term can be written as

$$\sum_{i=1}^{N} x_{ir}(\Lambda_2 - \Lambda_3)_{ii}(\Lambda_1 + 2\Lambda_2 - \Lambda_3)_{ii}(m_{ij} + v_{ij})u_i^2$$

$(j = 1,2,\ldots,K_1 + g \;;\; r = 1,2,\ldots,K)$,

i.e. $$\sum_{i=1}^{N} x_{ir}(\Lambda_2 - \Lambda_3)_{ii}(\Lambda_1 + 2\Lambda_2 - \Lambda_3)_{ii}m_{ij}u_i^2$$

$$+ \sum_{i=1}^{N} x_{ir}(\Lambda_2 - \Lambda_3)_{ii}(\Lambda_1 + 2\Lambda_2 - \Lambda_3)_{ii}\psi_j u_i^3$$

$$+ \sum_{i=1}^{N} x_{ir}(\Lambda_2 - \Lambda_3)_{ii}(\Lambda_1 + 2\Lambda_2 - \Lambda_3)_{ii}\, u_i^2 e_{ij} \,.$$

Using equations (B.20) and (B.30), by Theorem I

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} x_{ir}(\Lambda_2 - \Lambda_3)_{ii}(\Lambda_1 + 2\Lambda_2 - \Lambda_3)_{ii}m_{ij}u_i^2 = 0 \,,$$

provided $E(u_i^2)$ is finite. By the same argument

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} x_{ir}(\Lambda_2 - \Lambda_3)_{ii}(\Lambda_1 + 2\Lambda_2 - \Lambda_3)_{ii}\psi_j u_i^3 = 0,$$

provided $E(u_i^3)$ is finite. Finally, since u_i and e_{ij} are uncorrelated random variables with mean zero it follows, using Theorem I, that the remaining term in the above expression (multiplied by 1/N) converges in probability to zero.

Now consider the limiting form of the second term in expression (B.32), viz

$$Z'M_X(\Lambda_2 - \Lambda_3)(\tfrac{1}{N}\,\underline{u}\,\underline{u}')(\Lambda_1 + 2\Lambda_2 - \Lambda_3)Z \ .$$

In view of equation (B.33) we are concerned with the limiting form of

$$N^{-1}.X'(\Lambda_2 - \Lambda_3)\underline{u}\,\underline{u}'(\Lambda_1 + 2\Lambda_2 - \Lambda_3)Z \ .$$

Combining Assumption (v), Section 2.1.3., with equation (B.20) it follows that

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N} . X'(\Lambda_2 - \Lambda_3)\underline{u} = \underline{0} \ ,$$

which when combined with equation (B.31) ensures that

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N} \left[\frac{1}{N} . Z'M_X(\Lambda_2 - \Lambda_3)(\underline{u}\,\underline{u}')(\Lambda_1 + 2\Lambda_2 - \Lambda_3)Z\right] = 0 \ .$$

Thus the second, and hence the third, term in equation (B.28) has been shown to converge in probability to zero.

The fourth term in equation (B.28) can be written as

$$Z'X(X'X)^{-1}X'(\Lambda_2 - \Lambda_3)(\text{diag }\underline{u}\,\underline{u}')(\Lambda_2 - \Lambda_3)X(X'X)^{-1}X'Z$$

$$+ \ Z'X(X'X)^{-1}X'(\Lambda_2 - \Lambda_3)(N^{-1}.\underline{u}\,\underline{u}')(\Lambda_2 - \Lambda_3)X(X'X)^{-1}X'Z \ . \qquad \text{(B.34)}$$

Again, combining Assumption (v), Section 2.1.3, with equations (B.20) and (B.33) we have shown that

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N} \left[\frac{1}{N} . Z'M_X(\Lambda_2 - \Lambda_3)(\underline{u}\,\underline{u}')(\Lambda_2 - \Lambda_3)M_XZ\right] = 0 \ .$$

Consider the expression

$$X'(\Lambda_2 - \Lambda_3)(\text{diag }\underline{u}\,\underline{u}')(\Lambda_2 - \Lambda_3)X \ ,$$

which has rsth term

$$\sum_{i=1}^{N} x_{ir}x_{is}(\Lambda_2 - \Lambda_3)^2_{ii}u_i^2 \,. \qquad (r,s = 1,2,\ldots,K) \tag{B.35}$$

From equation (B.21)

$$\underset{N\to\infty}{\text{plim}}\ (\Lambda_2 - \Lambda_3)^2_{ii} = 1$$

for all $i(i = 1,2,\ldots,N)$, from which it follows that

$$\underset{N\to\infty}{\text{plim}}\ \max_{1\leq i\leq N} \left\{1 - (\Lambda_2 - \Lambda_3)^2_{ii}\right\} = 0 \,.$$

Thus, using Theorem I, it follows that

$$\underset{N\to\infty}{\text{plim}}\ \frac{1}{N}\sum_{i=1}^{N} x_{ir}x_{is}\left[1 - (\Lambda_2 - \Lambda_3)^2_{ii}\right](u_i^2 - \sigma^2) = 0 \,,$$

and hence we deduce that

$$\underset{N\to\infty}{\text{plim}}\ \frac{1}{N}\sum_{i=1}^{N} x_{ir}x_{is}u_i^2 = \sigma^2\,\frac{1}{N}\sum_{i=1}^{N} x_{ir}x_{is} \,.$$

Substituting this result back into equation (B.34), and using equation (B.33) we have shown that the fourth term in equation (B.28), and hence the entire expression, converges in probability to

$$\sigma^2\Sigma_{\underset{\sim}{P}} \,.$$

APPENDIX C

THE TORONTO FUNCTION

Copson [10] has shown that, for large $|x|$,

$$ {}_1F_1\ (\alpha;\gamma;x) \sim \frac{\Gamma(\gamma)}{\Gamma(\alpha)}\frac{e^x}{x^{\gamma-\alpha}}\ {}_2F_0(\gamma-\alpha,\ 1-\alpha;\ \ ;\frac{1}{x})\ , $$

where $ {}_2F_0(\gamma-\alpha;\ 1-\alpha;\ \ ;\ 1/x) = \sum_{r=0}^{\infty} \frac{(\gamma-\alpha)_r(1-\alpha)_r}{r!}\left(\frac{1}{x}\right)^r $,

and Pochhammer's symbol means

$$ (\gamma-\alpha)_r = \frac{\Gamma(\gamma-\alpha+r)}{\Gamma(\gamma-\alpha)}\ . $$

If $\gamma = \alpha + 1$ then

$$ {}_1F_1(\alpha;\alpha+1;\ x) \sim \frac{\alpha e^x}{x}\sum_{r=0}^{\infty}(1-\alpha)_r\left(\frac{1}{x}\right)^r\ , \qquad (C.1) $$

which has a finite number of terms if α is a positive integer.

Equation (C.1) is required for evaluation of the first order moment of the 2SLS estimator. For second and higher order moments $\gamma = \alpha + k$ (where k is the order of the moment under consideration), but can be expressed in terms of equation (C.1) by utilizing the recurrence relations for the confluent hypergeometric function (see Slater [64] ; p.19).

The Toronto function was developed by Heatley [18] and is defined as

$$ T(2\alpha-1,\gamma-1,x^{\frac{1}{2}}) = x^{(\gamma-\alpha)}\ e^{-x}\ \frac{\Gamma(\alpha)}{\Gamma(\gamma)}\ {}_1F_1(\alpha;\gamma;\ x) \qquad (C.2) $$

(N.B. Slater [64] gives this formula with an incorrect sign, p.99).

This function is characterized by convergence to unity as x increases indefinitely.

If $\gamma = \alpha + 1$, then equation (C.2) can be rewritten as

$$T(2\alpha - 1, \alpha, x^{\frac{1}{2}}) = \frac{x}{\alpha} e^{-x} {}_1F_1 (\alpha ; \alpha + 1; x) \quad . \tag{C.3}$$

We state two special forms of the Toronto function that are required in the forthcoming analysis:

$$T(1, 1, x^{\frac{1}{2}}) = 1 - e^{-x} , \tag{C.4}$$

and $$T(1, 2, x^{\frac{1}{2}}) = 1 - (1 + x)e^{-x} . \tag{C.5}$$

In addition, we require two of the recurrence formulae for the Toronto function (see Heatley [18; p.17]):

$$T(\nu + 2, \alpha + 1, x^{\frac{1}{2}}) = \frac{(\nu-2\alpha-1)}{2x} T(\nu, \alpha + 1, x^{\frac{1}{2}}) + T(\nu, \alpha, x^{\frac{1}{2}}), \tag{C.6}$$

and

$$T(\nu + 4, \alpha + 2, x^{\frac{1}{2}}) = \frac{(\nu+1)}{2x} T(\nu, \alpha, x^{\frac{1}{2}}) - \frac{2(\alpha+1-x)}{2x} T(\nu+2, \alpha+1, x^{\frac{1}{2}}), \tag{C.7}$$

where $\nu = 2\alpha - 1$. Thus all values of α can be evaluated with ease.

If e^{-x} is assumed to be zero, the Toronto functions in equations (C.4) and (C.5) will both be unity. Thus, by setting $\alpha = 1$, initial values for the recurrence formulae can be determined, and it is then possible to evaluate the Toronto function for all integer values of α by repeated application of equations (C.6) and (C.7).

Following the above procedure, equation (C.3) can be rewritten as $${}_1F_1(\alpha;\alpha+1;x) = \frac{\alpha}{x} e^{x} \sum_{r=0}^{\infty} (1-\alpha)_r \left(\frac{1}{x}\right)^r ,$$

which is identical to the asymptotic approximation to the confluent hypergeometric function given in equation (C.1). Thus the error incurred in utilizing the asymptotic approximation for finite x is simply the error caused by assuming e^{-x} to be zero in the Toronto function.

It is easy to show that this error can be expressed as

$$\frac{\Gamma(\alpha+1)}{(-x)^{\alpha}} \ , \qquad (C.8)$$

thus as x increases indefinitely (for α fixed), the error of approximation tends to zero.

This analysis is only valid when α is an integer, a condition which will not always be upheld (e.g. when considering the moments of the 2SLS estimator, even values of K_2 will yield integer values for α, whereas odd values of K_2 will yield non-integer values for α).

For α non-integer, equation (C.1) is an infinite series, although it can be truncated after (say) n terms. If this is done the error involved by truncating the infinite series after the nth term will not exceed the (n+1)th term, and will be of the same sign as the (n+1)th term (e.g. see Luke [25; p.127]).

BIBLIOGRAPHY

1. Arvesen, J.N., Jackknifing U-Statistics, Annals of Mathematical Statistics, Vol. 40, 1969, pp. 2076-2100.

2. Bartlett, M.S., An Inverse Matrix Adjustment Arising in Discriminant Analysis, Annals of Mathematical Statistics, Vol. 22, 1951, pp. 107-111.

3. Basmann, R.L., A Generalised Classical Method of Linear Estimation of Coefficients in a Structural Equation, Econometrica, Vol. 25, 1957, pp. 77-83.

4. Basmann, R.L., A Note on the Exact Finite Sample Frequency Function of Generalized Classical Linear Estimators in Two Leading Over-Identified Cases, Journal of the American Statistical Association, Vol. 56, 1961, pp. 619-636.

5. Basmann, R.L., A Note on the Exact Finite Sample Frequency Functions of Generalized Classical Linear Estimators in a Leading Three Equation Case, Journal of the American Statistical Association, Vol. 58, 1963, pp. 161-171.

6. Box, G.E.P. and Muller, M.E., A Note on the Generation of Normal Deviates, Annals of Mathematical Statistics, Vol. 29, 1958, pp. 610-611.

7. Brillinger, D.R., The Asymptotic Behaviour of Tukey's General Method of Setting Approximate Confidence Limits (The Jackknife) When Applied to Maximum Likelihood Estimates, Review of the International Statistical Institute, Vol. 32, 1964, pp. 202-206.

8. Brundy, J. and Jorgenson, D., Efficient Estimation of Simultaneous Equations by Instrumental Variables, Review of Economics and Statistics, Vol. 53, 1971, pp. 207-224.

9. Chay, S.C., Fardo, R.D. and Mazumbar, M., On Using the Box-Muller Transformation with Congruential Pseudo-Random Number Generators, Applied Statistics, Vol. 24, 1975, pp. 132-135.

10. Copson, E.T., Theory of Functions of a Complex Variable, London, Oxford University Press, 1957.

11. Devine, C.M., Scalecholeski (Algorithm 12), Computer Journal, Vol. 9, 1966-1967, pp. 321-322.

12. Dhrymes, P.J., Alternative Asymptotic Tests of Significance and Related Aspects of 2SLS and 3SLS Estimated Parameters, Review of Economic Studies, Vol. 36, 1969, pp. 213-226.

13. Dhrymes, P.J., Restricted and Unrestricted Reduced Forms: Asymptotic Distribution and Relative Efficiency, Econometrica, Vol. 41, 1973, pp. 119-134.

14. Durbin, J., A Note on the Application of Quenouille's Method of Bias Reduction to the Estimation of Ratios, Biometrika, Vol. 46, 1959, pp. 477-480.

15. Fisher, W.D. and Wadycki, W.J., Estimating a Structural Equation in a Large System, Econometrica, Vol. 39, 1971, pp. 461-465.

16. Hammersley, J.M. amd Handscomb, D.C., Monte Carlo Methods, London, Methuen, 1964.

17. Hatanaka, M., On the Existence and the Approximation Formulae for the Moments of k-Class Estimators, The Economic Studies Quarterly, Vol. XXIV, 1973, pp. 1-15.

18. Heatley, A.H., A Short Table of the Toronto Function, Royal Society of Canada, Transactions, Section III, 1943, pp. 13-19.

19. Hoeffding, W., A Class of Statistics With Asymptotically Normal Distribution, Annals of Mathematical Statistics, Vol. 19, 1948, pp. 293-325.

20. Johnston, J., Econometric Methods, Second Edition, New York, McGraw-Hill, 1972.

21. Kabe, D.G., A Note on the Exact Distribution of the G.C.L. Estimators in Two Leading Overidentified Cases, Journal of the American Statistical Association, Vol. 58, 1963, pp. 535-537.

22. Kabe, D.G., On the Exact Distribution of the G.C.L. Estimators in a Leading Three Equation Case, Journal of the American Statistical Association, Vol. 59, 1964, pp. 881-894.

23. Kadane, J.B., Comparisons of k-Class Estimators when the Disturbances are Small, Econometrica, Vol. 39, 1971, pp. 723-737

24. Kendall, M.G. and Stuart, A., The Advanced Theory of Statistics, Volume II, London, Griffin, 1967.

25. Luke, Y.L., The Special Functions and Their Approximations, Volume I, New York, Academic Press, 1969.

26. Maddala, G.S., Some Small Sample Evidence on Tests of Significance in Simultaneous Equations Models, Econometrica, Vol. 42, 1974, pp. 841-851.

27. Malinvaud, E., Statistical Methods of Econometrics, Amsterdam, North Holland, 1966.

28. Mariano, R.S., The Existence of Moments of the Ordinary Least Squares and Two-Stage Least Squares Estimators, Econometrica, Vol. 40, 1972, pp. 643-652.

29. Mariano, R.S. and Ramage, J.G., Finite Sample Effects of Misspecification on Structural Coefficient Estimators, Discussion Paper, No. 320. The Warton School, Department of Economics, University of Pennsylvania, 1975.

30. Mariano, R.S. and Sawa, T., Exact Finite-Sample Distribution of the Limited Information Maximum Likelihood Estimator for Two Included Endogenous Variables, Journal of the American Statistical Society, Vol. 67, 1972, pp. 159-163.

31. Marriott, F.H.C. and Pope, J.A., Bias in Estimation of Auto-correlations, Biometrika, Vol. 41, 1954, pp. 390-402.

32. McCarthy, M.D., A Note on the Forecasting Properties of 2SLS Restricted Reduced Forms, International Economic Review, Vol. 13, 1972, pp. 757-761.

33. Mikhail, W.M., Simulating the Small-Sample Properties of Econometric Estimators, Journal of the American Statistical Association, Vol. 67, 1972, pp. 620-624.

34. Miller, R.G., A Trustworthy Jackknife, Annals of Mathematical Statistics, Vol. 35, 1964, pp. 1594-1605.

35. Miller, R.G., The Jackknife - A Review, Biometrika, Vol. 61, 1974, pp. 1-15.

36. Miller, R.G., An Unbalanced Jackknife, Annals of Statistics, Vol. 2, 1974, pp. 880-891.

37. Nagar, A.L., The Bias and Moment Matrix of the General k-Class Estimators of the Parameters in Simultanecus Equations, Econometrica, Vol. 27, 1959, pp. 575-595.

38. Neave, H.R., On Using the Box-Muller Transformation with Multiplicative Congruential Pseudo-Random Number Generators, Applied Statistics, Vol. 22, 1973, pp. 92-97.

39. Orcutt, G.H. and Winokur, H.S., First Order Autoregression: Inference, Estimation and Prediction, Econometrica, Vol. 37, 1969, pp. 1-14.

40. Owen, A.D., A Proof that Both the Bias and the Mean Square Error of the Two-Stage Least Squares Estimator are Monotonically Non-Increasing Functions of Sample Size, Econometrica, Vol. 44, 1976, pp. 409-411.

41. Owen A.D. and Knight, J.L., The Exact Finite Sample Properties of Simultaneous Equation Estimators and Associated Statistics: A Synthesis and Classified Bibliography, Unpublished Manuscript, 1976.

42. Phillips, G.D.A., Recursions for the Two Stage Least Squares Estimators, Journal of Econometrics Vol. 6, 1977, pp. 65-77.

43. Pike, M.C. and Hill, I.D., Pseudo-Random Numbers (Algorithm 266), Communications of the Association for Computing Machinery, Vol. 8, 1965, pp. 605-606.

44. Quenouille, M.H., Approximate Tests of Correlation in Time-Series, Journal of the Royal Statistical Society, Series B, Vol. 11, 1949, pp. 68-84.

45. Quenouille, M.H., Notes on Bias in Estimation, Biometrika, Vol. 43, 1956, pp. 353-360.

46. Rao, C.R., Linear Statistical Inference and its Applications, New York, Wiley, 1965.

47. Rao, J.N.K., A Note on the Estimation of Ratios by Quenouille's Method, Biometrika, Vol. 52, 1965, pp. 647-649.

48. Rao, J.N.K. and Webster, J., On Two Methods of Bias Reduction in the Estimation of Ratios, Biometrika, Vol. 53, 1966, pp. 571-577.

49. Relph, A.P., Confluent Hypergeometric (Algorithm 192), Communications of the Association for Computing Machinery, Vol. 6, 1963, p. 388.

50. Revankar, N.S., The Two-Stage Least Squares Estimators as Aitken Estimators and the Asymptotic Tests of Significance, Discussion Paper No. 174, Department of Economics, State University of New York at Buffalo.

51. Revankar, N.S. and Hartley, M.J., An Independence Tests and Conditional Unbiased Predictions in the Context of Simultaneous Equation Systems, International Economic Review, Vol. 14, 1973, pp. 625-631.

52. Richardson, D.H., The Exact Distribution of a Structural Coefficient Estimator, Journal of the American Statistical Association, Vol. 63, 1968, pp. 1214-1226.

53. Richardson, D.H. and Rohr, R.J., Distribution of a Structural t Statistic for the Case of Two Included Endogenous Variables, Journal of the American Statistical Association, Vol. 66, 1971, pp. 375-382.

54. Richardson, D.H. and Rohr, R.J., An Experimental Study of Two Stage Least Squares Distributions in Equations Containing Three Endogenous Variables, Unpublished Manuscript, 1975.

55. Richarson, D.H. and Wu, De-Min, A Note on the Comparison of Ordinary and Two-Stage Least Squares Estimators, Econometrica, Vol. 39, 1971, pp. 973-981.

56. Robson, D.S. and Whitlock, J.H., Estimation of a Truncation Point, Biometrika, Vol. 51, 1964, pp. 33-39.

57. Rohr, R.J., On the Distribution of Two Alternative Structural t Statistics, Journal of the American Statistical Association, Vol. 71, 1976, pp. 731-736.

58. Sawa, T., The Exact Sampling Distribution of Ordinary Least Squares and Two-Stage Least Squares Estimators, Journal of the American Statistical Association, Vol. 64, 1969, pp. 923-937.

59. Sawa, T., Finite-Sample Properties of the k-Class Estimators, Econometrica, Vol. 40, 1972, pp. 653-680.

60. Sawa, T., Almost Unbiased Estimator in Simultaneous Equations Systems, International Economic Review, Vol. 14, 1973, pp. 97-106.

61. Sawa, T., The Mean Square Error of a Combined Estimator and Numerical Comparison with the TSLS Estimator, Journal of Econometrics, Vol. 1, 1973, pp. 115-132.

62. Schucany, W.R., Gray, H.L. and Owen, D.B., On Bias Reduction in Estimation, Journal of the American Statistical Association, Vol. 66, 1971, pp. 524-533.

63. Siegel, S., Nonparametric Statistics for the Behavioural Sciences, New York, McGraw-Hill, 1956.

64. Slater, L.J., Confluent Hypergeometric Functions, London, Cambridge University Press, 1960.

65. Sullins, W.L., Certification of Algorithm 266, Communications of the Association for Computing Machinery, Vol. 15, 1972, pp. 1072-1073.

66. Summers, R., A Capital Intensive Approach to the Small Sample Properties of Various Simultaneous Equation Estimators, Econometrica, Vol. 33, 1965, pp. 1-41.

67. Takeuchi, K., Exact Sampling Moments of the Ordinary Least Squares, Instrumental Variable and Two-Stage Least Squares Estimators, International Economic Review, Vol. 11, 1970, pp. 1-12.

68. Teekens, R., Predictive Methods in Multiplicative Models, Rotterdam, Rotterdam University Press, 1972.

69. Thacher, H.C., Certification of Algorithm 191 and 192 Hypergeometric Confluent Hypergeometric, Communications of the Association for Computing Machinery, Vol. 7, 1964, p. 244.

70. Theil, H., Economic Forecasts and Policy, Second Edition, Amsterdam, North-Holland, 1961.

71. Theil, H., Principles of Econometrics, New York, Wiley, 1971.

72. Tukey, J.W., Bias and Confidence in Not-Quite Large Samples (Abstract), Annals of Mathematical Statistics, Vol. 29, 1958, p. 614.

73. Walsh, J.E., Concerning the Effect of Intraclass Correlation on Certain Significance Tests, Annals of Mathematical Statistics, Vol. 18, 1947, pp. 88-96.

THE JACKKNIFE STATISTIC:
AN APPLICATION IN ECONOMETRICS

by

Anthony David Owen
B.A. (Leicester)
M.A. (Kent)

Thesis submitted to the University of Kent in fulfilment of
the requirements for the degree of
Doctor of Philosophy
1977

I hereby declare that the research embodied in this thesis is my own work and that it has not previously been submitted for a degree at this or any other University.

A.D. Owen

ABSTRACT

Quenouille has developed a procedure, later termed the jackknife by Tukey, for reducing the bias of a consistent estimator of an unknown parameter. A measure of the variance of the resulting estimator can be obtained and used to provide approximate confidence intervals and tests of significance. Thus the jackknife technique may be especially interesting when the estimator under consideration is biased but consistent and mathematically intractable distribution theory prevents the construction of exact confidence intervals.

Considerable research has been devoted to studying the jackknife technique, predominantly in the fields of biometrics, statistics and numerical analysis. So far the use of the jackknife method in econometrics has been negligible, although one very important class of econometric estimators, the simultaneous equation estimators, is biased in finite samples and, in general, has a mathematically intractable distribution.

In this thesis we investigate the application of the jackknife technique to the Two-Stage Least Squares (2SLS) structural parameter estimator in a simultaneous equation system. The bias reducing property was found to be present in the majority of cases considered in an investigation of the effects of jackknifing on the exact bias of the 2SLS estimator in a two equation model. Conditions are given for which it is unlikely that jackknifing will reduce the bias of the 2SLS estimator.

Since the exact variance of the jackknifed 2SLS estimator is unknown, an examination of the effect on the variance of 2SLS of applying the jackknife had to be made by a simulation experiment. Whilst the 2SLS estimator always had a smaller mean square error than

the jackknifed 2SLS estimator, a comparison of absolute errors rarely produced a significant difference between them.

Finally, it was observed that t statistics formed using the 2SLS estimator may not be distributed according to the Student t distribution. The actual distribution may be highly skewed and serious errors could result if the postulated theoretical distribution was used for statistical inference. In general, this feature was less noticeable for the J2SLS estimator which appeared to have a reasonably symmetric distribution, and consequently there is less likelihood of serious errors being made if the postulated theoretical distribution is used for the purpose of statistical inference.

ACKNOWLEDGEMENTS

Certain parts of the work contained in this thesis have been presented at

The Third World Econometric Congress (with G.D.A. Phillips)(1975)

and The Third Australian Statistical Conference (1976).

I am grateful for comments received at these two conferences, and for comments received from a Staff Seminar in Statistics at the Australian National University (1976).

Material extracted from Chapter 5 has been published in Econometrica (1976), whilst material extracted from Chapter 4 and Appendix C is forthcoming in the Journal of Statistical Computation and Simulation. My thanks are due to the editors and anonymous referees of both journals for their constructive comments.

For advice, comments and criticism during the course of the research work on which this thesis is based, I am indebted to my supervisor, Mr G.D.A. Phillips.

To my parents, to whom I dedicate this work, I owe a great debt for their encouragement throughout my University career. Finally, to my wife, Jackie, who has been a source of continual encouragement I extend my most affectionate thanks.

TABLE OF CONTENTS

Page

LIST OF TABLES

"Grown-ups love figures. When you tell them you have made a new friend, they never ask you any questions about essential matters. They never say to you, 'What does his voice sound like? What games does he love best? Does he collect butterflies?' Instead, they demand: 'How old is he? How many brothers has he? How much does he weigh? How much money does his father make?' Only from these figures do they think they have learned anything about him."

Antoine de Saint-Exupéry, The Little Prince

CHAPTER 1

INTRODUCTION

Quenouille [45] has developed a technique, later termed the jackknife by Tukey [72], for reducing the bias which may be present in an (otherwise consistent) estimator of an unknown parameter. Quenouille's original justification for using the technique was based upon the assumption of the existence of a Taylor series expansion for the bias of an estimator whereupon, by applying the jackknife technique, the bias term to order ($^1/_N$) could be removed. In addition to its bias reducing properties, the jackknife technique can also be used to provide approximate confidence intervals and tests of significance. Thus the jackknife technique is a viable proposition where the estimator under consideration is biased, but consistent, and/or where mathematically intractable distribution theory prohibits the formation of exact confidence intervals.

Considerable research has been devoted to studying the jackknife technique, predominantly in the fields of biometrics, statistics and numerical analysis. Its use in econometrics has been negligible, yet a class of consistent econometric estimators possess both bias and intractable distribution theory in finite samples, which would suggest that application of the jackknife technique may be a fruitful exercise. This class of estimators is the class of simultaneous equation estimators.

This thesis considers the effects of applying the jackknife technique to one of this class of estimators, the Two-Stage Least Squares (2SLS) estimator.

2SLS is a "limited information" estimator in the sense that it estimates the equations comprising a simultaneous economic system one at a time. In order to estimate any one equation, 2SLS only requires a specification of the equation being estimated and a list of the other predetermined variables appearing in the system. It does not therefore take account of contemporaneous correlation between the disturbances of the equations in the system. Neither does it use the information contained in the overidentifying restrictions on the other equations in the system. Consequently, if the entire system has been specified, 2SLS may not make the most effective use of all the available information and a "full-information" estimator may be preferred. Under the assumption that the hypothesized model is correctly specified, the most efficient method of estimation would be one of the full information methods. Most economists, however, would consider this assumption rather heroic and would select one of the limited information estimators in order to isolate the deleterious effects of any specification errors to the equations in which they arise.

There are two reasons for selecting the 2SLS estimator from such a wide class of estimators.

Firstly, the exact bias (and higher order moments) of the 2SLS estimator have been derived for a two equation model and this allows an exact investigation of the jackknife's bias reducing ability vis-a-vis 2SLS, albeit under rather restrictive assumptions.

Secondly, the other limited information simultaneous equation estimators of any importance are the Ordinary Least Squares (OLS) and the Limited Information Maximum Likelihood (LIML) estimators. OLS is not a candidate for jackknifing since it contravenes Quenouille's assumption of a consistent estimator, whilst the non-finite moments of the LIML estimator (see Mariano and Sawa [30]) precludes any examination of the effects of the jackknife technique on its "bias". In addition, within the class of limited information simultaneous equation estimators, on the basis of numerous Monte Carlo results (the major works are summarized in Johnston [20], Chapter 13, Section 8) 2SLS is generally preferred on the grounds of "all-round" performance and computational efficiency and simplicity.

"Full information" methods of estimation were not considered as possible candidates for jackknifing as this would seem to be the logical step forward after the limited information estimators had been considered. This point is discussed further in Chapter 8.

The general form of the simultaneous equation system which will be used throughout this thesis, together with the relevant notation and assumptions, is defined in Chapter 2. The 2SLS estimator and its asymptotic properties are derived for the parameters of any single equation in the system. Conditions and assumptions under which the exact finite sample results of the 2SLS estimator have been derived are also stated.

Chapter 2 continues with a description of the jackknife statistic, its bias reducing properties, and its use in formulating approximate confidence intervals and tests of significance. The literature on the jackknife and its applications is so extensive that only (what the author considers to be) the more relevant works are cited, although a bibliographical reference is given.

The asymptotic properties of the jackknife 2SLS (J2SLS) estimator are investigated in Chapter 3. A proof of the asymptotic equivalence of the J2SLS and 2SLS estimators is given, and a t ratio formed using the J2SLS estimator is shown to be asymptotically distributed as the standardized normal distribution.

The small sample properties of the J2SLS estimator are investigated by a series of simulation experiments in Chapters 5, 6 and 7. The computer algorithms used in the experiments are described in Chapter 4 together with results of verification where they do not already exist. A formula given in Chapter 3 reduces the computational burden involved in calculating J2SLS parameter estimates, and should reduce the probability of significant inaccuracies due to the build-up of rounding errors resulting from repeated use of the matrix inversion algorithm. Chapter 4 also contains a method for evaluating the accuracy of the asymptotic approximations to the exact moments of the 2SLS estimator.

For an equation containing just two endogenous variables the exact first and second order moments of the 2SLS estimator have been derived. It is relatively easy to adapt the exact bias of the 2SLS estimator to obtain the exact bias of the J2SLS estimator, but the exact mean square error of the J2SLS estimator has not, as yet, been derived. In Chapter 5 the exact relative biases of the 2SLS and J2SLS estimators are compared, under conditions which prevail for "exact" theory, by means of a simulation experiment. This experiment gives exact results on the ability of the jackknife to reduce the bias of the 2SLS estimator. For the general model, however, this form of analysis is not possible, and the author has only been able to derive a rather weak conditon under which jackknifing is "unlikely" to reduce the bias of the 2SLS estimator.

Chapter 6 presents the results of a Monte Carlo experiment into the properties of the two estimators. Comparisons of relative bias, mean square error and mean absolute error are made using a two equation model. The use of the jackknife statistic to form approximate confidence intervals and tests of significance using the 2SLS estimator is also investigated and the results are presented in Chapter 7. It is well known that standardized normal ratios and t ratios formed using the 2SLS estimator are only valid asymptotically, and that in small samples they could diverge significantly from their postulated theoretical distributions. A comparison of the small sample distributions of test statistics using both 2SLS and J2SLS estimators is made.

Concluding remarks are contained in Chapter 8.

CHAPTER 2

THE TWO-STAGE LEAST SQUARES ESTIMATOR AND THE BIAS REDUCING PROPERTIES OF THE JACKKNIFE STATISTIC

2.1 The General Linear Simultaneous Equations Model

2.1.1 Specification of the Model

The analysis in this thesis is concerned with a simultaneous economic system of G linear stochastic equations relating G endogenous (or jointly-dependent) variables and K exogenous variables, which can be written as

$$\bar{Y}B + X\Gamma = U \quad . \tag{2.1}$$

We are interested in the estimation of just one equation from this system, (say) the jth, which can be written as

$$\underline{y}_j = Y_j\underline{\beta}_j + X_{1j}\underline{\gamma}_{1j} + X_{2j}\underline{\gamma}_{2j} + \underline{u}_j \ , \tag{2.2}$$

and we will refer to this equation as the jth structural equation $(j = 1,2,\ldots,G)$. For notational simplicity we will generally omit the j subscript.

2.1.2 Notation

$\bar{Y}$ is a matrix of N observations on the G endogenous variables in the entire system;

$\underline{y}$ is a vector of N observations on the "dependent" endogenous variable;

Y is a matrix of N observations on the other g endogenous variables included in the jth equation. In the unlikely event that all G endogenous variables appear in the jth equation then $g = G-1$ and

$[\underline{y} : Y] = \overline{Y}$, otherwise $g < G-1$;

X is a matrix of N observations on K exogenous variables partitioned as $X = [X_1 : X_2]$;

X_1 is a matrix of observations on the K_1 exogenous variables included in the jth equation;

X_2 is a matrix of observations on the K_2 exogenous variables excluded from the jth equation (i.e. $K = K_1 + K_2$);

U is a matrix of N unobservable disturbances for each of the G equations, with jth column denoted by $\underline{u}_j$;

B is a $G \times G$ matrix of unknown structural coefficients;

$\underline{\beta}$ is an unknown g component sub-vector of B with non-zero elements.

Γ is a $K \times G$ matrix of unknown structural coefficients;

$\underline{\gamma}_1$ is a K_1 component sub-vector of Γ with non-zero elements;

$\underline{\gamma}_2$ is a K_2 component sub-vector of Γ with zero elements.

2.1.3 Basic Assumptions

The following conventional assumptions are made for the system (2.1), and for the jth structural equation (2.2):

(i) B is non-singular;

(ii) the jth structural equation, (2.2), is just- or over-identified by zero restrictions on the structural coefficients, i.e. $K_2 \geq g$;

(iii) the matrix X consists of non-stochastic elements and is of full rank, K. Further, as $N \to \infty$ the matrix $N^{-1}(X'X)$ converges to a finite matrix, denoted by

$$\lim_{N \to \infty} \frac{1}{N} \cdot (X'X) = \Sigma_{XX} ,$$

where Σ_{XX} is a finite positive definite matrix ;

(iv) the sample size (N) is greater than the total number of exogenous variables (K) in the system;

(v) the N rows of U are independently and identically distributed with zero mean vector and unknown finite covariance matrix, Σ. In addition, the analysis in Chapter 3 requires that the structural disturbances have finite fourth order moment.

Postmultiplying equation (2.1) by B^{-1} we obtain the reduced form of the system, which can be written as

$$\bar{Y} = X\bar{\Pi} + \bar{V} , \qquad (2.3)$$

where $\bar{\Pi} = -\Gamma B^{-1}$,

and $\bar{V} = U B^{-1}$.

The reduced form equation for the jth "dependent" endogenous variable and the reduced form equations for the g "explanatory" endogenous variables can be written as

$$\underline{y}_j = X\underline{\pi}_j + \underline{v}_j$$

and

$$Y_j = X\Pi_j + V_j \qquad (2.4)$$

respectively. Since, for notational convenience, we are omitting the j subscripts, this explains the necessity to write equation (2.3) in the above form rather than in the more common form which would coincide with equation (2.4) when the subscripts are omitted.

2.1.4 The Two Endogenous Variables Case

The majority of results on the exact properties of the 2SLS estimator have been derived under the assumption that $g = 1$, i.e. the equation being estimated contains only two endogenous variables. In addition, it is assumed that the matrices X_1 and X_2 contain no lagged endogenous variables.

Under the above conditions, the first structural equation can be written as

$$\underline{y}_1 = \underline{y}_2\beta + X_1\underline{\gamma}_1 + X_2\underline{\gamma}_2 + \underline{u}_1 , \qquad (2.5)$$

with reduced form equations

$$\underline{y}_1 = X_1\underline{\pi}_{11} + X_2\underline{\pi}_{12} + \underline{v}_1$$

$$\underline{y}_2 = X_1\underline{\pi}_{21} + X_2\underline{\pi}_{22} + \underline{v}_2 ,$$

where $\underline{\pi}_{11}$, $\underline{\pi}_{12}$, and $\underline{\pi}_{21}$, $\underline{\pi}_{22}$ are vectors of constant coefficients. The random vector $(\underline{v}_1' : \underline{v}_2')$ is assumed to be distributed as bivariate normal with zero mean and positive definite covariance matrix $\Omega \otimes I_N$, where $\Omega = \omega_{ij}$ $(i,j = 1,2)$ is a matrix of reduced form parameters.

2.2 The Two-Stage Least Squares Estimator

It is well known that OLS is, in general, an inconsistent estimator of the parameters in the structural equation (2.2). This inconsistency is due to the correlation between the explanatory endogenous variables (Y) and the vector of structural disturbances ($\underline{u}$). Basmann [3] and Theil [70] derived, independently, an alternative estimator which "purges" Y of the stochastic component associated with the disturbance term, and then estimates the revised equation by OLS. This "alternative" estimator is called the Two-Stage Least Squares Estimator.

From equation (2.2) we write the jth structural equation as

$$\underline{y} = Y\underline{\beta} + X_1\underline{\gamma}_1 + \underline{u} .$$

If we rewrite the above equation as

$$\underline{y} = (Y - V)\underline{\beta} + X_1\underline{\gamma}_1 + \underline{u} + V\underline{\beta} ,$$

then using equation (2.4), $(Y - V) = X\Pi$ is uncorrelated with $(\underline{u} + V\underline{\beta})$ since X is non-stochastic by assumption (iii).

Since V is unobservable we must use its estimated value $\hat{V}$, where $\hat{V} = Y - X\hat{\Pi}$. Provided $\operatorname{plim}_{N\to\infty} \hat{\Pi} = \Pi$, it follows that $\operatorname{plim}_{N\to\infty} (Y - \hat{V}) = (Y - V)$ and hence $(Y - \hat{V})$ and $(\underline{u} + \hat{V}\underline{\beta})$ are asymptotically uncorrelated.

Thus if the least squares estimator is applied to

$$\underline{y} = (Y - \hat{V})\underline{\beta} + X_1\underline{\gamma}_1 + \underline{u} + \hat{V}\underline{\beta} ,$$

we can obtain consistent estimates of $\underline{\beta}$ and $\underline{\gamma}_1$. Since this process of estimation involves two successive applications of least squares it is known as Two-Stage Least Squares (2SLS).

In this thesis we shall work with the Instrumental Variables type formulation of the 2SLS estimator, viz:

$$\hat{\underline{\theta}} = \left[Z'X(X'X)^{-1}X'Z\right]^{-1} Z'X(X'X)^{-1} X'\underline{y} , \tag{2.6}$$

where $Z = \left[Y : X_1\right]$ and $\hat{\underline{\theta}}' = \left[\hat{\underline{\beta}}' : \hat{\underline{\gamma}}_1'\right]$.

In order to apply tests of significance, knowledge of the distribution of the 2SLS estimator is required. The finite sample distribution of 2SLS is only known for a few specific cases, thus reliance is usually placed upon its asymptotic distribution.

Substituting for $\underline{y}$ in equation (2.6) we obtain

$$\hat{\underline{\theta}} - \underline{\theta} = \left[Z'X(X'X)^{-1}X'Z\right]^{-1} Z'X(X'X)^{-1} X'\underline{u} ,$$

and we require the limiting distribution of the sequence

$$\sqrt{N}(\hat{\underline{\theta}} - \underline{\theta}) = \left[\frac{1}{N} . Z'X\left(\frac{1}{N} . X'X\right)^{-1} \frac{1}{N} . X'Z\right]^{-1} \frac{1}{N} . Z'X\left(\frac{1}{N} . X'X\right)^{-1} \frac{1}{\sqrt{N}} . X'\underline{u} .$$

Since X is (by assumption) non-stochastic, it follows that

$$\frac{1}{N} . X'Z = \frac{1}{N} . X'[Y : X_1] = \frac{1}{N} . X'[X\Pi + V : X_1]$$

converges in probability to a finite limit, denoted by

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N} . X'Z = \Sigma_{XZ} \quad .$$

We have already assumed the existence of a finite limit for $N^{-1} . X'X$, and thus we can denote its inverse by

$$\lim_{N\to\infty} \left(\frac{1}{N} . X'X \right)^{-1} = \Sigma_{XX}^{-1} \; .$$

Under assumption (V), modified application of the Lindeberg-Levy theorem (see e.g. Theil [71; pp.498-499]) using the above results will yield

$$\sqrt{N}\,(\underline{\hat{\theta}} - \underline{\theta}) \sim N \left[0, \sigma^2 \operatorname*{plim}_{N\to\infty} \left\{ \frac{1}{N} . Z'X \left(\frac{1}{N} . X'X \right)^{-1} \frac{1}{N} . X'Z \right\}^{-1} \right] , \qquad (2.7)$$

where σ^2 denotes the variance of the jth structural disturbance, i.e. the jjth component of Σ.

A consistent estimator of σ^2 is given by

$$\hat{\sigma}^2 = \underline{\hat{u}}'\underline{\hat{u}} \,/ (N - K_1 - g) \; ,$$

where $\underline{\hat{u}} = [\underline{y} - Y\underline{\hat{\beta}} - X_1 \underline{\hat{\gamma}}_1]$.

Since the asymptotic covariance matrix of the 2SLS estimator coincides with the Cramer-Rao bound (when the structural disturbances are normally distributed), 2SLS is an efficient estimator in its class of limited information simultaneous equation estimators. Its relative (small sample) efficiency however has not, in general, been ascertained.

2.3 The Jackknife Statistic

2.3.1 Definition

Let α be an unknown parameter, and let $X_1, X_2, \ldots, X_N$ be N independently and identically distributed observations from the cumulative distribution function F_α. Further, let $\hat{\alpha}$ be a biased estimator of α such that

$$E(\hat{\alpha} - \alpha) = \frac{a_1}{N} + \frac{a_2}{N^2} + \ldots . + \frac{a_r}{N^r} + \ldots . , \qquad (2.8)$$

where a_1, a_2, ..., a_r are constants and not dependent upon N. If the N observations can be divided into n groups, each of r observations (i.e. $N = nr$), then the estimator

$$J_i(\hat{\alpha}) = n\hat{\alpha} - (n-1)\hat{\alpha}_i , \qquad (i = 1,2,\ldots,n)$$

where $\hat{\alpha}_i$ denotes the estimate of α obtained with the ith group of observations omitted, removes the term in 1/N from equation (2.8).

Applying the technique to equation (2.8) gives

$$E[J_i(\hat{\alpha})] = n\alpha + \frac{a_1}{r} + \frac{a_2}{r^2 n} + \frac{a_3}{r^3 n^2} + \ldots\ldots$$

$$- (n-1)\alpha - \frac{a_1}{r} - \frac{a_2}{r^2(n-1)} - \frac{a_3}{r^3(n-1)^2} - \ldots\ldots ;$$

$$\text{i.e. } E[J_i(\hat{\alpha})] = \alpha - \frac{a_2}{r^2 n(n-1)} - \frac{(2n-1)a_3}{r^3 n^2 (n-1)^2} - \ldots . .$$

$(i = 1,2,\ldots,n)$.

Tukey,[1] in unpublished work, has named $J_i(\hat{\alpha})$ the pseudo-jackknife estimator. He defined the jackknife estimator, $J(\hat{\alpha})$, as the average of the i pseudo-jackknife values ($i = 1,2,\ldots, n$), i.e.

$$J(\hat{\alpha}) = \frac{1}{n} \sum_{i=1}^{n} J_i(\hat{\alpha}) = n\hat{\alpha} - \frac{(n-1)}{n} \sum_{i=1}^{n} \hat{\alpha}_i . \qquad (2.9)$$

---

1. The definition that follows is taken from Brillinger [7] who cites an unpublished paper and an abstract [72] of a conference paper by Tukey.

$J(\hat{\alpha})$ will have the same expected value as $J_i(\hat{\alpha})$, but a smaller variance. The term jackknife was coined for this procedure since it shared two characteristics with a boy scout's jackknife:

(i) wide applicability to many different problems;

(ii) inferiority to special tools for those problems for which special tools have been designed.

In most problems however the property of removing bias would not be sufficient to recommend the use of the jackknife. A comparison of the dispersion of the original estimator with that of the jackknife estimator is needed. Tukey noted that not only are the pseudo-jackknife estimates nearly unbiased, but their average sum of squares of deviations is nearly N(N-1) times the variance of their means. He proposed that in many instances the $J_i(\hat{\alpha})$ are approximately independently and identically distributed and hence an approximate estimate of the variance of $J(\hat{\alpha})$ is given by

$$\sum_{i=1}^{n} \frac{[J_i(\hat{\alpha}) - J(\hat{\alpha})]^2}{n(n-1)}, \tag{2.10}$$

whilst

$$\frac{[J(\hat{\alpha}) - \alpha]}{\sqrt{\sum_{i=1}^{n} \frac{[J_i(\hat{\alpha}) - J(\hat{\alpha})]^2}{n(n-1)}}} \tag{2.11}$$

is approximately distributed as a t variate with (n-1) degrees of freedom.

The jackknife can be re-applied in order to remove the bias term of order $1/N^2$ which remains after the initial application. Quenouille [45] and Kendall and Stuart [24] give a formula to achieve this further bias reduction, but if $a_k = 0$ for all $k > 2$ then the second application of the jackknife does not yield an <u>exactly</u> unbiased statistic as one would have desired. Schucany, Gray and Owen [62] give a higher order transformation which provides an algorithm for eliminating, exactly, bias terms of higher order.

This thesis considers the jackknife technique when $r = 1$ (i.e. $N = n$). Thus each of the N pseudo-jackknife estimates is calculated from the total number of observations less one, and the jackknife statistic is defined as

$$J(\hat{\alpha}) = N\hat{\alpha} - \frac{(N-1)}{N} \sum_{i=1}^{N} \hat{\alpha}_i \ . \qquad (2.12)$$

Intuitively, choosing $r = 1$ is appealing since problems of dividing up samples and being left with awkward remainders are avoided. In addition, several studies involving applications of the jackknife have found $r = 1$ to be the "optimal" value of r (e.g. see Robson and Whitlock [56] and Rao [47]).

2.3.2 The Generalized Jackknife

Schucany et al. [62] provide a general method for bias reduction which includes the jackknife as a special case. Suppose that there are $k + 1$ biased estimators of α, viz: $\hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_{k+1}$, defined over the N(=n) observations, and further suppose that the biases of these $k + 1$ estimators can be written as

$$E(\hat{\alpha}_i) - \alpha = \sum_{j=1}^{k} f_{ij}(N) b_j(\alpha) \ , \qquad (i = 1,2,\ldots,k+1)$$

then the estimator

$$\tilde{\alpha}^{(k)} = \frac{\begin{vmatrix} \hat{\alpha}_1 & \hat{\alpha}_2 & \cdots & \hat{\alpha}_{k+1} \\ f_{11} & f_{12} & \cdots & f_{1,k+1} \\ \vdots & \vdots & & \vdots \\ f_{k1} & f_{k2} & & f_{k,k+1} \end{vmatrix}}{\begin{vmatrix} 1 & 1 & \cdots & 1 \\ f_{11} & f_{12} & \cdots & f_{1,k+1} \\ f_{21} & f_{22} & \cdots & f_{2,k+2} \\ \vdots & \vdots & & \vdots \\ f_{k1} & f_{k2} & \cdots & f_{k,k+1} \end{vmatrix}} \qquad (2.13)$$

reduces the order of bias to terms of order (k + 1) in 1/N, i.e.

$$E\left[\tilde{\alpha}^{(k)}\right] - \alpha = O\left[N^{-(k+1)}\right] ,$$

where the argument of the f_{ij} functions has been suppressed for notational convenience, and these functions are assumed to be known. Further, it is assumed that $1 \le k \le N-1$ and that the denominator of equation (2.13) is non-zero.

If $k = 1$, then $\hat{\alpha}_1 = \hat{\alpha}$, $\hat{\alpha}_2 = \frac{1}{N} \sum_{i=1}^{N} \hat{\alpha}_i$,

$$f_{11}(N) = \frac{1}{N} , \quad \text{and} \quad f_{12}(N) = \frac{1}{(N-1)} ,$$

and equation (2.13) reduces to the "regular" jackknife as defined by equation (2.12).

The formula given by equation (2.13) is exact, in the sense that if the bias of the original estimator takes the form of equation (2.8) with only the first k terms non-zero, application of $\tilde{\alpha}^{(k)}$ will remove all bias.

Schucany et al. only considered the problem of bias reduction. The effect of their higher order transformation on the variance of $\tilde{\alpha}^{(k)}$ was not investigated for the general case.

2.4 Previous Applications of the Jackknife Technique in Econometrics

2.4.1 Partial Correlation Coefficient

If the estimated value of the partial correlation coefficient is used as an approximate test for serial correlation in time series, Quenouille [44] has shown that the bias of the estimator is inversely proportional to the sample size, N. He suggested using (what later became known as) the jackknife technique with $n = 2$, i.e. the sample was split in half, in order to remove the bias term of order $(1/N)$. In a later paper, Quenouille [45] generalized this procedure by noting that the same amount of bias reduction could be achieved by splitting the sample into n groups each of size r (where $N = nr$).

2.4.2 Autoregressive Processes

Quenouille's [44] original method of jackknifing (i.e. $n = 2$) was later applied by Orcutt and Winokur [39] to the least squares estimator in an attempt to reduce the bias of $\hat{\beta}$ (the least squares estimator of β) in the autoregressive process

$$y_t' = \alpha + \beta y_{t-1} + \varepsilon_t \qquad (t = 1,2,\ldots,N)$$

(ε_t normally and independently distributed).

Using a Monte Carlo study they compared sample means and mean square errors of three estimators of β: least squares, jackknife least squares, and an estimator based upon correcting the bias of least squares using an expression derived by Marriott and Pope [31]. Whilst both modified least squares estimators reduced bias, the jackknife estimator exhibited a larger mean square error than the other two estimators and consequently was not to be preferred.

2.5 Use of the Jackknife Technique in Other Disciplines

A substantial body of literature on the application of the jackknife technique in various disciplines has evolved since Tukey's [72] initial conjecture. A survey of these applications, together with a comprehensive bibliography, has been compiled by Miller [35]. With the exception of the two papers cited in the previous section, few of the applications have any direct relevance to econometrics.

Perhaps the most successful area in which the jackknife has been used to date is that of ratio estimation. Given a bivariate sample (X_i, Y_i) $(i = 1,2,\ldots,T)$ from a population of size N $(T < N)$ with means μ and η respectively, we are interested in estimation of the ratio $R = \mu/\eta$. In many instances the classical ratio estimator $r = \overline{X}/\overline{Y}$ (i.e. the ratio of sample means) with $\overline{X}$ known, may exhibit a large bias compared to its standard error in surveys with many strata and small samples within strata. Durbin [14] suggested the jackknife with $n = 2$ as a bias reducing tool and investigated its properties under two distributional assumptions on the stochastic error term in the general linear model. Under both assumptions the jackknife not only reduced the bias of the ratio estimator, but also reduced the mean square error. Rao [47] and Rao and Webster [48] showed that the optimal choice of n under both of Durbin's [14] distributional assumptions is $n = N$.

Subsequent research investigated the performance of the jackknife in ratio estimation as compared with several other estimators. In general, the jackknife appeared to rank close behind the most efficient estimators but had the disadvantage of being more complicated to compute.

An application of the jackknife with direct relevance to econometrics is Miller's [36] proof that the jackknife OLS estimator of the vector of parameters in the general linear model is asymptotically normally distributed under conditions that do not require the vector of stochastic disturbances to be normally distributed. He conjectured that his proof extended to the case of non-linear least squares.

The jackknife has also been applied in the areas of maximum likelihood estimation, functions of a U-statistic, stochastic processes, inference on variances, and multivariate analysis. This list is far from exhaustive and the interested reader is referred to Miller's [35] bibliography for additional areas of application, and his synthesis for a review of the performance of the jackknife statistic over the many disciplines in which it has been used.

2.6 Alternative Methods of Bias Reduction Using the 2SLS Estimator

2.6.1 General Remarks

Methods designed to reduce the bias of the 2SLS estimator, without increasing the mean square error, have been devised by Nagar [37] and Sawa [60, 61]. Strictly speaking neither author "manipulates" the 2SLS estimator specifically, but since both proposed estimators converge in distribution to the 2SLS estimator as the sample size increases indefinitely, they could offer themselves as alternatives to the J2SLS estimator, at least on a bias reduction criterion.

2.6.2 Nagar's Unbiased k-Class Estimator

Nagar [37] has derived an expression for the bias to order 1/N of a distribution approximating the distribution of the k-class estimators. He noted that for $k = 1 + \nu/N$, where ν is the degree of overidentification of the equation being estimated, the bias vanishes to order $1/N$. Asymptotically, Nagar's unbiased estimator is clearly equivalent to the 2SLS estimator.

Using Klein's model I, Nagar showed that whilst this choice of k certainly exhibited a smaller "bias" than the corresponding 2SLS estimator, 2SLS dominated on a "mean square error" criterion. Sawa [59], however, has shown (for a two endogenous variables model) that if $k > 1$ and nonstochastic then no moments of the k-class estimators are finite; hence Nagar's "unbiased" k-class estimator does not possess a finite first order, or any other order, moment.

2.6.3 Sawa's Combined Estimator

On the basis of an asymptotic expansion of the exact bias of the k-class estimators in a two endogenous variables model, Sawa [60] proposed an estimator which uses a weighted combination of the 2SLS and OLS estimators in order to remove the leading term of the asymptotic expansion. The weights are such that, asymptotically, Sawa's combined estimator converges to 2SLS.

In a series of experiments, the combined estimator dominated the 2SLS estimator (on a mean square error criterion) when the number of exogenous variables excluded from the equation being estimated was very large. The reduction in bias (over 2SLS) obtained by using the combined estimator was always evident and frequently substantial.

The experiments were only conducted for an equation containing just two endogenous variables. Sawa [61] justified the extension of his combined estimator to equations containing an arbitrary number of endogenous variables by using Kadane's [23] small σ approximations. As yet, however, no Monte Carlo results have been published on the relative merits of the combined estimator vis-a-vis other limited information estimators. Clearly if the combined estimator dominates other limited information estimators on a mean square error criterion only for a large number of excluded exogenous variables, a Monte Carlo study may be impracticable, or at least very expensive.

2.7 Justification for Applying the Jackknife Technique to the Two-Stage Least Squares Estimator

The author has been unable to produce a rigorous justification for applying the jackknife technique to the 2SLS estimator, as he cannot express the bias of 2SLS as a Taylor series expansion in terms of increasing powers of 1/N. Nagar [37], however, has shown that the bias of the 2SLS estimator can be approximated by an expression involving terms of increasing powers of order $(1/N^{\frac{1}{2}})$ in probability. In addition, using Kadane's [23] approximation to the bias of the 2SLS estimator, the author has been able to derive a condition under which the jackknife is "unlikely" to reduce the bias of the 2SLS estimator. This analysis is contained in Chapter 5.

Whilst these results cannot provide a rigorous justification for using the jackknife technique as a bias reducing tool, it suggests that its application may be worth pursuing.

CHAPTER 3

ASYMPTOTIC THEORY

3.1 Derivation of the Computing Formula for the J2SLS Estimator

From equation (2.6) the 2SLS estimator can be written, in instrumental variable form, as

$$\hat{\underline{\theta}} = [Z'X(X'X)^{-1}X'Z]^{-1}Z'X(X'X)^{-1}X'\underline{y} \quad . \tag{3.1}$$

We denote the 2SLS estimator of $\underline{\theta}$ based upon (N-1) observations as

$$\hat{\underline{\theta}}_i = [Z_i'X_i(X_i'X_i)^{-1}X_i'Z_i]^{-1} Z_i'X_i(X_i'X_i)^{-1}X_i'\underline{y}_i \ , \tag{3.2}$$

where the i subscript denotes that the ith observation $(i = 1,2,\ldots,N)$ has been removed from the relevant data matrix. Using Appendix A we can show that

$$(X_i'X_i)^{-1} = (X'X - \underline{x}_i\underline{x}_i')^{-1} = (X'X)^{-1} + \frac{(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}}{1 - \underline{x}_i'(X'X)^{-1}\underline{x}_i} \ , \tag{3.3}$$

where $\underline{x}_i$ (a K dimensional column vector) denotes the ith row of X; i.e. the ith observation on X.

Using equation (3.3), we can rewrite equation (3.2) as

$$\hat{\underline{\theta}}_i = \left\{ [Z'X-\underline{z}_i\underline{x}_i'] \left[(X'X)^{-1} + \frac{(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}}{1 - \underline{x}_i'(X'X)^{-1}\underline{x}_i} \right] [X'Z-\underline{x}_i\underline{z}_i'] \right\}^{-1}$$

$$\times \left\{ [Z'X-\underline{z}_i\underline{x}_i'] \left[(X'X)^{-1} + \frac{(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}}{1 - \underline{x}_i'(X'X)^{-1}\underline{x}_i} \right] [X'\underline{y}-\underline{x}_i y_i] \right\}, \tag{3.4}$$

where $\underline{z}_i$ (a $K_1 + g$ dimensional column vector) and y_i (scalar) denote the ith observations on Z and $\underline{y}$ respectively.

Consider the term to be inverted in equation (3.4). Upon expansion we obtain

$$\Bigg[Z'X(X'X)^{-1}X'Z + \underline{z}_i\underline{x}_i'(X'X)^{-1}\underline{x}_i\underline{z}_i' - Z'X(X'X)^{-1}\underline{x}_i\underline{z}_i' - \underline{z}_i\underline{x}_i'(X'X)^{-1}X'Z$$

$$+ \frac{Z'X(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}X'Z}{1-\underline{x}_i'(X'X)^{-1}\underline{x}_i} + \frac{\underline{z}_i\underline{x}_i'(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}\underline{x}_i\underline{z}_i'}{1-\underline{x}_i'(X'X)^{-1}\underline{x}_i}$$

$$- \frac{Z'X(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}\underline{x}_i\underline{z}_i'}{1-\underline{x}_i'(X'X)^{-1}\underline{x}_i} - \frac{\underline{z}_i\underline{x}_i'(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}X'Z}{1-\underline{x}_i'(X'X)^{-1}\underline{x}_i}\Bigg]^{-1} . \quad (3.5)$$

Let $P = Z'X(X'X)^{-1}X'Z$,

$s_i = \underline{x}_i'(X'X)^{-1}\underline{x}_i$, (scalar)

$b_i = \underline{x}_i'(X'X)^{-1}X'\underline{y} = \underline{x}_i'\hat{\underline{\pi}}$, (scalar)

and $\underline{a}_i = Z'X(X'X)^{-1}\underline{x}_i$,

then equation (3.5) can be rewritten as

$$\left[P + s_i\underline{z}_i\underline{z}_i' + \frac{\underline{a}_i\underline{a}_i'}{(1-s_i)} + \frac{s_i^2\underline{z}_i\underline{z}_i'}{(1-s_i)} - \underline{a}_i\underline{z}_i' - \underline{z}_i\underline{a}_i' - \frac{s_i\underline{a}_i\underline{z}_i'}{(1-s_i)} - \frac{s_i\underline{z}_i\underline{a}_i'}{(1-s_i)}\right]^{-1}$$

$$= \left[P - \underline{z}_i\underline{z}_i' + \frac{1}{(1-s_i)}(\underline{z}_i-\underline{a}_i)(\underline{z}_i-\underline{a}_i)'\right]^{-1} . \quad (3.6)$$

Let $\left[P + \frac{1}{(1-s_i)}(\underline{z}_i-\underline{a}_i)(\underline{z}_i-\underline{a}_i)'\right] = C$,

then, using Appendix I, equation (3.6) can be rewritten as

$$(C - \underline{z}_i\underline{z}_i')^{-1} = C^{-1} + \frac{C^{-1}\underline{z}_i\underline{z}_i'\,C^{-1}}{1-\underline{z}_i'\,C^{-1}\underline{z}_i} , \quad (3.7)$$

and using the same expansion, it follows that

$$C^{-1} = P^{-1} - \frac{P^{-1}(\underline{z}_i-\underline{a}_i)(\underline{z}_i-\underline{a}_i)'P^{-1}}{(1-s_i)+(\underline{z}_i-\underline{a}_i)'P^{-1}(\underline{z}_i-\underline{a}_i)} . \quad (3.8)$$

Combining equations (3.7) and (3.8) and simplifying we obtain

$$\left[P^{-1} - \underline{z}_i\underline{z}_i' + \frac{1}{(1-s_i)}(\underline{z}_i - \underline{a}_i)(\underline{z}_i - \underline{a}_i)'\right]^{-1} =$$

$$P^{-1} - \left\{\frac{P^{-1}(\underline{z}_i-\underline{a}_i)(\underline{z}_i-\underline{a}_i)'}{(1 - s_i + d_i)} + \frac{1}{k_i}\left[P^{-1}\underline{z}_i\underline{z}_i' - \frac{P^{-1}(\underline{z}_i-\underline{a}_i)(\underline{z}_i-\underline{a}_i)'P^{-1}\underline{z}_i\underline{z}_i'}{(1 - s_i + d_i)}\right.\right.$$

$$\left.\left. - \frac{P^{-1}\underline{z}_i\underline{z}_i'P^{-1}(\underline{z}_i-\underline{a}_i)(\underline{z}_i-\underline{a}_i)'}{(1 - s_i + d_i)} + \frac{P^{-1}(\underline{z}_i-\underline{a}_i)(\underline{z}_i-\underline{a}_i)'P^{-1}\underline{z}_i\underline{z}_i'P^{-1}(\underline{z}_i-\underline{a}_i)(\underline{z}_i-\underline{a}_i)'}{(1 - s_i + d_i)^2}\right]\right\}P^{-1} , \tag{3.9}$$

where $d_i = (\underline{z}_i-\underline{a}_i)'P^{-1}(\underline{z}_i-\underline{a}_i)$, (scalar)

and $k_i = 1 - \underline{z}_i'P^{-1}\underline{z}_i + \dfrac{\underline{z}_i'P^{-1}(\underline{z}_i-\underline{a}_i)(\underline{z}_i-\underline{a}_i)'P^{-1}\underline{z}_i}{(1 - s_i + d_i)}$. (scalar)

The last term in equation (3.9) can be rewritten as

$$P^{-1}[\underline{z}_i-\underline{a}_i][\underline{z}_i-\underline{a}_i]'\left[\frac{k_i - (1-\underline{z}_i'P^{-1}\underline{z}_i)}{(1 - s_i + d_i)}\right] ,$$

and combining this with the first term in curly brackets gives

$$P^{-1}[\underline{z}_i-\underline{a}_i][\underline{z}_i-\underline{a}_i]'\left[-\frac{(1 - \underline{z}_i'P^{-1}\underline{z}_i)}{k_i(1 - s_i + d_i)}\right] .$$

Thus equation (3.9) can be written as

$$P^{-1} + \frac{1}{k_i}\left[-\frac{(1-\underline{z}_i'\,P^{-1}\underline{z}_i)}{(1 - s_i + d_i)}P^{-1}(\underline{z}_i-\underline{a}_i)(\underline{z}_i-\underline{a}_i)' + P^{-1}\underline{z}_i\underline{z}_i'\right.$$

$$\left. - \frac{(\underline{z}_i-\underline{a}_i)'P^{-1}\underline{z}_i}{(1 - s_i + d_i)}P^{-1}(\underline{z}_i-\underline{a}_i)\underline{z}_i' - \frac{\underline{z}_i'P^{-1}(\underline{z}_i-\underline{a}_i)}{(1 - s_i + d_i)}P^{-1}\underline{z}_i(\underline{z}_i-\underline{a}_i)'\right]P^{-1} . \tag{3.10}$$

Now consider the second "term" in equation (3.4), viz:

$$\left\{ [Z'X - \underline{z}_i\underline{x}_i'] \left[(X'X)^{-1} + \frac{(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}}{1 - \underline{x}_i'(X'X)^{-1}\underline{x}_i} \right] [X'\underline{y} - \underline{x}_i y_i] \right\}$$

$$= \Big[Z'X(X'X)^{-1}X'\underline{y} + \underline{z}_i\underline{x}_i'(X'X)^{-1}\underline{x}_i y_i$$

$$+ \frac{1}{(1-s_i)} \cdot Z'X(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}X'\underline{y} + \frac{1}{(1-s_i)} \cdot \underline{z}_i\underline{x}_i'(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}\underline{x}_i y_i$$

$$- Z'X(X'X)^{-1}\underline{x}_i y_i - \underline{z}_i\underline{x}_i'(X'X)^{-1}X'\underline{y}$$

$$- \frac{1}{(1-s_i)} \cdot Z'X(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}\underline{x}_i y_i - \frac{1}{(1-s_i)} \cdot \underline{z}_i\underline{x}_i'(X'X)^{-1}\underline{x}_i\underline{x}_i'(X'X)^{-1}X'\underline{y} \Big] . \tag{3.11}$$

Let

$$\underline{q} = Z'X(X'X)^{-1}X'\underline{y} ,$$

then equation (3.11) can be written as

$$\underline{q} - y_i\underline{z}_i + \frac{1}{(1-s_i)} \left[(\underline{z}_i - \underline{a}_i)(y_i - \underline{x}_i'(X'X)^{-1}X'\underline{y}) \right] . \tag{3.12}$$

To obtain an expression for $\hat{\underline{\theta}}_i$ in (3.2), we must postmultiply equation (3.10) by expression (3.12). Postmultiplying equation (3.10) by $\underline{q}$ we obtain

$$\hat{\underline{\theta}} + \frac{1}{k_i} \Bigg[- \frac{(1 - \underline{z}_i'P^{-1}\underline{z}_i)}{(1 - s_i + d_i)} P^{-1}(\underline{z}_i - \underline{a}_i) \quad (\underline{z}_i - \underline{a}_i)' + P^{-1}\underline{z}_i\underline{z}_i'$$

$$- \frac{(\underline{z}_i - \underline{a}_i)'P^{-1}\underline{z}_i}{(1 - s_i + d_i)} P^{-1}(\underline{z}_i - \underline{a}_i)\underline{z}_i' - \frac{\underline{z}_i'P^{-1}(\underline{z}_i - \underline{a}_i)}{(1 - s_i + d_i)} P^{-1}\underline{z}_i(\underline{z}_i - \underline{a}_i)' \Bigg] \hat{\underline{\theta}} .$$

Postmultiplying equation (3.10) by $-y_i \underline{z}_i$ and simplifying we obtain

$$-\frac{1}{k_i}\left[P^{-1}\underline{z}_i - \frac{(\underline{z}_i-\underline{a}_i)'P^{-1}\underline{z}_i}{(1 - s_i + d_i)} P^{-1}(\underline{z}_i-\underline{a}_i)\right] y_i \; .$$

Postmultiplying equation (3.10) by the term in square brackets in expression (3.12) and simplifying we obtain

$$\frac{1}{k_i}\left[\frac{(1-\underline{z}_i'P^{-1}\underline{z}_i)}{(1 - s_i + d_i)} P^{-1}(\underline{z}_i-\underline{a}_i) + \frac{\underline{z}_i'P^{-1}(\underline{z}_i-\underline{a}_i)}{(1 - s_i + d_i)} P^{-1}\underline{z}_i\right]\left[y_i - \underline{x}_i'\hat{\underline{\pi}}\right] \; .$$

Then rearranging the above expressions we obtain

$$\hat{\underline{\theta}}_i = \hat{\underline{\theta}} + \left[\frac{(\underline{z}_i-\underline{a}_i)'P^{-1}\underline{z}_i}{k_i(1-s_i+d_i)} P^{-1}(\underline{z}_i-\underline{a}_i) - \frac{1}{k_i} P^{-1}\underline{z}_i\right]\left[y_i-\underline{z}_i'\hat{\underline{\theta}}\right]$$

$$+ \left[\frac{(1 - \underline{z}_i'P^{-1}\underline{z}_i)}{k_i(1-s_i+d_i)} P^{-1}(\underline{z}_i-\underline{a}_i) + \frac{\underline{z}_i'P^{-1}(\underline{z}_i-\underline{a}_i)}{k_i(1-s_i+d_i)} P^{-1}\underline{z}_i\right] \times$$

$$\left[(y_i - \underline{x}_i'\hat{\pi}) - (\underline{z}_i-\underline{a}_i)'\hat{\underline{\theta}}\right]. \qquad (i = 1,2,\ldots,N) \qquad (3.13)$$

Note that $(y_i - \underline{x}_i'\hat{\pi})$ is the ith component of the reduced form residual vector $\hat{\underline{v}} = (I - M_X)\underline{y}$, we denote it therefore by $\hat{v}_i$. Similarly, $\hat{u}_i = (y_i - \underline{z}_i\hat{\underline{\theta}})$ is the ith component of the structural form residual vector $\hat{\underline{u}} = (\underline{y} - Z\hat{\underline{\theta}})$, and $\hat{w}_i = (y_i - \underline{a}_i'\hat{\theta})$ is the ith component of the "second-stage" residual vector $\hat{\underline{w}} = (\underline{y} - M_X Z\hat{\underline{\theta}})$ where $M_X = X(X'X)^{-1}X'$.

Equation (3.13) was used for computing the J2SLS estimator and its associated test statistic in the Monte Carlo study of Chapters 6 and 7.

For future analysis, it will be convenient to rewrite equation (3.13) as

$$\underline{\hat{\theta}}_i = \underline{\hat{\theta}} + P^{-1}\underline{g}_i \,, \qquad (i = 1,2,\ldots,N) \tag{3.14}$$

where $$\underline{g}_i = (\underline{h}_i + \underline{j}_i)\hat{u}_i + \underline{h}_i(\hat{v}_i - \hat{w}_i) \,, \tag{3.15}$$

$$\underline{h}_i = \frac{(1 - \underline{z}_i'P^{-1}\underline{z}_i)}{k_i(1-s_i+d_i)}(\underline{z}_i-\underline{a}_i) + \frac{\underline{z}_i'P^{-1}(\underline{z}_i-\underline{a}_i)}{k_i(1-s_i+d_i)}\underline{z}_i \,, \tag{3.16}$$

and $$\underline{j}_i = \frac{(\underline{z}_i-\underline{a}_i)'P^{-1}\underline{z}_i}{k_i(1-s_i+d_i)}(\underline{z}_i-\underline{a}_i) - \frac{1}{k_i}\underline{z}_i \,. \tag{3.17}$$

The result in (3.14) is given in Phillips [42].

3.2 An Expression for the J2SLS Estimator

To form the J2SLS estimator we are required to take the summation of equation (3.14) over all i $(i = 1,2,\ldots,N)$ omitted observations. Using equation (3.14) we can form the J2SLS estimator as

$$J(\underline{\hat{\theta}}) = N\underline{\hat{\theta}} - \frac{(N-1)}{N}\sum_{i=1}^{N}\underline{\hat{\theta}}_i$$

$$= \underline{\hat{\theta}} - \frac{(N-1)}{N} P^{-1}\sum_{i=1}^{N}\underline{g}_i \,,$$

and using equation (3.15) we then obtain

$$J(\underline{\hat{\theta}}) = \underline{\hat{\theta}} - \frac{(N-1)}{N}P^{-1}\left[\sum_{i=1}^{N}(\underline{h}_i + \underline{j}_i)\hat{u}_i + \sum_{i=1}^{N}\underline{h}_i(\hat{v}_i-\hat{w}_i)\right] . \tag{3.18}$$

Substituting for $\underline{h}_i$ from (3.16) we obtain

$$\sum_{i=1}^{N}\underline{h}_i(\hat{v}_i - \hat{w}_i) = \sum_{i=1}^{N}\frac{1-\underline{z}_i'P^{-1}\underline{z}_i}{k_i(1-s_i+d_i)}(\underline{z}_i-\underline{a}_i)(\hat{v}_i-\hat{w}_i)$$

$$+ \sum_{i=1}^{N}\frac{\underline{z}_i'P^{-1}(\underline{z}_i-\underline{a}_i)}{k_i(1-s_i+d_i)}\underline{z}_i(\hat{v}_i-\hat{w}_i) \,. \tag{3.19}$$

Since $(\underline{z}_i-\underline{a}_i)$ and $\underline{z}_i$ are the ith columns of $Z'(I-M_X)$ and Z' respectively, it follows that

$$\sum_{i=1}^{N} \frac{1-\underline{z}_i'P^{-1}\underline{z}_i}{k_i(1-s_i+d_i)} (\underline{z}_i-\underline{a}_i)(\hat{v}_i-\hat{w}_i) = Z'(I-M_X)\Lambda_3(\underline{\hat{v}}-\underline{\hat{w}}) \tag{3.20}$$

and

$$\sum_{i=1}^{N} \frac{(\underline{z}_i-\underline{a}_i)'P^{-1}\underline{z}_i}{k_i(1-s_i+d_i)} \underline{z}_i(\hat{v}_i-\hat{w}_i) = -Z'\Lambda_2(\underline{\hat{v}}-\underline{\hat{w}}) \quad ; \tag{3.21}$$

where Λ_3 is an $N \times N$ diagonal matrix with iith component

$$(\Lambda_3)_{ii} = \frac{1-\underline{z}_i'P^{-1}\underline{z}_i}{k_i(1-s_i+d_i)} , \qquad (i = 1,2,\ldots,N)$$

and Λ_2 is an $N \times N$ diagonal matrix with iith component

$$(\Lambda_2)_{ii} = \frac{-\underline{z}_i'P^{-1}(\underline{z}_i-\underline{a}_i)}{k_i(1-s_i+d_i)} . \qquad (i = 1,2,\ldots,N)$$

Substituting from equations (3.20) and (3.21) into equation (3.19) gives

$$\sum_{i=1}^{N} \underline{h}_i(\hat{v}_i-\hat{w}_i) = Z'(I-M_X)\Lambda_3(\underline{\hat{v}}-\underline{\hat{w}}) - Z'\Lambda_2(\underline{\hat{v}}-\underline{\hat{w}}) . \tag{3.22}$$

Similarly it can be shown that

$$\sum_{i=1}^{N} \underline{h}_i\hat{u}_i = Z'(I-M_X)\Lambda_3\underline{\hat{u}} - Z'\Lambda_2\underline{\hat{u}} . \tag{3.23}$$

Consider the term $\sum_{i=1}^{N} \underline{j}_i\hat{u}_i$ in equation (3.18). Substituting for $\underline{j}_i$ from equation (3.17) gives

$$\sum_{i=1}^{N} \underline{j}_i\hat{u}_i = \sum_{i=1}^{N} \frac{(\underline{z}_i-\underline{a}_i)'P^{-1}\underline{z}_i}{k_i(1-s_i+d_i)} (\underline{z}_i-\underline{a}_i)\hat{u}_i - \sum_{i=1}^{N} \frac{1}{k_i} \cdot \underline{z}_i\hat{u}_i$$

$$= -Z'(I-M_X)\Lambda_2\underline{\hat{u}} - Z'\Lambda_1\underline{\hat{u}} , \tag{3.24}$$

where Λ_1 is an $N \times N$ diagonal matrix with iith component

$$(\Lambda_1)_{ii} = \frac{1}{k_i} .$$

Substituting from equations (3.22), (3.23), and (3.24) into equation (3.18) we obtain

$$J(\underline{\hat{\theta}}) = \underline{\hat{\theta}} + \frac{(N-1)}{N}\Big[P^{-1}Z'(\Lambda_1 + 2\Lambda_2 - \Lambda_3)\underline{\hat{u}}$$

$$- P^{-1}Z'M_X(\Lambda_2 - \Lambda_3)\underline{\hat{u}}$$

$$+ P^{-1}Z'(\Lambda_2 - \Lambda_3)(\underline{\hat{v}} - \underline{\hat{w}})$$

$$+ P^{-1}Z'M_X\Lambda_3(\underline{\hat{v}} - \underline{\hat{w}})\Big] \ . \qquad (3.25)$$

The ensuing analysis is simplified by writing equation (3.25) in a slightly amended form. Recall that

$$\underline{\hat{u}} = [I - Z(Z'M_XZ)^{-1}Z'M_X]\underline{y} \ ,$$

$$\underline{\hat{v}} = [I - M_X]\underline{y} \ ,$$

and

$$\underline{\hat{w}} = [I - M_XZ(Z'M_XZ)^{-1}Z'M_X]\underline{y} \ ,$$

from which it follows that

$$Z'(\underline{\hat{v}} - \underline{\hat{w}}) = \underline{0} \ ,$$

$$Z'M_X\underline{\hat{u}} = \underline{0} \ ,$$

and $Z'M_X(\underline{\hat{v}} - \underline{\hat{w}}) = \underline{0}$.

Thus, if we define

$$\bar{\Lambda}_3 = I - \Lambda_3 \ ,$$

we can rewrite equation (3.25) as

$$J(\underline{\hat{\theta}}) = \underline{\hat{\theta}} + \frac{(N-1)}{N}. \ P^{-1}\Big[Z'(\Lambda_1 + 2\Lambda_2 - \Lambda_3)\ \underline{\hat{u}}$$

$$- Z'M_X(\Lambda_2 - \bar{\Lambda}_3)\underline{\hat{u}}$$

$$+ Z'(\Lambda_2 - \bar{\Lambda}_3)(\underline{\hat{v}} - \underline{\hat{w}})$$

$$+ Z'M_X\bar{\Lambda}_3(\underline{\hat{v}} - \underline{\hat{w}})\Big] \ . \qquad (3.26)$$

3.3 The Asymptotic Equivalence of J2SLS and 2SLS

3.3.1 Preliminary Results

In this section the asymptotic behaviour of the three diagonal matrices (viz: Λ_1, Λ_2, and $\bar{\Lambda}_3$) introduced earlier in this Chapter will be investigated.

Essentially we must consider the following terms:

$$k_i,\ (1 - s_i + d_i),\ \underline{z}_i' P^{-1} \underline{z}_i \text{ and } \underline{z}_i' P^{-1}(\underline{z}_i - \underline{a}_i),$$

where $s_i = \underline{x}_i'(X'X)^{-1}\,\underline{x}_i$,

$$d_i = (\underline{z}_i - \underline{a}_i)' P^{-1}\,(\underline{z}_i - \underline{a}_i) ,$$

and $$k_i = 1 - \underline{z}_i' P^{-1} \underline{z}_i + \frac{\underline{z}_i' P^{-1}(\underline{z}_i - \underline{a}_i)(\underline{z}_i - \underline{a}_i)' P^{-1} \underline{z}_i}{(1 - s_i + d_i)} .$$

The reader is reminded of the following results which were established in Chapter 2:

(a) $$\underset{N\to\infty}{\text{plim}} \left[\frac{1}{N} \cdot Z'X\left(\frac{1}{N} \cdot X'X\right)^{-1} \frac{1}{N} \cdot X'Z\right]^{-1} = \underset{N\to\infty}{\text{plim}}\ N.P^{-1} = \Sigma_P^{-1} ,$$

where Σ_P^{-1} is a finite positive definite matrix;

and

(b) $$\underset{N\to\infty}{\text{plim}}\ \frac{1}{N} \cdot X'Z = \Sigma_{XZ} ,$$

where Σ_{XZ} is a finite matrix.

For the ensuing analysis result (b) will be expressed in a different form. Since

$$X'Z = X'[X\Pi + V : X_1] ,$$

we can rewrite result (b) as

$$\underset{N\to\infty}{\text{plim}}\ \frac{1}{N} \cdot X'Z = \left[\Sigma_{XX}\Pi : \Sigma_{XX_1}\right] .$$

It follows from assumption (iii), section 2.1.3, that

$$\lim_{N\to\infty} Ns_i = \lim_{N\to\infty} \underline{x}_i'\left(\frac{1}{N} \cdot X'X\right)^{-1}\underline{x}_i = \underline{x}_i' \Sigma_{XX}^{-1} \underline{x}_i ,$$

which is a finite constant, and consequently

$$\lim_{N\to\infty} s_i = 0 . \tag{3.27}$$

Consider the vector $\underline{a}_i'$, where

$$\underline{a}_i' = \underline{x}_i'(X'X)^{-1}X'Z = \underline{x}_i'(X'X)^{-1}X' \cdot [Y : X_1] ;$$

i.e. $$\underline{a}_i' = [\underline{x}_i'\hat{\Pi} : \underline{x}_{1i}'] . \tag{3.28}$$

Using result (a) it follows that

$$\underset{N\to\infty}{\text{plim}}\, N \cdot \underline{a}_i'P^{-1}\underline{a}_i = [\underline{x}_i'\Pi : \underline{x}_{1i}']\Sigma_P^{-1}[\Pi'\underline{x}_i : \underline{x}_{1i}] , \tag{3.29}$$

a finite constant, where $\underline{x}_{1i}'$ is the ith row of X_1 .

This result can be shown as follows:

the matrix $(X'X)^{-1}X'X_1$

is a submatrix of $(X'X)^{-1}X'X = I_K$,

and thus consists of K_1 columns of the $K \times K$ identity matrix. By premultiplying these columns by $\underline{x}_i'$ we obtain $\underline{x}_{1i}'$.

It will be convenient to write equation (3.29) as

$$\underset{N\to\infty}{\text{plim}}\, N \cdot \underline{a}_i' P^{-1}\underline{a}_i = \bar{\underline{a}}_i' \Sigma_P^{-1} \bar{\underline{a}}_i , \tag{3.30}$$

where $\underset{N\to\infty}{\text{plim}}\ \underline{a}_i = \bar{\underline{a}}_i$, a finite constant vector.

We can conclude, therefore, that

$$\underset{N\to\infty}{\text{plim}}\ \underline{a}_i' P^{-1} \underline{a}_i = 0. \tag{3.31}$$

Consider the term

$$(\underline{z}_i - \underline{a}_i)' \, P^{-1} \, (\underline{z}_i - \underline{a}_i) \, ,$$

where the vector $(\underline{z}_i - \underline{a}_i)'$ is the ith observation on the matrix

$$[Z - X(X'X)^{-1}X'Z] \, .$$

Partitioning this matrix we obtain

$$[Y : X_1] - X(X'X)^{-1}X'[Y : X_1]$$

$$= [Y : X_1] - [X\hat{\Pi} : X_1]$$

$$= [\hat{V} : 0] \, ,$$

which will have ith observation denoted by

$$[\hat{\underline{v}}_i' : \underline{0}'] \, .$$

It follows that

$$(\underline{z}_i - \underline{a}_i)' \, P^{-1}(\underline{z}_i - \underline{a}_i) = \hat{\underline{v}}_i' P^{-1}\hat{\underline{v}}_i \, , \qquad (3.32)$$

$$(\underline{z}_i - \underline{a}_i)' P^{-1} \underline{z}_i = [\hat{\underline{v}}_i' : \underline{0}'] \, P^{-1}[\underline{v}_i + \Pi'\underline{x}_i : \underline{x}_{1i}] \, , \qquad (3.33)$$

and

$$\underline{a}_i' P^{-1}(\underline{z}_i - \underline{a}_i) = [\underline{x}_i'\hat{\Pi} : \underline{x}_{1i}'] \, P^{-1}[\hat{\underline{v}}_i : \underline{0}] \, . \qquad (3.34)$$

Since each element of the OLS reduced form residuals matrix converges in distribution to the corresponding element of the disturbance matrix, from equation (3.32) and using result (a) we can write

$$\underset{N \to \infty}{\text{plim}} \; N \, . \, \hat{\underline{v}}_i' \, P^{-1}\hat{\underline{v}}_i = \underline{v}_i' \, \Sigma_P^{-1} \, \underline{v}_i \, . \qquad (3.35)$$

Since Σ_P^{-1} is a finite positive definite matrix, and since the $\underline{v}_i'$ $(i = 1,2,\ldots,N)$ are independently and identically distributed with mean zero and finite covariance matrix (this fact follows from assumption (v), section 2.1.3, since the reduced form disturbances are just linear combinations of the structural form disturbances), it follows that

$$\underline{v}_i' \Sigma_P^{-1} \underline{v}_i$$

is a random variable with finite mean and variance. Hence

$$\underset{N\to\infty}{\text{plim}} \ \hat{\underline{v}}_i' P^{-1} \hat{\underline{v}}_i = 0 \ , \tag{3.36}$$

since $\frac{1}{N} \cdot v_i' \Sigma_P^{-1} v_i$ converges in probability to zero.

Combining equations (3.32), (3.33), and (3.34) we can write

$$(\underline{z}_i - \underline{a}_i)'P^{-1}\underline{z}_i = (\underline{z}_i - \underline{a}_i)'P^{-1}(\underline{z}_i - \underline{a}_i) + (\underline{z}_i - \underline{a}_i)'P^{-1}\underline{a}_i \ . \tag{3.37}$$

The probability limit of the last term in equation (3.37) can be written as

$$\underset{N\to\infty}{\text{plim}} \ N.(\underline{z}_i - \underline{a}_i)'P^{-1}\underline{a}_i = \underline{v}_i' \Sigma_P^{-1} \bar{\underline{a}}_i \ . \tag{3.38}$$

Since $\Sigma_P^{-1} \bar{\underline{a}}_i$ is a finite vector, and since the $\underline{v}_i'$ $(i = 1,2,\ldots,N)$ are independently and identically distributed with mean zero and finite covariance matrix, it follows that

$$\underline{v}_i' \Sigma_P^{-1} \bar{\underline{a}}_i$$

is a random variable with mean zero and finite variance. Thus

$$\underset{N\to\infty}{\text{plim}} \ (\underline{z}_i - \underline{a}_i)' P^{-1} \underline{a}_i = 0 \ ,$$

since $\frac{1}{N} \cdot \underline{v}_i' \Sigma_P^{-1} \bar{\underline{a}}_i$ converges in probability to zero.

Combining the above result with that given by equation (3.36), and substituting into equation (3.37), we have shown that

$$\underset{N\to\infty}{\text{plim}} \ (\underline{z}_i - \underline{a}_i)' P^{-1}\underline{z}_i = 0 \ . \tag{3.39}$$

We now consider the scalar $\underline{z}_i' P^{-1}\underline{z}_i$ which can be written as

$$\begin{aligned} \underline{z}_i'P^{-1}\underline{z}_i &= [(\underline{z}_i - \underline{a}_i) + \underline{a}_i]' P^{-1} [(\underline{z}_i - \underline{a}_i) + \underline{a}_i] \\ &= (\underline{z}_i - \underline{a}_i)'P^{-1}(\underline{z}_i - \underline{a}_i) + \underline{a}_i'P^{-1}\underline{a}_i \\ &\quad + 2\underline{a}_i'P^{-1}(\underline{z}_i - \underline{a}_i) \ . \end{aligned}$$

Using equations (3.29), (3.32), (3.36), and (3.39) it follows that

$$\operatorname*{plim}_{N\to\infty} \underline{z}_i' P^{-1} \underline{z}_i = 0 \; . \tag{3.40}$$

From equations (3.27), (3.32), and (3.36) we have shown that

$$\operatorname*{plim}_{N\to\infty} (1 - s_i + d_i) = 1 \; , \tag{3.41}$$

and from equations (3.39), (3.40), and (3.41) we have shown that

$$\operatorname*{plim}_{N\to\infty} k_i = 1 \; . \tag{3.42}$$

3.3.2 Proof of Asymptotic Equivalence

To prove that the 2SLS and J2SLS estimators are asymptotically equivalent, we are required to show that

$$\operatorname*{plim}_{N\to\infty} \sqrt{N}[J(\underline{\hat{\theta}}) - \underline{\hat{\theta}}] = \underline{0} \; .$$

From equation (3.26) we can write this requirement as

$$\operatorname*{plim}_{N\to\infty} \sqrt{N}\,[J(\underline{\hat{\theta}})-\underline{\hat{\theta}}] = \operatorname*{plim}_{N\to\infty}\left(\frac{1}{N} . P\right)^{-1}\left\{\operatorname*{plim}_{N\to\infty} \frac{1}{\sqrt{N}} . Z'(\Lambda_1 + 2\Lambda_2 - \Lambda_3)\underline{\hat{u}}\right.$$

$$- \operatorname*{plim}_{N\to\infty} \frac{1}{\sqrt{N}} . Z'M_X(\Lambda_2 - \overline{\Lambda}_3)\underline{\hat{u}} + \operatorname*{plim}_{N\to\infty} \frac{1}{\sqrt{N}} . Z'(\Lambda_2 - \overline{\Lambda}_3)(\underline{\hat{v}} - \underline{\hat{w}})$$

$$\left. + \operatorname*{plim}_{N\to\infty} \frac{1}{\sqrt{N}} . Z'M_X\overline{\Lambda}_3(\underline{\hat{v}} - \underline{\hat{w}}) \right\} = \underline{0} \; . \tag{3.43}$$

A term by term evaluation of equation (3.43) now follows.

Consider the first term in curly brackets in equation (3.43), viz:

$$Z'(\Lambda_1 + 2\Lambda_2 - \Lambda_3)\underline{\hat{u}} \; . \tag{3.44}$$

We know that:

the iith component of Λ_1 is $1/k_i$;

the iith component of Λ_2 is $\dfrac{-(\underline{z}_i - \underline{a}_i)'P^{-1}\underline{z}_i}{k_i(1 - s_i + d_i)}$;

and the iith component of Λ_3 is $\dfrac{(1 - \underline{z}_i'P^{-1}\underline{z}_i)}{k_i(1 - s_i + d_i)}$;

thus, after some algebraic manipulation, the iith component of the bracketed term in expression (3.44) can be written as

$$(\Lambda_1 + 2\Lambda_2 - \Lambda_3)_{ii} = -\frac{s_i}{k_i(1 - s_i + d_i)} + \frac{\underline{a}_i'\, P^{-1}\underline{a}_i}{k_i(1 - s_i + d_i)} .$$

Let $\underline{a}'_s P^{-1}\underline{a}_s$ be the largest of the $\underline{a}_i'\, P^{-1}\underline{a}_i$ $(i = 1, 2, \ldots, N)$, then it follows from equation (3.30) that

$$\underset{N\to\infty}{\text{plim}}\ N.\underline{a}'_s P^{-1}\underline{a}_s = \overline{\underline{a}}'_s\ \Sigma_P^{-1}\ \overline{\underline{a}}_s ,$$

a finite positive definite quadratic form. It follows, therefore, that

$$\underset{N\to\infty}{\text{plim}}\ \sqrt{N}\ \max_{1\leq i\leq N}\ \underline{a}_i'\, P^{-1}\underline{a}_i = 0 . \qquad (3.45)$$

Using a similar argument it can be shown that

$$\lim_{N\to\infty} \sqrt{N}\ \max_{1\leq i\leq N}\ s_i = 0 . \qquad (3.46)$$

Combining equations (3.41), (3.42), (3.45), and (3.46) we obtain

$$\underset{N\to\infty}{\text{plim}}\ \sqrt{N}\ \max_{1\leq i\leq N}\ (\Lambda_1 + 2\Lambda_2 - \Lambda_3)_{ii} = 0 . \qquad (3.47)$$

The jth component of the random vector (3.44) can be written as

$$\sum_{i=1}^{N} m_{ij}(\Lambda_1 + 2\Lambda_2 - \Lambda_3)_{ii}\ \hat{u}_i$$

$$+ \sum_{i=1}^{N} (\Lambda_1 + 2\Lambda_2 - \Lambda_3)_{ii}\hat{u}_i v_{ij} \qquad (3.48)$$

where the m_{ij} $(i = 1,2,\ldots,N;\ j = 1,2,\ldots,K_1 + g)$ represent the nonstochastic part of the z_{ij} (the ijth element of Z), and the $v_{ij}(i = 1,2,\ldots,N;\ \ j = 1,2,\ldots K_1 + g)$ represent the reduced form disturbance part of the z_{ij} (where appropriate). Without loss of generality we can assume that the observations on the (g) explanatory endogenous variables occur in the first g columns of Z; thus $v_{ij} = 0$ for all $j > g$ (for all i).

Consider the first term in expression (3.48), viz:

$$\sum_{i=1}^{N} m_{ij}(\Lambda_1 + 2\Lambda_2 - \Lambda_3)_{ii}\,\hat{u}_i \ .$$

Since the m_{ij} are nonstochastic, it follows from equation (3.47) that

$$\underset{N\to\infty}{\text{plim}}\ \sqrt{N}\ \max_{1\leq i\leq N}\left|m_{ij}(\Lambda_1 + 2\Lambda_2 - \Lambda_3)_{ii}\right| = 0\ . \qquad (3.49)$$

We now require the following theorem which is taken from Malinvaud [27; pp. 322-323] and is cited without proof.

THEOREM I.

Let x_{tT} $(t = 1,2,\ldots,T;\ \ T = 1,2,\ldots.)$ be random variables. If

$$\underset{T\to\infty}{\text{plim}}\ \max_{1\leq t\leq T}\left|x_{tT}\right| = 0\ ,$$

and if the u_t are mutually independent random variables identically distributed with zero mean, then:

$$\underset{T\to\infty}{\text{plim}}\ \frac{1}{T}\sum_{t=1}^{T} x_{tT} = 0 \quad \text{and} \quad \underset{T\to\infty}{\text{plim}}\ \frac{1}{T}\sum_{t=1}^{T} u_t x_{tT} = 0\ .$$

Since each component of the 2SLS residual vector converges in distribution to the corresponding element of the disturbance vector, using Theorem I together with equation (3.49) it follows that

$$\underset{N\to\infty}{\text{plim}} \frac{1}{N} \sum_{i=1}^{N} m_{ij} \sqrt{N}\, (\Lambda_1 + 2\Lambda_2 - \Lambda_3)_{ii} \hat{u}_i = 0 .$$

Now consider the second term in expression (3.48), viz:

$$\sum_{i=1}^{N} (\Lambda_1 + 2\Lambda_2 - \Lambda_3)_{ii}\, \hat{u}_i\, v_{ij} . \tag{3.50}$$

The reduced form disturbances associated with the g explanatory variables can be decomposed into a term ($\underline{u}\,\underline{\Psi}'$) which is proportional to the disturbance term in the ith structural equation, and a term (E) which is uncorrelated with $\underline{u}$ (e.g. see Nagar [37; p.577]), viz:

$$V = \underline{u}\,\underline{\Psi}' + E . \tag{3.51}$$

The ith row of v can be written as

$$\underline{v}_i' = u_i\, \underline{\Psi}' + \underline{e}_i' , \quad (i = 1,2,\ldots,N)$$

whereupon by substituting for v_{ij} in (3.50) we obtain

$$\sum_{i=1}^{N} (\Lambda_1 + 2\Lambda_2 - \Lambda_3)_{ii}\, \psi_j \hat{u}_i u_i$$

$$+ \sum_{i=1}^{N} (\Lambda_1 + 2\Lambda_2 - \Lambda_3)_{ii}\, \hat{u}_i e_{ij} ,$$

where ψ_j denotes the jth element of $\underline{\Psi}'$.

Let $E(u_i^2) = \sigma^2$ then, since $\hat{u}_i \to u_i$ as $N \to \infty$, it follows that the $(\hat{u}_i^2 - \sigma^2)$ are (asymptotically) independently and identically distributed random variables with mean zero.

Since the ψ_j are nonstochastic it follows, using equation (3.47), that

$$\underset{N\to\infty}{\text{plim}} \sqrt{N} \max_{1\leq i\leq N} \left| (\Lambda_1 + 2\Lambda_2 - \Lambda_3)_{ii}\, \psi_j \right| = 0 .$$

Combining this result with Theorem I it follows that

$$\underset{N\to\infty}{\text{plim}} \frac{1}{N} \sum_{i=1}^{N} \sqrt{N}\ (\Lambda_1 + 2\Lambda_2 - \Lambda_3)_{ii}\ \psi_j (\hat{u}_i u_i - \sigma^2) = 0\ .$$

This result implies that

$$\underset{N\to\infty}{\text{plim}} \frac{1}{N} \sum_{i=1}^{N} \sqrt{N}(\Lambda_1 + 2\Lambda_2 - \Lambda_3)_{ii}\ \psi_j \hat{u}_i u_i$$

$$= \underset{N\to\infty}{\text{plim}} \frac{1}{N} \sum_{i=1}^{N} \sqrt{N}(\Lambda_1 + 2\Lambda_2 - \Lambda_3)_{ii}\ \psi_j\ \sigma^2$$

$$= \sigma^2\ \underset{N\to\infty}{\text{plim}} \frac{1}{N} \sum_{i=1}^{N} \sqrt{N}(\Lambda_1 + 2\Lambda_2 - \Lambda_3)_{ii}\ \psi_j$$

$= 0$ from Theorem I .

Since $\hat{u}_i e_{ij} \to u_i e_{ij}$ (as $N\to\infty$) which are mutually independent random variables (i.e. $u_{i+1} e_{i+1,j}$ is independent of $u_i e_{ij}$) it follows from Theorem I that

$$\underset{N\to\infty}{\text{plim}} \frac{1}{N} \sum_{i=1}^{N} \sqrt{N}(\Lambda_1 + 2\Lambda_2 - \Lambda_3)_{ii} \hat{u}_i e_{ij} = 0.$$

This concludes the analysis on the first term in curly brackets in equation (3.43). To summarize, we have shown that

$$\underset{N\to\infty}{\text{plim}} \frac{1}{\sqrt{N}} \cdot Z'(\Lambda_1 + 2\Lambda_2 - \Lambda_3)\underline{\hat{u}} = \underline{0}\ .$$

Consider the second term in curly brackets in equation (3.43), viz:

$$- Z' M_X (\Lambda_2 - \overline{\Lambda}_3)\underline{\hat{u}}\ , \tag{3.52}$$

where the iith component of the term in brackets can be written as

$$(\Lambda_2 - \overline{\Lambda}_3)_{ii} = - \frac{(\underline{z}_i - \underline{a}_i)' P^{-1} \underline{z}_i}{k_i(1 - s_i + d_i)} - 1 + \frac{(1 - \underline{z}_i' P^{-1} \underline{z}_i)}{k_i(1 - s_i + d_i)}\ . \tag{3.53}$$

Expressing equation (3.53) in terms of a common denominator, the numerator can be written as

$$- (\underline{z}_i - \underline{a}_i)'P^{-1}\underline{z}_i + s_i - s_i\underline{z}_i' P^{-1}\underline{z}_i - d_i + d_i\underline{z}_i' P^{-1}\underline{z}_i$$
$$- \underline{z}_i' P^{-1}(\underline{z}_i - \underline{a}_i)(\underline{z}_i - \underline{a}_i)'P^{-1}\underline{z}_i \ .$$

The following probability limits can now be established:

$$\underset{N\to\infty}{\text{plim}}\ N(s_i\underline{z}_i'P^{-1}\underline{z}_i) = \underset{N\to\infty}{\text{plim}}\ \frac{1}{N} \cdot \underline{x}_i'\left(\frac{1}{N} \cdot X'X\right)^{-1} \underline{x}_i \cdot \underline{z}_i'\left(\frac{1}{N} \cdot P\right)^{-1} \underline{z}_i = 0 \ ,$$

using equation (3.40) and the knowledge that $\underline{x}_i' \Sigma_{XX}^{-1} \underline{x}_i$ is a finite constant;

$$\underset{N\to\infty}{\text{plim}}\left(\underline{z}_i' P^{-1}(\underline{z}_i - \underline{a}_i)(\underline{z}_i - \underline{a}_i)'P^{-1}\underline{z}_i\right) = \underset{N\to\infty}{\text{plim}}\left[\underline{z}_i'P^{-1}(\underline{z}_i - \underline{a}_i)\right]^2 = 0,$$

using equation (3.39);

and

$$\underset{N\to\infty}{\text{plim}}\ N(d_i\underline{z}_i'P^{-1}\underline{z}_i) = \underset{N\to\infty}{\text{plim}}\ \frac{1}{N}\left\{\hat{\underline{v}}_i' \left(\frac{1}{N} \cdot P\right)^{-1} \hat{\underline{v}}_i \cdot \underline{z}_i' \left(\frac{1}{N} \cdot P\right)^{-1} \underline{z}_i\right\} = 0 \ ,$$

using equations (3.35) and (3.40).

Combining the above three results with equations (3.41) and (3.42) we have shown that

$$\underset{N\to\infty}{\text{plim}}\ N(\Lambda_2 - \overline{\Lambda}_3)_{ii} = \underline{x}_i' \Sigma_{XX}^{-1} \underline{x}_i - 2\underline{v}_i' \Sigma_P^{-1} \underline{v}_i \ , \qquad (3.54)$$

and, by the same proof, that

$$\underset{N\to\infty}{\text{plim}}\ N(\overline{\Lambda}_3)_{ii} = - \underline{x}_i' \Sigma_{XX}^{-1} \underline{x}_i + \underline{v}_i' \Sigma_P^{-1} \underline{v}_i \ . \qquad (3.55)$$

Expression (3.52) can be written as

$$- \frac{1}{N} \cdot Z'X\left(\frac{1}{N} \cdot X'X\right)^{-1} \cdot X'(\Lambda_2 - \overline{\Lambda}_3)\hat{\underline{u}} \ . \qquad (3.56)$$

The rth element of

$$X'(\Lambda_2 - \overline{\Lambda}_3)\hat{\underline{u}}$$

can therefore be written as

$$\frac{1}{N} \sum_{i=1}^{N} x_{ir} \underline{x}_i' \left(\frac{1}{N} \cdot X'X\right)^{-1} \underline{x}_i \hat{u}_i$$

$$- \frac{2}{N} \sum_{i=1}^{N} x_{ir} \hat{\underline{v}}_i' \left(\frac{1}{N} \cdot P\right)^{-1} \hat{\underline{v}}_i \hat{u}_i , \tag{3.57}$$

where x_{ir} $(r = 1,2,\ldots,K)$ is the rith element of X'.

Rearranging the first term in expression (3.57) and taking its probability limit in the context of expression (3.56) gives

$$\Sigma_{ZX} \Sigma_{XX}^{-1} \sum_{i=1}^{N} \lim_{N\to\infty} \underline{x}_i' \left(\frac{1}{N} \cdot X'X\right)^{-1} \underline{x}_i \cdot \operatorname*{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} x_{ir}\hat{u}_i = \underline{0} .$$

This result is obtained by noting that the limit term is a finite constant, whilst the Law of Large Numbers (e.g. see Malinvaud [27; Proposition 12, p.322]) ensures that the probability limit term is zero.

Since each element of the 2SLS residual vector converges in distribution to the corresponding element of the disturbance vector, and since each element of the OLS reduced form residuals matrix converges in distribution to the corresponding element of the disturbance matrix, it follows that

$$\hat{\underline{v}}_i' \left(\frac{1}{N} \cdot P\right)^{-1} \hat{\underline{v}}_i \hat{u}_i$$

converges in probability to

$$\underline{v}_i' \Sigma_P^{-1} \underline{v}_i u_i . \tag{3.58}$$

Using equation (3.51), expression (3.58) can be written as

$$(u_i \underline{\Psi}' + \underline{e}_i') \Sigma_P^{-1} (\underline{\Psi} u_i + \underline{e}_i) u_i ,$$

which upon expansion gives

$$\underline{\Psi}'\Sigma_P^{-1}\,\underline{\Psi}\,u_i^3 + 2\underline{e}_i'\,\Sigma_P^{-1}\,\underline{\Psi}\,u_i^2 + \underline{e}_i'\,\Sigma_P^{-1}\,\underline{e}_i u_i \ . \qquad (3.59)$$

In the context of the second term in expression (3.57), the first term in expression (3.59) can be written as

$$-\,\underline{\Psi}'\,\Sigma_P^{-1}\,\underline{\Psi}\ \underset{N\to\infty}{\text{plim}}\ \frac{2}{N}\sum_{i=1}^{N} x_{ir}[u_i^3 - E(u_i^3)] = 0 \ .$$

Noting that the quadratic form in the above equation is a constant, the Law of Large Numbers ensures the result.

In the context of expression (3.56) this result implies that

$$-\ \Sigma_{ZX}\,\Sigma_{XX}^{-1}\,\underline{\Psi}'\,\Sigma_P^{-1}\,\underline{\Psi}\ \underset{N\to\infty}{\text{plim}}\ \frac{2}{N}\sum_{i=1}^{N} x_{ir}u_i^3$$

in a finite matrix, provided $E(u_i^3)$ is finite.

Since u_i and $\underline{e}_i$ are uncorrelated random variables (by assumption) it follows, using Theorem I, that the second and third terms in expression (3.59), in the context of expression (3.57), converge in probability to zero.

Collecting results, we have shown that

$$\underset{N\to\infty}{\text{plim}}\ Z'M_X(\Lambda_2 - \bar{\Lambda}_3)\hat{\underline{u}}$$

is a finite constant, and hence

$$\underset{N\to\infty}{\text{plim}}\ \frac{1}{\sqrt{N}}\ .\ Z'M_X(\Lambda_2 - \bar{\Lambda}_3)\hat{\underline{u}} = \underline{0} \ .$$

The third term in curly brackets in equation (3.43) is

$$Z'(\Lambda_2 - \bar{\Lambda}_3)(\underline{\hat{v}} - \underline{\hat{w}}) \,. \tag{3.60}$$

We know that

$$\underline{\hat{u}} = (y - Z\underline{\hat{\theta}}) \,,$$

$$\underline{\hat{v}} = (I - M_X)\underline{y} \,,$$

and

$$\underline{\hat{w}} = (\underline{y} - M_X Z\underline{\hat{\theta}}) \,,$$

from which it follows that

$$(\underline{\hat{v}} - \underline{\hat{w}}) = - M_X\underline{\hat{u}} \,. \tag{3.61}$$

Substituting from equation (3.61), expression (3.60) can be written as

$$Z'(\Lambda_2 - \bar{\Lambda}_3)X\left(\frac{1}{N}\,.\,X'X\right)^{-1}\left(\frac{1}{N}\,.\,X'\underline{\hat{u}}\right) \,. \tag{3.62}$$

Consider the term

$$Z'(\Lambda_2 - \bar{\Lambda}_3)X$$

which has jsth element given by

$$\frac{1}{N}\sum_{i=1}^{N} z_{ij}\underline{x}_i'\left(\frac{1}{N}\,.\,X'X\right)^{-1}\underline{x}_i\; x_{is}$$

$$- \frac{2}{N}\sum_{i=1}^{N} z_{ij}\underline{\hat{v}}_i'\left(\frac{1}{N}\,.\,P\right)^{-1}\underline{\hat{v}}_i\; x_{is} \quad , \tag{3.63}$$

where x_{is} (s = 1,2,...,K) is the isth element of X.

The first term in expression (3.63) can be expanded as

$$\frac{1}{N}\sum_{i=1}^{N} m_{ij}\underline{x}_i'\left(\frac{1}{N}\,.\,X'X\right)^{-1}\underline{x}_i x_{is}$$

$$+ \frac{1}{N}\sum_{i=1}^{N} \underline{x}_i'\left(\frac{1}{N}\,.\,X'X\right)^{-1}\underline{x}_i x_{is} v_{ij} \,,$$

whence the first term of the above expression converges to a finite constant and the second term converges in probability to zero by the

Law of Large Numbers.

The second term in expression (3.63) can be expanded as

$$- \frac{2}{N} \sum_{i=1}^{N} m_{ij} \; x_{is} \; \hat{\underline{v}}_i' \left(\frac{1}{N} \cdot P\right)^{-1} \hat{\underline{v}}_i$$

$$- \frac{2}{N} \sum_{i=1}^{N} x_{is} \; \hat{\underline{v}}_i' \left(\frac{1}{N} \cdot P\right)^{-1} \hat{\underline{v}}_i \; v_{ij} \; . \qquad (3.64)$$

Using the argument preceding expression (3.58), and substituting for $\underline{v}_i$ from equation (3.51), expression (3.64) can be rewritten as

$$- \frac{2}{N} \sum_{i=1}^{N} m_{ij} x_{is} \left[\underline{\Psi}' \Sigma_P^{-1} \underline{\Psi} u_i^2 + 2\underline{e}_i' \Sigma_P^{-1} \underline{\Psi} u_i + \underline{e}_i' \Sigma_P^{-1} \underline{e}_i \right]$$

$$- \frac{2}{N} \sum_{i=1}^{N} x_{is} \left[\underline{\Psi}' \Sigma_p^{-1} \underline{\Psi} u_i^2 + 2\underline{e}_i' \Sigma_p^{-1} \underline{\Psi} u_i + \underline{e}_i' \Sigma_p^{-1} \underline{e}_i \right] (\psi_j u_i + e_{ij}) \; .$$

From our analysis to date, it follows that both terms in the above expression converge in probability to finite constants.

Since, from our initial assumptions, the term to be inverted in expression (3.62) converges to a finite matrix and the term on its right converges in probability to a null vector, we have shown that

$$\operatorname*{plim}_{N \to \infty} \frac{1}{\sqrt{N}} \cdot Z'(\Lambda_2 - \bar{\Lambda}_3)(\hat{\underline{v}} - \hat{\underline{w}}) = \underline{0} \; .$$

The fourth term in curly brackets in equation (3.43) is

$$Z'M_X \bar{\Lambda}_3 (\hat{\underline{v}} - \hat{\underline{w}})$$

which, using equation (3.61), can be rewritten as

$$\left(\frac{1}{N} \cdot Z'X\right) \left(\frac{1}{N} \cdot X'X\right)^{-1} X' \bar{\Lambda}_3 X \left(\frac{1}{N} \cdot X'X\right)^{-1} \left(\frac{1}{N} \cdot X'\hat{\underline{u}}\right) .$$

From our initial assumptions, the first three terms in round brackets converge in probability to finite matrices, whilst the fourth converges in probability to a null vector.

Using equation (3.55) the rsth element of the "middle" term in the above expression can be written as

$$- \frac{1}{N} \sum_{i=1}^{N} x_{ir} \underline{x}_i' \Sigma_{XX}^{-1} \underline{x}_i x_{is}$$

$$+ \frac{1}{N} \sum_{i=1}^{N} x_{ir} \underline{v}_i' \Sigma_P^{-1} \underline{v}_i x_{is} ,$$

whence the first term converges to a finite limit whilst the second term converges in probability to a finite limit. Thus we have shown that

$$\underset{N\to\infty}{\text{plim}} \frac{1}{\sqrt{N}} \cdot Z'M_X \bar{\Lambda}_3 (\hat{\underline{v}} - \hat{\underline{w}}) = \underline{0} .$$

Using the above results, we have shown that

$$\underset{N\to\infty}{\text{plim}} \sqrt{N} [J(\hat{\underline{\theta}}) - \hat{\underline{\theta}}] = \underline{0} ;$$

i.e. the J2SLS and 2SLS estimators are asymptotically equivalent.

3.4 Asymptotic Normality of J2SLS t-Ratios

From equation (2.10) the variance of the J2SLS estimator of θ can be written as

$$V[J(\hat{\underline{\theta}})] = \frac{1}{N(N-1)} \sum_{i=1}^{N} \left[J_i(\hat{\underline{\theta}}) - J(\hat{\underline{\theta}}) \right] \left[J_i(\hat{\underline{\theta}}) - J(\hat{\underline{\theta}}) \right]' . \qquad (3.65)$$

Using the definition of the jackknife it can be shown that

$$J_i(\underline{\hat{\theta}}) - J(\underline{\hat{\theta}}) = \left[N.\underline{\hat{\theta}} - (N-1)\underline{\hat{\theta}}_i\right] - \left[N.\underline{\hat{\theta}} - \frac{(N-1)}{N}\sum_{i=1}^{N}\underline{\hat{\theta}}_i\right], \tag{3.66}$$

whereupon, if we let $\bar{\underline{\hat{\theta}}}_i = \frac{1}{N}\sum_{i=1}^{N}\underline{\hat{\theta}}_i$,

equation (3.65) can be rewritten as ,

$$V[J(\underline{\hat{\theta}})] = \frac{1}{N(N-1)}(N-1)^2\sum_{i=1}^{N}(\bar{\underline{\hat{\theta}}}_i - \underline{\hat{\theta}}_i)(\bar{\underline{\hat{\theta}}}_i - \underline{\hat{\theta}}_i)' . \tag{3.67}$$

Using equations (3.25) and (3.66) we can write

$$\begin{aligned}\bar{\underline{\hat{\theta}}}_i = \underline{\hat{\theta}} - N^{-1} . P^{-1}[&Z'(\Lambda_1 + 2\Lambda_2 - \Lambda_3)\underline{\hat{u}}\\ &- Z'M_X(\Lambda_2 - \Lambda_3)\underline{\hat{u}}\\ &+ Z'(\Lambda_2 - \Lambda_3)(\underline{\hat{v}} - \underline{\hat{w}})\\ &+ Z'M_X\Lambda_3(\underline{\hat{v}} - \underline{\hat{w}})] .\end{aligned} \tag{3.68}$$

Let $\quad \bar{\underline{g}} = \frac{1}{N}\sum_{i=1}^{N}\underline{g}_i$

represent the terms within the square brackets in equation (3.68), then equation (3.68) can be written as

$$\bar{\underline{\hat{\theta}}}_i = \underline{\hat{\theta}} - N^{-1} . P^{-1}\,\bar{\underline{g}} ,$$

and thus

$$\sum_{i=1}^{N}\left(\bar{\underline{\hat{\theta}}}_i - \underline{\hat{\theta}}_i\right)\left(\bar{\underline{\hat{\theta}}}_i - \underline{\hat{\theta}}_i\right)' = \sum_{i=1}^{N}\left(\underline{\hat{\theta}} - \underline{\hat{\theta}}_i + N^{-1}.P^{-1}\bar{\underline{g}}\right)\left(\underline{\hat{\theta}} - \underline{\hat{\theta}}_i + N^{-1}.P^{-1}\bar{\underline{g}}\right)' .$$

Expanding the right hand side of the above expression gives

$$\sum_{i=1}^{N}\left(\hat{\underline{\theta}}_i - \hat{\underline{\theta}}\right)\left(\hat{\underline{\theta}}_i - \hat{\underline{\theta}}\right)' - \frac{1}{N}\sum_{i=1}^{N}\left(\hat{\underline{\theta}}_i - \hat{\underline{\theta}}\right)\bar{\underline{g}}'\ P^{-1}$$

$$- \frac{1}{N}\sum_{i=1}^{N} P^{-1}\ \bar{\underline{g}}(\hat{\underline{\theta}}_i - \hat{\underline{\theta}})' + \frac{1}{N^2}\sum_{i=1}^{N} P^{-1}\bar{\underline{g}}\ \bar{\underline{g}}'\ P^{-1} \quad .$$

Since

$$\sum_{i=1}^{N}(\hat{\underline{\theta}}_i - \hat{\underline{\theta}}) = \sum_{i=1}^{N}\hat{\underline{\theta}}_i - N\hat{\underline{\theta}} = -P^{-1}\ \bar{\underline{g}} \quad ,$$

we can write

$$\sum_{i=1}^{N}(\hat{\underline{\theta}}_i - \hat{\underline{\theta}})\bar{\underline{g}}'P^{-1} = -\sum_{i=1}^{N} P^{-1}\ \bar{\underline{g}}\ \bar{\underline{g}}'\ P^{-1} \quad .$$

From equation (3.14) it follows that

$$\sum_{i=1}^{N}(\hat{\underline{\theta}} - \hat{\underline{\theta}}_i)(\hat{\underline{\theta}} - \hat{\underline{\theta}}_i)' = \sum_{i=1}^{N} P^{-1}\ \underline{g}_i\ \underline{g}_i'\ P^{-1} \quad ,$$

and using the definition of $\underline{g}_i$ as given by equations (3.15), (3.16), and (3.17) we can write

$$\sum_{i=1}^{N}\underline{g}_i\underline{g}_i' = \sum_{i=1}^{N}\left[(\hat{v}_i - \hat{w}_i + \hat{u}_i)\underline{h}_i + \hat{u}_i\underline{j}_i\right]\left[(\hat{v}_i - \hat{w}_i + \hat{u}_i)\underline{h}_i + \hat{u}_i\underline{j}_i\right]'$$

$$= \sum_{i=1}^{N}\Big[(\hat{v}_i - \hat{w}_i + \hat{u}_i)^2\,\underline{h}_i\underline{h}_i' + (\hat{v}_i - \hat{w}_i + \hat{u}_i)\hat{u}_i\underline{h}_i\underline{j}_i'$$

$$+ \hat{u}_i(\hat{v}_i - \hat{w}_i + \hat{u}_i)\underline{j}_i\underline{h}_i' + \hat{u}_i^2\ \underline{j}_i\ \underline{j}_i'\Big] \quad .$$

Letting $\hat{e}_i = (\hat{v}_i - \hat{w}_i + \hat{u}_i)$ and expanding the above terms individually we obtain the following four expressions:

$$\sum_{i=1}^{N} (\hat{v}_i - \hat{w}_i + \hat{u}_i)^2 \underline{h}_i \underline{h}_i' = \sum_{i=1}^{N} \hat{e}_i^2 (\Lambda_3)_{ii}^2 (\underline{z}_i - \underline{a}_i)(\underline{z}_i - \underline{a}_i)'$$

$$+ \sum_{i=1}^{N} \hat{e}_i^2 (\Lambda_2)_{ii}^2 \underline{z}_i \underline{z}_i'$$

$$+ \sum_{i=1}^{N} \hat{e}_i^2 (\Lambda_3)_{ii} (\Lambda_2)_{ii} (\underline{z}_i - \underline{a}_i) \underline{z}_i'$$

$$+ \sum_{i=1}^{N} \hat{e}_i^2 (\Lambda_2)_{ii} (\Lambda_3)_{ii} \underline{z}_i (\underline{z}_i - \underline{a}_i)' \; ;$$

$$\sum_{i=1}^{N} (\hat{v}_i - \hat{w}_i + \hat{u}_i) \hat{u}_i \underline{h}_i \underline{j}_i' = \sum_{i=1}^{N} \hat{e}_i \hat{u}_i (\Lambda_3)_{ii} (\Lambda_2)_{ii} (\underline{z}_i - \underline{a}_i)(\underline{z}_i - \underline{a}_i)'$$

$$- \sum_{i=1}^{N} \hat{e}_i \hat{u}_i (\Lambda_3)_{ii} (\Lambda_1)_{ii} (\underline{z}_i - \underline{a}_i) \underline{z}_i'$$

$$+ \sum_{i=1}^{N} \hat{e}_i \hat{u}_i (\Lambda_2)_{ii}^2 \underline{z}_i (\underline{z}_i - \underline{a}_i)'$$

$$- \sum_{i=1}^{N} \hat{e}_i \hat{u}_i (\Lambda_2)_{ii} (\Lambda_1)_{ii} \underline{z}_i \underline{z}_i' \; ;$$

$$\sum_{i=1}^{N} u_i (\hat{v}_i - \hat{w}_i + \hat{u}_i) \underline{j}_i \underline{h}_i' = \sum_{i=1}^{N} \hat{u}_i \hat{e}_i (\Lambda_3)_{ii} (\Lambda_2)_{ii} (\underline{z}_i - \underline{a}_i)(\underline{z}_i - \underline{a}_i)'$$

$$- \sum_{i=1}^{N} \hat{u}_i \hat{e}_i (\Lambda_3)_{ii} (\Lambda_1)_{ii} \underline{z}_i (\underline{z}_i - \underline{a}_i)'$$

$$+ \sum_{i=1}^{N} \hat{u}_i \hat{e}_i (\Lambda_2)_{ii}^2 (\underline{z}_i - \underline{a}_i) \underline{z}_i'$$

$$- \sum_{i=1}^{N} \hat{u}_i \hat{e}_i (\Lambda_2)_{ii} (\Lambda_1)_{ii} \underline{z}_i \underline{z}_i' \; ;$$

$$\sum_{i=1}^{N} \hat{u}_i^2 \, \underline{j}_i \underline{j}_i' = \sum_{i=1}^{N} \hat{u}_i^2 (\Lambda_2)_{ii}^2 (\underline{z}_i - \underline{a}_i)(\underline{z}_i - \underline{a}_i)'$$
$$- \sum_{i=1}^{N} \hat{u}_i^2 (\Lambda_2)_{ii} (\Lambda_1)_{ii} (\underline{z}_i - \underline{a}_i)\underline{z}_i'$$
$$- \sum_{i=1}^{N} \hat{u}_i^2 (\Lambda_2)_{ii} (\Lambda_1)_{ii} \, \underline{z}_i (\underline{z}_i - \underline{a}_i)'$$
$$+ \sum_{i=1}^{N} \hat{u}_i^2 (\Lambda_1)_{ii}^2 \, \underline{z}_i \underline{z}_i' \; ;$$

where, as before, the ii subscript on a matrix indicates the iith component of that matrix.

Gathering terms, we can write

$$\sum_{i=1}^{N} \underline{g}_i \underline{g}_i' = \sum_{i=1}^{N} \left[\hat{e}_i^2 (\Lambda_3)_{ii}^2 + 2\hat{e}_i \hat{u}_i (\Lambda_3)_{ii} (\Lambda_2)_{ii} + \hat{u}_i^2 \, (\Lambda_2)_{ii}^2\right] (\underline{z}_i - \underline{a}_i)(\underline{z}_i - \underline{a}_i)'$$
$$+ \sum_{i=1}^{N} \left[\hat{e}_i^2 (\Lambda_2)_{ii}^2 - 2\hat{e}_i \hat{u}_i (\Lambda_2)_{ii} (\Lambda_1)_{ii} + \hat{u}_i^2 (\Lambda_1)_{ii}^2\right] \underline{z}_i \underline{z}_i'$$
$$+ \sum_{i=1}^{N} \Big[\hat{e}_i^2 (\Lambda_3)_{ii} (\Lambda_2)_{ii} - \hat{e}_i \hat{u}_i (\Lambda_3)_{ii} (\Lambda_1)_{ii} + \hat{u}_i \hat{e}_i (\Lambda_2)_{ii}^2$$
$$- \hat{u}_i^2 (\Lambda_2)_{ii} (\Lambda_1)_{ii}\Big] (\underline{z}_i - \underline{a}_i)\underline{z}_i'$$
$$+ \sum_{i=1}^{N} \Big[\hat{e}_i^2 (\Lambda_2)_{ii} (\Lambda_3)_{ii} + \hat{e}_i \hat{u}_i (\Lambda_2)_{ii}^2 - \hat{e}_i \hat{u}_i (\Lambda_3)_{ii} (\Lambda_1)_{ii}$$
$$- \hat{u}_i^2 (\Lambda_2)_{ii} (\Lambda_1)_{ii}\Big] \underline{z}_i (\underline{z}_i - \underline{a}_i)' \, . \qquad (3.69)$$

We define the following matrices

$$\left.\begin{aligned} R_1 &= (\underline{y} - Z\hat{\underline{\theta}})(\underline{y} - Z\hat{\underline{\theta}})' = \hat{\underline{u}}\,\hat{\underline{u}}' \, , \\ R_2 &= \left[(I - M_X)(\underline{y} - Z\hat{\underline{\theta}})\right]\left[(I - M_X)(\underline{y} - Z\hat{\underline{\theta}})\right]' = (I - M_X)\hat{\underline{u}}\,\hat{\underline{u}}'(I - M_X) \, , \\ \text{and} \quad R_3 &= (y - Z\hat{\underline{\theta}})\left[(I - M_X)(y - Z\hat{\underline{\theta}})\right]' = \hat{\underline{u}}\,\hat{\underline{u}}'(I - M_X) \, , \end{aligned}\right\} \qquad (3.70)$$

which allows us to rewrite equation (3.69) as

$$\sum_{i=1}^{N} \underline{g}_i \underline{g}_i' = Z'(I - M_X)[\Lambda_2(\text{diag } R_1)\Lambda_2 + \Lambda_3 (\text{diag } R_2)\Lambda_3$$

$$+ 2\Lambda_2(\text{diag } R_3)\Lambda_3] (I - M_X)Z$$

$$+ Z'[\Lambda_2(\text{diag } R_2)\Lambda_2 - 2\Lambda_2(\text{diag } R_3)\Lambda_1 + \Lambda_1(\text{diag } R_1)\Lambda_1]Z$$

$$+ Z'(I-M_X)[\Lambda_3(\text{diag } R_2)\Lambda_2 - \Lambda_3(\text{diag } R_3)\Lambda_1$$

$$+ \Lambda_2(\text{diag } R_3)\Lambda_2 - \Lambda_2(\text{diag } R_1)\Lambda_1]Z$$

$$+ Z'[\Lambda_2(\text{diag } R_2)\Lambda_3 + \Lambda_2(\text{diag } R_3)\Lambda_2$$

$$- \Lambda_3(\text{diag } R_3)\Lambda_1 - \Lambda_2(\text{diag } R_1)\Lambda_1](I - M_X)Z , \qquad (3.71)$$

where (diag) denotes that the relevant matrix has all off-diagonal components equal to zero.

If we also define $S_1 = Z'\Lambda_1 + Z'(I - M_X)\Lambda_2$

and $S_2 = - Z'\Lambda_2 + Z'(I - M_X)\Lambda_3$,

then equation (3.71) can be rewritten as

$$\sum_{i=1}^{N} \underline{g}_i \underline{g}_i' = S_1(\text{diag } R_1)S_1' + S_2(\text{diag } R_2)S_2'$$

$$- S_1(\text{diag } R_3)S_2' - S_2(\text{diag } R_3)S_1' .$$

We also require $\bar{\underline{g}}$ which can be written, using equation (3.68), as

$$\bar{\underline{g}} = Z'(\Lambda_1 + 2\Lambda_2 - \Lambda_3)\hat{\underline{u}} - Z'M_X(\Lambda_2 - \Lambda_3)\hat{\underline{u}}$$

$$+ Z'(\Lambda_2 - \Lambda_3)(\hat{\underline{v}} - \hat{\underline{w}}) + Z'M_X\Lambda_3(\hat{\underline{v}} - \hat{\underline{w}}) ;$$

i.e. $\bar{\underline{g}} = S_1\hat{\underline{u}} - S_2(\hat{\underline{v}} - \hat{\underline{w}} + \hat{\underline{u}})$,

and hence

$$\bar{\underline{g}}\,\bar{\underline{g}}' = S_1R_1S_1' + S_2R_2S_2' - S_1R_3S_2' - S_2R_3S_1' .$$

Upon substituting the above results into equation (3.68), and then into equation (3.67), we obtain

$$V[J(\underline{\hat{\theta}})] = \frac{(N-1)}{N} \cdot P^{-1} \Big[S_1 \Big(\text{diag } R_1 + \frac{1}{N} \cdot R_1 \Big) S_1' + S_2 \Big(\text{diag } R_2 + \frac{1}{N} \cdot R_2 \Big) S_2' - S_1 \Big(\text{diag } R_3 + \frac{1}{N} \cdot R_3 \Big) S_2' - S_2 \Big(\text{diag } R_3 + \frac{1}{N} \cdot R_3 \Big) S_1' \Big] P^{-1} \ . \qquad (3.72)$$

It is shown in Appendix B that the expression in square brackets in equation (3.72) converges to

$$\sigma^2 \Sigma_P$$

in probability as $N \to \infty$.

It follows from equation (3.72) that

$$\operatorname*{plim}_{N \to \infty} V[J(\underline{\hat{\theta}})] = \sigma^2 \Sigma_P^{-1} \ ,$$

since $(N-1)/N \to 1$ as $N \to \infty$.

Since $J(\underline{\hat{\theta}})$ has been shown to be asymptotically equivalent to $\underline{\hat{\theta}}$ it follows that, asymptotically,

$$\frac{(J(\hat{\theta}_j) - \theta_j)}{\sqrt{V[J(\hat{\theta}_j)]}} \sim N(0,1) \ . \qquad (j = 1,2,\ldots, K_1 + g) \ .$$

CHAPTER 4

COMPUTATIONAL ASPECTS

4.1 Computer Algorithms and their Certification

From equation (2.6) the 2SLS estimator of $\underline{\theta}$ can be written as

$$\hat{\underline{\theta}} = [Z'X(X'X)^{-1}X'Z]^{-1}Z'X(X'X)^{-1}X'\underline{y} \qquad (4.1)$$

In all but the simplest cases, equation (4.1) must be evaluated using a computer. Matrix manipulations can be performed using either standard algorithms designed for a specific computer and usually incorporated in the software library, or machine independent algorithms published in computer programming journals. Alternatively one could write one's own algorithms although this might be inadvisable for the more complicated operations such as matrix inversion.

In all computational work in this thesis, matrix manipulations were performed with algorithms written by the author, except for the matrix inversion algorithm. To perform inversions an algorithm written by Devine [11], which inverts a symmetric positive definite matrix by the Choleski decomposition method was selected. All programs were written in Algol 60.

Certification of Devine's algorithm was carried out by the author. This was performed by multiplying the original matrix by its calculated inverse and then obtaining the maximum absolute deviation of elements from the unit matrix. These maximum absolute deviations are given in Table 4.1 for the eight different data matrices which are inverted during the Monte Carlo study in Chapters 6 and 7. The column headed K represents the dimensions of the matrix (i.e. the number of exogenous variables in the model), whilst the column headed λ

denotes the theoretical pairwise correlation between the K variables. The sample correlation matrices are given in Table 6.2.

Table 4.1: Maximum Absolute Deviations (M.A.D.) of $(X'X)^{-1}(X'X)$ from the Unit Matrix

K	λ	M.A.D.
5	0.00	5.46×10^{-12}
5	0.45	8.19×10^{-12}
8	0.00	2.32×10^{-11}
8	0.45	2.46×10^{-11}
8	0.90	1.36×10^{-11}
11	0.00	2.91×10^{-11}
11	0.45	2.18×10^{-11}
11	0.90	2.18×10^{-11}

The accuracy of the matrix inversion, as reflected by the maximum absolute deviations given in Table 4.1, is certainly satisfactory for our purposes.

For $K = 5$ and $\lambda = 0.90$, whilst the moment matrix of predetermined variables was inverted satisfactorily, a further inversion incorporating stochastic matrices which is required at each replication in the Monte Carlo experiment exhibited substantial "inversion errors" and consequently "inconsistent" results were obtained. This problem is discussed in Chapter 6.

A machine independent pseudo-random number generator devised by Pike and Hill [43] was used for generating uniformly distributed pseudo-random numbers for the experiments in Chapters 5, 6 and 7. Favourable evidence of randomness for this algorithm is given by serial and poker tests conducted by Pike and Hill, and by frequency tests in the certification by Sullins [65].

The Box and Muller [6] transformation for generating normally distributed pseudo-random variates is given by

$$\left.\begin{aligned} x_1 &= (-2 \log_e r_1)^{\frac{1}{2}} \sin 2\pi r_2 \\ x_2 &= (-2 \log_e r_1)^{\frac{1}{2}} \cos 2\pi r_2 \ , \end{aligned}\right\} \qquad (4.2)$$

where x_1 and x_2 are two uncorrelated pseudo-random standardized normal variates, and r_1 and r_2 are uniformly distributed pseudo-random variates defined on the [0,1] interval. This transformation produces exact results conditional upon the accuracy of evaluation of the sin and cos functions and the correct distribution of r_1 and r_2. When used in conjunction with a multiplicative congruential pseudo-random number generator however, Neave [38] has shown how the transformation may break down. Amendments to equation (4.2), as suggested by Chay, Fardo and Mazumbar [9], were used in this research, therefore, to avoid Neave's objections. With these amendments the transformation becomes

$$x_1 = (-2 \log_e r_2)^{\frac{1}{2}} \sin 2\pi r_1 \ ,$$

where it should be noted that only the sin transformation is used and the uniformly distributed variates have been interchanged.

The Monte Carlo study reported in Chapters 6 and 7 necessitated the generation of 4,400 pseudo-random standardized normal variates (this figure excludes the additional normally distributed variates required to calculate the power functions in Chapter 7). The Kolmogorov-Smirnov test was conducted to test for any significant divergence between the theoretical (standardized normal) and empirical distributions of the pseudo-random variates. The maximum absolute value of D (the difference between the two distributions) was 0.01306. At the 5% level of significance the hypothesis of equality cannot be rejected.

The pseudo-random normal variates were subsequently transformed into pseudo-random bivariate normal variates by using the transformation

$$Z_1 = \omega_{11}^{\frac{1}{2}} x_1$$

$$Z_2 = \omega_{22}^{\frac{1}{2}} (\delta x_1 + \sqrt{1 - \delta}\, x_2) ,$$

where Z_1 and Z_2 are correlated normal variates with coefficient of correlation equal to δ. ω_{11} and ω_{22} are the specified population variances of Z_1 and Z_2 respectively, and the covariance of Z_1 and Z_2 is given by $\delta\omega_{12}$.

4.2 Computing J2SLS Parameter Estimates

In order to apply the jackknife to the 2SLS estimator we must have some method by which the ith observation can be extracted from equation (4.1). Clearly one could calculate equation (4.1) N times using a 2SLS program and omitting a different observation on each occasion, but this would be a tedious and computationally expensive procedure especially for "large" N and/or K as it would require inverting both matrices in square brackets in equation (4.1) (minus one observation) at each iteration. In addition, rounding errors from the inversion algorithm may lead to a build-up of inaccuracies.

In Chapter 3 we derived equation (3.13) for calculating the 2SLS estimator with the ith observation removed which obviates the need to perform matrix inversions additional to those required for 2SLS with all N observations included. This formula was checked by calculating the J2SLS estimator both ways with a test program and noting that the parameter estimates were identical to <u>at least</u> the sixth decimal place.

4.3 Computing Exact Results

Calculation of the exact moments of the 2SLS estimator, and exact bias in the case of J2SLS, requires evaluation of the confluent hypergeometric function

$$ {}_1F_1\ (\alpha; \gamma\ ; x)\ . \qquad (4.3)$$

Although tables are available (e.g. see Slater [64]), relatively few values of α, γ and x have been tabulated. In general, therefore, the function must be calculated by direct summation of an infinite series or via an asymptotic approximation.

An algorithm for calculating the confluent hypergeometric function with complex parameters via the method of direct summation has been written by Relph [49]. Thacher [69] in his certification of this algorithm mentioned its inefficiency for real arguments.

A problem frequently encountered in this thesis was that of relatively small α and γ, but relatively large x, whence evaluation of equation (4.3) is characterized by slow convergence. When this problem arose it was resolved by using an asymptotic approximation to the confluent hypergeometric function, which for integer α and $\gamma = \alpha + 1$ contains a finite number of terms. A check on the error involved in using the approximation can be made if α is an integer and, if necessary, a correction made.

For a model containing just two endogenous variables, Richardson and Wu [55] have derived the bias of the 2SLS estimator ($\hat{\beta}$) of β in equation (2.5) as

$$E(\hat{\beta}-\beta) = -\frac{\omega_{22}\beta-\omega_{12}}{\omega_{22}}\, e^{-\mu^2/2}\ {}_1F_1\left(\frac{K_2}{2}-1\ ;\ \frac{K_2}{2}\ ;\ \frac{\mu^2}{2}\right), \qquad (4.4)$$

where $\mu^2 = \omega_{22}^{-1}\, \underline{\pi}_{22}'\, X_2'[I_N - X_1(X_1'X_1)^{-1}X_1']\, X_2\, \underline{\pi}_{22}$ is the concentration parameter so named because for every $\varepsilon > 0$

$$\lim_{\mu^2 \to \infty} \Pr(|\hat{\beta} - \beta| > \varepsilon) = 0 .$$

All other notation was explained in Chapter 2.

Clearly $\alpha = (K_2/2 - 1)$ is an integer if K_2 is even.

From Appendix C (equation (C.1)) the asymptotic (in μ^2) expansion of the confluent hypergeometric function (for $\gamma = \alpha + 1$) can be written as

$$_1F_1\ (\alpha; \alpha + 1; x) \sim \frac{\alpha}{x}\, e^x \sum_{r=0}^{\infty} (1 - \alpha)_r \left(\frac{1}{x}\right)^r , \qquad (4.5)$$

and thus the asymptotic approximation to the bias (4.4) is

$$E(\hat{\beta} - \beta) \sim - \frac{\omega_{22}\beta - \omega_{12}}{\omega_{22}} \frac{(K_2 - 2)}{\mu^2} \sum_{r=0}^{\infty} \left(2 - \frac{K_2}{2}\right)_r \left(\frac{2}{\mu^2}\right)^r . \qquad (4.6)$$

The error incurred by applying this approximation for finite μ^2 and integer α is given (from Appendix C (equation C.8)) by

$$\frac{\omega_{22}\beta - \omega_{12}}{\omega_{22}}\ e^{-\mu^2/2}\ \Gamma\left(\frac{K_2}{2}\right)\left(-\frac{2}{\mu^2}\right)^k , \qquad (4.7)$$

where $k = (K_2 - 2)/2$.

It is interesting to note from equation (4.6) that for "large" μ^2 and $K_2 = 2$ the 2SLS estimator is unbiased.

Thus provided the asymptotic approximation of the confluent hypergeometric function terminates after a finite number of terms, equations (4.5) and (4.7) will ensure <u>exact</u> evaluation of this function. The gain in computational efficiency will be particularly marked when the summation of the infinite series required for direct evaluation of the confluent hypergeometric function is slow to converge.

For α non-integer, equation (4.5) is an infinite series, although it can be truncated after (say) n terms. If this is done the error involved in truncating the infinite series after the nth term will not exceed the (n + 1)th term, and will be of the same sign as the (n + 1)th term (Luke [25; p.127]).

In this thesis, when α is not an integer the confluent hypergeometric function had to be truncated in such a way as to ensure that all values of bias and mean square error were correct to at least the number of decimal places given in the text. For integer α, all results are "exact".

CHAPTER 5

THE EXACT BIASES OF THE TWO-STAGE LEAST SQUARES AND JACKKNIFE TWO-STAGE LEAST SQUARES ESTIMATORS

5.1 Résumé of "Exact" Studies

In his pioneering work on the exact finite sample distribution function of the 2SLS estimator, Basmann [4] demonstrated analytically that for a two equation simultaneous equations model, under certain conditions, the moments may not exist (i.e. they may not be finite).

Prior to Basmann's [4] paper, Monte Carlo studies of the relative properties of simultaneous equations estimators had frequently used as their objective function the mean square error in order to compare the relative properties of the estimators. Basmann remarked that an objective function which involved moments of the estimators would have little significance if the moments of the estimators did not exist. In addition, non-finite moments could give rise to "outliers" when this form of objective function is used in Monte Carlo studies, and thus uncritical rejection of these outliers is not a valid procedure.

On the basis of his early work, Basmann [4] conjectured that the moments of the 2SLS estimator exist up to the order of over-identification of the equation being estimated. Basmann's proof was only valid for a two-equation model with $K_1 = K_2 = 2$ and $K_1 = 1$, $K_2 = 3$, although in a later paper (Basmann [5]) he extended it to a three equation model with $g = 2$, $K_1 = 1$ and $K_2 = 3$.

Kabe [21, 22] greatly simplified Basmann's derivations, and this was followed by analytical proofs of Basmann's conjecture for $g = 1$, $K_2 \geq 2$, by Richardson [52] and Sawa [58].

For the general case (i.e. g and K_2 both arbitrary) Mariano [28] has provided a proof of Basmann's conjecture for the even-ordered moments of the 2SLS estimator, whilst Hatanaka [17] has shown that the same conjecture provides a sufficient condition for the existence of the odd-ordered moments.

Sawa [58] and Richardson and Wu [55] derived, independently, the distribution function of the OLS estimator, and then showed how the distribution function of the 2SLS estimator could be derived as a corollary to the derivation of the OLS estimator. For g = 1 the exact moments of the coefficient (β) of the right-hand side endogenous variable in equation (2.5) have been calculated by Sawa [58], Takeuchi [67], and Richardson and Wu [55] for both estimators. From Richardson and Wu [55], the first order moment of the 2SLS estimator can be written as

$$E(\hat{\beta} - \beta) = -\frac{\omega_{22}\beta - \omega_{12}}{\omega_{22}} e^{-\mu^2/2} \, {}_1F_1\left(\frac{K_2}{2} - 1 \, ; \, \frac{K_2}{2} \, ; \, \frac{\mu^2}{2}\right) \qquad (5.1)$$

Second and higher order moments take a more complicated form and the interested reader is referred to the literature previously cited.

The fundamental parameter in all "exact" studies is the concentration parameter μ^2, and not the sample size which does not enter equation (5.1) explicitly, although it is implicit in μ^2.

As μ^2 increases indefinitely, the 2SLS estimator of β converges to its true parameter value (i.e. it is a consistent estimator). A sufficient, but not a necessary, condition for μ^2 to increase indefinitely is for the sample size to increase indefinitely.

In general, the concentration parameter for the jth equation is defined by

$$\mu^2_j = \text{trace } (M_j \textstyle\sum_*^{-1}) \quad ,$$

where $M_j = \Pi_{22}' X_{2j}' [I - X_{1j}(X_{1j}'X_{1j})^{-1} X_{1j}'] X_{2j}\Pi_{22}$

and $\sum_*^{-1}$ is the covariance matrix of non-normalized endogenous variables included in the structural equation.

Essentially, therefore, the moments of the 2SLS estimator are derived in terms of "nuisance" parameters. Sawa [58] assigned "reasonable" values to these nuisance parameters in order to ascertain the relative importance of N, $\rho = \frac{\omega_{12}}{\omega_{22}}$ and K_2. He observed that the bias of 2SLS is an increasing function of $|\rho|$ and that frequently it "is not negligible". In addition, he found that the distribution of the 2SLS estimator was often considerably asymmetric.

Mariano and Ramage [29] considered the effects on 2SLS of excluding relevant exogenous variables and including extraneous exogenous variables in the equation to be estimated. Mathematical complexity precludes useful analysis of the former specification error, but under the latter type of misspecification both the concentration parameter and the degrees of freedom are smaller than for a correctly specified model. The decrease in the concentration parameter increases the bias and mean square error of both estimators, whilst the effect of the decrease in the degrees of freedom is indefinite and depends on other unknown parameters in the model.

5.2 The Concentration Parameter and a Change in Sample Size

Let μ_N^2 and μ_{N-1}^2 denote the concentration parameter based upon N and (N-1) observations respectively, then

$$\mu_N^2 = \omega_{22}^{-1} \underline{\pi}_{22}' X_2' \left[I - X_1(X_1' X_1)^{-1}X_1'\right]X_2 \underline{\pi}_{22} \tag{5.2}$$

and

$$\mu^2_{N-1} = \omega_{22}^{-1} \underline{\pi}_{22}' X_2^{*'} \left[I - X_1^*(X_1^{*'} X_1^*)^{-1} X_1^{*'}\right]X_2^* \underline{\pi}_{22} \quad , \tag{5.3}$$

where the asterisk superscript refers to the relevant data matrix with one observation removed. Without loss of generality we assume that the Nth observation has been removed, i.e.

$$X_1 = \begin{bmatrix} X_1^* \\ \\ \underline{x}_1' \end{bmatrix} \qquad \text{and} \qquad X_2 = \begin{bmatrix} X_2^* \\ \\ \underline{x}_2' \end{bmatrix}$$

where $\underline{x}_1$ and $\underline{x}_2$ are K_1 and K_2 dimensional column vectors representing the omitted observation from X_1 and X_2 respectively.

Noting that

$$(X_1^{*'} X_1^*) = (X_1' X_1 - \underline{x}_1 \underline{x}_1') ,$$

$$(X_2^{*'} X_2^*) = (X_2' X_2 - \underline{x}_2 \underline{x}_2') ,$$

and $$(X_1^{*'} X_2^*) = (X_1' X_2 - \underline{x}_1 \underline{x}_2') ;$$

equation (5.3) can be written as

$$\mu^2_{N-1} = \omega_{22}^{-1} \; \underline{\pi}_{22}' \left\{ (X_2' X_2 - \underline{x}_2 \underline{x}_2') - (X_2' X_1 - \underline{x}_2 \underline{x}_1')(X_1' X_1 - \underline{x}_1 \underline{x}_1')^{-1}(X_1' X_2 - \underline{x}_1 \underline{x}_2') \right\} \underline{\pi}_{22}$$

It can be shown (see Appendix A) that

$$(X_1' X_1 - \underline{x}_1 \underline{x}_1')^{-1} = (X_1' X_1)^{-1} + \frac{(X_1' X_1)^{-1} \underline{x}_1 \underline{x}_1' (X_1' X_1)^{-1}}{1 - \underline{x}_1' (X_1' X_1)^{-1} \underline{x}_1} .$$

Using this result, and after considerable algebraic manipulation, equation (5.3) can be written as

$$\mu^2_{N-1} = \mu^2_N - \frac{1}{(1-c)} \omega_{22}^{-1} \; \underline{\pi}_{22}' (\underline{x}_2 - \underline{d})(\underline{x}_2 - \underline{d})' \underline{\pi}_{22} , \qquad (5.4)$$

where $c = \underline{x}_1' (X_1' X_1)^{-1} \underline{x}_1$, $\quad 0 < c < 1$

and $d = (X_2' X_1)(X_1' X_1)^{-1} \underline{x}_1$.

Since $\underline{\pi}_{22}'(\underline{x}_2 - \underline{d})(\underline{x}_2 - \underline{d})'\underline{\pi}_{22}$ is a positive semi-definite quadratic form, and as ω_{22} and $(1 - c)$ are both greater than zero, it follows that

$$\mu^2_N \geqslant \mu^2_{N-1} \quad ,$$

i.e. the concentration parameter is a monotonically non-decreasing function of sample size.

5.3 The Exact Bias of the Jackknife Two-Stage Least Squares Estimator

5.3.1 Introduction

Since only μ^2 is dependent upon changes in N, the bias of the 2SLS estimator of β with the ith observation omitted $(\hat{\beta}_i)$ can be written, using equation (5.1), as

$$E(\hat{\beta}_i - \beta) = - \frac{\omega_{22}\beta - \omega_{12}}{\omega_{22}} \exp\left(-\frac{\mu^2_{N-1}}{2}\right) {}_1F_1\left(\frac{K_2}{2} - 1 \; ; \; \frac{K_2}{2} \; ; \; \frac{\mu^2_{N-1}}{2}\right) . \quad (5.5)$$

Thus, when the exact bias of the 2SLS estimator can be calculated, it is relatively easy to calculate the exact bias of the J2SLS estimator.

Differentiating the <u>absolute bias</u> with respect to $\mu^2_N/2$, and utilizing the contiguity relations of the confluent hypergeometric function (e.g. see Slater [64; p.19] gives

$$\frac{d|E(\hat{\beta} - \beta)|}{d\,\mu^2_N/2} = - \frac{\omega_{22}\beta - \omega_{12}}{\omega_{22}} \cdot \frac{2}{K_2} \cdot e^{-\mu^2_N/2} \; {}_1F_1\left(\frac{K_2}{2} - 1; \; \frac{K_2}{2} + 1; \; \frac{\mu^2_N}{2}\right) . \quad (5.6)$$

From equation (5.6) is is apparent that the absolute value of the bias is a monotonically decreasing function of the concentration parameter μ^2_N, provided $\beta > \omega_{12}/\omega_{22}$. If $\beta = \omega_{12}/\omega_{22}$ no bias exists, whilst if $\beta < \omega_{12}/\omega_{22}$ it follows that the <u>actual bias</u> is a monotonically decreasing function of μ^2. Similarly, the mean square error of the

2SLS estimator can be shown to be a monotonically decreasing function of the concentration parameter (see Owen [40]).

Earlier in this Chapter is was shown that the concentration parameter is a monotonically non-decreasing function of sample size. Thus, combining these two results, it has been shown that the bias (and the mean square error) of the 2SLS estimator are monotonically non-increasing functions of the sample size; conditional, of course, on the exogenous variables.

We have already seen that the bias of the J2SLS estimator can be written as

$$E(\underline{\hat{\theta}} - \underline{\theta}) + (N-1)\left[E(\underline{\hat{\theta}} - \underline{\theta}) - \frac{1}{N} E\left\{\sum_{i=1}^{N} (\underline{\hat{\theta}}_i - \underline{\theta})\right\}\right]. \quad (5.7)$$

It follows from the above result that the term in square brackets in equation (5.7) will be either zero or opposite in sign to $E(\underline{\hat{\theta}} - \underline{\theta})$. Consequently, application of the jackknife will have one of three possible effects on the bias of the 2SLS estimator:

1. The absolute bias decreases but its sign remains unchanged;
2. The absolute bias decreases and its sign changes;
3. The absolute bias increases and its sign changes.

If the bias decreases slowly or approximately linearly with sample size, then it seems reasonable to expect possibilities 1. or 2. to occur. When the bias is decreasing rapidly with sample size however, there could be a tendency for the jackknife to "over-correct" for bias and possibility 3. could occur.

Since the above eventualities are somewhat vague, we turn from heuristic analysis to consider an analytical investigation of the conditions under which jackknifing is unlikely to decrease the bias of the 2SLS estimator. First we consider the exact bias of the 2SLS estimator of β as given by equation (5.1) for the special case of $K_2 = 2$, then we consider a more general approach using Kadane's [23] approximation to the bias of the 2SLS estimator.

5.3.2 Effect of Jackknifing on the Exact Bias of 2SLS when $K_2 = 2$

From equation (5.1), if $K_2 = 2$ the exact bias of the 2SLS estimator of β degenerates to

$$E(\hat{\beta} - \beta) = - \frac{(\omega_{22}\beta - \omega_{12})}{\omega_{22}} e^{-\mu^2/2} \quad , \tag{5.8}$$

since $\;_1F_1(0,1,\mu^2/2) = 1$.

Expanding the exponential term in equation (5.8) and setting $\rho = \frac{\omega_{12}}{\omega_{22}}$ gives

$$E(\hat{\beta} - \beta) = -(\beta-\rho)\left[1 + \left(-\frac{\mu^2}{2}\right) + \left(-\frac{\mu^2}{2}\right)^2 . \frac{1}{2!} + \ldots . + \left(-\frac{\mu^2}{2}\right)^r . \frac{1}{r!} + \ldots\right] \tag{5.9}$$

Since μ^2 is of order N, when $K_2 = 2$ the bias of the 2SLS estimator is clearly a function of terms (with alternating signs) of increasing powers of order N. Whilst alternating signs will not weaken the jackknife's bias reducing properties, equation (5.9) clearly contravenes Quenouille's basic assumption regarding the application of the jackknife, viz: that the bias can be expressed as an expansion in terms of increasing powers of order $(\frac{1}{N})$. This suggests that application of the jackknife technique is unlikely to be successful if $K_2 = 2$.

When $K_2 > 2$

$$_1F_1\left(\frac{K_2}{2} - 1; \frac{K_2}{2} ; \frac{\mu^2}{2}\right)$$

takes the form of an infinite series and the bias cannot be expanded into an expression such as equation (5.9). We can however fall back on equation (4.6) which gives a "large" μ^2 expansion of the bias, equation (5.1), in terms of increasing powers of order $\frac{1}{N}$, provided $K_2 > 2$ (although if K_2 is an integer this expansion will terminate after a finite number of terms). This suggests that for "large" μ^2 and $K_2 > 2$, application of the jackknife technique could reduce the bias of the 2SLS estimator.

Both of the above observations will be investigated by means of a simulation experiment in Section 5.5.

5.4 Jackknifing the Approximate Bias of the 2SLS Estimator

Kadane [23] has derived the leading terms of the first two moments of a distribution approximating the exact distribution of the 2SLS estimator, although it should be emphasized that the moments of approximate distributions are not necessarily identical to approximations to the moments of the exact distribution.

Nagar's [37] work in this field carries a similar interpretation.

Kadane's approximations are based on a "small" σ asymptotic expansion of the moments of the k-class estimators (N.B. in our notation $\sigma = \omega_{11} - \omega_{12}\rho + \omega_{22}(\beta - \rho)^2$ and is not to be confused with the σ used elsewhere in this thesis. The definition of σ given here is restricted solely to this Section). For N fixed, $\mu^2 \to \infty$ if $\sigma \to 0$ and it can be shown (see Sawa [59; Appendix C]) that Kadane's (and Nagar's) expansion coincides with "large" μ^2 expansions of the exact moments, provided the latter exist.

Kadane [23] has approximated the bias of the 2SLS estimator by

$$E(\underline{\hat{\theta}} - \underline{\theta}) = \sigma^2(L-1)Q\underline{q} + O(\sigma^3), \quad (5.10)$$

where $L = K_2 - g$, i.e. the degree of overidentification of the equation being estimated ,

$$W = [X\Pi : X_1],$$

$$Q = (W'W)^{-1}$$

$$\underline{q} = \frac{1}{N} E[Y : X_1]'\underline{u} ,$$

and $$\underline{\theta} = \begin{bmatrix} \underline{\beta} \\ .. \\ \underline{\gamma}_1 \end{bmatrix}$$

Let $\underline{\hat{\theta}}_i$ denote the 2SLS estimator of $\underline{\theta}$ with the ith observation removed then,

$$E(\underline{\hat{\theta}}_i - \underline{\theta}) = \sigma^2(L-1)Q_i\underline{q} + O(\sigma^3) , \quad (5.11)$$

where $Q_i = (W'W - \underline{w}_i\underline{w}_i')^{-1}$ and $\underline{w}_i$ is a $K_1 + g$ dimensional column vector representing the omitted observation from W.

From Appendix A, it can be shown that

$$Q_i = (W'W)^{-1} + \frac{(W'W)^{-1}\underline{w}_i\underline{w}_i'(W'W)^{-1}}{1 - \underline{w}_i'(W'W)^{-1}\underline{w}_i} = Q + \frac{Q\underline{w}_i\underline{w}_i'Q}{1 - \underline{w}_i'Q\underline{w}_i} ,$$

and hence

$$E(\underline{\hat{\theta}}_i - \underline{\theta}) = \sigma^2(L-1)Q\underline{q} + \sigma^2(L-1)\frac{Q\underline{w}_i\underline{w}_i'Q}{1 - \underline{w}_i'Q\underline{w}_i}\underline{q}$$

$$= E(\underline{\hat{\theta}} - \underline{\theta}) + \sigma^2(L-1)\frac{Q\underline{w}_i\underline{w}_i'Q}{1 - \underline{w}_i'Q\underline{w}_i}\underline{q} , \quad (5.12)$$

where terms of higher order in σ have been neglected.

From the definition of the jackknife, and using equation (5.12), we obtain

$$E[J(\underline{\hat{\theta}}) - \underline{\theta}] = NE(\underline{\hat{\theta}} - \underline{\theta}) - \frac{(N-1)}{N}\sum_{i=1}^{N} E(\underline{\hat{\theta}}_i - \underline{\theta})$$

$$= NE(\hat{\underline{\theta}} - \underline{\theta}) - (N-1)E(\hat{\underline{\theta}} - \underline{\theta}) - \sigma^2 \frac{(N-1)}{N} \sum_{i=1}^{N} (L-1) \frac{Q\underline{w}_i\underline{w}_i' Q}{1-\underline{w}_i' Q\underline{w}_i} \underline{q}$$

$$= E(\hat{\underline{\theta}} - \underline{\theta}) - \sigma^2 (L-1) \frac{(N-1)}{N} \sum_{i=1}^{N} \frac{Q\underline{w}_i\underline{w}_i' Q}{1-\underline{w}_i' Q\underline{w}_i} \underline{q} \ . \qquad (5.13)$$

For jackknifing <u>not to increase</u> the absolute value of the bias of the 2SLS estimator over all parameters being estimated, we require

$$E[J(\hat{\underline{\theta}}) - \underline{\theta}]\ E[J(\hat{\underline{\theta}}) - \underline{\theta}]' - E[\hat{\underline{\theta}} - \underline{\theta}]\ E[\hat{\underline{\theta}} - \underline{\theta}]' \qquad (5.14)$$

to have all main diagonal components ≤ 0.

Consider the last term in equation (5.13) which can be rewritten as

$$\sigma^2 (L-1) \frac{(N-1)}{N} \sum_{i=1}^{N} \frac{Q\underline{w}_i\underline{w}_i' Q}{1-\underline{w}_i' Q\underline{w}_i} \underline{q} = \sigma^2 (L-1) \frac{(N-1)}{N} Q \sum_{i=1}^{N} \left[\frac{\underline{w}_i\underline{w}_i'}{1-\underline{w}_i' Q\underline{w}_i} \right] Q\underline{q} \ .$$

Let Λ be an $N \times N$ diagonal matrix with iith component equal to $\underline{w}_i' Q \underline{w}_i$, then

$$[I - \Lambda]^{-1}$$

is an $N \times N$ diagonal matrix with iith component equal to

$$\frac{1}{1 - \underline{w}_i' Q\underline{w}_i} \ ,$$

and hence

$$\sum_{i=1}^{N} \frac{\underline{w}_i\underline{w}_i'}{1-\underline{w}_i' Q\underline{w}_i} = W'[I - \Lambda]^{-1} W \ . \qquad (5.15)$$

Thus equation (5.13) can be rewritten as

$$E[J(\hat{\underline{\theta}}) - \underline{\theta}] = E(\hat{\underline{\theta}} - \underline{\theta}) - \sigma^2 (L-1) \frac{(N-1)}{N} QW'[I - \Lambda]^{-1} WQ\underline{q} \ , \qquad (5.16)$$

and upon substituting into equation (5.14) we obtain

$$E[J(\hat{\underline{\theta}}) - \underline{\theta}]E[J(\hat{\underline{\theta}}) - \underline{\theta}]' = E[\hat{\underline{\theta}} - \underline{\theta}]E[\hat{\underline{\theta}} - \underline{\theta}]'$$

$$- \sigma^2(L-1)\frac{(N-1)}{N} QW' \; [I - \Lambda]^{-1} \; WQ\underline{q} \; E[\hat{\underline{\theta}} - \underline{\theta}]'$$

$$- \sigma^2(L-1)\frac{(N-1)}{N} E[\hat{\underline{\theta}} - \underline{\theta}]\underline{q}' \; QW' \; [I - \Lambda]^{-1}WQ$$

$$+ \sigma^4(L-1)^2\frac{(N-1)^2}{N^2} QW'[I - \Lambda]^{-1}WQ\underline{q}\,\underline{q}' QW'[I - \Lambda]^{-1}WQ \quad ,$$

which can be rewritten as

$$E[J(\hat{\underline{\theta}})-\underline{\theta}]E[J(\hat{\underline{\theta}})-\underline{\theta}]' = \sigma^4(L-1)^2 QW'\left[I-\frac{(N-1)}{N}(I-\Lambda)^{-1}\right] WQ\underline{q}\,\underline{q}' QW' \cdot\left[I - \frac{(N-1)}{N} (I - \Lambda)^{-1}\right]WQ \quad ,$$

where $\sigma^2(L-1)Q\underline{q}$ has been substituted for $E[\hat{\underline{\theta}} - \underline{\theta}]$.

Thus, for the jackknife not to increase the bias of the 2SLS estimator, we are required to show that

$$\sigma^4(L-1)^2 QW'\left[I - \frac{(N-1)}{N} (I-\Lambda)^{-1}\right] WQ\underline{q}\,\underline{q}' QW' \left[I - \frac{(N-1)}{N} (I - \Lambda)^{-1}\right] WQ - \sigma^4(L-1)^2 Q\underline{q}\,\underline{q}' Q \qquad (5.17)$$

has all main diagonal components ≤ 0.

If we denote the iith component of Λ by λ_i, then Teekens [68; pp.103-106] has shown that, in general,

$$\frac{1}{N} \leq \lambda_i \leq 1, \qquad (i = 1, 2, \ldots , N)$$

and it follows that

$$\left[I - \frac{(N-1)}{N} (I - \Lambda)^{-1}\right]_{ii} < 0 \;, \; (i=1, 2, \ldots , N) \qquad (5.18)$$

where the ii subscript refers to the iith component of the matrix formed by those terms in the square brackets.

Thus, when equation (5.18) holds,

$$- QW' \left[I - \frac{(N-1)}{N} (I - \Lambda)^{-1} \right] WQ$$

must be positive definite.

We now require the following theorem from Rao [46; p.37]:

THEOREM

Let A and B be real $m \times m$ symmetric matrices of which B is positive definite. Then there exists a matrix R such that

$$A = R'^{-1} \Delta R^{-1} \quad \text{and} \quad B = R'^{-1} R^{-1}$$

where Δ is a diagonal matrix.

Using this theorem, there exists a matrix R such that

$$-R'QW' \left[I - \frac{(N-1)}{N} (I - \Lambda)^{-1} \right] WQR = \Delta$$

and $\quad R'QR = I$,

where Δ is a diagonal matrix whose main diagonal components are positive and equal to the roots of the equation

$$\left| - QW' \left[I - \frac{(N-1)}{N} (I - \Lambda)^{-1} \right] WQ - \lambda Q \right| = 0 ,$$

or $\quad \left| - Q^{\frac{1}{2}}W' \left[I - \frac{(N-1)}{N} (I - \Lambda)^{-1} \right] WQ^{\frac{1}{2}} - \lambda I \right| = 0 .$

Thus, from equation (5.17), for the jackknife not to increase bias we require

$$\sigma^4 (L-1)^2 (R')^{-1} R'QRR^{-1} \underline{q}\underline{q}' (R')^{-1} R'QRR^{-1}$$

$$- \sigma^4 (L-1)^2 (R')^{-1} R'QW \left[I - \frac{(N-1)}{N} (I - \Lambda)^{-1} \right] W QRR^{-1} \underline{q}\underline{q}' R'^{-1}$$

$$.R'QW' \left[I - \frac{(N-1)}{N} (I - \Lambda)^{-1} \right] WQRR^{-1}$$

$$= \sigma^4 (L-1)^2 \left[(R')^{-1} R^{-1} \underline{q}\underline{q}' R'^{-1} (R^{-1}) - (R')^{-1} \Delta R^{-1} \underline{q}\underline{q}' R'^{-1} \Delta R^{-1} \right]$$

to have non-negative main diagonal components. This cannot be shown but the sum of squared biases will be reduced in the general case and, in the case of two included endogenous variables, the squared bias of the endogenous coefficient estimator will be reduced if the roots λ_i, of the matrix

$$-Q^{\frac{1}{2}}W'\left[I - \frac{(N-1)}{N}(I-\Lambda)^{-1}\right]WQ^{\frac{1}{2}}$$

are such that

$$0 \leq \lambda_i \leq 1 \quad . \qquad (i = 1, 2, \ldots, K_1 + g) \tag{5.20}$$

Since this condition is dependent upon W it is not possible to give a general statement concerning its existence. However, a sufficient condition for equation (5.20) to hold is that

$$\left[I - \frac{(N-1)}{N}(I-\Lambda)^{-1}\right]_{ii} \geq -1 , \qquad (i = 1, 2, \ldots, N)$$

i.e. $\frac{(N-1)}{N} \frac{1}{1 - \underline{w}_i' Q\underline{w}_i} - 1 \leq 1$.

or $\underline{w}_i' Q\underline{w}_i \leq \frac{N+1}{2N}$. $\qquad (i = 1, 2, \ldots, N)$

It is known that

$$\sum_{i=1}^{N} \underline{w}_i' Q\underline{w}_i = \text{trace } W(W'W)^{-1}W' = K_1 + g ,$$

and so the "average" value of $\underline{w}_i' Q\underline{w}_i$ is $(K_1 + g)/N$.

But for

$$\frac{K_1 + g}{N} \geq \frac{N+1}{2N} , \quad \text{or identically} \quad K_1 + g \geq \frac{N+1}{2} ,$$

the sufficient condition cannot hold.

This suggests that when the number of observations is not at least twice the number of included variables, the jackknife should not be used.

5.5 A Comparison of the Exact Bias of the 2SLS and J2SLS Estimators

The analytical results derived in this Chapter can be summarized as follows:

(i) for a structural equation containing just two endogenous variables, if $K_2 = 2$ jackknifing is unlikely to be successful;

(ii) in general, even when $K_2 > 2$ and μ^2 is "reasonably large", jackknifing is unlikely to be successful unless the number of observations is at least twice the number of variables included in the equation being estimated.

It is apparent from these results that analytical guidelines on criteria for applying the jackknife to the 2SLS estimator are rather vague. A series of experiments was conducted therefore to observe circumstances in which the jackknife is successful in reducing the bias of the 2SLS estimator.

The experiments compare the exact biases of 2SLS and J2SLS as given by equations (5.1) and (5.7) (using equation (5.5)) respectively, but take no account of any resulting change in variance.

The exogenous variables were generated as pseudo- random numbers from the uniform distribution in the range 0 to 100. A specified level of theoretical multicollinearity (λ) was applied such that the theoretical pairwise correlation between exogenous variables was the same for each experiment. λ took values from 0.0 to 0.8 in steps of 0.2.

The relative biases of the 2SLS and J2SLS estimators were calculated exactly for specified values of N, K_1, K_2, ω_{12}, ω_{22}, and the sub-vector of reduced form coefficients, π_{22}.

The values of ω_{12} and ω_{22} were set at 0.0 and 1000.0 respectively for all experiments. From equations (5.1) and (5.5) it can be seen that ω_{12} and ω_{22} enter the expressions for bias only through ρ. Consequently a change in either or both of these parameters only has a simple multiplicative effect on the biases and can be ignored without loss of generality.

K_1 was fixed at 2 for the majority of the experiments, whilst K_2 took on values of 2, 4 and 6. N took values of 10, 20 and 30.

Tables 5.1 - 5.7 give the results of the experiments. The relative bias of both estimators is given, together with the corresponding value of the concentration parameter, μ_N^2.

Table 5.7 gives the results of experiments designed to test the conclusion derived in Section 5.4, viz: if the number of observations is not at least twice the number of included variables the jackknife should not be used. For the purpose of these experiments N and K_2 were fixed at 20 and 4 respectively, whilst K_1 took values of 4, 6 and 8.

An asterisk indicates experiments where the jackknife did not reduce the bias of the 2SLS estimator.

It was suggested in Section 5.3.2 that if $K_2 = 2$ jackknifing may not be successful in reducing bias. From Tables 5.1 and 5.4 it is apparent that jackknifing is indeed generally unsuccessful. In addition, in Section 4.3 it was shown that for "large" μ^2 and $K_2 = 2$ the 2SLS estimator is "nearly" unbiased. The results in Table 5.4 indicate the deleterious effects of using the jackknife under such conditions, even though μ_N^2 is not very "large".

For $K_2 > 2$ application of the jackknife, in general, produces a fairly substantial reduction in the bias of the 2SLS estimator. Note that for fixed N, J2SLS does not exhibit a consistent pattern of bias as λ increases, whereas the bias of 2SLS always increases with increasing λ.

In general, except for very small values of μ_N^2, jackknifing changes the sign of the 2SLS bias.

The results in Table 5.7 indicate that it would be unwise to apply the jackknife to the 2SLS estimator when the number of observations is not at least twice the number of included variables. For "small" μ_N^2 the jackknife produces a substantial reduction in bias, but the ensuing Monte Carlo study will indicate that there is likely to be a substantial increase in the variance of the J2SLS estimator when μ_N^2 is "small". However, since μ_N^2 is never known in practice, it would be unwise to use the jackknife when this condition prevails.

These exact results suggest that the jackknife can be most useful in reducing bias when the equation being estimated is "well" over-identified. It would certainly be unwise to use the jackknife when $K_2 = 2$ or when the number of observations is not at least twice the number of included variables.

Table 5.1: Exact Relative Biases of the 2SLS and J2SLS Estimators

$K_2 = 2$ $\pi_{22} = (0.5, -0.5)$

	N = 10			N = 20			N = 30		
λ	μ_N^2	Relative Bias		μ_N^2	Relative Bias		μ_N^2	Relative Bias	
		2SLS	J2SLS		2SLS	J2SLS		2SLS	J2SLS
0.0	5.8775	-0.0529	+0.1873*	8.9645	-0.0113	+0.0592*	12.0464	-0.0024	+0.0179*
0.2	4.7927	-0.0910	+0.2156*	7.4405	-0.0242	+0.0924*	10.3583	-0.0056	+0.0317*
0.4	3.5482	-0.1696	+0.2259*	5.6726	-0.0586	+0.1418*	8.3496	-0.0154	+0.0611*
0.6	2.2820	-0.3195	+0.1421	3.7775	-0.1513	+0.1732*	6.0012	-0.0498	+0.1189*
0.8	1.0954	-0.5783	-0.1826	1.8411	-0.3983	-0.0028	3.2606	-0.1959	+0.1490

$K_1 = 2$ $\omega_{12} = 0.0$ $\omega_{22} = 1000.0$

Table 5.2: Exact Relative Biases of the 2SLS and J2SLS Estimators

$K_2 = 4$ $\pi_{22} = (0.5, -0.5, 0.5, -0.5)$

	N = 10			N = 20			N = 30		
λ	μ_N^2	Relative Bias		μ_N^2	Relative Bias		μ_N^2	Relative Bias	
		2SLS	J2SLS		2SLS	J2SLS		2SLS	J2SLS
0.0	4.8967	-0.3731	-0.0161	15.5754	-0.1284	+0.0194	29.7400	-0.0672	+0.0071
0.2	3.6992	-0.4556	-0.0906	12.3167	-0.1620	+0.0241	21.4978	-0.0930	+0.0101
0.4	2.8151	-0.5366	-0.1842	9.4623	-0.2095	+0.0242	15.0562	-0.1328	+0.0143
0.6	2.0354	-0.6275	-0.3146	6.6289	-0.2907	+0.0011	9.6338	-0.2059	+0.0131
0.8	1.1894	-0.7538	-0.5275	3.5345	-0.4692	-0.1395	4.6532	-0.3879	-0.0745

$K_1 = 2$ $\omega_{12} = 0.0$ $\omega_{22} = 1000.0$

Table 5.3: Exact Relative Biases of the 2SLS and J2SLS Estimators

$K_2 = 6$ $\pi_{22} = (0.5, -0.5, 0.5, -0.5, 0.5, -0.5)$

	N = 20			N = 30		
λ	μ_N^2	Relative Bias		μ_N^2	Relative Bias	
		2SLS	J2SLS		2SLS	J2SLS
0.0	32.8269	-0.1144	+0.0101	52.3695	-0.0735	+0.0046
0.2	23.0495	-0.1585	+0.0070	36.8341	-0.1027	+0.0039
0.4	15.3959	-0.2261	-0.0054	24.5283	-0.1498	+0.0001
0.6	9.2195	-0.3407	-0.0544	14.6172	-0.2362	-0.0191
0.8	4.2646	-0.5502	-0.2455	6.6967	-0.4252	-0.1300

$K_1 = 2$ $\omega_{12} = 0.0$ $\omega_{22} = 1000.0$

Table 5.4: Exact Relative Biases of the 2SLS and J2SLS Estimators

$K_2 = 2$ $\Pi_{22} = (1.0, -1.0)$

	N = 10			N = 20			N = 30		
		Relative Bias			Relative Bias			Relative Bias	
λ	μ_N^2	2SLS	J2SLS	μ_N^2	2SLS	J2SLS	μ_N^2	2SLS	J2SLS
0.0	23.5099	0.0000	+0.0008*	35.8579	0.0000	0.0000*	48.1854	0.0000	-0.0005*
0.2	19.1710	-0.0001	+0.0027*	29.7621	0.0000	0.0000*	41.4331	0.0000	0.0000*
0.4	14.1927	-0.0008	+0.0141*	22.6905	0.0000	+0.0003*	33.3983	0.0000	0.0000*
0.6	9.1281	-0.0104	+0.0777*	15.1101	-0.0005	+0.0058*	24.0050	0.0000	+0.0001*
0.8	4.3816	-0.1118	+0.2585	7.3645	-0.0252	+0.0918*	13.0424	-0.0015	+0.0107*

$K_1 = 2$ $\omega_{12} = 0.0$ $\omega_{22} = 1000.0$

Table 5.5: Exact Relative Biases of the 2SLS and J2SLS Estimators

$K_2 = 4$ π_{22} = (1.0, -1.0, 1.0, -1.0)

	N = 10			N = 20			N = 30		
		Relative Bias			Relative Bias			Relative Bias	
λ	μ_N^2	2SLS	J2SLS	μ_N^2	2SLS	J2SLS	μ_N^2	2SLS	J2SLS
0.0	19.5866	-0.1021	+0.0409	62.3016	-0.0321	+0.0051	118.9602	-0.0168	+0.0018
0.2	14.7967	-0.1351	+0.0578	49.2668	-0.0406	+0.0070	85.9910	-0.0233	+0.0026
0.4	11.2603	-0.1770	+0.0778	37.8492	-0.0528	+0.0098	60.2250	-0.0332	+0.0036
0.6	8.1416	-0.2415	+0.0849	26.5155	-0.0754	+0.0145	38.5351	-0.0519	+0.0061
0.8	4.7575	-0.3814	+0.0017	14.1380	-0.1413	+0.0248	18.6126	-0.1074	+0.0127

$K_1 = 2$ $\omega_{12} = 0.0$ $\omega_{22} = 1000.0$

Table 5.6: Exact Relative Biases of the 2SLS and J2SLS Estimators

$K_2 = 6$ $\pi_{12} = (1.0, -1.0, 1.0, -1.0, 1.0, -1.0)'$

	N = 20			N = 30		
λ	μ_N^2	Relative Biases		μ_N^2	Relative Bias	
		2SLS	J2SLS		2SLS	J2SLS
0.0	131.3077	-0.0300	+0.0046	209.4781	-0.0189	+0.0019
0.2	92.1980	-0.0424	+0.0058	147.3363	-0.0268	+0.0024
0.4	61.5836	-0.0628	+0.0081	98.1132	-0.0399	+0.0031
0.6	36.8780	-0.1026	+0.0102	58.4687	-0.0661	+0.0040
0.8	17.0585	-0.2070	+0.0024	26.7869	-0.1382	+0.0016

$K_1 = 2$ $\omega_{12} = 0.0$ $\omega_{22} = 1000.0$

Table 5.7: Exact Relative Biases of the 2SLS and J2SLS Estimators ($K_1 + g$ "large" relative to N)

$K_2 = 4$ $\pi_{22} = (0.5, -0.5, 0.5, -0.5)$

	$K_1 = 8$	N = 20		$K_1 = 10$	N = 20		$K_1 = 12$	N = 30	
		Relative Bias			Relative Bias			Relative Bias	
λ	μ_N^2	2SLS	J2SLS	μ_N^2	2SLS	J2SLS	μ_N^2	2SLS	J2SLS
0.0	10.5786	-0.1881	+0.1529	11.7998	-0.1690	+0.2040*	4.4360	-0.4018	+0.5649*
0.2	8.6514	-0.2281	+0.1626	8.8120	-0.2242	+0.2356*	2.9796	-0.5199	+0.3831
0.4	6.7927	-0.2846	+0.1608	6.2274	-0.3069	+0.2373	1.8676	-0.6500	+0.0091
0.6	4.8248	-0.3774	+0.1174	3.8990	-0.4399	+0.1422	1.0282	-0.7819	-0.2801
0.8	2.5090	-0.5698	-0.1007	1.8195	-0.6566	-0.1904	0.4652	-0.8922	-0.6428

$\omega_{12} = 0.0$ $\omega_{22} = 1000.0$

CHAPTER 6

MONTE CARLO STUDY

6.1 Design of Experiments

An evaluation of the effects of applying the jackknife technique to the 2SLS estimator necessitates the use of Monte Carlo methods. Although the exact finite sample distribution and exact moments (where they exist) have been derived for several simultaneous equation estimators in the context of the model used in the ensuing study (e.g. see the bibliographical paper compiled by Owen and Knight [41]), neither the exact finite sample distribution nor exact second and higher order moments of the J2SLS estimator have been derived. Consequently, a Monte Carlo analysis is our only method of evaluating the effects of applying the jackknife technique to the 2SLS estimator.

The model used for one-third of the experiments was

$$\underline{y}_1 = \beta_{12}\underline{y}_2 + \gamma_{10} + \gamma_{11}\underline{X}_1 + \underline{u}_1 \tag{6.1}$$

$$\underline{y}_2 = \beta_{21}\underline{y}_1 + \gamma_{20} + \gamma_{22}\underline{X}_2 + \gamma_{23}\underline{X}_3 + \gamma_{24}\underline{X}_4 + \underline{u}_2 , \tag{6.2}$$

whilst for the remaining experiments equation (6.2) was augmented by an additional three or six exogenous variables.

The reduced form of this two-equation model is given by

$$\underline{y}_1 = \underline{\pi}_{10} + \sum_{i=1}^{4} \underline{X}_i \, \pi_{1i} + \underline{v}_1 \tag{6.3}$$

$$\underline{y}_2 = \underline{\pi}_{20} + \sum_{i=1}^{4} \underline{X}_i \, \pi_{2i} + \underline{v}_2 , \tag{6.4}$$

where both equations should be augmented by the relevant additional terms when $K_2 = 6$ and $K_2 = 9$.

The set of parameter values used in the experiments is given in Table 6.1.

Table 6.1: Parameter Values Used in Monte Carlo Experiments

β_{12}	0.8	β_{21}	-0.7	γ_{25}	-1.0	γ_{28}	1.2
γ_{10}	50.0	γ_{20}	50.0	γ_{26}	1.9	γ_{29}	-1.5
γ_{11}	1.2	γ_{22}	1.3	γ_{27}	-1.1	γ_{210}	0.9
		γ_{23}	1.6				
		γ_{24}	-2.0				

The exogenous variables were generated as rectangularly and independently distributed pseudo-random variables in the range 0 to 100, but were then transformed in order to obtain a specified theoretical pairwise correlation (λ) between them. The sample correlations are given in Table 6.2. Values for experiments using less than the full set of exogenous variables (i.e. less than 10, excluding the constant) should be read-off from the upper left corner of the table.

All experiments were based on a sample size of 20.

The reduced form disturbances, the v_{it} $(i = 1,2)$, were generated as bivariate normal variates with zero mean and covariance matrix

$$\Omega = \begin{bmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{bmatrix} = \begin{bmatrix} 1600 & 1520\delta \\ 1520\delta & 1444 \end{bmatrix},$$

where the coefficient δ was given the value 0.19 in half of the experiments, and 0.76 in the other half.

Each estimate of the parameters in equation (6.1) (i.e. the first equation only) was calculated as the mean of 100 replications of the

Table 6.2: Matrix of Sample Correlations Between the Exogenous Variables

		X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
X_1	0.00	+1.0000									
	0.45	+1.0000									
	0.90	+1.0000									
X_2	0.00	-0.2977	+1.0000								
	0.45	+0.1317	+1.0000								
	0.90	+0.5001	+1.0000								
X_3	0.00	+0.0628	+0.0971	+1.0000							
	0.45	+0.1084	+0.5244	+1.0000							
	0.90	+0.3874	+0.7815	+1.0000							
X_4	0.00	-0.0818	-0.1558	-0.0404	+1.0000						
	0.45	-0.0215	+0.0791	+0.3923	+1.0000						
	0.90	+0.2987	+0.5786	+0.7979	+1.0000						
X_5	0.00	+0.1821	-0.3951	-0.0084	-0.0513	+1.0000					
	0.45	+0.1606	-0.2720	+0.0516	+0.3636	+1.0000					
	0.90	+0.3811	+0.3987	+0.6353	+0.8106	+1.0000					
X_6	0.00	-0.3281	+0.0148	+0.1738	-0.0418	-0.2519	+1.0000				
	0.45	-0.2641	-0.2447	+0.1183	+0.1618	+0.1851	+1.0000				
	0.90	+0.1171	+0.1936	+0.5301	+0.6734	+0.7453	+1.0000				
X_7	0.00	+0.2261	-0.1912	-0.2425	+0.1720	+0.1684	-0.3381	+1.0000			
	0.45	+0.1211	-0.2045	-0.2060	+0.1037	+0.2572	+0.1454	+1.0000			
	0.90	+0.2562	+0.1639	+0.3414	+0.5461	+0.6784	+0.7272	+1.0000			
X_8	0.00	+0.1431	+0.0475	+0.3820	-0.2319	+0.0195	+0.0832	-0.0321	+1.0000		
	0.45	+0.1829	+0.0039	+0.2346	+0.0241	+0.1246	+0.1440	+0.4693	+1.0000		
	0.90	+0.2746	+0.2069	+0.4450	+0.5172	+0.6199	+0.6784	+0.8658	+1.0000		
X_9	0.00	+0.0028	+0.3188	+0.1081	-0.0222	-0.3941	+0.1093	-0.1402	-0.4943	+1.0000	
	0.45	+0.0979	+0.3869	+0.3884	+0.1222	-0.3117	+0.0447	+0.0636	-0.0715	+1.0000	
	0.90	+0.2714	+0.3866	+0.6075	+0.6528	+0.5936	+0.6996	+0.8171	+0.7982	+1.0000	
X_{10}	0.00	+0.1320	+0.1307	+0.2076	-0.1996	+0.4577	-0.0002	-0.0124	-0.0414	+0.3131	+1.0000
	0.45	+0.1426	+0.3024	+0.3545	+0.0024	+0.2026	+0.1517	+0.0708	-0.0348	+0.6672	+1.0000
	0.90	+0.2598	+0.3838	+0.5766	+0.5465	+0.6131	+0.6751	+0.7360	+0.6869	+0.9206	+1.0000

relevant estimator. All experiments were devised to ensure that at least the first two integer moments of the 2SLS estimator were finite. This would not be so with the second equation, (6.2).

6.2 Exact Results

Although exact values for the bias and mean square error (MSE) of the 2SLS estimator, and for the bias of the J2SLS estimator, in equation (6.1) are known (and for β_{12} are given in Table 6.3), for compatibility reasons comparison of variance and MSE must necessarily be based upon a Monte Carlo study.

The values in Table 6.3 can serve as a guide to the accuracy of the experiments which follow.

It should be noted that when $\delta = 0.76$, the 2SLS and J2SLS estimators of β_{12} are both unbiased.

From equation (5.1) it can be seen that the 2SLS estimator of β_{12} is unbiased if $\beta_{12} = \rho$ $(= \omega_{12}/\omega_{22})$. In the experiments conducted here, $\beta_{12} = 0.8$ whilst

$$\rho = \frac{1520\delta}{1444} = 0.8 \text{ (if } \delta = 0.76) .$$

It follows from equation (5.5) and the definition of the jackknife that the J2SLS estimator of β_{12} will also be unbiased under the same conditions.

Richardson and Wu [55, pp.977-978] have shown that if the 2SLS estimator of β_{12} is unbiased, then the 2SLS estimator of the coefficients of the exogenous variables must also be unbiased.

If $\beta_{12} = \rho$, then it follows that $\underline{y}_1$ is independent of $\underline{u}_1$, and hence estimation of equation (6.1) becomes a mixed stochastic regression problem. In these circumstances ordinary least squares would be an unbiased estimator and would be the appropriate method of estimation.

Table 6.3: Exact Values of Relative Bias and M.S.E. (β_{12} only)

K_2	λ	δ	μ^2	Relative Bias (2SLS)	M.S.E. (2SLS)	Relative Bias (J2SLS)
3	0.00	0.19	41.2725	-0.01865	0.03552	+0.00631
	0.45	0.19	29.1234	-0.02675	0.05094	+0.00968
	0.00	0.76	41.2725	0.0	0.01163	0.0
	0.45	0.76	29.1234	0.0	0.01668	0.0
6	0.00	0.19	95.7945	-0.03066	0.01483	+0.01093
	0.45	0.19	56.3108	-0.05138	0.02514	+0.01633
	0.90	0.19	8.4440	-0.27237	0.01576	-0.04342
	0.00	0.76	95.7945	0.0	0.00478	0.0
	0.45	0.76	56.3108	0.0	0.00801	0.0
	0.90	0.76	8.4440	0.0	0.04246	0.0
9	0.00	0.19	118.3348	-0.04252	0.01156	+0.01082
	0.45	0.19	61.3857	-0.07889	0.02405	+0.01527
	0.90	0.19	9.1349	-0.35015	0.15709	-0.12009
	0.00	0.76	118.3349	0.0	0.00379	0.0
	0.45	0.76	61.3857	0.0	0.00701	0.0
	0.90	0.76	9.1349	0.0	0.03113	0.0

For an equation containing an arbitrary number (g) of explanatory endogenous variables, Revankar and Hartley [51] have generalized the above result. An F test was derived by Revankar and Hartley for testing the hypothesis of equality of β_{12} and ρ.

The selection of δ to be 0.76 for half of the experiments allowed a comparison of test statistics to be made (see Chapter 7) without the added complications of bias and skewness entering the comparisons.

6.3 Computational Considerations

6.3.1 The Problem of "Outliers"

The satisfactory inversion of all moment matrices for all sets of exogenous variables was commented upon in Chapter 4. At each replication of the experiments however, it was necessary to invert the matrix

$$Z'X(X'X)^{-1}X'Z$$

and to check against singularity (or near-singularity) caused by the build-up of rounding errors. If singularity was found to be present, the relevant sample values were disgarded and an additional replication performed.

For experiments involving $K_2 = 3$ and $\lambda = 0.9$, although no replication was rejected, the 2SLS and J2SLS parameter estimates were grossly in error as compared with their exact values for β_{12}. Rather than design an *ad hoc* procedure to allow rejection of "unrepresentative" sample values, or outliers, in order to achieve "reasonable" parameter estimates, it was decided to reject this particular experiment completely.

It is difficult to justify the rejection of "outliers" since any cut-off point obviously suffers from a great degree of arbitrariness. Indeed, one could very well be rejecting "true" sample values as well as "rounding error" sample values by applying such a procedure.

6.3.2 Antithetic Variates

The technique of antithetic variates was used in an attempt to reduce (to an unknown degree) the sampling error of the Monte Carlo study when estimating the biases of both estimators (see Hammersley and Handscomb [16] for a description of the technique).

Whilst the antithetic method produced estimates of β_{12} which were marginally closer (than direct simulation) to their exact values for the majority of experiments, there was little to choose between the two methods for estimating the MSE of the 2SLS estimator of β_{12}. This latter feature was noticed by Mikhail [33] in a similar experiment, although he managed to achieve a substantial reduction in sampling error when estimating the bias of the 2SLS estimator.

The additional computer time and storage required to calculate parameter estimates using antithetic variates is minimal, as it merely requires a sign change at an advanced stage in the calculations. However, there is a considerable increase in computer time and storage involved in constructing, storing and sorting twice as many test statistics as were generated by direct simulation. Since this study was already facing computer time and storage constraints using direct simulation, the author did not feel that the small decrease in sampling error justified the increased computer time and storage.

6.4 Results of Monte Carlo Study

Tables 6.4, 6.5 and 6.6 (which are situated at the end of this Chapter) summarize the Monte Carlo results on relative bias, variance, MSE and mean absolute error (MAE) for the three parameters of interest; viz β_{12}, γ_{10} and γ_{11}. Values of the standardized normal statistic for the Wilcoxon Matched-Pairs Signed-Ranks test (e.g. see Siegel [63;

pp.47-52]) under the hypothesis of equality of absolute errors of the two estimators are given in the final column.

Each of these three tables is subdivided into two parts, (a) and (b). Results for $\delta = 0.19$ are given in part (a) of each table, whilst part (b) contains the results for the situation where both estimators are unbiased, i.e. $\delta = 0.76$.

We now consider, in turn, four criteria for discriminating between the two estimators.

6.4.1 Bias

The "large" relative bias of 2SLS which was evident in the exact study (Chapter 5) for high levels of multicollinearity was also apparent in the Monte Carlo study when $\delta = 0.19$. For these experiments the jackknife never failed to reduce the bias of the 2SLS estimator, although this reduction was more marked for β_{12} than for the coefficients of the (2) exogenous variables, γ_{10} and γ_{11}.

All estimates of relative bias had the correct sign. From Table 6.3 it can be seen that the exact relative bias of β_{12} for both 2SLS and J2SLS were very close to the simulation results when $K_2 = 6$. For $K_2 = 3$ and $K_2 = 9$, however, the degree of agreement between the simulated and exact results was not as good.

For $\delta = 0.76$ (i.e. both estimators unbiased) the "relative bias" figures obtained from the experiments must be due to sampling and rounding errors. These errors are particularly noticeable when the level of multicollinearity (λ) is high.

We can be reasonably pleased with the degree of agreement between the exact and experimental results on bias. It is interesting to note that in Summer's [66] experiments 1A - 4A and 1B - 4B, with a model which only differed from the one used in this study by the inclusion of fewer exogenous variables, the mean of the 2SLS estimator of β_{12} over 50 replications had an incorrect bias sign on four (of the eight) occasions.

6.4.2 Variance

In general, 2SLS exhibited a smaller variance than J2SLS for all three parameter estimates, and this was particularly noticeable as the degree of multicollinearity increased. Where the jackknife produced a smaller variance, its superiority was never significant. As K_2 increased, the discrepancy between the 2SLS variance and the larger J2SLS variance widened for all parameter estimates.

6.4.3 Mean Square Error

In general, the reduction in bias due to the application of the jackknife was not of sufficient size to offset the smaller variance of 2SLS. In most cases (for both estimators) the square of the bias was small and had little additional effect when added to the variance. Consequently, in common with the variance, 2SLS was generally superior (for all parameters) on a MSE criterion.

It should be noted, however, that this superiority was particularly marked for "small" values of μ^2 (e.g. when $\mu^2 = 8.440$ and $\mu^2 = 9.1349$). For "larger" values of μ^2, the MSEs of the two estimators did not differ greatly. Frequently, the Wilcoxon test picks up this substantial difference between the two estimators for "small" μ^2, but this statistic is based on testing absolute errors.

With only one exception, the MSE of the 2SLS estimator of β_{12} obtained from the experiments underestimated the exact MSE. Despite this, the exact and experimental values were very close for all values of K_2 and λ.

6.4.4 Mean Absolute Error

In general, 2SLS was superior on a MAE criterion, although its superiority was not as marked as for the MSE criterion. Again, "small" values of μ^2 lead to a great discrepancy between the MAEs of the 2SLS and J2SLS estimators.

6.5 Difference of Absolute Errors

At each replication the absolute error of both estimators was calculated. Let $\hat{b}_i$ and $\tilde{b}_i$ be the absolute errors at the ith replication of the 2SLS and J2SLS estimators of β_{12} respectively, then the difference score is defined as

$$d_i = \hat{b}_i - \tilde{b}_i \,. \qquad (i = 1,2,\ldots,R)$$

We wish to test the hypothesis of equality of $\hat{b}_i$ and $\tilde{b}_i$ over all R replications.

The usual parametric technique for handling such a problem is Student's t distribution, but this requires the assumption that the difference scores (the d_i) are normally and independently distributed in the population from which the sample was drawn. Since this assumption has no theoretical justification for the case being considered here, the Wilcoxon Matched-Pairs Signed-Ranks test (e.g. see Siegel [63; pp.75-83]) was used to test the hypothesis of equality of absolute errors. If the assumptions of the parametric t test are in fact met, the asymptotic efficiency near the null hypothesis of the Wilcoxon test compared with the t test is 95.5%.

Under the stated hypothesis, the Wilcoxon test was conducted for all three parameters being estimated, and the resulting Z statistics are given in the last column of Tables IV, V, and VI. Negative values favour 2SLS.

At a 5% level of significance, the hypothesis of equality of absolute errors is rejected only twice over all parameters when $\delta = 0.19$. Both rejections are in favour of the 2SLS estimator, and both occur for $K_2 = 9$ and $\lambda = 0.90$ (i.e. when μ^2 is "small").

When $\delta = 0.76$, however, the hypothesis is rejected on four occasions for β_{12} alone, all four rejections in favour of the 2SLS estimator. Surprisingly, this result did not carry over to the 2SLS estimates of γ_{10} and γ_{11}.

6.6 Conclusion

The results of the Monte Carlo study are not encouraging for proponents of the jackknife technique. Whilst 2SLS was clearly superior when there existed a high degree of multicollinearity, application of the jackknife technique, in general, could not produce superior results using either a MSE or MAE criterion. In view of the increased complexity and computation time involved in applying the jackknife, its use cannot be recommended on the basis of the above results alone.

On the basis of the above results, the following statements can be made:

(i) for a relatively high degree of overidentification (i.e. $K_2 = 6$ or $K_2 = 9$ in these experiments), application of the jackknife technique produces a substantial reduction in the bias of the 2SLS estimator;

(ii) over all experiments 2SLS is superior on a MSE criterion, this superiority being particularly marked when μ^2 is "small";

(iii) when $\delta = 0.19$, there appears to be little significant difference between the two estimators over all parameters, using the absolute error criterion, on the basis of the Wilcoxon Matched-Pairs Signed-Ranks test;

(iv) over all experiments, differences between 2SLS and J2SLS estimates of β_{12} using MAE, MSE, and variance criteria are far less marked than the same differences for γ_{10} and γ_{11};

(v) when the 2SLS estimator is unbiased (i.e. $\delta = 0.76$), application of the jackknife is clearly unwarranted and its application in error is likely to have a detrimental effect on the parameter estimates. Clearly, to avoid this possibility, Revankar and Hartley's [51] test should be used prior to estimation.

Table 6.4(a): Results of Monte Carlo Experiments

Parameter = β_{12}

			RELATIVE BIAS		VARIANCE		M.S.E.		M.A.E.		WILCOXON
K_2	λ	μ^2	2SLS	J2SLS	2SLS	J2SLS	2SLS	J2SLS	2SLS	J2SLS	TEST
3	0.00	41.2725	-0.02739	-0.00151	0.03014	0.03044	0.03062	0.03044	0.14138	0.14030	+0.7324
	0.45	29.1234	-0.04351	-0.00052	0.04022	0.04433	0.04143	0.04433	0.16927	0.17156	+0.1977
6	0.00	95.7945	-0.02911	+0.01092	0.01329	0.01473	0.01382	0.01480	0.09423	0.09717	-0.4487
	0.45	56.3108	-0.05176	+0.01428	0.01963	0.02315	0.02134	0.02327	0.11690	0.12211	-0.6223
	0.90	8.4440	-0.26531	-0.03903	0.09687	0.32332	0.14192	0.32429	0.30989	0.38605	-0.4986
9	0.00	118.3349	-0.03588	+0.02243	0.01123	0.01327	0.01205	0.01359	0.08642	0.09118	-0.4590
	0.45	61.3857	-0.06731	+0.03637	0.01977	0.02815	0.02267	0.02900	0.12234	0.13239	-0.4590
	0.90	9.1349	-0.30998	-0.03315	0.08634	0.37555	0.14784	0.37626	0.32287	0.44477	-2.3088

Sample size = 20 δ = 0.19

Table 6.4(b): Results of Monte Carlo Experiments

Parameter = β_{12}

			RELATIVE BIAS		VARIANCE		M.S.E.		M.A.E.		WILCOXON
K_2	λ	μ^2	2SLS	J2SLS	2SLS	J2SLS	2SLS	J2SLS	2SLS	J2SLS	TEST
3	0.00	41.2725	-0.00629	-0.00215	0.00939	0.00946	0.00942	0.00946	0.08139	0.08130	+0.4109
	0.45	29.1234	-0.00814	-0.00136	0.01326	0.01431	0.01330	0.01431	0.09392	0.09762	-1.2705
6	0.00	95.7945	-0.00524	-0.00328	0.00438	0.00490	0.00439	0.00490	0.05223	0.05582	-1.9444
	0.45	56.3108	-0.00668	-0.00305	0.00655	0.00799	0.00657	0.00800	0.06450	0.07098	-2.5616
	0.90	8.4440	-0.00844	+0.01385	0.03462	0.10562	0.03466	0.10573	0.14633	0.22304	-5.1540
9	0.00	118.3349	-0.00235	+0.00143	0.00371	0.00445	0.00371	0.00445	0.04988	0.05334	-1.5576
	0.45	61.3857	-0.00191	+0.00588	0.00624	0.00880	0.00624	0.00883	0.06383	0.07461	-2.7232
	0.90	9.1349	-0.00188	+0.03911	0.02810	0.09701	0.02811	0.09799	0.12742	0.22593	-6.4812

Sample size = 20 $\delta = 0.76$

Table 6.5(a): Results of Monte Carlo Experiments

Parameter = γ_{10}

			RELATIVE BIAS		VARIANCE		M.S.E.		M.A.E.		WILCOXON
K_2	λ	μ^2	2SLS	J2SLS	2SLS	J2SLS	2SLS	J2SLS	2SLS	J2SLS	TEST
3	0.00	41.2725	+0.05783	+0.03567	494.50	509.77	502.86	512.95	17.80	18.16	-0.8218
	0.45	29.1234	+0.05958	+0.03691	532.11	552.15	540.98	555.55	18.52	18.69	-0.3335
6	0.00	95.7945	+0.07142	+0.01475	521.61	552.85	534.36	553.40	18.30	18.89	-0.8011
	0.45	56.3108	+0.07951	+0.01679	554.61	590.48	570.42	591.19	18.87	19.14	-0.1169
	0.90	8.4440	+0.21776	+0.07183	1712.25	2487.70	1830.80	2500.60	34.24	37.75	-1.7811
9	0.00	118.3349	+0.08609	-0.00401	524.09	560.12	542.62	560.16	18.60	18.63	+0.1994
	0.45	61.3857	+0.10527	-0.00937	568.82	636.35	596.53	636.57	19.52	19.55	+0.2571
	0.90	9.1349	+0.26538	+0.02130	1717.59	3005.34	1893.66	3006.47	35.04	39.50	-1.3169

Sample size = 20 $\delta = 0.19$

Table 6.5(b): Results of Monte Carlo Experiments

Parameter = γ_{10}

			RELATIVE BIAS		VARIANCE		M.S.E.		M.A.E.		WILCOXON
K_2	λ	μ^2	2SLS	J2SLS	2SLS	J2SLS	2SLS	J2SLS	2SLS	J2SLS	TEST
3	0.00	41.2725	+0.02198	+0.01713	159.76	161.79	160.97	162.52	10.10	10.19	-0.2372
	0.45	29.1234	+0.02128	+0.01546	175.59	178.15	176.73	178.75	10.66	10.74	-0.6292
6	0.00	95.7945	+0.02206	+0.01941	174.43	177.16	175.65	178.10	10.58	10.69	-0.0688
	0.45	56.3108	+0.02064	+0.01770	188.26	190.72	189.32	191.50	10.95	10.89	+0.6464
	0.90	8.4440	+0.01336	+0.00231	664.16	789.39	664.61	789.40	20.49	21.88	-0.4573
9	0.00	118.3349	+0.01833	+0.01135	176.50	182.49	177.34	182.81	10.68	10.86	-0.5123
	0.45	61.3857	+0.01673	+0.00717	194.44	205.75	195.14	205.88	11.10	11.35	-0.7427
	0.90	9.1349	+0.01557	-0.01026	642.99	798.66	643.59	798.92	20.24	22.14	-1.2103

Sample size = 20 $\delta = 0.76$

Table 6.6(a): Results of Monte Carlo Experiments

Parameter = γ_{11}

			RELATIVE BIAS		VARIANCE		M.S.E.		M.A.E.		WILCOXON
K_2	λ	μ^2	2SLS	J2SLS	2SLS	J2SLS	2SLS	J2SLS	2SLS	J2SLS	TEST
3	0.00	41.2725	-0.01578	-0.00552	0.12200	0.12560	0.12236	0.12564	0.27814	0.27567	+1.1175
	0.45	29.1234	-0.01522	-0.00789	0.13579	0.14115	0.13612	0.14124	0.29161	0.29209	+0.0499
6	0.00	95.7945	-0.02943	+0.00569	0.13652	0.13946	0.13777	0.13951	0.29307	0.29189	+0.5794
	0.45	56.3108	-0.03245	+0.00222	0.14783	0.15137	0.14934	0.15138	0.30267	0.29797	+1.1346
	0.90	8.4440	-0.11355	-0.03921	0.56781	0.79627	0.58637	0.79849	0.61435	0.67577	-1.7501
9	0.00	118.3349	-0.03078	+0.01239	0.12644	0.13457	0.12780	0.13479	0.27999	0.28116	+0.2201
	0.45	61.3857	-0.03260	+0.01124	0.14321	0.15645	0.14474	0.15664	0.29747	0.30058	-0.1221
	0.90	9.1349	-0.06151	+0.02542	0.55360	0.79839	0.55905	0.79932	0.59897	0.67042	-2.1954

Sample size = 20 δ = 0.19

Table 6.6(b): Results of Monte Carlo Experiments

Parameter = γ_{11}

			RELATIVE BIAS		VARIANCE		M.S.E.		M.A.E.		WILCOXON
K_2	λ	μ^2	2SLS	J2SLS	2SLS	J2SLS	2SLS	J2SLS	2SLS	J2SLS	TEST
3	0.00	41.2725	-0.00167	+0.00144	0.04010	0.04035	0.04010	0.04036	0.15781	0.15610	+0.4968
	0.45	29.1234	-0.00082	+0.00282	0.04498	0.04577	0.04498	0.04578	0.16653	0.16748	-0.6533
6	0.00	95.7945	-0.00307	-0.00161	0.04666	0.04552	0.04667	0.04552	0.16974	0.16665	+1.0762
	0.45	56.3108	-0.00165	+0.00019	0.05082	0.04981	0.05083	0.04981	0.17583	0.17240	+1.2808
	0.90	8.4440	+0.00305	+0.00767	0.21889	0.26339	0.21890	0.26348	0.36742	0.40131	-1.5163
9	0.00	118.3349	+0.00022	+0.00334	0.04273	0.04391	0.04273	0.04393	0.16162	0.16309	-0.1083
	0.45	61.3857	+0.00155	+0.00551	0.04836	0.05005	0.04836	0.05010	0.17096	0.17022	+0.6017
	0.90	9.1349	+0.00250	+0.00808	0.19668	0.22882	0.19670	0.22892	0.34628	0.36884	-0.7375

Sample size = 20 $\delta = 0.76$

CHAPTER 7

INFERENCE

7.1 Tests of Significance

7.1.1 Conventional Tests of Significance

So far we have only considered point estimation of the parameters in a simultaneous equation system. In applied economics however it is usual to test for significance of the parameter estimates, or (identically) to formulate interval estimates.

From equation (2.6), the 2SLS estimator of $\underline{\theta}$ is written as

$$\underline{\hat{\theta}} = \left[Z'X(X'X)^{-1}X'Z \right]^{-1} Z'X(X'X)^{-1} X'\underline{y} \, ,$$

and from equation (2.7) the limiting distribution of the sequence $\sqrt{N}\,(\underline{\hat{\theta}} - \underline{\theta})$ is given by

$$\sqrt{N}\,(\underline{\hat{\theta}} - \underline{\theta}) \sim N \left[0, \; \sigma^2 \operatorname*{plim}_{N \to \infty} \left[\frac{1}{N} \cdot Z'X \left(\frac{1}{N} \cdot X'X \right)^{-1} \frac{1}{N} \cdot X'Z \right]^{-1} \right], \tag{7.1}$$

provided $\lim_{N \to \infty} \left(\frac{1}{N} \cdot X'X \right)^{-1}$ exists .

The correct asymptotic test of significance therefore is the standardized normal test statistic, and a consistent estimator of σ^2 is given by

$$\tilde{\sigma}^2 = \underline{\hat{u}}'\underline{\hat{u}} \Big/ N \, , \tag{7.2}$$

where $\underline{\hat{u}} = \underline{y} - Y\underline{\hat{\beta}} - X_1 \underline{\hat{\gamma}}_1$.

It has become common practice however to adjust the estimator of σ^2 for loss of degrees of freedom and use the t statistic, rather than the standardized normal, when dealing with finite samples (e.g. see

Johnston [20; p.384]. Thus in finite samples a consistent estimator of σ^2 is given by

$$\hat{\sigma}^2 = \underline{\hat{u}}'\underline{\hat{u}} \Big/ (N - K_1 - g) .$$

From equations (7.1) and (7.2) it follows that, asymptotically,

$$\frac{\sqrt{N}\,(\hat{\theta}_k - \theta_k)}{\tilde{\sigma}\sqrt{\overline{S}_k}} \sim N(0,1) , \qquad (7.3)$$

where $\hat{\theta}_k$ and θ_k are the kth components of $\underline{\hat{\theta}}$ and $\underline{\theta}$ respectively $(k = 1,2,\ldots,K_1 + g)$, and $\overline{S}_k$ is the kkth component of

$$\left[\frac{1}{N} . Z'X \left(\frac{1}{N} . X'X\right) \frac{1}{N} . X'Z\right]^{-1} .$$

Let S_k denote the kkth component of $[Z'X(X'X)^{-1} X'Z]^{-1}$, then $\overline{S}_k = NS_k$ and expression (7.3) can be rewritten as

$$\frac{(\hat{\theta}_k - \theta_k)}{\tilde{\sigma}\sqrt{S_k}} \sim N(0,1) . \qquad (7.4)$$

The conventional finite sample counterpart of expression (7.4) is the statistic

$$\frac{(\hat{\theta}_k - \theta_k)}{\hat{\sigma}\sqrt{S_k}} , \qquad (7.5)$$

which is tested as though it is distributed as Student t with $N - K_1 - g$ degrees of freedom.

7.1.2 Dhrymes' Alternative Test of Significance

An alternative asymptotic test of significance based on Student's t distribution has been proposed by Dhrymes [12]. Use of the t statistic is customary for testing the significance of 2SLS parameter estimates yet, until Dhrymes showed the asymptotic validity of his test, no theory existed to justify the practice. On the basis of the asymptotic distribution of the 2SLS estimator, the relevant test of significance should have been based on the standardized normal distribution as described by expression (7.4).

Rewrite the equation being estimated as

$$\underline{y} = Z\underline{\theta} + \underline{u} \tag{7.6}$$

where $Z = [Y : X_1]$ and $\underline{\theta}' = [\underline{\beta}' : \underline{\gamma}_1']$, then define a square, non-singular matrix R of order K such that $RR' = X'X$. Further define $P = R^{-1}X'$, then premultiplying equation (7.6) by P gives

$$\underline{w} = Q\underline{\theta} + \underline{e} ,$$

where $\underline{w} = P\underline{y}$, $Q = PZ$ and $\underline{e} = P\underline{u}$. Dhrymes showed that the 2SLS estimator of $\underline{\theta}$ in equation (7.6) is the OLS estimator of $\underline{\theta}$ in this transformed system. Further, by analogy with least squares, Dhrymes showed that, asymptotically,

$$\frac{(\hat{\theta}_k - \theta_k)}{\bar{\sigma}\sqrt{S_k}} \sim t_{K_2 - g} , \tag{7.7}$$

where an asymptotically unbiased, but inconsistent, estimator of σ^2 is given by

$$\bar{\sigma}^2 = \underline{\hat{e}}'\underline{\hat{e}}/(K_2 - g) = \underline{\hat{u}}'X(X'X)^{-1}X'\underline{\hat{u}}/(K_2 - g).$$

Thus the test is only valid if the structural equation in question is over-identified.

Revankar [50], however, has shown that information is lost when a dimension reducing transformation is used as a basis for testing, thus Dhrymes' test could be expected to be inefficient compared to the conventional test based on the standardized normal distribution.

In a Monte Carlo study, Maddala [26] observed that the Dhrymes test had low power compared with the conventional tests in a two equation model. Richardson and Rohr [54] came to the same conclusion on the basis of a Monte Carlo study using a three equation model.

7.2 The Exact Distribution of a t Statistic

The exact finite sample distribution functions of several t statistics for hypothesis testing and the construction of confidence intervals on 2SLS parameter estimates have been studied by Richardson and Rohr [53] and Rohr [57]. As with many other finite sample studies into the properties of the 2SLS estimator, the results were derived for a model with just two jointly-dependent variables.

Richardson and Rohr [53] considered the finite sample distribution of Dhrymes' t statistic, expression (7.7), which Dhrymes had already shown to be asymptotically distributed as Student t with $K_2 - g$ degrees of freedom. However, since the sample size does not appear explicitly in their finite sample derivations, convergence of the t statistic to Student's t distribution was analysed for μ^2 (the concentration parameter) increasing indefinitely.

The moments of the exact distribution were found not to exist to order $K_2 - g$ and higher, but where they did exist they converged to the moments of Student's t distribution with $K_2 - g$ degrees of freedom as $\mu^2 \to \infty$. On the basis of their results Richardson and Rohr conjectured that, for large μ^2, the exact distribution function of the t statistic can be adequately approximated by Student's t distribution with $K_2 - g$ degrees of freedom.

Richardson and Rohr investigated their conjecture for one degree of freedom and for several values of μ and β. On the basis of their computations they concluded that the actual probability of Type I error (for a significance level of 5%) will be less than 5% if β is positive, and greater than 5% if β is negative. If $\mu > 3$ the exact t statistic was found to be a good approximation to the Student t, but for small β and $\mu \lesssim 3$ differences between the two could lead to serious errors.

Richardson and Rohr also tabulated the exact value of the second moment and the exact absolute values for the first and third moments of the t statistic for various values of degrees of freedom, β, and μ^2, from which they concluded that the density function is highly skewed and that often the moments differ considerably from those of Student's t distribution with $K_2 - g$ degrees of freedom.

Rohr [57] has derived the exact distribution of two "more conventional" test statistics, only one of which is used in this study, viz:

$$\frac{(\hat{\theta}_k - \theta_k)}{\hat{\sigma}\sqrt{S_k}} ,$$

which is identical to expression (7.5).

Rohr showed that asymptotically (in μ^2) expression (7.5) converges to Student's t distribution with $N - K_1 - g$ degrees of freedom, but that in finite samples the moments of the statistic (7.5) exist only up to order $N - K_1 - g + 1$.

It should be noted, however, that mathematical complexity in the derivation of the moments of expression (7.5) forced Rohr to consider only the special case where $\beta = \sigma_{12}/\sigma_{22}$; i.e. 2SLS unbiased. Under this restriction, expression (7.5) has all odd moments (where they exist) equal to zero, and 2SLS and OLS are equivalent.

Rohr also showed that the variance of expression (7.5) is always less than or equal to the variance of its limiting distribution.

7.3 Student's t Distribution and its use with the Two-Stage Least Squares Estimator

The ratio

$$\frac{w}{\sqrt{v/r}} = \frac{\hat{\theta}_k - E(\hat{\theta}_k)}{S.E.\ (\hat{\theta}_k)} \qquad (k = 1,2,\ldots,K_1 + g) \qquad (7.8)$$

is distributed as Student t if w is normally distributed with zero mean and unit variance and if v has a χ^2 distribution with r degrees of freedom, provided that v and w are stochastically independent.

For 2SLS, in general, $E(\hat{\theta}_k) \neq \theta_k$ and $\hat{\sigma}^2$ is a consistent, but not unbiased, estimator of σ^2. Consequently the denominator of expression (7.8) only approximates a χ^2 distribution. In addition, $\hat{\theta}_k - E(\hat{\theta}_k)$ is not stochastically independent of its standard error (S.E.) in finite samples. It should be noted that $E(\hat{\theta}_k)$ may not even be finite, although in the ensuing Monte Carlo analysis the experiments were designed in such a way as to ensure that the first two moments of the 2SLS estimator were always finite.

7.4 An Approximate t Statistic constructed using the Jackknife Technique

Tukey [72] has suggested that the N pseudo-jackknife estimates could be treated as approximately independent, identically distributed observations from which an approximate t statistic could be constructed as

$$\frac{\sqrt{N}\ [J(\hat{\theta}_k) - \theta_k]}{\left\{(N-1)^{-1}\sum_{i=1}^{N}\left[J_i(\hat{\theta}_k) - J(\hat{\theta}_k)\right]^2\right\}^{\frac{1}{2}}} \qquad (7.9)$$

We have already shown (in Chapter 3) that expression (7.9) is asymptotically distributed as the standardized normal distribution in the context of the J2SLS estimator.

Although in general $E[J(\hat{\theta}_k)] \neq \theta_k$, in many instances it will exhibit a smaller deviation from θ_k than 2SLS, as was observed in Chapter 4. In common with 2SLS, the numerator and the denominator of expression (7.9) will not be stochastically independent in finite samples.

Miller [34] gives several counterexamples to Tukey's conjecture, but Arvesen [1] gives a wide class of situations where this suggestion is valid, i.e. when $J_i(\hat{\alpha}_k)$ and $J(\hat{\alpha}_k)$ are U statistics (see Hoeffding [19]) or functions of U statistics.

Recently, Miller [36] provided an asymptotic justification of Tukey's conjecture for a function of the regression parameters in a general linear model.

7.5 Independence of the Pseudo-Jackknife Estimates

Walsh [73] has demonstrated the deleterious effects of using correlated samples for the construction of certain significance tests. If the N pseudo-jackknife estimates could be considered as a single observation of a normal multivariate population, for which the N variables have common mean μ and variance σ^2, the effect on the t statistic of a common level of pairwise correlation between the pseudo-jackknife estimates would be to raise or lower the true confidence coefficient depending on whether the correlation was positive or negative. Thus if the pairwise correlation (r) was positive, a test result which would be significant for a random sample need no longer be so. To correct the t statistic the multiplying factor

$$\sqrt{\frac{(1-r)}{1+(N-1)r}}$$

is required.

Walsh illustrated the error incurred in assuming $r = 0$ by tabulating the true value of the confidence coefficient for varying values of N and r. Even for small r the deleterious effect of correlation was very marked; e.g. for $N = 8$ and $r = 0.1$ the true value of the 95% confidence coefficient is 86.5%, and for $N = 32$ and $r = 0.1$ the true value falls to 68%. Thus the dangers of ignoring the possibility of $|r|>0$ are evident.

Miller [34], using different initial assumptions, has also shown the deleterious effect on the t statistic of correlation among the pseudo-jackknife estimates.

Three statistics were selected, therefore, to test for the "approximate" independence of the pseudo-jackknife 2SLS estimates, and for this purpose the pseudo-jackknife estimates were expressed as deviations from their mean, viz:

$$d_{ik} = J_i(\hat{\theta}_k) - J(\hat{\theta}_k), \ (i = 1,2, \ldots, N)$$

for all k $(k = 1,2, \ldots, K_1 + g)$. The three tests used for this purpose are well known tests for departures from randomness, and a detailed explanation of all three (the Swed-Eisenhart One Sample Runs Test, the Fisher Exact Probability Test, and Spearman's Rank Correlation Coefficient) is given in Siegel [63].

The Swed-Eisenhart test (denoted by SE in Table 7.1) was used to ascertain whether the sequence of signs of the d_{ik} was random. The Fisher test (denoted by FI) was also based on sign sequences. A 2 x 2 contingency table was set-up for each value of k and scores allotted according to the sequence of the signs of successive d_{ik} over the i observations. Spearman's Rank Correlation Coefficient (denoted by SR) was used to test for association between the natural ordering of the d_{ik} and their ranked ordering. All three tests were repeated over all replications.

The problem with using these aforementioned tests is that no general statement can be made about the efficiency of any of them. In the context in which they are used in this study, each of these three tests will produce a different "measure" of randomness. All three reject a certain amount of relevant information and therefore, at best, the test results can only be used as an approximate guide to departures from randomness of the pseudo-jackknife 2SLS estimates.

The number of times the hypothesis of randomness was rejected for each test over the 100 replications is given in Table 7.1. A visual appraisal of the results indicates that the hypothesis of randomness is upheld "approximately" 95% of the time. These results appear to offer some support to Tukey's conjecture for this particular application.

7.6 Validity of Test Statistics

It is essential to examine the validity of the standard tests of significance to ensure that the test statistics do not diverge significantly from their postulated theoretical distribution. To this end, the Kolmogorov-Smirnov One-Sample Test (see e.g. Siegel [63; pp.47-52]) was employed to test five hypotheses:

$$\frac{\hat{\theta}_k - \theta_k}{\hat{\sigma}\sqrt{S_k}} \sim N(0,1) \quad , \qquad (7.10a)$$

$$\frac{\hat{\theta}_k - \theta_k}{\hat{\sigma}\sqrt{S_k}} \sim t_{N-K_1-g} \quad , \qquad (7.10b)$$

$$\frac{\hat{\theta}_k - \theta_k}{\bar{\sigma}\sqrt{S_k}} \sim t_{K_2-g} \quad , \qquad (7.10c)$$

$$\frac{\sqrt{N}\left[J(\hat{\theta}_k)-\theta_k\right]}{\left[(N-1)^{-1}\sum_{i=1}^{N}\left(J_i(\hat{\theta}_k) - J(\hat{\theta}_k)\right)^2\right]^{\frac{1}{2}}} \sim N(0,1) \ , \qquad (7.10d)$$

and

$$\frac{\sqrt{N}\left[J(\hat{\theta}_k)-\theta_k\right]}{\left[(N-1)^{-1}\sum_{i=1}^{N}\left(J_i(\hat{\theta}_k) - J(\hat{\theta}_k)\right)^2\right]^{\frac{1}{2}}} \sim t_{N-1} \qquad (7.10e)$$

$(k = 1,2,\ldots,K_1 + g)$.

Tables 7.2(a) and 7.2(b) set out the values of the maximum deviation, D, between the relevant empirical and theoretical distributions for each of these five hypotheses. The distributional assumptions are rejected at the 5% level for $D > 0.13403$.

Over all experiments 48 "sets" of values for D were obtained, i.e. 24 sets for each value of δ. The lowest D value in each set was designated "1st", the second lowest "2nd", and so on. Table 7.3 summarizes the number of firsts, seconds, etc., for each test statistic over all parameters and all values of K_2, for $\delta = 0.19$ and for $\delta = 0.76$. The following abbreviations are used:

CT1 - "Conventional Test No. 1", formula (7.10a);

CT2 - "Conventional Test No. 2", formula (7.10b);

DT - "Dhrymes Test", formula (7.10c);

JT1 - "Jackknife Test No. 1", formula (7.10d);

JT2 - "Jackknife Test No. 2", formula (7.10e).

Care must be taken in interpreting these figures, as the postulated theoretical distribution differs across each set.

When 2SLS was biased (i.e. $\delta = 0.19$) the jackknife-based test statistics always dominated the others for β_{12}, and γ_{10}, and for six out of the eight sets of values for γ_{11}. The t statistic based upon the Dhrymes derivation (DT) consistently produced the poorest fit.

Table 7.1: Tests of Independence of Pseudo-Jackknife Estimates
(Number of rejections at 5% level of significance)

β_{12}, $\delta = 0.19$

K_2	λ	FI	SR	SE
3	0.00	4	1	2
	0.45	2	3	2
6	0.00	7	5	5
	0.45	6	3	4
	0.90	1	3	0
9	0.00	5	6	4
	0.45	5	4	2
	0.90	4	5	1

β_{12}, $\delta = 0.76$

K_2	λ	FI	SR	SE
3	0.00	9	6	7
	0.45	5	3	5
6	0.00	4	7	4
	0.45	3	4	3
	0.90	3	7	3
9	0.00	4	10	4
	0.45	4	4	3
	0.90	7	8	2

γ_{10}, $\delta = 0.19$

K_2	λ	FI	SR	SE
3	0.00	3	8	2
	0.45	2	6	1
6	0.00	6	6	7
	0.45	6	5	5
	0.90	6	7	6
9	0.00	5	7	4
	0.45	4	9	2
	0.90	2	6	2

γ_{10}, $\delta = 0.76$

K_2	λ	FI	SR	SE
3	0.00	2	7	3
	0.45	2	5	1
6	0.00	1	7	2
	0.45	1	7	4
	0.90	3	7	2
9	0.00	3	9	1
	0.45	5	8	3
	0.90	2	5	2

γ_{11}, $\delta = 0.19$

K_2	λ	FI	SR	SE
3	0.00	6	4	3
	0.45	4	4	0
6	0.00	7	3	4
	0.45	9	4	2
	0.90	5	5	3
9	0.00	2	2	2
	0.45	1	4	1
	0.90	5	4	4

γ_{11}, $\delta = 0.76$

K_2	λ	FI	SR	SE
3	0.00	1	4	0
	0.45	5	3	3
6	0.00	4	3	2
	0.45	4	5	3
	0.90	3	4	2
9	0.00	6	2	4
	0.45	3	5	4
	0.90	2	3	2

Table 7.2(a): Kolmogorov-Smirnov D Statistic

			2SLS		Dhrymes	J2SLS	
	K_2	λ	Normal	t	t	Normal	t
β_{12}	3	0.00	0.1336	0.1329	0.1342	0.0929	0.0941
		0.45	0.1478	0.1471	0.1406	0.0878	0.0871
	6	0.00	0.1454	0.1429	0.1290	0.0912	0.0944
		0.45	0.1888	0.1853	0.1743	0.1014	0.1055
		0.90	0.3608	0.3578	0.3410	0.1381	0.1385
	9	0.00	0.1406	0.1384	0.1546	0.0900	0.0925
		0.45	0.1996	0.1987	0.2039	0.1009	0.1032
		0.90	0.3965	0.3896	0.4000	0.1589	0.1530
γ_{10}	3	0.00	0.1062	0.1068	0.1166	0.0907	0.0918
		0.45	0.1115	0.1083	0.1047	0.0952	0.0965
	6	0.00	0.1085	0.1051	0.1206	0.0439	0.0455
		0.45	0.1222	0.1229	0.1199	0.0601	0.0603
		0.90	0.1560	0.1561	0.1562	0.1052	0.1065
	9	0.00	0.1224	0.1190	0.1300	0.0532	0.0551
		0.45	0.1340	0.1343	0.1345	0.0540	0.0543
		0.90	0.1968	0.1947	0.1892	0.1040	0.1039
γ_{11}	3	0.00	0.0996	0.0987	0.1066	0.0680	0.0682
		0.45	0.0778	0.0771	0.0969	0.0783	0.0790
	6	0.00	0.0893	0.0881	0.0901	0.0851	0.0893
		0.45	0.1225	0.1235	0.1121	0.0832	0.0851
		0.90	0.1087	0.1082	0.1183	0.1200	0.1200
	9	0.00	0.0957	0.0948	0.1156	0.0790	0.0840
		0.45	0.1060	0.1061	0.1086	0.0843	0.0881
		0.90	0.0761	0.0764	0.0768	0.0780	0.0779

Sample size = 20 $\delta = 0.19$

Table 7.2(b): Kolmogorov-Smirnov D Statistic

			2SLS		Dhrymes	J2SLS	
	K_2	λ	Normal	t	t	Normal	t
β_{12}	3	0.00	0.1206	0.1184	0.0792	0.0790	0.0863
		0.45	0.0710	0.0696	0.0623	0.0967	0.1023
	6	0.00	0.0424	0.0391	0.0573	0.0477	0.0525
		0.45	0.0576	0.0617	0.0474	0.0489	0.0532
		0.90	0.0761	0.0818	0.0754	0.0594	0.0614
	9	0.00	0.0843	0.0820	0.0787	0.0650	0.0620
		0.45	0.0654	0.0629	0.0551	0.0778	0.0724
		0.90	0.0558	0.0607	0.0530	0.1252	0.1204
γ_{10}	3	0.00	0.0683	0.0707	0.0971	0.0501	0.0513
		0.45	0.0798	0.0792	0.0747	0.0594	0.0614
	6	0.00	0.0764	0.0748	0.0809	0.0497	0.0513
		0.45	0.0683	0.0702	0.0709	0.0829	0.0841
		0.90	0.0602	0.0608	0.0583	0.0494	0.0549
	9	0.00	0.0693	0.0681	0.0747	0.0571	0.0559
		0.45	0.0613	0.0628	0.0611	0.0383	0.0400
		0.90	0.0500	0.0515	0.0500	0.0511	0.0570
γ_{11}	3	0.00	0.0536	0.0596	0.0633	0.0777	0.0811
		0.45	0.0799	0.0796	0.0706	0.0876	0.0913
	6	0.00	0.0523	0.0598	0.0545	0.0905	0.0957
		0.45	0.0629	0.0642	0.0584	0.0843	0.0877
		0.90	0.0646	0.0681	0.0527	0.0751	0.0793
	9	0.00	0.0546	0.0620	0.0508	0.0827	0.0871
		0.45	0.0564	0.0592	0.0493	0.0746	0.0786
		0.90	0.0536	0.0588	0.0505	0.0705	0.0757

Sample size = 20 $\delta = 0.76$

Table 7.3: Ranking of D Statistic over the Five Tests of Significance

$\delta = 0.19$

RANK	1st*	2nd*	3rd*	4th*	5th
CT1	1	1	5	11	6
CT2	1	2	9	10	2
DT	0	0	9	0	15
JT1	19	3	1	0	1
JT2	4	17	1	2	0

$\delta = 0.76$

RANK	1st*	2nd*	3rd	4th	5th
CT1	4	7	4	6	3
CT2	1	5	9	5	4
DT	11	2	7	0	4
JT1	7	3	2	10	2
JT2	2	6	2	3	11

Sample size = 20

* Denotes that column total does not sum to 24 because of ties (to 4 decimal places).

For 2SLS unbiased, however, the superiority of the jackknife-based test statistics was less marked. This was particularly noticeable for γ_{11} where the two jackknife-based test statistics always produced the poorest fit.

The number of rejections, at the 5% level of significance, of the hypothesis that each sample was drawn from the specified theoretical distribution is given in Table 7.4 In any one cell the total possible number of rejections is 24; percentages of rejections are given next to the absolute figures.

Table 7.4: Number of Rejections of the Null Hypothesis

δ =	0.19		0.76	
CT1	9	37.5%	0	-
CT2	10	41.7%	0	-
DT	10	41.7%	0	-
JT1	2	8.3%	0	-
JT2	2	8.3%	0	-

Sample size = 20

Clearly when $\delta = 0.19$ the distribution of the t statistic formed using the 2SLS estimator gives a poor approximation to both Student's t distribution and the standardized normal distribution. Thus if the bias of the 2SLS estimator is significantly different from zero, the distribution of 2SLS-based test statistics may be a poor approximation to their postulated theoretical distributions.

7.7 Inference

7.7.1 Tests of Significance

In the preceding section it was shown that the distributions of expressions (7.10a), (7.10b) and (7.10c) show a substantial divergence from their postulated theoretical distributions when $\delta = 0.19$, even for relatively large values of μ^2. It is important to ascertain the effect of this divergence on statistical inference.

In this section we consider the degree of accuracy afforded by using the relevant theoretical distributions as approximations for making statistical inference.

The hypotheses that the biases of the 2SLS and J2SLS estimators were not significantly different from zero were tested at both the 5% and 10% levels of significance. The proportion of samples falling in the .05 and .95 percentiles of the relevant theoretical distributions are given in Tables 7.5, 7.6 and 7.7. These tables are further divided into parts (a) and (b), the former for results when $\delta = 0.19$, the latter for $\delta = 0.76$.

In these tables each cell contains three values. The number of "rejections" are tabulated according to whether they were rejected in the lower or upper tail of the relevant distribution, and are given by the figures in parentheses on the left and right respectively at the top of each cell. The total number of "rejections" is given below these two figures.

For the parameter β_{12}, both JT1 and JT2 show a number of "rejections" nearer the nominal level of significance than CT1 and CT2 in just over half of the experiments for $\delta = 0.19$. There is little to choose between these two jackknife-based test statistics, although JT2 (i.e. the t statistic given by formula (7.10e)) was marginally closer to the nominal level of significance for $K_2 = 6$ and 9 and $\lambda = 0.45$ and 0.90. CT2 is to be preferred to CT1 as the number of "rejections" were, in general, nearer the nominal level of significance. Using the same criterion, CT2 is to be preferred to JT1 but not to JT2.

Dhrymes' test statistic (DT) gave a similar pattern of "rejections" to the other t statistics, but it should be noted that approximate confidence intervals using DT will be much wider than those using either CT2 or JT2.

The striking feature about these results however is the distribution of "rejections" between the tails of the relevant distributions. The downward bias of the 2SLS estimator of β_{12} ensured that virtually all rejections for CT1, CT2 and DT fell in the lower tail, this being most noticeable when K_2 was relatively large.

The constant term, γ_{10}, gave a fairly even spread of "rejections" between the tails for all tests, whereas γ_{11} showed a similar, but less marked, pattern to that for β_{12}.

For all three parameters, the three t statistics (CT2, DT and JT2) are to be preferred to those tests based on the normal distribution, although this preference is most marked for β_{12}.

The skewness of the foregoing statistics, which is particularly noticeable for the 2SLS-based statistics, can have important consequences when the postulated distributions are used as a basis for constructing approximate critical regions for one-sided tests of hypotheses. From Tables 7.5(a), 7.6(a) and 7.7(a), it can be seen that if the lower tail of the CT1, CT2 and DT distributions is used to construct an approximate test for β_{12}, the estimate of the level of significance is generally considerably higher than the postulated level of either 2.5% or 5%, i.e. the level of significance is underestimated. Conversely, if the upper tail is used then the level of significance will be overestimated. Moreover, in general, the degree of error is larger the higher the level of multicollinearity and the greater the degree of overidentification.

By comparison, test statistics for β_{12} based on the jackknife statistics JT1 and JT2 give a more even spread of rejections and consequently a smaller error of over- or under-estimation when performing one-sided tests of hypothesis. Even here, however, for large λ and $K_2 = 6$ or $K_2 = 9$ the lower tail was markedly larger than its nominal level, although generally very much less than for CT1, CT2 or DT.

For $\delta = 0.76$ all five tests generally only differ to a small degree over the three parameters, although for β_{12}, JT1, JT2 and DT tend to overestimate the total nominal level of significance in both tails by approximately the same margin as CT1 and CT2 tend to underestimate it. For $K_2 = 9$ and $\lambda = 0.9$ the jackknife-based tests produced a "wayward" result.

When $\delta = 0.76$ the 2SLS estimator is not only unbiased, but the odd order moments (those that exist) of both CT2 and DT are zero (see Section 7.2) in the model under consideration. Thus selecting $\delta = 0.76$ has not only removed the location problem but also the problem of skewness of the distribution of CT2 and DT, provided that the first three moments exist (which they do for $K_2 = 6$ and $K_2 = 9$). It is hardly surprising, therefore, that the jackknife-based test statistics cannot show superiority over CT1, CT2 and DT under such circumstances.

From the preceding results it can be concluded that the t statistic based on the J2SLS estimator (JT2) will, in general, produce confidence intervals which are at least as accurate as those produced using test statistics based on the 2SLS estimator.

7.7.2 Power of the Tests

Finally we consider the power of the alternative tests under the hypothesis that $\beta_{12} = \beta_{12}^*$, where β_{12}^* was specified to be 0.8.

Tables 7.8 (a-c) present power functions for the five tests when $\delta = 0.19$. The computational expense involved in computing power functions prohibited further calculations. The significance level for all tests was 5%.

Table 7.5(a): Distribution of Rejections of Hypothesis that the Bias of the Relevant Estimator is not Significantly Different from Zero.

PARAMETER: β_{12}

K_2	λ	CT1	CT2	DT	JT1	JT2
3	0.00	(3) (1) 4	(2) (1) 3	(3) (2) 5	(2) (2) 4	(2) (2) 4
	0.45	(4) (0) 4	(3) (0) 3	(1) (3) 4	(2) (2) 4	(2) (1) 3
6	0.00	(6) (2) 8	(6) (2) 8	(5) (3) 8	(5) (2) 7	(5) (2) 7
	0.45	(6) (2) 8	(4) (1) 5	(5) (3) 8	(6) (2) 8	(3) (2) 5
	0.90	(15) (0) 15	(12) (0) 12	(7) (3) 10	(8) (5) 13	(6) (3) 9
9	0.00	(10) (2) 12	(7) (1) 8	(6) (2) 8	(5) (4) 9	(5) (3) 8
	0.45	(7) (2) 9	(7) (1) 8	(6) (2) 8	(4) (3) 7	(4) (3) 7
	0.90	(19) (0) 19	(17) (0) 17	(16) (0) 16	(13) (7) 20	(11) (7) 18

α = 5% δ = 0.19

PARAMETER: β_{12}

K_2	λ	CT1	CT2	DT	JT1	JT2
3	0.00	(8) (4) 12	(5) (4) 9	(5) (3) 8	(2) (2) 4	(2) (2) 4
	0.45	(8) (3) 11	(7) (1) 8	(4) (5) 9	(4) (3) 7	(2) (2) 4
6	0.00	(8) (3) 11	(6) (3) 9	(7) (4) 11	(6) (2) 8	(6) (2) 8
	0.45	(11) (4) 15	(9) (3) 12	(5) (4) 9	(6) (5) 11	(6) (4) 10
	0.90	(21) (1) 22	(2) (0) 20	(14) (3) 17	(11) (7) 18	(11) (6) 17
9	0.00	(10) (2) 12	(10) (1) 11	(10) (2) 12	(6) (6) 12	(6) (5) 11
	0.45	(10) (2) 12	(9) (2) 11	(10) (3) 13	(6) (8) 14	(6) (6) 12
	0.90	(26) (0) 26	(23) (0) 23	(23) (0) 23	(17) (10) 27	(16) (8) 24

α = 10% δ = 0.19

Sample size = 20

Table 7.5(b) : Distribution of Rejections of Hypothesis that the Bias of the Relevant Estimator is not Significantly Different from Zero.

PARAMETER: β_{12}

K_2	λ	CT1	CT2	DT	JT1	JT2
3	0.00	(1) (3) 4	(1) (1) 2	(3) (3) 6	(1) (2) 3	(1) (2) 3
	0.45	(2) (1) 3	(1) (0) 1	(2) (3) 5	(2) (2) 4	(1) (1) 2
6	0.00	(4) (2) 6	(2) (2) 4	(4) (4) 8	(4) (1) 5	(3) (1) 4
	0.45	(4) (2) 6	(2) (2) 4	(3) (2) 5	(4) (3) 7	(3) (1) 4
	0.90	(3) (1) 4	(2) (0) 2	(2) (4) 6	(4) (4) 8	(4) (4) 8
9	0.00	(5) (3) 8	(4) (3) 7	(2) (2) 4	(3) (6) 9	(3) (4) 7
	0.45	(4) (4) 8	(2) (2) 4	(1) (3) 4	(3) (5) 8	(3) (5) 8
	0.90	(3) (3) 6	(3) (2) 5	(4) (4) 8	(5) (8) 13	(3) (6) 9

α = 5% $\delta = 0.76$

PARAMETER: β_{12}

K_2	λ	CT1	CT2	DT	JT1	JT2
3	0.00	(3) (6) 9	(3) (5) 8	(7) (6) 13	(3) (4) 7	(1) (2) 3
	0.45	(4) (3) 7	(4) (2) 6	(4) (6) 10	(3) (4) 7	(2) (3) 5
6	0.00	(7) (4) 11	(6) (3) 9	(6) (5) 11	(5) (5) 10	(5) (4) 9
	0.45	(6) (4) 10	(5) (3) 8	(5) (5) 10	(6) (6) 12	(5) (5) 10
	0.90	(5) (2) 7	(5) (2) 7	(4) (4) 8	(5) (7) 12	(5) (7) 12
9	0.00	(5) (4) 9	(5) (4) 9	(7) (4) 11	(6) (7) 13	(6) (7) 13
	0.45	(5) (4) 9	(5) (4) 9	(6) (6) 12	(6) (5) 11	(5) (5) 10
	0.90	(4) (5) 9	(4) (5) 9	(5) (6) 11	(9) (12) 21	(7) (9) 16

α = 10% $\delta = 0.76$

Sample size = 20

Table 7.6(a) : Distribution of Rejections of Hypothesis that the Bias of the Relevant Estimator is not Significantly Different from Zero.

PARAMETER: γ_{10}

K_2	λ	CT1	CT2	DT	JT1	JT2
3	0.00	(3) (4) 7	(2) (4) 6	(4) (3) 7	(1) (4) 5	(1) (4) 5
	0.45	(3) (4) 7	(2) (4) 6	(3) (3) 6	(1) (4) 5	(1) (4) 5
6	0.00	(2) (5) 7	(1) (4) 5	(2) (4) 6	(1) (6) 7	(1) (5) 6
	0.45	(3) (4) 7	(1) (4) 5	(2) (3) 5	(1) (7) 8	(1) (4) 5
	0.90	(1) (5) 6	(1) (5) 6	(1) (3) 4	(2) (6) 8	(2) (4) 6
9	0.00	(3) (5) 8	(3) (4) 7	(1) (4) 5	(2) (4) 6	(1) (4) 5
	0.45	(3) (4) 7	(2) (4) 6	(1) (5) 6	(2) (4) 6	(0) (4) 4
	0.90	(1) (5) 6	(1) (5) 6	(1) (3) 4	(6) (6) 12	(5) (4) 9

$\alpha = 5\%$ $\delta = 0.19$

PARAMETER: γ_{10}

K_2	λ	CT1	CT2	DT	JT1	JT2
3	0.00	(5) (8) 13	(4) (4) 8	(4) (5) 9	(2) (7) 9	(2) (6) 8
	0.45	(5) (7) 12	(4) (7) 11	(3) (7) 10	(2) (8) 10	(2) (8) 10
6	0.00	(5) (8) 13	(4) (6) 10	(2) (8) 10	(3) (7) 10	(3) (7) 10
	0.45	(5) (9) 14	(4) (7) 11	(4) (8) 12	(3) (7) 10	(3) (7) 10
	0.90	(4) (10) 14	(3) (8) 11	(4) (8) 12	(4) (11) 15	(3) (10) 13
9	0.00	(5) (5) 10	(4) (5) 9	(4) (8) 12	(4) (8) 12	(3) (7) 10
	0.45	(6) (6) 12	(5) (5) 10	(4) (7) 11	(5) (7) 12	(4) (6) 10
	0.90	(5) (5) 10	(3) (4) 7	(6) (5) 11	(6) (6) 12	(6) (5) 11

$\alpha = 10\%$ $\delta = 0.19$

Sample size = 20

Table 7.6(b) : Distribution of Rejections of Hypothesis that the Bias of the Relevant Estimator is not Significantly Different from Zero.

PARAMETER : γ_{10}

K_2	λ	CT1	CT2	DT	JT1	JT2
3	0.00	(1) (4) 5	(1) (3) 4	(4) (4) 8	(2) (4) 6	(0) (3) 3
	0.45	(2) (4) 6	(1) (3) 4	(3) (5) 8	(3) (4) 7	(1) (4) 5
6	0.00	(1) (4) 5	(1) (4) 5	(2) (4) 6	(1) (5) 6	(0) (4) 4
	0.45	(1) (4) 5	(1) (4) 5	(2) (4) 6	(2) (7) 9	(0) (4) 4
	0.90	(2) (4) 6	(1) (4) 5	(2) (3) 5	(2) (5) 7	(1) (3) 4
9	0.00	(3) (4) 7	(2) (3) 5	(1) (2) 3	(1) (5) 6	(1) (3) 4
	0.45	(2) (4) 6	(1) (3) 4	(1) (2) 3	(2) (5) 7	(1) (4) 5
	0.90	(3) (3) 6	(1) (3) 4	(2) (2) 4	(4) (5) 9	(3) (4) 7

$\alpha = 5\%$ $\delta = 0.76$

PARAMETER: γ_{10}

K_2	λ	CT1	CT2	DT	JT1	JT2
3	0.00	(5) (4) 9	(4) (4) 8	(6) (7) 13	(3) (5) 8	(3) (4) 7
	0.45	(4) (4) 8	(4) (4) 8	(5) (9) 14	(3) (5) 8	(3) (4) 7
6	0.00	(5) (6) 11	(3) (5) 8	(4) (7) 11	(3) (8) 11	(3) (8) 11
	0.45	(6) (6) 12	(3) (5) 8	(4) (7) 11	(3) (7) 10	(2) (7) 9
	0.90	(3) (5) 8	(3) (4) 7	(3) (7) 10	(5) (6) 11	(4) (5) 9
9	0.00	(5) (5) 10	(4) (5) 9	(4) (8) 12	(4) (8) 12	(3) (7) 10
	0.45	(6) (6) 12	(5) (5) 10	(4) (7) 11	(5) (7) 12	(4) (6) 10
	0.90	(5) (5) 10	(3) (4) 7	(6) (5) 11	(6) (6) 12	(6) (5) 11

$\alpha = 10\%$ $\delta = 0.76$

Sample size = 20

Table 7.7(a) : Distribution of Rejections of Hypothesis that the Bias of the Relevant Estimator is not Significantly Different from Zero.

PARAMETER : γ_{11}

K_2	λ	CT1	CT2	DT	JT1	JT2
3	0.00	(3) (1) 4	(2) (1) 3	(4) (2) 6	(5) (1) 6	(3) (1) 4
	0.45	(2) (1) 3	(2) (1) 3	(4) (1) 5	(3) (2) 5	(3) (1) 4
6	0.00	(5) (2) 7	(4) (2) 6	(1) (2) 3	(4) (2) 6	(3) (2) 5
	0.45	(3) (2) 5	(3) (1) 4	(1) (3) 4	(3) (2) 5	(3) (1) 4
	0.90	(4) (1) 5	(3) (1) 4	(2) (2) 4	(5) (2) 7	(4) (2) 6
9	0.00	(4) (2) 6	(3) (2) 5	(2) (1) 3	(4) (3) 7	(3) (2) 5
	0.45	(4) (2) 6	(2) (2) 4	(4) (1) 5	(4) (3) 7	(4) (1) 5
	0.90	(3) (1) 4	(3) (1) 4	(3) (1) 4	(4) (6) 10	(3) (4) 7

α = 5% δ = 0.19

PARAMETER: γ_{11}

K_2	λ	CT1	CT2	DT	JT1	JT2
3	0.00	(5) (4) 9	(5) (4) 9	(6) (6) 12	(5) (2) 7	(5) (2) 7
	0.45	(4) (5) 9	(4) (2) 6	(5) (5) 10	(4) (4) 8	(4) (2) 6
6	0.00	(7) (3) 10	(5) (2) 7	(8) (4) 12	(7) (5) 12	(5) (4) 9
	0.45	(6) (3) 9	(6) (2) 8	(6) (4) 10	(5) (5) 10	(4) (5) 9
	0.90	(7) (2) 9	(7) (1) 8	(6) (3) 9	(8) (5) 13	(6) (4) 10
9	0.00	(5) (5) 10	(4) (3) 7	(6) (6) 12	(5) (4) 9	(5) (4) 9
	0.45	(5) (4) 9	(4) (3) 7	(5) (3) 8	(5) (6) 11	(5) (4) 9
	0.90	(8) (4) 12	(4) (2) 6	(8) (3) 11	(5) (8) 13	(5) (7) 12

α = 10% δ = 0.19

Sample size = 20

Table 7.7(b) : Distribution of Rejections of Hypothesis that the Bias of the Relevant Estimator is not Significantly Different from Zero.

PARAMETER: γ_{11}

K_2	λ	CT1	CT2	DT	JT1	JT2
3	0.00	(3) (2) 5	(2) (1) 3	(4) (4) 8	(3) (2) 5	(3) (1) 4
	0.45	(2) (1) 3	(2) (1) 3	(4) (3) 7	(3) (1) 4	(3) (1) 4
6	0.00	(3) (3) 6	(2) (2) 4	(0) (2) 2	(4) (4) 8	(3) (3) 6
	0.45	(3) (2) 5	(2) (2) 4	(2) (3) 5	(3) (3) 6	(3) (3) 6
	0.90	(2) (1) 3	(2) (1) 3	(3) (2) 5	(5) (2) 7	(3) (2) 5
9	0.00	(3) (3) 6	(3) (3) 6	(2) (3) 5	(4) (4) 8	(4) (3) 7
	0.45	(4) (3) 7	(2) (3) 5	(3) (2) 5	(4) (3) 7	(4) (3) 7
	0.90	(2) (2) 4	(2) (1) 3	(2) (4) 6	(3) (3) 6	(2) (3) 5

α = 5% δ = 0.76

PARAMETER : γ_{11}

K_2	λ	CT1	CT2	DT	JT1	JT2
3	0.00	(5) (5) 10	(5) (5) 10	(5) (6) 11	(5) (5) 10	(5) (5) 10
	0.45	(4) (4) 8	(4) (3) 7	(7) (5) 12	(4) (4) 8	(4) (4) 8
6	0.00	(6) (5) 11	(6) (3) 9	(4) (4) 8	(7) (5) 12	(6) (4) 10
	0.45	(6) (5) 11	(6) (4) 10	(4) (4) 8	(5) (6) 11	(5) (4) 9
	0.90	(6) (3) 9	(6) (3) 9	(5) (3) 8	(6) (5) 11	(5) (3) 8
9	0.00	(4) (6) 10	(4) (4) 8	(4) (5) 9	(4) (4) 8	(4) (4) 8
	0.45	(4) (6) 10	(4) (4) 8	(5) (7) 12	(4) (5) 9	(4) (4) 8
	0.90	(4) (5) 9	(4) (4) 8	(4) (7) 11	(7) (5) 12	(7) (4) 11

α = 10% δ = 0.76

Sample size = 20

For each value of β_{12}, ranging from 0.0 to 1.6 in steps of 0.2, a new set of 100 replications was generated and the power of each test was evaluated.

When $\beta_{12} = 0.8$, the power reported in the tables is of course equivalent to the probability of a Type I error.

Strictly speaking the term "power" is not appropriate as one cannot compare the powers of a number of tests when the probability of Type I errors are clearly not equal. Perhaps "probabilities of rejection" would be a more appropriate term.

Tests based on the standardized normal distribution (CT1 and JT1) showed greater "power" than their counterparts based on Student's t distribution, although we have already noted that the former produce higher Type I errors. Of the two tests based on the standardized normal distribution, CT1 generally had higher "power" than JT1 except for "large" K_2(6 or 9) and high levels of multicollinearity ($\lambda = 0.9$). A similar pattern was evident for comparisons of "power" between CT2 and JT2. On the other hand, the Type I errors associated with CT1 and CT2 were often greater than those of JT1 and JT2 (which themselves were generally greater than the nominal level of significance).

Dhrymes' test (DT) consistently exhibits the lowest "power" of the five tests, a result also noted by Maddala [26], although the Type I errors associated with this test are frequently nearer the nominal level than those associated with the other tests.

As one would expect, high levels of multicollinearity reduce the "power" of all five tests.

In conclusion, CT1 and CT2 dominate JT1 and JT2 respectively (i.e. they have higher "probabilities of rejection") although rarely over the entire range of values of β_{12}. This superiority however will be offset by the lower Type I errors which JT1 and JT2 frequently exhibit. No definitive statement can be made, therefore, concerning the relative powers of these four tests. The substantially lower "power" which is generally exhibited by DT suggests that this test is not a practical proposition, despite its accuracy for estimating the level of significance.

Table 7.8(a) Power of the Test Statistics

λ	β_{12}=	0.0	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6
0.00	CT1	1.00	0.92	0.62	0.11	0.04	0.30	0.70	0.87	0.96
	CT2	0.99	0.91	0.52	0.05	0.03	0.24	0.68	0.86	0.95
	DT	0.52	0.27	0.15	0.07	0.05	0.14	0.38	0.43	0.71
	JT1	0.96	0.82	0.48	0.11	0.04	0.19	0.62	0.77	0.92
	JT2	0.95	0.79	0.44	0.09	0.04	0.18	0.60	0.77	0.90
0.45	CT1	0.96	0.82	0.43	0.06	0.04	0.24	0.65	0.77	0.88
	CT2	0.93	0.75	0.36	0.03	0.03	0.21	0.62	0.74	0.85
	DT	0.40	0.29	0.13	0.05	0.04	0.12	0.33	0.38	0.62
	JT1	0.93	0.68	0.37	0.10	0.04	0.18	0.55	0.64	0.85
	JT2	0.89	0.61	0.30	0.07	0.03	0.15	0.52	0.59	0.82

$K_2 = 3$ $N = 20$ $\delta = 0.19$

Table 7.8(b): Power of the Test Statistics

λ	β_{12}=	0.0	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6
0.00	CT1	1.00	1.00	0.90	0.35	0.08	0.50	0.94	1.00	1.00
	CT2	1.00	0.99	0.88	0.29	0.08	0.47	0.93	1.00	1.00
	DT	1.00	0.98	0.80	0.23	0.08	0.37	0.90	0.97	1.00
	JT1	1.00	0.99	0.83	0.32	0.07	0.25	0.77	0.97	0.99
	JT2	1.00	0.98	0.78	0.31	0.07	0.24	0.74	0.95	0.99
0.45	CT1	1.00	0.96	0.70	0.14	0.08	0.45	0.82	1.00	1.00
	CT2	1.00	0.95	0.65	0.12	0.05	0.41	0.79	0.98	0.98
	DT	0.99	0.87	0.54	0.11	0.08	0.31	0.69	0.91	0.99
	JT1	1.00	0.88	0.70	0.20	0.08	0.28	0.65	0.85	0.94
	JT2	1.00	0.85	0.64	0.15	0.05	0.28	0.60	0.85	0.94
0.90	CT1	0.48	0.19	0.08	0.04	0.15	0.30	0.58	0.76	0.85
	CT2	0.41	0.15	0.04	0.03	0.12	0.24	0.50	0.71	0.84
	DT	0.35	0.14	0.08	0.06	0.10	0.24	0.37	0.56	0.74
	JT1	0.45	0.28	0.16	0.12	0.13	0.19	0.27	0.47	0.64
	JT2	0.45	0.26	0.13	0.09	0.09	0.16	0.24	0.38	0.60

$K_2 = 6$ $N = 20$ $\delta = 0.19$

Table 7.8(c): Power of the Test Statistics

λ	β_{12}=	0.0	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6
0.00	CT1	1.00	1.00	0.95	0.26	0.12	0.73	0.97	1.00	1.00
	CT2	1.00	1.00	0.94	0.21	0.08	0.63	0.95	1.00	1.00
	DT	1.00	1.00	0.92	0.15	0.08	0.58	0.96	1.00	1.00
	JT1	1.00	1.00	0.92	0.29	0.09	0.35	0.87	1.00	1.00
	JT2	1.00	1.00	0.91	0.25	0.08	0.32	0.86	0.99	1.00
0.45	CT1	1.00	0.98	0.76	0.11	0.09	0.59	0.88	0.99	0.99
	CT2	1.00	0.98	0.72	0.08	0.08	0.55	0.84	0.99	0.99
	DT	1.00	0.93	0.62	0.07	0.08	0.44	0.82	0.97	0.99
	JT1	1.00	0.91	0.73	0.16	0.07	0.28	0.75	0.94	0.96
	JT2	1.00	0.89	0.72	0.15	0.07	0.27	0.74	0.90	0.96
0.90	CT1	0.36	0.17	0.07	0.10	0.19	0.48	0.75	0.88	0.96
	CT2	0.27	0.15	0.05	0.08	0.17	0.44	0.74	0.87	0.96
	DT	0.20	0.17	0.05	0.07	0.16	0.36	0.68	0.81	0.96
	JT1	0.41	0.34	0.13	0.11	0.20	0.29	0.56	0.60	0.75
	JT2	0.36	0.31	0.12	0.09	0.18	0.27	0.53	0.57	0.72

$K_2 = 9$ $N = 20$ $\delta = 0.19$

CHAPTER 8

CONCLUSION

8.1 General Remarks

In Chapter 3 it was shown that, asymptotically, the 2SLS and J2SLS estimators are equivalent. Thus, one would expect the MSEs of the two estimators not to be significantly different from each other for "large" values of μ^2. If this were indeed so, the superiority of the jackknife technique for constructing confidence intervals and performing tests of significance would justify its use in applied economics.

From the preceding Monte Carlo study it is evident that the jackknife technique, whilst reducing the bias of the 2SLS estimator is not to be recommended for "small" μ^2 if the criterion for selection of an estimator is either minimum MSE or MAE. For "large" values of μ^2 there was little difference between the MSEs and MAEs of the 2SLS and J2SLS estimators, whilst the Wilcoxon Matched Pairs Signed Ranks test indicated significant differences between the two estimators only for small μ^2.

It was then observed (Chapter 7) that t(and z) statistics formed using the 2SLS estimator were not distributed according to the Student t or standardized normal distributions when $\delta = 0.19$. The actual distributions are highly skewed and serious errors could result if these postulated distributions were used for statistical inference. In general, this feature was less noticeable for the J2SLS estimator which, on the basis of Kolmogorov Smirnov tests, appears to have a reasonably symmetric distribution, and consequently there is less likelihood of serious errors being made if the postulated theoretical distributions are used for the purpose of statistical inference.

Even under "ideal" conditions (i.e. $\delta = 0.76$), test statistics based on the 2SLS estimator cannot show superior (to J2SLS) fit to their postulated theoretical distributions for the parameter β_{12}.

Finally, the "power" functions of the alternative tests were calculated over a range of values for β_{12}. The problem involved in comparing the "power" of two or more statistics when the Type I errors are not equal was recognized, but even making allowance for this problem Dhrymes' t statistic showed considerably lower "power" than the other statistics considered. This latter result confirms Maddala's [26] conclusions.

Clearly, therefore, a decision on circumstances under which application of the jackknife would be fruitful, hinges on one's definition of "large" in the context of the concentration parameter, μ^2.

8.2 When is the Concentration Parameter "Large"?

Whilst selection by "informed guesswork" of a value of μ^2 which could be taken as "large" is a somewhat haphazard procedure, two other problems of greater magnitude present themselves:

(i) can a value of the concentration parameter which is designated as "large" for an equation containing just two endogenous variables also be designated as "large" for an equation containing three (or more) endogenous variables?

(ii) how can the value of the concentration parameter be calculated?

To date, most of our knowledge concerning μ^2 is in the context of an equation containing just two endogenous variables, but preliminary work by Richardson and Rohr [54] appears to indicate that a value of μ^2 which is considered "large" in the context of an equation containing two endogenous variables may be "small" in the context of an equation containing three endogenous variables.

With regard to the second problem Rohr [57] has proposed that μ^2 be estimated from the sample and that this value be used to indicate whether μ^2 was "large" or "small" (he was interested in determining if μ^2 was large enough to enable the limiting distribution function (Student's t distribution) to be used as an approximation to the conventional t statistic without involving appreciable error).

Unfortunately, in the absence of knowledge of the sampling distribution of μ^2, when σ_{22} and $\underline{\pi}_{22}$ are replaced by their estimated values it would not be possible to obtain any measure of the reliability (i.e. the sampling variance) of our estimate. It should also be noted that there would be a conflict regarding the optimal method for estimating σ_{22} and $\underline{\pi}_{22}$. The Unrestricted Least Squares estimator would, intuitively, seem to be inefficient relative to the 2SLS induced Restricted Reduced Form estimator (although Dhrymes [13] has shown that, asymptotically, this may not be so), but the latter estimator may not possess moments of any order (see McCarthy [32]).

Clearly, therefore, considerably more knowledge concerning both the distribution of μ^2 and the properties of reduced form estimators is required before Rohr's [57] proposal can be properly evaluated.

8.3 Extension of the Results

The Monte Carlo experiments did not investigate the effects of an increase in sample size on the two estimators, although a proof that both the bias and the MSE of the 2SLS estimator are monotonically non-increasing functions of the sample size was given in Chapter 5. As the sample size increases, other variables being constant, the concentration parameter will, in general, increase in size and hence one would expect the MSE of the J2SLS estimator to tend towards (perhaps not monotonically) that of the 2SLS estimator. Conversely, a decrease in sample size might be expected to have the opposite effect on the J2SLS estimator.

The estimation of "large" (e.g. economy-wide) models may present a problem if use of the J2SLS estimator is contemplated. In such circumstances, the computing time and storage requirements will increase more rapidly for J2SLS than for 2SLS as the size of the model increases. It is unlikely however that 2SLS (and hence J2SLS) would be a feasible proposition anyway in large models, since it is probable that K would exceed N and consequently 2SLS would degenerate to OLS (see Fisher and Wadycki [15]). The jackknife could however be applied to an Instrumental Variables estimator which only considered a sub-set of the excluded predetermined variables when estimating any one structural equation, thus ensuring that $K < N$. Although such a procedure may yield inconsistent (perhaps of a minor nature) parameter estimates and would thus contravene Quenouille's original assumption that a consistent estimator is necessary for the jackknife to be successfully applied, Brundy and Jorgenson [8] cite conditions under which Instrumental Variables estimators based on sub-sets of the predetermined variables retain the property of consistency.

8.4 Extension to Three-Stage Least Squares

The foregoing analysis suggests that an extension of the jackknife technique to the Three-Stage Least Squares (3SLS) estimator may be an extremely tedious procedure. Having obtained 2SLS estimates of all structural coefficients in the system, the 3SLS estimator can be calculated by applying Generalized Least Squares to the entire system (where the equations are written in stacked form) to obtain

$$\left(Z'\left[\Omega^{-1} \otimes X(X'X)^{-1}X'\right]^{-1}Z\right)^{-1} Z'\left[\Omega^{-1} \otimes X(X'X)^{-1}X'\right]y, \qquad (8.1)$$

where $\otimes$ denotes the Kronecker product.

In general, Ω will be unknown and must be replaced by $\hat{\Omega}$, the matrix of mean squares and products of the 2SLS residuals. With Ω replaced by $\hat{\Omega}$ we obtained the 3SLS estimator.

If the jackknife were applied to the 3SLS estimator, Ω would have to be replaced by the matrix of mean squares and products of the J2SLS residuals, and (8.1) would have to be estimated N times with the ith observation omitted at each (of the N) replications.

It is the author's contention that this would not be a very fruitful exercise, especially as no exact results on the moments of the 3SLS estimator are available to provide an exact analysis of the jackknife's bias reducing potential. In addition, it is unlikely that the "simplifying" formula developed for J2SLS could be extended to J3SLS without considerable difficulty and, even then, the additional (to 3SLS) computer run-time involved would probably be substantial.

8.5 The Final Word

In this thesis we have demonstrated the value of the jackknife statistic for forming "accurate" confidence intervals and tests of significance when μ^2 is "large". The bias reducing property of the jackknife is generally present in the context of the 2SLS estimator, although it would certainly be unwise to jackknife the 2SLS estimator if the sample size is less than twice the number of variables included in the equation being estimated.

In applied economics, if the above condition is met <u>and</u> provided the degree of multicollinearity is not excessive, it is the author's contention that the true (unknown) value of the concentration parameter would, in general, be large enough to enable the jackknife technique to be used on the 2SLS estimator.

APPENDIX A

LEMMA

Proof of the following lemma is due to Bartlett [2].

LEMMA:

If A is a k x k non-singular matrix, and $\underline{c}$ and $\underline{d}$ are two k dimensional column vectors, then

$$(A + \underline{c}\,\underline{d}')^{-1} = A^{-1} - \frac{A^{-1}\underline{c}\,\underline{d}'A^{-1}}{1 + \underline{d}'A^{-1}\underline{c}} .$$

PROOF:

$$\begin{aligned}(A + \underline{c}\,\underline{d}')^{-1} &= A^{-1}\left(I + \underline{c}\,\underline{d}'A^{-1}\right)^{-1} \\ &= A^{-1}(I - \underline{c}\,\underline{d}'A^{-1} + \underline{c}\,\underline{d}'A^{-1}\,\underline{c}\,\underline{d}'A^{-1} - \ldots\ldots) \\ &= A^{-1} - A^{-1}\underline{c}\,\underline{d}'A^{-1}\,\{I - \underline{d}'A^{-1}\underline{c} + (\underline{d}'A^{-1}\underline{c})^2 - \ldots\ldots\} \\ &= A^{-1} - \frac{A^{-1}\underline{c}\,\underline{d}'A^{-1}}{1 + \underline{d}'A^{-1}\underline{c}} .\end{aligned}$$

APPENDIX B

DERIVATION OF A RESULT ON ASYMPTOTIC NORMALITY

Consider the first term in square brackets in equation (3.72), viz:

$$S_1(\text{diag } R_1 + N^{-1}. R_1)S_1'$$

$$= Z'(\Lambda_1 + \Lambda_2)(\text{diag } R_1 + N^{-1}. R_1)(\Lambda_1 + \Lambda_2)Z$$

$$+ Z'M_X\Lambda_2(\text{diag } R_1 + N^{-1}. R_1)\Lambda_2 M_X Z$$

$$- Z'(\Lambda_1 + \Lambda_2)(\text{diag } R_1 + N^{-1}. R_1)\Lambda_2 M_X Z$$

$$- Z'M_X\Lambda_2(\text{diag } R_1 + N^{-1}. R_1)(\Lambda_1 + \Lambda_2)Z \ , \qquad \text{(B.1)}$$

where $R_1 = \underline{\hat{u}}\,\underline{\hat{u}}' = [\underline{u} - Z(\underline{\hat{\theta}} - \underline{\theta})][\underline{u} - Z(\underline{\hat{\theta}} - \underline{\theta})]'$.

The ith element of $\underline{\hat{u}}$ can be written as

$$\hat{u}_i = u_i - \underline{z}_i'(\underline{\hat{\theta}} - \underline{\theta}) \ ,$$

and consequently the jkth element of the first term in equation (B.1) can be written as

$$\sum_{i=1}^{N} z_{ij} z_{ik}(\Lambda_1 + \Lambda_2)^2_{ii} \ [u_i - \underline{z}_i'(\underline{\hat{\theta}} - \underline{\theta})]^2 \ , \qquad \text{(B.2)}$$

ignoring, for the present, the term incorporating $N^{-1}. R_1$.

Upon expansion, equation (B.2) can be written as

$$\sum_{i=1}^{N} z_{ij} z_{ik}(\Lambda_1 + \Lambda_2)^2_{ii} \ u_i^2 - 2\sum_{i=1}^{N} z_{ij} z_{ik}(\Lambda_1 + \Lambda_2)^2_{ii} \ \underline{z}_i'(\underline{\hat{\theta}} - \underline{\theta})u_i$$

$$+ \sum_{i=1}^{N} z_{ij} z_{ik}(\Lambda_1 + \Lambda_2)^2_{ii} \ [\underline{z}_i'(\underline{\hat{\theta}} - \underline{\theta})]^2 \ . \qquad \text{(B.3)}$$

In the forthcoming analysis we will assume, without loss of generality, that the observations on the (g) explanatory endogenous variables occur in the first g elements of $\underline{z}_i'$.

Expanding the second term in equation (B.3) we obtain

$$- 2\sum_{i=1}^{N} z_{ij}z_{ik}(\Lambda_1 + \Lambda_2)^2_{ii} \Big[z_{i1}(\hat{\theta}_1 - \theta_1) + z_{i2}(\hat{\theta}_2 - \theta_2) + \ldots\ldots$$

$$\ldots\ldots + z_{i,K_1+g}(\hat{\theta}_{K_1+g} - \theta_{K_1+g})\Big] u_i \ . \quad \text{(B.4)}$$

Consider the first (of the K_1+g) term of the above expansion:

$$- 2\sum_{i=1}^{N} z_{ij}z_{ik}(\Lambda_1 + \Lambda_2)^2_{ii}\ z_{i1}(\hat{\theta}_1 - \theta_1)u_i \ . \quad \text{(B.5)}$$

We can partition z_{ij} and z_{ik} as

$z_{ij} = m_{ij} + v_{ij}$ and $z_{ik} = m_{ik} + v_{ik}$, (for $j,k \leq g$) where m_{ij} and m_{ik} represent the nonstochastic part of z_{ij} and z_{ik} respectively ($j,k = 1,2,\ldots,K_1 + g$), and v_{ij} and v_{ik} represent the reduced form disturbance part of z_{ij} and z_{ik} respectively (for $j,k > g$ this will of course be zero).

Expression (B.5) can therefore be written as

$$- 2(\hat{\theta}_1 - \theta_1)\Big[\sum_{i=1}^{N}(\Lambda_1 + \Lambda_2)^2_{ii}m_{ij}m_{ik}m_{i1}u_i + \sum_{i=1}^{N}(\Lambda_1 + \Lambda_2)^2_{ii}m_{ij}m_{ik}v_{i1}u_i$$

$$+ \sum_{i=1}^{N}(\Lambda_1+\Lambda_2)^2_{ii}m_{ij}v_{ik}v_{i1}u_i + \sum_{i=1}^{N}(\Lambda_1 + \Lambda_2)^2_{ii}m_{ij}m_{i1}v_{ik}u_i$$

$$+ \sum_{i=1}^{N}(\Lambda_1 + \Lambda_2)^2_{ii}v_{ij}v_{ik}m_{i1}u_i + \sum_{i=1}^{N}(\Lambda_1 + \Lambda_2)^2_{ii}v_{ij}m_{ik}v_{i1}u_i$$

$$+ \sum_{i=1}^{N}(\Lambda_1 + \Lambda_2)^2_{ii}v_{ij}m_{ik}m_{i1}u_i + \sum_{i=1}^{N}(\Lambda_1 + \Lambda_2)^2_{ii}v_{ij}v_{ik}v_{i1}u_i\Big] \ .$$

(B.6)

Recall the decomposition introduced as equation (3.51), viz:

$V = \underline{u}\,\Psi' + E$,

the ijth element of which can be written as

$$v_{ij} = u_i \, \psi_j + e_{ij} \; . \tag{B.7}$$

Substituting for v_{ij}, v_{ik} and v_{il} in expression (B.6) we obtain

$$\begin{aligned}
- 2(\hat{\theta}_1 - \theta_1) \Bigg[& \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{ij} m_{ik} m_{il} u_i \\
& + \left(\sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{ij} m_{ik} \psi_l u_i^2 + \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{ij} m_{ik} e_{il} u_i \right) \\
& + \left(\sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{ij} \psi_k \psi_l u_i^3 + \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{ij} e_{ik} e_{il} u_i \right) \\
& + \left(\sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{ij} m_{il} \psi_k u_i^2 + \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{ij} m_{il} e_{ik} u_i \right) \\
& + \left(\sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{il} \psi_j \psi_k u_i^3 + \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{il} e_{ij} e_{ik} u_i \right) \\
& + \left(\sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{ik} \psi_j \psi_l u_i^3 + \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{ik} e_{ij} e_{il} u_i \right) \\
& + \left(\sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{ik} m_{il} \psi_j u_i^2 + \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{ik} m_{il} e_{ij} u_i \right) \\
& + \left(\sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} \psi_j \psi_k \psi_l u_i^4 + \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} e_{ij} e_{ik} e_{il} u_i \right) \Bigg] .
\end{aligned} \tag{B.8}$$

Consider the first term in expression (B.8) and note that

$$\max_{1 \leq i \leq N} \left| (\hat{\theta}_1 - \theta_1)(\Lambda_1 + \Lambda_2)^2_{ii} m_{ij} m_{ik} m_{il} \right|$$

$$\leq \max_{1 \leq i \leq N} \left\{ \left| (\hat{\theta}_1 - \theta_1) \right| \left| (\Lambda_1 + \Lambda_2)^2_{ii} m_{ij} m_{ik} m_{il} \right| \right\}$$

$$= \left| \hat{\theta}_1 - \theta_1 \right| \max_{1 \leq i \leq N} \left| (\Lambda_1 + \Lambda_2)^2_{ii} m_{ij} m_{ik} m_{il} \right| .$$

Since $$\operatorname*{plim}_{N\to\infty} (\hat{\theta}_j - \theta_j) = 0 , \qquad (j = 1,2,\ldots,K_1 + g) \qquad \text{(B.9)}$$

it follows that

$$\operatorname*{plim}_{N\to\infty} (|\hat{\theta}_1 - \theta_1|) \operatorname*{plim}_{N\to\infty} \max_{1\leq i\leq N} \left| (\Lambda_1+\Lambda_2)^2_{ii} m_{ij} m_{ik} m_{il} \right| = 0 . \qquad \text{(B.10)}$$

Then using Theorem I (from Chapter 3) it follows that

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N} \left[- 2(\hat{\theta}_1 - \theta_1) \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{ij} m_{ik} m_{il} u_i \right] = 0 .$$

Consider the second term in expression (B.8). Using the above logic it follows that

$$\operatorname*{plim}_{N\to\infty} (|\hat{\theta}_1 - \theta_1|) \operatorname*{plim}_{N\to\infty} \max_{1\leq i\leq N} \left| (\Lambda_1 + \Lambda_2)^2_{ii} m_{ij} m_{ik} \psi_1 \right| = 0 .$$

Then using Theorem I it follows that

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N} \left[- 2(\hat{\theta}_1 - \theta_1) \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{ij} m_{ik} \psi_1 (u_i^2 - \sigma^2) \right] = 0 ,$$

where $E(u_i^2) = \sigma^2$(finite) .

This result implies that

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N} \left[- 2(\hat{\theta}_1 - \theta_1) \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{ij} m_{ik} \psi_1 u_i^2 \right]$$

$$= \operatorname*{plim}_{N\to\infty} \frac{1}{N} \left[- 2(\hat{\theta}_1 - \theta_1) \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{ij} m_{ik} \psi_1 \sigma^2 \right]$$

$$= \sigma^2 \operatorname*{plim}_{N\to\infty} \frac{1}{N} \left[- 2(\hat{\theta}_1 - \theta_1) \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{ij} m_{ik} \psi_1 \right]$$

$= 0$, from Theorem I. (B.11)

With minor variations, this analysis can be used to show that, in the probability limit, the fourth, sixth, eighth, tenth, twelfth and fourteenth terms in expression (B.8) are all zero. Since u_i and e_{ij} are, by assumption, uncorrelated random variables with mean zero, it follows using Theorem I that the remaining terms in expression (B.8) all converge in probability to zero.

Expression (B.5) was analysed under the assumption that $j, k \lesssim g$. If both j and k are greater than g then no partitioning of z_{ij} and z_{ik} is necessary as they only contain nonstochastic (corresponding to X_1) elements. Under such circumstances, the resulting expansion of expression (B.5) is limited to the first two terms of expression (B.6) and thus the first three terms of expression (B.8). We have already argued that, in the probability limit, these three terms are zero.

If either j or k is less than g then one partitioning of z_{ij} (or z_{ik}) is necessary. The subsequent expansion of expression (B.5) will be limited to just four terms of expression (B.6) and we have already argued that the corresponding terms in expression (B.8) converge in probability to zero.

This concludes the analysis of the first term in expression (B.4). The remaining $K_1 + g - 1$ terms can be dealt with in an analogous manner noting, once more, that the last K_1 values of z_{ij} $(j = 1,2,\ldots, K_1 + g)$ contain no stochastic component.

Returning to expression (B.3) we have shown that the second term converges in probability to zero.

Consider the third term in expression (B.3), viz:

$$\sum_{i=1}^{N} z_{ij} z_{ik} (\Lambda_1 + \Lambda_2)^2_{ii} [\underline{z}_i' (\underline{\hat{\theta}} - \underline{\theta})]^2$$

$$= \sum_{i=1}^{N} z_{ij} z_{ik} (\Lambda_1 + \Lambda_2)^2_{ii} \Big[z_{i1}(\hat{\theta}_1 - \theta_1) + z_{i2}(\hat{\theta}_2 - \theta_2) + \ldots\ldots$$

$$\ldots\ldots + z_{i,K_1+g}(\hat{\theta}_{K_1+g} - \theta_{K_1+g}) \Big]^2 .$$

Upon squaring the term in square brackets we obtain

$$\sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} z_{ij} z_{ik} \Big[z^2_{i1}(\hat{\theta}_1 - \theta_1)^2 + z^2_{i2}(\hat{\theta}_2 - \theta_2)^2 + \ldots\ldots + z^2_{i,K_1+g}(\hat{\theta}_{K_1+g} - \theta_{K_1+g})^2$$

$$+ z_{i1} z_{i2} (\hat{\theta}_1 - \theta_1)(\hat{\theta}_2 - \theta_2) + z_{i1} z_{i3} (\hat{\theta}_1 - \theta_1)(\hat{\theta}_3 - \theta_3) + \ldots\ldots$$

$$+ z_{i1} z_{i,K_1+g} (\hat{\theta}_1 - \theta_1)(\hat{\theta}_{K_1+g} - \theta_{K_1+g})$$

$$+ z_{i2} z_{i1} (\hat{\theta}_2 - \theta_2)(\hat{\theta}_1 - \theta_1) + z_{i2} z_{i3} (\hat{\theta}_2 - \theta_2)(\hat{\theta}_3 - \theta_3) + \ldots\ldots$$

$$+ z_{i2} z_{i,K_1+g} (\hat{\theta}_2 - \theta_2)(\hat{\theta}_{K_1+g} - \theta_{K_1+g})$$

$$+ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$+ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$+ z_{i,K_1+g} z_{i1} (\hat{\theta}_{K_1+g} - \theta_{K_1+g})(\hat{\theta}_1 - \theta_1) + z_{i,K_1+g} z_{i2} (\hat{\theta}_{K_1+g} - \theta_{K_1+g})(\hat{\theta}_2 - \theta_2) + \ldots\ldots$$

$$+ z_{i,K_1+g} z_{i,K_1+g-1} (\hat{\theta}_{K_1+g} - \theta_{K_1+g})(\hat{\theta}_{K_1+g-1} - \theta_{K_1+g-1}) \Big] . \qquad \text{(B.12)}$$

Consider the first term of the above expression, viz:

$$\sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} z_{ij} z_{ik} z^2_{i1} (\hat{\theta}_1 - \theta_1)^2 , \qquad \text{(B.13)}$$

where the z_{ij} can again be partitioned only now

$$z^2_{i1} = (m_{i1} + v_{i1})^2 = m^2_{i1} + v^2_{i1} + 2m_{i1} v_{i1} . \quad (j,k \leq g)$$

Upon expansion, expression (B.13) can be written as

$$(\hat{\theta}_1 - \theta_1)^2 \Bigg[\sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{ij} m_{ik} m^2_{il} + \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{ij} m_{ik} v^2_{il}$$

$$+ 2 \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{ij} m_{ik} m_{il} v_{il} + \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{ij} v_{ik} m^2_{il}$$

$$+ \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{ij} v_{ik} v^2_{il} + 2 \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{ij} v_{ik} m_{il} v_{il}$$

$$+ \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} v_{ij} m_{ik} m^2_{il} + \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} v_{ij} m_{ik} v^2_{il}$$

$$+ 2 \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} v_{ij} m_{ik} m_{il} v_{il} + \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} v_{ij} v_{ik} m^2_{il}$$

$$+ \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} v_{ij} v_{ik} v^2_{il} + 2 \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} v_{ij} v_{ik} m_{il} v_{il} \Bigg] . \qquad \text{(B.14)}$$

Using the decomposition given by equation (B.7), expression (B.14) can be evaluated in a similar term by term manner to the analysis used for evaluating expression (B.6).

Consider the first term in expression (B.14). Since

$$\operatorname*{plim}_{N\to\infty} (\hat{\theta}_j - \theta_j)^2 = \operatorname*{plim}_{N\to\infty} (\hat{\theta}_j - \theta_j) \operatorname*{plim}_{N\to\infty} (\hat{\theta}_j - \theta_j) = 0, \text{ for all } j ,$$

and since m^2_{il} is a constant, it follows by an analogous proof to that used in deriving equation (B.10) that

$$\operatorname*{plim}_{N\to\infty} \max_{1\leq i\leq N} \left| (\hat{\theta}_1 - \theta_1)^2 (\Lambda_1 + \Lambda_2)^2_{ii} m_{ij} m_{ik} m_{il} \right| = 0 .$$

Using equation (B.7) the second term in expression (B.14) can be written as

$$(\hat{\theta}_1 - \theta_1)^2 \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{ij} m_{ik} \psi^2_1 u^2_i$$

$$+ (\hat{\theta}_1 - \theta_1)^2 \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{ij} m_{ik} e^2_{il}$$

$$+ (\hat{\theta}_1 - \theta_1)^2 \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)^2_{ii} m_{ij} m_{ik} \psi_1 u_i e_{i1} \quad . \qquad \text{(B.15)}$$

Since m_{ij}, m_{ik} and ψ_1 are constants the first two terms in expression (B.15), multiplied by $^1/N$, converge in probability to zero by the same proof used to derive the result given by equation (B.11), assuming $E(e_{i1}^2)$ is finite. Further, since u_i and e_{i1} are uncorrelated random variables with mean zero, it follows from Theorem I that the third term in expression (B.15) (multiplied by $^1/N$) converges in probability to zero.

Evaluation of the remaining terms in expression (B.14) follows a similar pattern, all converging in probability to zero.

Returning to expression (B.13) if either j or k (or both) are greater than g then the above analysis involves fewer terms in expression (B.14), as was shown when dealing with the second term in expression (B.3). The analysis, however, is identical.

Returning to expression (B.12), a similar analysis can be used to show that the remaining terms in the first line of this expression all converge in probability to zero. The same result holds for the terms in the remaining lines of expression (B.12), although the analysis is more tedious due to the introduction of another (the fourth) term in z.

To summarize, we have shown that the second and third terms in expression (B.3) converge in probability to zero. Thus we have shown that

$$\operatorname*{plim}_{N \to \infty} \frac{1}{N} \cdot Z'(\Lambda_1 + \Lambda_2)(\text{diag } R_1)(\Lambda_1 + \Lambda_2)Z$$

$$= \operatorname*{plim}_{N \to \infty} \frac{1}{N} \cdot Z'(\Lambda_1 + \Lambda_2)(\text{diag } \underline{u}\,\underline{u}')(\Lambda_1 + \Lambda_2)Z \; .$$

Noting that

$$\underset{N\to\infty}{\text{plim}}\ (\Lambda_2)_{ii} = -\ \underset{N\to\infty}{\text{plim}} \frac{\underline{z}_i'\ P^{-1}(\underline{z}_i - \underline{a}_i)}{k_i(1 - s_i + d_i)} = 0\ , \tag{B.16}$$

using equations (3.39), (3.41) and (3.42), the remaining terms in (diag R_1) in equation (B.1) can be analysed in an analogous manner to the first term. We have shown therefore that

$$\underset{N\to\infty}{\text{plim}}\ \frac{1}{N}\ .\ S_1(\text{diag } R_1)S_1' = \underset{N\to\infty}{\text{plim}}\ \frac{1}{N}\ .\ S_1(\text{diag } \underline{u}\ \underline{u}')S_1'\ .$$

We now consider the terms in $N^{-1}\ .\ R_1$ in expression (B.1). The first term can be expanded as follows:

$$\begin{aligned} & N^{-1}\ .\ Z'(\Lambda_1 + \Lambda_2)R_1(\Lambda_1 + \Lambda_2)Z \\ = \ & N^{-1}\ .\ Z'(\Lambda_1 + \Lambda_2)\hat{\underline{u}}\ \hat{\underline{u}}'(\Lambda_1 + \Lambda_2)Z \\ = \ & N^{-1}.\ Z'(\Lambda_1 + \Lambda_2)\ \underline{u}\ \underline{u}'(\Lambda_1 + \Lambda_2)Z \\ + \ & N^{-1}\ .\ Z'(\Lambda_1 + \Lambda_2)Z(\hat{\underline{\theta}} - \underline{\theta})(\hat{\underline{\theta}} - \underline{\theta})'Z'(\Lambda_1 + \Lambda_2)Z \\ - \ & N^{-1}\ .\ Z'(\Lambda_1 + \Lambda_2)Z(\hat{\underline{\theta}} - \underline{\theta})\underline{u}'(\Lambda_1 + \Lambda_2)Z \\ - \ & N^{-1}\ .\ Z'(\Lambda_1 + \Lambda_2)\underline{u}(\hat{\underline{\theta}} - \underline{\theta})'Z'(\Lambda_1 + \Lambda_2)Z\ . \end{aligned}$$

Since 2SLS is a consistent estimator we know that

$$\underset{N\to\infty}{\text{plim}}\ (\hat{\underline{\theta}} - \underline{\theta}) = \underline{0}\ ,$$

and from the preceeding analysis it is easy to show that

$$\underset{N\to\infty}{\text{plim}}\ \frac{1}{N}\ .\ Z'(\Lambda_1 + \Lambda_2)Z$$

is a finite matrix. Now consider the term

$$Z'(\Lambda_1 + \Lambda_2)\ \underline{u}\ . \tag{B.17}$$

The jth component of this random vector can be written as

$$\sum_{i=1}^{N} z_{ij}(\Lambda_1 + \Lambda_2)_{ii} u_i .$$

Partitioning z_{ij} into its stochastic and nonstochastic components, and using the decomposition of v_{ij} given by equation (B.7) we obtain

$$\sum_{i=1}^{N} m_{ij}(\Lambda_1 + \Lambda_2)_{ii} u_i + \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)_{ii} \psi_j u_i^2$$

$$+ \sum_{i=1}^{N} (\Lambda_1 + \Lambda_2)_{ii} u_i e_{ij} . \qquad \text{(B.18)}$$

Since the m_{ij} are constants, and using the result that

$$\underset{N\to\infty}{\text{plim}} (\Lambda_1 + \Lambda_2)_{ii} = 1 ,$$

it follows that

$$\underset{N\to\infty}{\text{plim}} \frac{1}{N} \sum_{i=1}^{N} m_{ij}(\Lambda_1 + \Lambda_2)_{ii} u_i = 0$$

by the Law of Large Numbers. The same Law ensures that the second term in equation (B.18) (multiplied by $^1/N$) converges in probability to a finite constant, provided $E(u_i^2)$ is finite, and that, since u_i and e_{ij} are uncorrelated random variables, the third term (multiplied by $^1/N$) converges in probability to zero.

Combining the above results, we have shown that

$$\underset{N\to\infty}{\text{plim}} \frac{1}{N} \left[\frac{1}{N} \cdot Z'(\Lambda_1 + \Lambda_2)R_1(\Lambda_1 + \Lambda_2)Z\right]$$

$$= \underset{N\to\infty}{\text{plim}} \frac{1}{N} \left[\frac{1}{N} \cdot Z'(\Lambda_1 + \Lambda_2)\underline{u}\,\underline{u}'(\Lambda_1 + \Lambda_2)Z\right] .$$

Using the result given by equation (B.16), in addition to the above results, it can be shown that the remaining terms containing $N^{-1}.R_1$ in equation (B.1) can be analysed in an analogous manner. All three remaining terms converge in probability to zero.

Thus we have shown that the first term in square brackets in equation (3.72) can be written as

$$\underset{N\to\infty}{\text{plim}} \frac{1}{N} . S_1(\text{diag } R_1 + N^{-1}.R_1)S_1' = \underset{N\to\infty}{\text{plim}} \frac{1}{N} . S_1(\text{diag } \underline{u}\,\underline{u}' + N^{-1}.\underline{u}\,\underline{u}')S_1' \ .$$

Consider the second term in square brackets in equation (3.72), viz: $S_2(\text{diag } R_2 + N^{-1}.R_2)S_2'$

$$= Z'(\Lambda_2 - \Lambda_3)(\text{diag } R_2 + N^{-1}.R_2)(\Lambda_2 - \Lambda_3)Z$$

$$+ Z'M_X\Lambda_3(\text{diag } R_2 + N^{-1}.R_2)\Lambda_3 M_X Z$$

$$+ Z'(\Lambda_2 - \Lambda_3)(\text{diag } R_2 + N^{-1}.R_2)\Lambda_3 M_X Z$$

$$+ Z'M_X\Lambda_3(\text{diag } R_2 + N^{-1}.R_2)(\Lambda_2 - \Lambda_3)Z \ , \tag{B.19}$$

where
$$R_2 = (I - M_X)\hat{\underline{u}}\,\hat{\underline{u}}'(I - M_X)$$

$$= (I - M_X)[\underline{u} - Z(\hat{\underline{\theta}} - \underline{\theta})][\underline{u} - Z(\hat{\underline{\theta}} - \underline{\theta})]'(I - M_X) \ .$$

The following results can be easily derived:

$$\underset{N\to\infty}{\text{plim}} (\Lambda_2 - \Lambda_3)_{ii} = -1 \ , \tag{B.20}$$

$$\underset{N\to\infty}{\text{plim}} (\Lambda_2 - \Lambda_3)^2_{ii} = 1 \ , \tag{B.21}$$

$$\underset{N\to\infty}{\text{plim}} (\Lambda_3)^2_{ii} = 1 \ , \tag{B.22}$$

and
$$\underset{N\to\infty}{\text{plim}} (\Lambda_2 - \Lambda_3)_{ii}(\Lambda_3)_{ii} = 1 \ . \tag{B.23}$$

The first term in equation (B.19) can be written as

$$Z'(\Lambda_2 - \Lambda_3)(\text{diag } R_2)(\Lambda_2 - \Lambda_3)Z$$

$$+ Z'(\Lambda_2 - \Lambda_3)(N^{-1}.R_2)(\Lambda_2 - \Lambda_3)Z \text{ ,} \tag{B.24}$$

whereupon, using the definition of R_2, the first term in equation (B.24) can be written as

$$Z'(\Lambda_2 - \Lambda_3)(\text{diag } \underline{\hat{u}}\, \underline{\hat{u}}')(\Lambda_2 - \Lambda_3)Z$$

$$- Z'(\Lambda_2 - \Lambda_3)M_X(\text{diag } \underline{\hat{u}}\, \underline{\hat{u}}')(\Lambda_2 - \Lambda_3)Z$$

$$- Z'(\Lambda_2 - \Lambda_3)(\text{diag } \underline{\hat{u}}\, \underline{\hat{u}}')M_X(\Lambda_2 - \Lambda_3)Z$$

$$+ Z'(\Lambda_2 - \Lambda_3)M_X(\text{diag } \underline{\hat{u}}\, \underline{\hat{u}}')M_X(\Lambda_2 - \Lambda_3)Z \text{ .} \tag{B.25}$$

The jkth element of the first term in equation (B.25) can be written as

$$\sum_{i=1}^{N} z_{ij} z_{ik} (\Lambda_2 - \Lambda_3)^2_{ii} \,[u_i - \underline{z}_i'(\underline{\hat{\theta}} - \underline{\theta})]^2 \text{ .}$$

Since

$$\underset{N\to\infty}{\text{plim}} \max_{1\leq i\leq N} (\Lambda_1 + \Lambda_2)^2_{ii} = \underset{N\to\infty}{\text{plim}} \max_{1\leq i\leq N} (\Lambda_2 - \Lambda_3)^2_{ii} = 1 \text{ ,}$$

this expression does not differ, asymptotically, from expression (B.2) and can therefore be analysed in an analogous manner. It follows that

$$\underset{N\to\infty}{\text{plim}} \frac{1}{N} . Z'(\Lambda_2 - \Lambda_3)(\text{diag } \underline{\hat{u}}\, \underline{\hat{u}}')(\Lambda_2 - \Lambda_3)Z$$

$$= \underset{N\to\infty}{\text{plim}} \frac{1}{N} . Z'(\Lambda_2 - \Lambda_3)(\text{diag } \underline{u}\, \underline{u}')(\Lambda_2 - \Lambda_3)Z \text{ .}$$

Consider the remaining terms in expression (B.25). We have already shown (equation (3.27)) that

$$\lim_{N\to\infty} (M_X)_{ii} = 0 \text{ ,} \qquad (i = 1,2,\ldots,N) \tag{B.26}$$

and thus the remaining terms must all converge in probability to zero by an analogous proof to that employed in analysing the first term.

Further, equations (B.22), (B.23) and (B.26) allow the remaining terms containing (diag R_2) in equation (B.19) to be analysed in an identical manner to their corresponding terms containing (diag R_1) in equation (B.1). It follows that

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N} \,.\, S_2(\text{diag } R_2)S_2' = \operatorname*{plim}_{N\to\infty} \frac{1}{N} \,.\, S_2(\text{diag } \underline{u}\,\underline{u}')S_2' \,.$$

The terms containing $N^{-1}.R_2$ in equation (B.19) can also be analysed in a similar manner to their counterparts in equation (B.1). Consider the first term containing $N^{-1}R_2$ in equation (B.19), <u>viz</u>

$$N^{-1}.\, Z'\,(\Lambda_2 - \Lambda_3)R_2(\Lambda_2 - \Lambda_3)Z$$

$$= N^{-1}.Z'(\Lambda_2 - \Lambda_3)(I - M_X)\underline{\hat{u}}\,\underline{\hat{u}}'(I - M_X)(\Lambda_2 - \Lambda_3)Z$$

$$= N^{-1}.Z'(\Lambda_2 - \Lambda_3)(I - M_X)\underline{u}\,\underline{u}'(I - M_X)(\Lambda_2 - \Lambda_3)Z$$

$$+ N^{-1}.Z'(\Lambda_2 - \Lambda_3)(I - M_X)Z(\underline{\hat{\theta}} - \underline{\theta})(\underline{\hat{\theta}} - \underline{\theta})'Z'(I - M_X)(\Lambda_2 - \Lambda_3)Z$$

$$- N^{-1}.Z'(\Lambda_2 - \Lambda_3)(I - M_X)\underline{u}(\underline{\hat{\theta}} - \underline{\theta})'Z'(I - M_X)(\Lambda_2 - \Lambda_3)Z$$

$$- N^{-1}.Z'(\Lambda_2 - \Lambda_3)(I - M_X)Z(\underline{\hat{\theta}} - \underline{\theta})\underline{u}'(I - M_X)(\Lambda_2 - \Lambda_3)Z \,. \qquad \text{(B.27)}$$

The first term in equation (B.27) can be written as

$$N^{-1}.Z'(\Lambda_2 - \Lambda_3)\underline{u}\,\underline{u}'(\Lambda_2 - \Lambda_3)Z - N^{-1}.Z'(\Lambda_2 - \Lambda_3)M_X\underline{u}\,\underline{u}'(\Lambda_2 - \Lambda_3)Z$$

$$- N^{-1}.Z'(\Lambda_2 - \Lambda_3)\underline{u}\,\underline{u}'M_X(\Lambda_2 - \Lambda_3)Z + N^{-1}.Z'(\Lambda_2 - \Lambda_3)M_X\,\underline{u}\,\underline{u}'\,M_X(\Lambda_2 - \Lambda_3)Z \,.$$

From our initial assumptions (specifically, Assumption (iii) in Section 2.1.3) it follows that

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N} \,.\, Z'(\Lambda_2 - \Lambda_3)M_X\underline{u} = \underline{0} \,,$$

and hence, asymptotically, the first term in equation (B.27) can be written as

$$\underset{N\to\infty}{\text{plim}}\ \frac{1}{N}\left[\frac{1}{N}\,.\,Z'(\Lambda_2 - \Lambda_3)\underline{u}\,\underline{u}'(\Lambda_2 - \Lambda_3)Z\right].$$

Since $(I - M_X)$ is a nonstochastic matrix, using equation (B.20) it follows that the last three terms in equation (B.27) (multiplied by $\frac{1}{N}$) all converge in probability to zero. Thus we have shown that

$$\underset{N\to\infty}{\text{plim}}\ \frac{1}{N}\left[\frac{1}{N}\,.\,Z'(\Lambda_2 - \Lambda_3)R_2(\Lambda_2 - \Lambda_3)Z\right]$$

$$= \underset{N\to\infty}{\text{plim}}\ \frac{1}{N}\left[\frac{1}{N}\,.\,Z'(\Lambda_2 - \Lambda_3)\underline{u}\,\underline{u}'(\Lambda_2 - \Lambda_3)Z\right].$$

Clearly the third and fourth terms in equation (3.72) can be analysed in an identical manner since the relevant results have already been derived.

To conclude, we have shown that

$$\underset{N\to\infty}{\text{plim}}\ \frac{1}{N}\,.\,S_1(\text{diag } R_1 + N^{-1}.R_1)S_1' = \underset{N\to\infty}{\text{plim}}\ \frac{1}{N}\,.\,S_1(\text{diag } \underline{u}\,\underline{u}' + N^{-1}.\,\underline{u}\,\underline{u}')S_1',$$

$$\underset{N\to\infty}{\text{plim}}\ \frac{1}{N}\,.\,S_2(\text{diag } R_2 + N^{-1}.R_2)S_2' = \underset{N\to\infty}{\text{plim}}\ \frac{1}{N}\,.\,S_2(\text{diag } \underline{u}\,\underline{u}' + N^{-1}.\,\underline{u}\,\underline{u}')S_2',$$

and

$$\underset{N\to\infty}{\text{plim}}\ \frac{1}{N}\,.\,S_1(\text{diag } R_3 + N^{-1}.R_3)S_2' = \underset{N\to\infty}{\text{plim}}\ \frac{1}{N}\,.\,S_1(\text{diag } \underline{u}\,\underline{u}' + N^{-1}.\,\underline{u}\,\underline{u}')S_2'.$$

Consider the following summation

$$S_1(\text{diag } \underline{u}\,\underline{u}' + N^{-1}.\,\underline{u}\,\underline{u}')S_1' + S_2(\text{diag } \underline{u}\,\underline{u}' + N^{-1}.\,\underline{u}\,\underline{u}')S_2'$$

$$- S_1(\text{diag } \underline{u}\,\underline{u}' + N^{-1}.\,\underline{u}\,\underline{u}')S_2' + S_2(\text{diag } \underline{u}\,\underline{u}' + N^{-1}.\,\underline{u}\,\underline{u}')S_1' ,$$

which can be simplified to

$$(S_1 - S_2)(\text{diag } \underline{u}\,\underline{u}' + N^{-1}\underline{u}\,\underline{u}')(S_1 - S_2)'$$

$$= Z'(\Lambda_1 + 2\Lambda_2 - \Lambda_3)(\text{diag } \underline{u}\,\underline{u}' + N^{-1}\underline{u}\,\underline{u}')(\Lambda_1 + 2\Lambda_2 - \Lambda_3)Z$$

$$- Z'M_X(\Lambda_2 - \Lambda_3)(\text{diag } \underline{u}\,\underline{u}' + N^{-1}\underline{u}\,\underline{u}')(\Lambda_1 + 2\Lambda_2 - \Lambda_3)Z$$

$$- Z'(\Lambda_1 + 2\Lambda_2 - \Lambda_3)(\text{diag } \underline{u}\,\underline{u}' + N^{-1}\underline{u}\,\underline{u}')(\Lambda_2 - \Lambda_3)M_XZ$$

$$+ Z'M_X(\Lambda_2 - \Lambda_3)(\text{diag } \underline{u}\,\underline{u}' + N^{-1}\underline{u}\,\underline{u}')(\Lambda_2 - \Lambda_3)M_XZ \ . \qquad (B.28)$$

The jkth element of the first term containing (diag $\underline{u}\,\underline{u}'$) in equation (B.28) can be written as

$$\sum_{i=1}^{N} m_{ij}m_{ik}(\Lambda_1 + 2\Lambda_2 - \Lambda_3)^2_{ii}u_i^2 + \sum_{i=1}^{N} v_{ij}m_{ik}(\Lambda_1 + 2\Lambda_2 - \Lambda_3)^2_{ii}u_i^2$$

$$+ \sum_{i=1}^{N} m_{ij}v_{ik}(\Lambda_1 + 2\Lambda_2 - \Lambda_3)^2_{ii}u_i^2 + \sum_{i=1}^{N} v_{ij}v_{ik}(\Lambda_1 + 2\Lambda_2 - \Lambda_3)^2_{ii}u_i^2 \ . \qquad (B.29)$$

We have already shown (equation (3.47)) that

$$\underset{N\to\infty}{\text{plim}} \sqrt{N} \max_{1\le i\le N} (\Lambda_1 + 2\Lambda_2 - \Lambda_3)_{ii} = 0, \qquad (B.30)$$

from which it follows that

$$\underset{N\to\infty}{\text{plim}} \sqrt{N} \max_{1\le i\le N} \left| m_{ij}m_{ik}(\Lambda_1 + 2\Lambda_2 - \Lambda_3)_{ii} \right| = 0 \ .$$

Using Theorem I it follows that

$$\underset{N\to\infty}{\text{plim}} \frac{1}{N} \sum_{i=1}^{N} m_{ij}m_{ik}(\Lambda_1 + 2\Lambda_2 - \Lambda_3)^2_{ii}u_i^2 = 0 \ ,$$

provided $E(u_i^{\,2})$ is finite.

The decomposition given by equation (B.7) is required in order to evaluate the three remaining terms in equation (B.29). The second term can be written as

$$\sum_{i=1}^{N} (\Lambda_1 + 2\Lambda_2 - \Lambda_3)^2_{ii}m_{ik}\psi_j u_i^3 + \sum_{i=1}^{N} (\Lambda_1 + 2\Lambda_2 - \Lambda_3)^2_{ii}m_{ik}e_{ij}u_i^2 \ .$$

Equation (B.29) and Theorem I ensure that the first term in this expression (multiplied by 1/N) converges in probability to zero; and since the e_{ij} and u_i are uncorrelated random variables with mean zero, Theorem I ensures that the second term in this expression (multiplied by 1/N) also converges in probability to zero. Similarly the third term in equation (B.29) (multiplied by 1/N) converges in probability to zero.

The fourth term in expression (B.29) can be written as

$$\sum_{i=1}^{N} (\Lambda_1 + 2\Lambda_2 - \Lambda_3)^2_{ii}\psi_j\psi_k u_i^4 + \sum_{i=1}^{N} (\Lambda_1 + 2\Lambda_2 - \Lambda_3)^2_{ii}\psi_j e_{ik} u_i^3$$

$$+ \sum_{i=1}^{N} (\Lambda_1 + 2\Lambda_2 - \Lambda_3)^2_{ii}\psi_k e_{ij} u_i + \sum_{i=1}^{N} (\Lambda_1 + 2\Lambda_2 - \Lambda_3)^2_{ii} e_{ij} e_{ik} u_i^2 .$$

The first term of this expression (multiplied by 1/N) converges in probability to zero by virtue of equation (B.30) and Theorem I. The three remaining terms (multiplied by 1/N) also converge in probability to zero, using Theorem I and the assumption that e_{ij} and u_i are uncorrelated random variables with mean zero.

The first term in $N^{-1}.\ \underline{u}\ \underline{u}'$ in equation (B.28) is

$$N^{-1}.Z'(\Lambda_1 + 2\Lambda_2 - \Lambda_3)\underline{u}\ \underline{u}'(\Lambda_1 + 2\Lambda_2 - \Lambda_3)Z.$$

We have already shown in Chapter 3 that

$$\underset{N\to\infty}{\text{plim}}\ \frac{1}{\sqrt{N}} . Z'(\Lambda_1 + 2\Lambda_2 - \Lambda_3)\underline{\hat{u}} = \underline{0} .$$

Since each element of the 2SLS residual vector converges in distribution to the corresponding element of the disturbance vector, this implies that

$$\underset{N\to\infty}{\text{plim}}\ \frac{1}{\sqrt{N}} . Z'(\Lambda_1 + 2\Lambda_2 - \Lambda_3)\underline{u} = \underline{0} . \tag{B.31}$$

It follows that

$$\underset{N\to\infty}{\text{plim}}\ \frac{1}{N}\ [\frac{1}{N} . Z'(\Lambda_1 + 2\Lambda_2 - \Lambda_3)\underline{u}\ \underline{u}'(\Lambda_1 + 2\Lambda_2 - \Lambda_3)Z] = 0 .$$

Thus we have shown that the first term in equation (B.28) converges in probability to zero.

Consider the second term in equation (B.28), viz:

$$- Z'M_X(\Lambda_2 - \Lambda_3)(\text{diag } \underline{u}\,\underline{u}')(\Lambda_1 + 2\Lambda_2 - \Lambda_3)Z$$

$$- Z'M_X(\Lambda_2 - \Lambda_3)(N^{-1}. \underline{u}\,\underline{u}')(\Lambda_1 + 2\Lambda_2 - \Lambda_3)Z \; , \qquad \text{(B.32)}$$

and we will analyse the term in (diag $\underline{u}\,\underline{u}'$) first.

$$\text{Since } \operatorname*{plim}_{N\to\infty} \left(\frac{1}{N} . X'X\right)^{-1} \left(\frac{1}{N} . X'Z\right) = \Sigma_{XX}^{-1} \Sigma_{XZ} \; , \qquad \text{(B.33)}$$

we need only consider the limiting form of

$$X'(\Lambda_2 - \Lambda_3)(\text{diag } \underline{u}\,\underline{u}')(\Lambda_1 + 2\Lambda_2 - \Lambda_3)Z \; .$$

The rjth element of this term can be written as

$$\sum_{i=1}^{N} x_{ir}(\Lambda_2 - \Lambda_3)_{ii}(\Lambda_1 + 2\Lambda_2 - \Lambda_3)_{ii}(m_{ij} + v_{ij})u_i^2$$

$(j = 1,2,\ldots,K_1 + g \; ; \; r = 1,2,\ldots,K)$,

i.e. $$\sum_{i=1}^{N} x_{ir}(\Lambda_2 - \Lambda_3)_{ii}(\Lambda_1 + 2\Lambda_2 - \Lambda_3)_{ii}m_{ij}u_i^2$$

$$+ \sum_{i=1}^{N} x_{ir}(\Lambda_2 - \Lambda_3)_{ii}(\Lambda_1 + 2\Lambda_2 - \Lambda_3)_{ii}\psi_j u_i^3$$

$$+ \sum_{i=1}^{N} x_{ir}(\Lambda_2 - \Lambda_3)_{ii}(\Lambda_1 + 2\Lambda_2 - \Lambda_3)_{ii} \, u_i^2 e_{ij} \; .$$

Using equations (B.20) and (B.30), by Theorem I

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} x_{ir}(\Lambda_2 - \Lambda_3)_{ii}(\Lambda_1 + 2\Lambda_2 - \Lambda_3)_{ii}m_{ij}u_i^2 = 0 \; ,$$

provided $E(u_i^2)$ is finite. By the same argument

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} x_{ir}(\Lambda_2 - \Lambda_3)_{ii}(\Lambda_1 + 2\Lambda_2 - \Lambda_3)_{ii}\psi_j u_i^3 = 0,$$

provided $E(u_i^3)$ is finite. Finally, since u_i and e_{ij} are uncorrelated random variables with mean zero it follows, using Theorem I, that the remaining term in the above expression (multiplied by 1/N) converges in probability to zero.

Now consider the limiting form of the second term in expression (B.32), viz

$$Z'M_X(\Lambda_2 - \Lambda_3)(\tfrac{1}{N}\,\underline{u}\,\underline{u}')(\Lambda_1 + 2\Lambda_2 - \Lambda_3)Z \ .$$

In view of equation (B.33) we are concerned with the limiting form of

$$N^{-1}.X'(\Lambda_2 - \Lambda_3)\underline{u}\,\underline{u}'(\Lambda_1 + 2\Lambda_2 - \Lambda_3)Z \ .$$

Combining Assumption (v), Section 2.1.3., with equation (B.20) it follows that

$$\underset{N\to\infty}{\text{plim}}\ \frac{1}{N}\ .\ X'(\Lambda_2 - \Lambda_3)\underline{u} = \underline{0} \ ,$$

which when combined with equation (B.31) ensures that

$$\underset{N\to\infty}{\text{plim}}\ \frac{1}{N}\left[\frac{1}{N}\ .\ Z'M_X(\Lambda_2 - \Lambda_3)(\underline{u}\,\underline{u}')(\Lambda_1 + 2\Lambda_2 - \Lambda_3)Z\right] = 0 \ .$$

Thus the second, and hence the third, term in equation (B.28) has been shown to converge in probability to zero.

The fourth term in equation (B.28) can be written as

$$Z'X(X'X)^{-1}X'(\Lambda_2 - \Lambda_3)(\text{diag}\ \underline{u}\,\underline{u}')(\Lambda_2 - \Lambda_3)X(X'X)^{-1}X'Z$$

$$+\ Z'X(X'X)^{-1}X'(\Lambda_2 - \Lambda_3)(N^{-1}.\underline{u}\,\underline{u}')(\Lambda_2 - \Lambda_3)X(X'X)^{-1}X'Z \ . \qquad \text{(B.34)}$$

Again, combining Assumption (v), Section 2.1.3, with equations (B.20) and (B.33) we have shown that

$$\underset{N\to\infty}{\text{plim}}\ \frac{1}{N}\left[\frac{1}{N}\ .\ Z'M_X(\Lambda_2 - \Lambda_3)(\underline{u}\,\underline{u}')(\Lambda_2 - \Lambda_3)M_XZ\right] = 0 \ .$$

Consider the expression

$$X'(\Lambda_2 - \Lambda_3)(\text{diag}\ \underline{u}\,\underline{u}')(\Lambda_2 - \Lambda_3)X \ ,$$

which has rsth term

$$\sum_{i=1}^{N} x_{ir}x_{is}(\Lambda_2 - \Lambda_3)^2_{ii}u_i^2 \; . \qquad (r,s = 1,2,\ldots,K) \qquad (B.35)$$

From equation (B.21)

$$\operatorname*{plim}_{N\to\infty} (\Lambda_2 - \Lambda_3)^2_{ii} = 1$$

for all $i(i = 1,2,\ldots,N)$, from which it follows that

$$\operatorname*{plim}_{N\to\infty} \max_{1\leq i\leq N} \left\{1 - (\Lambda_2 - \Lambda_3)^2_{ii}\right\} = 0 \; .$$

Thus, using Theorem I, it follows that

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} x_{ir}x_{is} \left[1 - (\Lambda_2 - \Lambda_3)^2_{ii}\right] (u_i^2 - \sigma^2) = 0 \; ,$$

and hence we deduce that

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} x_{ir}x_{is}u_i^2 = \sigma^2 \frac{1}{N} \sum_{i=1}^{N} x_{ir}x_{is} \; .$$

Substituting this result back into equation (B.34), and using equation (B.33) we have shown that the fourth term in equation (B.28), and hence the entire expression, converges in probability to

$$\sigma^2\Sigma_p \; .$$

APPENDIX C

THE TORONTO FUNCTION

Copson [10] has shown that, for large $|x|$,

$$ {}_1F_1\,(\alpha;\gamma;x) \sim \frac{\Gamma(\gamma)}{\Gamma(\alpha)}\,\frac{e^x}{x^{\gamma-\alpha}}\;{}_2F_0(\gamma-\alpha,\,1-\alpha;\;;\frac{1}{x})\,, $$

where ${}_2F_0(\gamma-\alpha;\,1-\alpha;\;;\,1/x) = \sum_{r=0}^{\infty} \frac{(\gamma-\alpha)_r(1-\alpha)_r}{r!}\left(\frac{1}{x}\right)^r$,

and Pochhammer's symbol means

$$ (\gamma-\alpha)_r = \frac{\Gamma(\gamma-\alpha+r)}{\Gamma(\gamma-\alpha)}\,. $$

If $\gamma = \alpha + 1$ then

$$ {}_1F_1(\alpha;\alpha+1;\,x) \sim \frac{\alpha e^x}{x}\sum_{r=0}^{\infty}(1-\alpha)_r\left(\frac{1}{x}\right)^r\,, \tag{C.1} $$

which has a finite number of terms if α is a positive integer.

Equation (C.1) is required for evaluation of the first order moment of the 2SLS estimator. For second and higher order moments $\gamma = \alpha + k$ (where k is the order of the moment under consideration), but can be expressed in terms of equation (C.1) by utilizing the recurrence relations for the confluent hypergeometric function (see Slater [64] ; p.19).

The Toronto function was developed by Heatley [18] and is defined as

$$ T(2\alpha-1,\gamma-1,x^{\frac{1}{2}}) = x^{(\gamma-\alpha)}\;e^{-x}\,\frac{\Gamma(\alpha)}{\Gamma(\gamma)}\;{}_1F_1(\alpha;\gamma;\,x) \tag{C.2} $$

(N.B. Slater [64] gives this formula with an incorrect sign, p.99).

This function is characterized by convergence to unity as x increases indefinitely.

If $\gamma = \alpha + 1$, then equation (C.2) can be rewritten as

$$T(2\alpha - 1, \alpha, x^{\frac{1}{2}}) = \frac{x}{\alpha} e^{-x} \, {}_1F_1(\alpha; \alpha + 1; x) \quad . \tag{C.3}$$

We state two special forms of the Toronto function that are required in the forthcoming analysis:

$$T(1, 1, x^{\frac{1}{2}}) = 1 - e^{-x}, \tag{C.4}$$

and $$T(1, 2, x^{\frac{1}{2}}) = 1 - (1 + x)e^{-x} . \tag{C.5}$$

In addition, we require two of the recurrence formulae for the Toronto function (see Heatley [18; p.17]):

$$T(\nu + 2, \alpha + 1, x^{\frac{1}{2}}) = \frac{(\nu - 2\alpha - 1)}{2x} T(\nu, \alpha + 1, x^{\frac{1}{2}}) + T(\nu, \alpha, x^{\frac{1}{2}}), \tag{C.6}$$

and

$$T(\nu + 4, \alpha + 2, x^{\frac{1}{2}}) = \frac{(\nu + 1)}{2x} T(\nu, \alpha, x^{\frac{1}{2}}) - \frac{2(\alpha + 1 - x)}{2x} T(\nu + 2, \alpha + 1, x^{\frac{1}{2}}), \tag{C.7}$$

where $\nu = 2\alpha - 1$. Thus all values of α can be evaluated with ease.

If e^{-x} is assumed to be zero, the Toronto functions in equations (C.4) and (C.5) will both be unity. Thus, by setting $\alpha = 1$, initial values for the recurrence formulae can be determined, and it is then possible to evaluate the Toronto function for all integer values of α by repeated application of equations (C.6) and (C.7).

Following the above procedure, equation (C.3) can be rewritten as $${}_1F_1(\alpha; \alpha+1; x) = \frac{\alpha}{x} e^{x} \sum_{r=0}^{\infty} (1-\alpha)_r \left(\frac{1}{x}\right)^r ,$$

which is identical to the asymptotic approximation to the confluent hypergeometric function given in equation (C.1). Thus the error incurred in utilizing the asymptotic approximation for finite x is simply the error caused by assuming e^{-x} to be zero in the Toronto function.

It is easy to show that this error can be expressed as

$$\frac{\Gamma(\alpha+1)}{(-x)^{\alpha}}, \qquad \text{(C.8)}$$

thus as x increases indefinitely (for α fixed), the error of approximation tends to zero.

This analysis is only valid when α is an integer, a condition which will not always be upheld (e.g. when considering the moments of the 2SLS estimator, even values of K_2 will yield integer values for α, whereas odd values of K_2 will yield non-integer values for α).

For α non-integer, equation (C.1) is an infinite series, although it can be truncated after (say) n terms. If this is done the error involved by truncating the infinite series after the nth term will not exceed the (n+1)th term, and will be of the same sign as the (n+1)th term (e.g. see Luke [25; p.127]).

BIBLIOGRAPHY

1. Arvesen, J.N., Jackknifing U-Statistics, Annals of Mathematical Statistics, Vol. 40, 1969, pp. 2076-2100.

2. Bartlett, M.S., An Inverse Matrix Adjustment Arising in Discriminant Analysis, Annals of Mathematical Statistics, Vol. 22, 1951, pp. 107-111.

3. Basmann, R.L., A Generalised Classical Method of Linear Estimation of Coefficients in a Structural Equation, Econometrica, Vol. 25, 1957, pp. 77-83.

4. Basmann, R.L., A Note on the Exact Finite Sample Frequency Function of Generalized Classical Linear Estimators in Two Leading Over-Identified Cases, Journal of the American Statistical Association, Vol. 56, 1961, pp. 619-636.

5. Basmann, R.L., A Note on the Exact Finite Sample Frequency Functions of Generalized Classical Linear Estimators in a Leading Three Equation Case, Journal of the American Statistical Association, Vol. 58, 1963, pp. 161-171.

6. Box, G.E.P. and Muller, M.E., A Note on the Generation of Normal Deviates, Annals of Mathematical Statistics, Vol. 29, 1958, pp. 610-611.

7. Brillinger, D.R., The Asymptotic Behaviour of Tukey's General Method of Setting Approximate Confidence Limits (The Jackknife) When Applied to Maximum Likelihood Estimates, Review of the International Statistical Institute, Vol. 32, 1964, pp. 202-206.

8. Brundy, J. and Jorgenson, D., Efficient Estimation of Simultaneous Equations by Instrumental Variables, Review of Economics and Statistics, Vol. 53, 1971, pp. 207-224.

9. Chay, S.C., Fardo, R.D. and Mazumbar, M., On Using the Box-Muller Transformation with Congruential Pseudo-Random Number Generators, Applied Statistics, Vol. 24, 1975, pp. 132-135.

10. Copson, E.T., Theory of Functions of a Complex Variable, London, Oxford University Press, 1957.

11. Devine, C.M., Scalecholeski (Algorithm 12), Computer Journal, Vol. 9, 1966-1967, pp. 321-322.

12. Dhrymes, P.J., Alternative Asymptotic Tests of Significance and Related Aspects of 2SLS and 3SLS Estimated Parameters, Review of Economic Studies, Vol. 36, 1969, pp. 213-226.

13. Dhrymes, P.J., Restricted and Unrestricted Reduced Forms: Asymptotic Distribution and Relative Efficiency, Econometrica, Vol. 41, 1973, pp. 119-134.

14. Durbin, J., A Note on the Application of Quenouille's Method of Bias Reduction to the Estimation of Ratios, Biometrika, Vol. 46, 1959, pp. 477-480.

15. Fisher, W.D. and Wadycki, W.J., Estimating a Structural Equation in a Large System, Econometrica, Vol. 39, 1971, pp. 461-465.

16. Hammersley, J.M. amd Handscomb, D.C., Monte Carlo Methods, London, Methuen, 1964.

17. Hatanaka, M., On the Existence and the Approximation Formulae for the Moments of k-Class Estimators, The Economic Studies Quarterly, Vol. XXIV, 1973, pp. 1-15.

18. Heatley, A.H., A Short Table of the Toronto Function, Royal Society of Canada, Transactions, Section III, 1943, pp. 13-19.

19. Hoeffding, W., A Class of Statistics With Asymptotically Normal Distribution, Annals of Mathematical Statistics, Vol. 19, 1948, pp. 293-325.

20. Johnston, J., Econometric Methods, Second Edition, New York, McGraw-Hill, 1972.

21. Kabe, D.G., A Note on the Exact Distribution of the G.C.L. Estimators in Two Leading Overidentified Cases, Journal of the American Statistical Association, Vol. 58, 1963, pp. 535-537.

22. Kabe, D.G., On the Exact Distribution of the G.C.L. Estimators in a Leading Three Equation Case, Journal of the American Statistical Association, Vol. 59, 1964, pp. 881-894.

23. Kadane, J.B., Comparisons of k-Class Estimators when the Disturbances are Small, Econometrica, Vol. 39, 1971, pp. 723-737

24. Kendall, M.G. and Stuart, A., The Advanced Theory of Statistics, Volume II, London, Griffin, 1967.

25. Luke, Y.L., The Special Functions and Their Approximations, Volume I, New York, Academic Press, 1969.

26. Maddala, G.S., Some Small Sample Evidence on Tests of Significance in Simultaneous Equations Models, Econometrica, Vol. 42, 1974, pp. 841-851.

27. Malinvaud, E., Statistical Methods of Econometrics, Amsterdam, North Holland, 1966.

28. Mariano, R.S., The Existence of Moments of the Ordinary Least Squares and Two-Stage Least Squares Estimators, Econometrica, Vol. 40, 1972, pp. 643-652.

29. Mariano, R.S. and Ramage, J.G., Finite Sample Effects of Misspecification on Structural Coefficient Estimators, Discussion Paper, No. 320. The Warton School, Department of Economics, University of Pennsylvania, 1975.

30. Mariano, R.S. and Sawa, T., Exact Finite-Sample Distribution of the Limited Information Maximum Likelihood Estimator for Two Included Endogenous Variables, Journal of the American Statistical Society, Vol. 67, 1972, pp. 159-163.

31. Marriott, F.H.C. and Pope, J.A., Bias in Estimation of Autocorrelations, Biometrika, Vol. 41, 1954, pp. 390-402.

32. McCarthy, M.D., A Note on the Forecasting Properties of 2SLS Restricted Reduced Forms, International Economic Review, Vol. 13, 1972, pp. 757-761.

33. Mikhail, W.M., Simulating the Small-Sample Properties of Econometric Estimators, Journal of the American Statistical Association, Vol. 67, 1972, pp. 620-624.

34. Miller, R.G., A Trustworthy Jackknife, Annals of Mathematical Statistics, Vol. 35, 1964, pp. 1594-1605.

35. Miller, R.G., The Jackknife - A Review, Biometrika, Vol. 61, 1974, pp. 1-15.

36. Miller, R.G., An Unbalanced Jackknife, Annals of Statistics, Vol. 2, 1974, pp. 880-891.

37. Nagar, A.L., The Bias and Moment Matrix of the General k-Class Estimators of the Parameters in Simultaneous Equations, Econometrica, Vol. 27, 1959, pp. 575-595.

38. Neave, H.R., On Using the Box-Muller Transformation with Multiplicative Congruential Pseudo-Random Number Generators, Applied Statistics, Vol. 22, 1973, pp. 92-97.

39. Orcutt, G.H. and Winokur, H.S., First Order Autoregression: Inference, Estimation and Prediction, Econometrica, Vol. 37, 1969, pp. 1-14.

40. Owen, A.D., A Proof that Both the Bias and the Mean Square Error of the Two-Stage Least Squares Estimator are Monotonically Non-Increasing Functions of Sample Size, Econometrica, Vol. 44, 1976, pp. 409-411.

41. Owen A.D. and Knight, J.L., The Exact Finite Sample Properties of Simultaneous Equation Estimators and Associated Statistics: A Synthesis and Classified Bibliography, Unpublished Manuscript, 1976.

42. Phillips, G.D.A., Recursions for the Two Stage Least Squares Estimators, Journal of Econometrics Vol. 6, 1977, pp. 65-77.

43. Pike, M.C. and Hill, I.D., Pseudo-Random Numbers (Algorithm 266), Communications of the Association for Computing Machinery, Vol. 8, 1965, pp. 605-606.

44. Quenouille, M.H., Approximate Tests of Correlation in Time-Series, Journal of the Royal Statistical Society, Series B, Vol. 11, 1949, pp. 68-84.

45. Quenouille, M.H., Notes on Bias in Estimation, Biometrika, Vol. 43, 1956, pp. 353-360.

46. Rao, C.R., Linear Statistical Inference and its Applications, New York, Wiley, 1965.

47. Rao, J.N.K., A Note on the Estimation of Ratios by Quenouille's Method, Biometrika, Vol. 52, 1965, pp. 647-649.

48. Rao, J.N.K. and Webster, J., On Two Methods of Bias Reduction in the Estimation of Ratios, Biometrika, Vol. 53, 1966, pp. 571-577.

49. Relph, A.P., Confluent Hypergeometric (Algorithm 192), Communications of the Association for Computing Machinery, Vol. 6, 1963, p. 388.

50. Revankar, N.S., The Two-Stage Least Squares Estimators as Aitken Estimators and the Asymptotic Tests of Significance, Discussion Paper No. 174, Department of Economics, State University of New York at Buffalo.

51. Revankar, N.S. and Hartley, M.J., An Independence Tests and Conditional Unbiased Predictions in the Context of Simultaneous Equation Systems, International Economic Review, Vol. 14, 1973, pp. 625-631.

52. Richardson, D.H., The Exact Distribution of a Structural Coefficient Estimator, Journal of the American Statistical Association, Vol. 63, 1968, pp. 1214-1226.

53. Richardson, D.H. and Rohr, R.J., Distribution of a Structural t Statistic for the Case of Two Included Endogenous Variables, Journal of the American Statistical Association, Vol. 66, 1971, pp. 375-382.

54. Richardson, D.H. and Rohr, R.J., An Experimental Study of Two Stage Least Squares Distributions in Equations Containing Three Endogenous Variables, Unpublished Manuscript, 1975.

55. Richarson, D.H. and Wu, De-Min, A Note on the Comparison of Ordinary and Two-Stage Least Squares Estimators, Econometrica, Vol. 39, 1971, pp. 973-981.

56. Robson, D.S. and Whitlock, J.H., Estimation of a Truncation Point, Biometrika, Vol. 51, 1964, pp. 33-39.

57. Rohr, R.J., On the Distribution of Two Alternative Structural t Statistics, Journal of the American Statistical Association, Vol. 71, 1976, pp. 731-736.

58. Sawa, T., The Exact Sampling Distribution of Ordinary Least Squares and Two-Stage Least Squares Estimators, Journal of the American Statistical Association, Vol. 64, 1969, pp. 923-937.

59. Sawa, T., Finite-Sample Properties of the k-Class Estimators, Econometrica, Vol. 40, 1972, pp. 653-680.

60. Sawa, T., Almost Unbiased Estimator in Simultaneous Equations Systems, International Economic Review, Vol. 14, 1973, pp. 97-106.

61. Sawa, T., The Mean Square Error of a Combined Estimator and Numerical Comparison with the TSLS Estimator, Journal of Econometrics, Vol. 1, 1973, pp. 115-132.

62. Schucany, W.R., Gray, H.L. and Owen, D.B., On Bias Reduction in Estimation, Journal of the American Statistical Association, Vol. 66, 1971, pp. 524-533.

63. Siegel, S., Nonparametric Statistics for the Behavioural Sciences, New York, McGraw-Hill, 1956.

64. Slater, L.J., Confluent Hypergeometric Functions, London, Cambridge University Press, 1960.

65. Sullins, W.L., Certification of Algorithm 266, Communications of the Association for Computing Machinery, Vol. 15, 1972, pp. 1072-1073.

66. Summers, R., A Capital Intensive Approach to the Small Sample Properties of Various Simultaneous Equation Estimators, Econometrica, Vol. 33, 1965, pp. 1-41.

67. Takeuchi, K., Exact Sampling Moments of the Ordinary Least Squares, Instrumental Variable and Two-Stage Least Squares Estimators, International Economic Review, Vol. 11, 1970, pp. 1-12.

68. Teekens, R., Predictive Methods in Multiplicative Models, Rotterdam, Rotterdam University Press, 1972.

69. Thacher, H.C., Certification of Algorithm 191 and 192 Hypergeometric Confluent Hypergeometric, Communications of the Association for Computing Machinery, Vol. 7, 1964, p. 244.

70. Theil, H., Economic Forecasts and Policy, Second Edition, Amsterdam, North-Holland, 1961.

71. Theil, H., Principles of Econometrics, New York, Wiley, 1971.

72. Tukey, J.W., Bias and Confidence in Not-Quite Large Samples (Abstract), Annals of Mathematical Statistics, Vol. 29, 1958, p. 614.

73. Walsh, J.E., Concerning the Effect of Intraclass Correlation on Certain Significance Tests, Annals of Mathematical Statistics, Vol. 18, 1947, pp. 88-96.