RAMENOFSKY, Samuel David, 1944-
A MONTE CARLO STUDY OF A TWO EQUATION OVERIDENTIFIED MODEL WITH TWO SPECIFICATION ERRORS--AUTOCORRELATION AND MULTICOLLINEARITY.

The University of Oklahoma, Ph.D., 1972
Economics, general

THE UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

A MONTE CARLO STUDY OF A TWO EQUATION OVERIDENTIFIED MODEL WITH TWO SPECIFICATION ERRORS--AUTOCORRELATION AND MULTICOLLINEARITY

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

DOCTOR OF PHILOSOPHY

BY

SAMUEL DAVID RAMENOFSKY

Norman, Oklahoma

1972

A MONTE CARLO STUDY OF A TWO EQUATION OVERIDENTIFIED MODEL
WITH TWO SPECIFICATION ERRORS--AUTOCORRELATION
AND MULTICOLLINEARITY

APPROVED BY

Ed F Crim Jr

John C. Bruyer

James E. Hibdon

J. Kirker Stephens

DISSERTATION COMMITTEE

ACKNOWLEDGEMENTS

This study was undertaken in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the University of Oklahoma.

The author would like to extend his appreciation to a number of persons without whose assistance this study would not have been possible. The author sincerely thanks Dr. Ed F. Crim for his many efforts on behalf of the author in the preparation of this study. The author also thanks Dr. J. Kirker Stephens, Dr. James E. Hibdon, Dr. Alex Kondonassis, and Dr. J. Brixey for their helpful comments in preparing the manuscript.

For their cooperation to the author while doing research, appreciation is extended to Bob Newlin and the University of Missouri, Kansas City, Computer staff. The author also thanks Mrs. Dinah Henderson for her contribution in the typing of the manuscript.

The author expresses his appreciation to his wife, Cynthia, and daughter, Carrie, for their contributions and patience during the preparation of the manuscript. The author expresses his appreciation of the financial and moral contribution of his parents during this period.

The author alone accepts full responsibility for any errors or inconsistencies in this work.

TABLE OF CONTENTS

LIST OF TABLES

A MONTE CARLO STUDY OF A TWO EQUATION OVERIDENTIFIED MODEL WITH TWO SPECIFICATION ERRORS--AUTOCORRELATION AND MULTICOLLINEARITY

CHAPTER I

INTRODUCTION

Economic models are concerned with the general relationship between economic variables. Many times this interdependence is in the form of simultaneous equations. The degree of interdependence is different as compared to the classical analysis in that dependent variables of one equation usually are independent variables of another equation within the system. Econometricians have developed estimators to obtain empirical parameter estimates in simultaneous equations under specified assumptions. Each estimator has large sample properties and/or asymptotic properties deemed desirable for estimators based on large sample sizes (Christ, 1966; Theil, 1971). Knowing these properties does not eliminate the large sample problem of choosing the best estimator because economic models usually have specification errors.[1] Economists should be alert

[1]Specification error is due to the violation of one or more specified assumptions which were used to derive the large sample properties.

of these specification errors because they may give rise to empirical results that are erroneous.

The choice of estimators becomes more involved when the sample of economic data for a given model is small. Generally, the properties of estimators in small samples with no specification errors have not been derived. The theoretical studies of small sample estimators are very limited with regard to studying only the k-class estimators under specified assumptions within a given model. These theoretical studies are a conclusive first step in trying to solve for the distribution of estimates in finite samples. Because of the mathematical complexities in deriving distributions of estimates in small sample models, econometricians have resorted to empirical comparisons of estimators.

A Monte Carlo experiment is an empirical method of obtaining the probability distribution of a small sample estimator. Monte Carlo experiments consist of postulating a model, specifying parameters, generating values for exogenous variables and disturbance terms from assumed density functions, and solving for values of the endogenous variables from the reduced form equations. The econometrician then returns to the original model and applies estimators to the data under no _a priori_ information. This process is repeated a large number of times. The resultant frequency distribution of each estimator along with the _a priori_ information about the parameters is used to determine the small sample properties of each estimator.

In studying individual parameters a loss function must be chosen to determine which estimator performs "best" given this particular model. In many Monte Carlo studies, mean-square error has been chosen as the loss function. Robert Basmann has criticized this choice of a loss function because of lack of knowledge about the existence of low order moments (1961). Therefore, in this Monte Carlo study, the loss functions which will be chosen always exist and are measures of central location and dispersion.

Most studies of small sample sizes have been attempted observing only one specification error within the model. Even when more than one specification error has been observed, the Monte Carlo study has not generally been used to isolate the effects of each specification error upon the accuracy of each estimator. Major Monte Carlo studies of these types will be summarized in Chapter III.

As stated above, Chapter II and III will be a review of major research involved in searching for the best estimator that economists could use in empirically investigating economic models. As economists, we should be aware of these contributions and use them to guide us in the empirical investigations of economic phenomena.

The purpose of this Monte Carlo experiment is to study the effects of two specification errors on the accuracy of each estimator.[2] The specification errors introduced will be

---

[2]The estimators in this study are ordinary least squares, two stage least squares and three stage least squares.

autocorrelation of the disturbance terms and multicollinearity of the exogenous variables.[3] The Monte Carlo study will be carefully outlined and analyzed in Chapter IV. Due to the criticism by Hobson Thornber (1967), this study will vary both specification errors. Although this will not be a continuous variation, it may be sufficient to acknowledge the validity of the above criticism on past Monte Carlo studies. Secondly, it may allow the isolation of the effects of each specification error variation on the estimators involved.

The final chapter will outline and analyze an appropriate demand and supply model. The model is a demand and supply equation for pork which seems to have similar specification errors to those discussed in Chapter IV. This study will be contrasted and integrated into the store of knowledge already obtained from theoretical and empirical research summarized in Chapters II and III and the results of Chapter IV.

The goals of this paper are to observe the knowledge accumulated by economists in their search for improved understanding of estimators used in empirical economic research and to add to this store of knowledge by empirically observing and isolating the effects of two specification errors on each estimator within the model studied. It must be noted that generally, results in small sample size research are applicable only to the specified model studied. One hopes that the model

---

[3]The introduction of the autocorrelated disturbance terms will consider both serial and contemporaneous correlation.

chosen is general enough to give information to econometricians working with these specification errors in similar models under similar circumstances.

CHAPTER II

MAJOR THEORETICAL CONTRIBUTIONS TO THE STUDY OF ESTIMATORS IN SIMULTANEOUS EQUATIONS OF FINITE SAMPLE SIZES

Major theoretical contributions in deriving the distribution of different estimators of the structural parameters for finite samples have been limited, this limitation being due primarily to the complex mathematics involved. Simplifying assumptions, small models, and single equation estimators usually are requirements for deriving any theoretical distribution of structural parameters in finite samples. Even with the shortcomings of these theoretical works, econometricians have shown that research in this area is feasible. Theoretical research already completed has given insights concerning the properties of estimators in small samples. These theoretical contributions have helped develop better Monte Carlo studies. The major theoretical works will be summarized by authors.

Robert Basmann (1961) studied the finite frequency function of generalized classical linear estimators (GCL) in the following two equation, overidentified model:

$$-Y_{t1} + B_{12}Y_{t2} + \gamma_{11}X_{t1} + \gamma_{12}X_{t2} + \gamma_{13}X_{t3} + \gamma_{14}X_{t4}$$

$$+\gamma_{10} + e_{t1} = 0 \qquad t = 1,\ldots,n \qquad 2.1.1$$

$$-Y_{t1} + B_{22}Y_{t2} + \gamma_{21}X_{t1} + \gamma_{22}X_{t2} + \gamma_{23}X_{t3} + \gamma_{24}X_{t4}$$

$$+ \gamma_{20} + e_{t2} = 0 \qquad 2.1.2$$

where $E(e_{ti}) = 0$

$$E(e_{ti}, e_{tj}) = w_{ij} \qquad i,j = 1,2$$

$$e_{ti} \sim N(O, \sigma_{ii}). \qquad 2.1.3$$

Estimation was achieved on the basis of the following hypotheses:

$$H_1: \gamma_{11} = 0 \quad \gamma_{12} = 0 ; \quad \gamma_{13} \neq 0 \quad \gamma_{14} \neq 0 \qquad 2.1.4$$

$$H_2: \gamma_{23} = 0 \quad \gamma_{24} = 0 ; \quad \gamma_{21} \neq 0 \quad \gamma_{21} \neq 0. \qquad 2.1.5$$

The reduced form disturbance terms denoted by η_{t1} and η_{t2} were identically and independently distributed with

$$E(\eta_{ti}) = 0 \qquad 2.1.6$$

and $E(\eta_{ti}\eta_{tj}) = 0 \qquad i \neq j, i,j=1,2. \qquad 2.1.7$

The exogenous variables were independently distributed, i.e., no exogenous variable was a lagged endogenous variable.

$$\sum_{t=1}^{n} X_{tK} X_{tm} \begin{cases} = 0 & m \neq K \\ = 1 & m = K. \end{cases} \qquad 2.1.8$$

The restriction imposed on the exogenous variables by equation 2.1.8 was equivalent to orthogonalizing the exogenous sample data. The parameters of the reduced form were appropriately adjusted to the orthogonalization.

Dr. Basmann studied two cases. Case I was divided into two sub-cases. The first sub-case assumed hypothesis 2.1.4. The necessary condition for overidentification was met, $K^{**} - G^{\Delta} + 1 = 1$ (Johnston, 1963), where K** was the number of excluded exogenous variables and G^{Δ} was the number of included endogenous variables within the equation to be estimated. The exact finite sample frequency function of GCL estimators of B_{12} was

$$h_1(v_1) = \frac{e^{-\lambda^2/2 + b_2^2/4}}{\sqrt{\pi}(1+v_1)^{3/2}} \{\Gamma\left(\frac{3}{2}\right) I_0\left(\frac{b_2^2}{4}\right) {}_1F_1\left(\frac{3}{2};\ 1;\ \frac{X_1^2}{2}\right)$$

$$+ 2 \sum_{n=1}^{\infty} (-1)^n \frac{\Gamma(n+\frac{3}{2}) \left[\frac{X_1^2}{2}\right]^n}{\Gamma(2n+1)} I_n\left(\frac{b_2^2}{4}\right) {}_1F_1\left(n+\frac{3}{2};\ 2n+1;\ \frac{X_1^2}{2}\right)$$

$$-\infty < v_1 < +\infty \qquad 2.1.9$$

where $\lambda = (1+B_{12}^2)\ (\pi_{12}^2 + \pi_{22}^2)$ 2.1.10

$$b_2 = \sqrt{\pi_{21}^2 + \pi_{22}^2}\ B_{12} \qquad 2.1.11$$

${}_1F_1$ denoted the confluent hypergeometric function, I_n denoted a modified Bessel function, π_{ij} were reduced form coefficients, and $B_{12} \neq 0$, $X_1 = \dfrac{b_1}{\sqrt{1+v_1^2}}$, and $b_1 = \sqrt{\pi_{21}^2 + \pi_{22}^2}\ (B_{12}v_1 + 1)$.

The GCL estimators of B_{12} possessed a finite first moment but not a finite second moment. The second sub-case of case I assumed that $B_{12} = 0$ but the econometrician was unaware of this when applying the estimators. The exact frequency function reduced to

$$h_1(v_1) = \frac{e^{-\lambda^2/2}}{2(1 + v_1^2)^{3/2}} \quad {}_1F_1\left(\frac{3}{2};\ 1;\ \frac{X_1^2}{2}\right). \qquad 2.1.12$$

Once again the GCL estimators of B_{12} possessed a finite first moment and no finite second moment. These results showed the inadequacy of the mean square error as a loss function.

Case II assumed $\gamma_{11} = 0$, $\gamma_{12} = 0$ and $B_{12} = 0$. B_{12} was assumed to be zero but not known by the econometrician when applying the GCL estimators to the model, $K^{**} - G^{\Delta} - 1 = 2$ in this case. The frequency distribution of GCL for B_{12} was

$$h_1(v_1) = \frac{2}{\pi(1+v_1^2)^2}\ . \qquad 2.1.13$$

Finite first and second moments existed for Case II.

The above results hold if the structural normally distributed disturbance terms were serially correlated assuming the exogenous variables were fixed and not lagged endogenous variables. In this instance, the unrestricted maximum likelihood estimates of the reduced form coefficients were normally distributed but their variance was dependent upon the correlograms of the reduced form disturbance terms. Therefore, GCL estimates could be approximated by the normal distribution or by some other distribution function affected by the correlogram and the sample serial covariance of the exogenous variables.

Basmann's (1963) second article showed the generality in the transformation involved in solving for the exact distribution of GCL estimators. The model used was a general three equation model denoted as follows:

$$B' Y'_{t.} + \Gamma' X'_{t.} + e_{t.} = 0 \qquad 2.1.14$$

where B' was 3x3 matrix, $|B| \neq 0$, $B_{11} = -1$, Γ' was a 3xK matrix, $t = 1, \ldots, n$, and all the assumptions of his previous article held true (Basmann, 1961).

The equation of concern was:

$$-y_{t1} + B_{12} Y_{t2} + \gamma_{11} X_{t1} + \gamma_{12} X_{t2} + \ldots + \gamma_{1K} X_{tK} + e_{t1} = 0. \qquad 2.1.15$$

The identifying hypothesis was $\gamma_{11} = \gamma_{12} = \gamma_{13} = 0$, therefore $K^{**} - G^{\Delta} + 1 = 1$ in this case. The GCL estimators' density functions of B_{12} and B_{13}, denoted by $h_2(v_2)$ and $h_3(v_3)$ respectively, were derived under the assumption that $B_{12} = B_{13} = 0$. The marginal frequency functions of v_2 and v_3 were

$$h_2(v_2) = \frac{1}{2(1+v_2^2)^{3/2}} \qquad -\infty < v_2 < \infty \qquad 2.1.16$$

$$\text{and } h_3(v_3) = \frac{1}{2(1+v_3^2)^{3/2}} \sum_{n=0}^{\infty} \frac{{}_1F_1\{n+1;\ n+\frac{3}{2};\ \frac{-\pi_{33}^2}{2}\}}{n!} \left[\frac{\pi_{33}^2}{2(1+v_3^2)}\right]^n \qquad 2.1.17$$

The first moments of the GCL estimators existed but the second moments did not.

From these two studies, Robert Basmann conjectured that the existence of finite moments holds only to the order of $K^{**} - G^{\Delta} + 1$. If $K^{**} - G^{\Delta} + 1$ was equal to one, under Basmann's conjecture the GCL estimators would not possess moments higher than one. The second contribution of these studies was the solving of complex functions to find the exact frequency distribution of the estimates from the GCL estimators. But until the other estimators' distributions are also solved, the results can not be applied to the determination of the "best" estimator in finite samples. The third conclusion was that the finite sample frequency distribution of the estimates of the GCL estimators had, under similar assumptions, a general form in two and three equation models.

A. L. Nagar was another pioneer in research concerned with theoretical small sample properties of simultaneous equation estimators. In his first article, Nagar (1959) studied the approximate bias and moment matrix of the general k-class estimators.[4] The equation estimated was overidentified

---

[4]The approximate bias was due to the expansion of equation 2.2.4 where terms of higher order than T-1 were ignored.

and was assumed to exist in an M-equation model.

$$y = Y\gamma + X_1 B + u \qquad 2.2.1$$

where y was a column vector of T observations on the jointly dependent variable to be explained, Y was a Txm matrix of the explanatory jointly dependent variables, X_1 was a Txℓ matrix of the exogenous variables, and u was the disturbance term vector.

The total number of exogenous variables in the system was Λ. The exogenous variables were nonstochastic. Assumptions beyond the identifying assumptions were that the value k of the k-class estimators was nonstochastic, 1-k was of the order T^{-1}, and the disturbance terms were independent random drawings from a normal M-dimensional distribution with zero mean.

The first result was the bias of the k-class estimators in equation 2.2.1.

$$E(e_k) = \{-\chi + L - 1\}\ Qq \qquad 2.2.2$$

where $$k = 1 + \frac{\chi}{T} \qquad 2.2.3$$

e_k was the sampling error of the k-class estimators,

$$e_k = \binom{c}{b}_k - \binom{\gamma}{B} \qquad 2.2.4$$

and L was the number of exogenous variables in the model in excess of explanatory variables in equation 2.2.1.

$$L = \Lambda - (\ell + m) \qquad 2.2.5$$

$$Q = \begin{bmatrix} \bar{Y}'\bar{Y} & \bar{Y}'X_1 \\ X_1'\bar{Y} & \bar{X}_1'X_1 \end{bmatrix}^{-1} \qquad 2.2.6$$

where $\bar{Y}$ was the solution of the reduced form jointly dependent variables given that the reduced form error terms were distributed normally.

$$q = \begin{bmatrix} \text{cov} & (y_1, u) \\ & \vdots \\ \text{cov} & (y_m, u) \\ \text{cov} & (x_1, u) \\ & \vdots \\ \text{cov} & (x_\ell, u) \end{bmatrix} = \sigma_u^2 \begin{bmatrix} \pi \\ o \end{bmatrix} \qquad 2.2.7$$

where π was a column vector of m components that transformed the structural equation disturbance terms into the reduced form equation disturbance terms $\bar{V}$.

$$\bar{V} = u\,\pi' + W. \qquad 2.2.8$$

W was normally distributed and independent of the error terms in equation 2.2.1.

The bias of two stage least squares reduced to $(L-1)Qq$ because k=1 implied X=0. The bias vanished to the order T^{-1} when X=L-1 in equation 2.2.3.

The moment matrix to the order T^{-2} of the estimates

$\binom{C}{b}_k$ around the parameter vector $\binom{\gamma}{B}$ was given by

$$E(e_k e_k') = \sigma_u^2 Q(I + A^*) \qquad 2.2.9$$

where A^* was a matrix of order T^{-1}.

$$A^* = \{(2X - 2L + 3)\ \mathrm{tr}\ (C_1 Q) + \mathrm{tr}\ (C_2 Q)\}\ I +$$

$$\{(X - L + 2)^2 + 2(X + 1)\}\ C_1 Q +$$

$$(2X - L + 2)\ C_2 Q \qquad 2.2.10$$

where $$C_1 = \begin{Bmatrix} C_1^* & 0 \\ 0 & 0 \end{Bmatrix} = \frac{1}{\sigma_u^2} q\, q' = \frac{1}{\sigma_u^2} \pi\, \pi' \qquad 2.2.11$$

$$C_2 = \begin{Bmatrix} C_2^* & 0 \\ 0 & 0 \end{Bmatrix} = \frac{1}{T} E(W\, W') \qquad 2.2.12$$

and $$C = C_1 + C_2 = \frac{1}{T} E\ (\bar{V}'\bar{V}) = \begin{Bmatrix} C^* & 0 \\ 0 & 0 \end{Bmatrix}. \qquad 2.2.13$$

For the choice of k, the criteria was to minimize the value of the determinant of 2.2.9. This was the same as minimizing 2.2.14.

$$[E(e_k e_k')] = \sigma_u^2\ [Q]\ [1 + \mathrm{tr}\ (A^*)]. \qquad 2.2.14$$

The χ value which minimized 2.2.1 was

$$\chi = \Lambda - 2\ (m + \ell) - 3 - \frac{\mathrm{tr}(C_2 Q)}{\mathrm{tr}(C_1 Q)} \qquad 2.2.15$$

This minimization determined the optimum value of k which was usually less than one.[5]

In following up this conclusion, Dr. Nagar empirically studied Klein's model I. He estimated parameters letting k be optimal, k=1, k=0, $k=1+\frac{L+1}{T}$, and also solved for limited information and full information maximum likelihood estimates. For the k-class estimators, the bias and standard error were calculated according to 2.2.2 and 2.2.9, respectively. The results concurred with the theoretical results. The optimal value of k was below one for all three equations. The unbiased estimates had larger standard errors than the estimates corresponding to the optimal k.

Dr. Nagar's second article (1962), dealt with double k-class estimators. The model and assumptions were the same as in the previous article. The double k-class estimators were

$$k_1 = 1 + \frac{\chi_1}{T} \quad \text{and } k_2 = 1 + \frac{\chi_2}{T} \qquad 2.2.16$$

where χ_1 and χ_2 were non-stochastic real numbers and independent of the size of the samples. The results were as follows. The bias of the sampling error was

$$E(e) = (-\chi_2 + L-1)\ Qq + (\chi_1 - \chi_2)\ QC\binom{\gamma}{B}. \qquad 2.2.17$$

---

[5]The value of χ should usually be less than 0 under the conditions of 2.2.14 and 2.2.15. Therefore the above results followed from equation 2.2.3. χ could be positive if Λ was large. Since we are concerned with small finite samples, the T observations will limit Λ and generally cause χ to be negative.

The moment matrix, to the order T^{-2}, of the double k-class estimators was

$$E(ee') = \sigma_u^2 \{(1+B_0)\, Q + B_1 Q\, C_1 Q + B_2 Q\, C_2 Q$$

$$+ B_3\, (Qqr' + rq'Q) + B_4 rr'\} \qquad 2.2.18$$

where $$B_0 = -2(\chi_1 - \chi_2)\frac{r'q}{\sigma_u^2} - \{-2\chi_2 + 2L - 3$$

$$+ 2(\chi_1 - \chi_2)\, \frac{\begin{bmatrix}\gamma\\ B\end{bmatrix}' q}{\sigma_u^2}\}\ \mathrm{tr}\ C_1 Q + \mathrm{tr}\ C_2 Q \qquad 2.2.19$$

$$B_1 = 2(\chi_1 + 1) + \{(\chi_1 - \chi_2)\, \frac{\begin{bmatrix}\gamma\\ B\end{bmatrix}' q}{\sigma_u^2} - \chi_2 + L - 2\}^2 \qquad 2.2.20$$

$$B_2 = 2\,\chi_1 - L + 2 \qquad 2.2.21$$

$$B_3 = \frac{\chi_1 - \chi_2}{\sigma_u^2}\ \{(\chi_1 - \chi_2)\, \frac{\begin{bmatrix}\gamma\\ B\end{bmatrix}' q}{\sigma_u^2} - \chi_2 + L - 2\} \qquad 2.2.22$$

$$B_4 = \frac{(\chi_1 - \chi_2)^2}{\sigma_u^2} \qquad 2.2.23$$

and $$r = QC_2 \binom{\gamma}{B}. \qquad 2.2.24$$

Finally, the bias to the order T^{-1} of σ_u^2 of the double k-class estimators was

$$\frac{1}{T}\, E(u'u) - \sigma_u^2 = -\, \sigma_u^2 \Bigg[2\{ -\chi_2 + L - 1 + (\chi_1 - \chi_2)$$

$$\frac{\begin{bmatrix}\gamma\\ B\end{bmatrix}' q}{\sigma_u^2}\}\ \mathrm{tr}\ C_1 Q - \mathrm{tr}\ CQ + \frac{m+\ell}{T} + 2(\chi_1 - \chi_2)\frac{q'r}{\sigma_u^2} \Bigg] \qquad 2.2.25$$

All the results of the double k-class estimators were identical with the k-class estimators when $k_1 = k_2 = k$.

In summary, Nagar in both articles succeeded in deriving approximate distributions for k and double k-class estimators in models which were less restrictive than the ones studied by Basmann. Because of the generality of the models and the use of a family of estimators, this article has greatly broadened the theoretical knowledge of small sample estimators.

Dr. Gerhard Kabe presented two articles (1963; 1964) similar to each of Robert Basmann's articles to illustrate an alternative derivation of Basmann's results when using GCL estimators. Kabe's alternative approach used the properties of the non-central Wishart distribution. The second article will be discussed here because it was a generalization of the first article. In this article, Kabe (1964) specified a three equation model having three endogenous variables and three exogenous variables. Following Basmann's approach, the assumptions and identifying restrictions were given by 2.1.14-2.1.15. Using the non-central Wishart distribution, Kabe derived the same results as Basmann. Both authors simplified the problem by using identifying restrictions which reduced the rank of the reduced form coefficients to one. Kabe generalized this treatment by changing the identifying restriction so that the rank of the reduced form coefficient would be two. The results were given in his paper (p. 893, eq. 4.14, Kabe, 1964).

Kabe's articles showed that the sampling distributions of GCL estimates in two and three equation models followed a similar pattern. This concurred with Basmann's results. In extending Basmann's work, Kabe decreased the restrictions on the reduced form parameters and still derived the exact GCL estimators distribution. The feasibility of using the non-central Wishart distribution in deriving the distribution of the GCL estimators in a model with more than three equations seemed doubtful due to the mathematical complexities. Therefore Kabe, like Basmann, could only conjecture the similarity of the distribution of GCL estimators in a model as the number of equations increases.

A. R. Bergstrom (1962), solved the exact sampling distributions of ordinary least squares and maximum likelihood estimators of the marginal propensity to consume in a simple Keynesian Model.

$$C_t = \alpha + BY_t + U_t \qquad t=1,\ldots,T$$

$$Y_t = C + I_t. \qquad 2.4.1$$

Where $U_t \sim N(0,\sigma^2)$ 2.4.2

$I_1, \ldots, I_t$ are non-stochastic 2.4.3

and $$\frac{1}{T} \sum_{t=1}^{T} (I_t - \bar{I})^2 = 1. \qquad 2.4.4$$

The frequency distribution of the least squares estimator for finite samples of four and larger when T was even was:

$$f(b) = (-1)^{\frac{T-4}{2}} \; k \; \frac{4\sqrt{T}(1-B)}{\sigma Z^2 (1-b)^2} \; e^q \sum_{r=0}^{\frac{T-4}{2}} \frac{(\frac{T-4}{2})\,!\;(T-3-r)!}{r\,!\;(\frac{T-4-2r}{2})\,!}$$

$$(\frac{2}{Z})^{T-4-2} \quad (\frac{4\sqrt{T}}{\sigma Z} + 2)^r \quad (-1)^{T-4-r}$$

$$\{e^{-q} - \sum_{s=0}^{T-3-r} (\frac{-q}{s!})^s\} \qquad \text{for } 1 \neq b \neq \frac{B+1}{2} \qquad 2.4.5$$

where $$Z = \frac{\sqrt{T}\;(2b-B-1)}{(1-b)} \qquad 2.4.6$$

$$q = \frac{T(1+B-2b)(1-B)}{2\;\sigma^2 (1-b)^2} \qquad 2.4.7$$

and $$k = \{2^{\frac{T-1}{2}} \;\sqrt{\pi}\; \Gamma(\frac{T-2}{2}) \; e^{\frac{T}{2\sigma^2}}\}^{-1}. \qquad 2.4.8$$

The frequency distribution of the marginal propensity to consume for the maximum likelihood estimator was

$$\theta(b) = \frac{\sqrt{T}(1-B)}{\sqrt{2\pi}\;\sigma\;(1-b)^2} \; e^{\frac{-T(b-B)^2}{2\sigma^2(1-b)^2}} \qquad b \neq 1,\; T \geq 4 \text{ and even.} \qquad 2.4.9$$

From the above frequency distribution, a basic Keynesian model with no specification errors or identifying restrictions would be estimated by the maximum likelihood method for samples of size ten or more. For any sample size, it seemed doubtful that the ordinary least squares estimator could ever be deemed "better" than the maximum likelihood estimator.

In a more recent article, David H. Richardson (1968) studied the identical problem that Robert Basmann (1961; 1963) and Gerhard Kabe (1963; 1964) researched in the early sixties. Richardson adopted Kabe's approach, deriving the exact distribution of the GCL estimators from a system of equations based on the non-central Wishart distribution. The number of exogenous variables excluded from the equation to be estimated was two or more. The number of endogenous variables in the equation being estimated was two. The equation estimated was as follows:

$$y_1 = y_2 B_1^* + Z_1 \gamma_1^* + Z_2 \gamma_2^* + e_1 \qquad 2.5.1$$

where the assumptions were the same as those made in Basmann's two previous articles (1961; 1963).

Making substitutions in the non-central Wishart distribution, the author showed that the number of excluded exogenous variables, (n), appeared as a parameter in the density function of the GCL estimator of B_1^*.[6] The marginal density function of the transformed data for B_1^*, denoted by $h(v_1)$, was:

[6]Let K** represent the number of excluded exogenous variables in 2.5.1. n was also the sample size. To derive the following results, Richardson assumed n = K**. This was easily accomplished once the model was formulated. The loss due to the assumption was the loss of degrees of freedom. This could not happen when the model was small, but if the model was large, n = K** seemed quite feasible.

$$h(v_1) = \frac{1}{B(\frac{1}{2},\frac{n}{2})} \quad \frac{e^{-\frac{u}{2}(1+B_1^2)}}{(1+v_1^2)^{(n+1)/2}}$$

$$\sum_{j=0}^{\infty} \frac{(\frac{n+1}{2})_j \; (\frac{X^2}{2})^j}{(\frac{n}{2})_j \; j!} \qquad {}_1F_1 \left(\frac{n-1}{2} \; ; \; j + \frac{n}{2} \; ; \; 2z\right) \qquad 2.5.2$$

where $X^2 = \dfrac{u^2(1+B_1v_1)^2}{1 + v_1^2}$ 2.5.3

$$2z = \frac{u^2B_1^2}{2} \qquad 2.5.4$$

$$u^2 = \pi_{22}' \frac{S}{\sigma_{22}} \pi_{22} \qquad 2.5.5$$

B(a,b) was a Beta function,

and $S = Z_2'Z_2 - Z_2'Z_1(Z_1'Z_1)^{-1} Z_1'Z_2$. 2.5.6

Once the density function of the GCL estimators was derived, Richardson showed that the relative bias of v_1 was

$$\frac{E(v_1)-B_1}{B_1} = \frac{2}{n} g \left(\frac{u^2}{2}\right) e^{-\frac{u^2}{2}} -1. \qquad 2.5.7$$

Note as $\dfrac{u^2}{2} \to \infty$, $g(\dfrac{u^2}{2}) = \dfrac{n}{2}$; 2.5.8

therefore, $0 < g(\dfrac{u^2}{2}) \leq \dfrac{n}{2}$ 2.5.9

Richardson concluded that the sign of the bias was opposite to the sign of B_1 and that the size of the relative bias was between zero and minus one. Asymptotic expansion of 2.5.7 yielded $E(v_1)=B_1$. The biases were computed in the article for given u and n (p. 1222, Table 1, Richardson, 1968). Richardson, in expressing the second moment of the GCL estimators, showed that the existence of higher moments concurred with Basmann's (1959) hypothesis that moments exist to the order $K^{**} - G^{\Delta} + 1$. The final observation was that the estimates of the GCL estimators converged to the parameter B_1 when the sample size was fixed. This convergence was due to an indefinite increase of one of the parameters of the distribution $-v_1$, B_1 or u^2.

Richardson's study also indicated that the number of excluded exogenous variables appeared as a parameter in the density function of the GCL estimators.

Takamitsu Sawa (1969) derived results similar to Richardson's using equation 2.5.1. The difference in the two studies were:

1. The distribution of the two stage least squares estimator was derived as a corollary to the distribution of the ordinary least squares estimator. Thereby it was shown by Sawa that the distributions of the two estimators were similar in functional form.

2. In Richardson's paper, the variance-covariance matrix of the endogenous variables was assumed to be the identity matrix. Sawa assumed:

$$(y_{1t}y_{2t})'(y_{1t}y_{2t}) = \Sigma = \begin{bmatrix} \pi_{10} + \sum_{j=1}^{k} \pi_{1j}Z_{jt} \\ \pi_{20} + \sum_{j=1}^{k} \pi_{2j}Z_{jt} \end{bmatrix} \begin{bmatrix} \pi_{10} + \sum_{j=1}^{k} \pi_{1j}Z_{jt} \\ \pi_{20} + \sum_{j=1}^{k} \pi_{2j}Z_{jt} \end{bmatrix} \qquad 2.6.1$$

which emphasized the dependency in the difference between the values of the structural parameters and of the regression coefficients of the reduced form equations. The relationship was $B_{1j} = \pi_{2j}$, $j = 1,\ldots,k$.

3. A numerical evaluation for the comparison of the two estimators was shown by Sawa. The equation studied was

$$y_{2t} = \alpha + By_{1t} + u_t \qquad t = 1,\ldots,N. \qquad 2.6.2$$

The assumptions in the study were identical to the assumptions made by Richardson in his study (1968). The density function of the ordinary least squares estimator of B, $B \neq \rho$, was

$$f(\hat{B}) = C_2 \sum_{\alpha=0}^{\infty} \frac{1}{\alpha!\Gamma\left[\frac{N-1}{2} + \alpha\right]} \left[\frac{\tau^2(B-\rho)^2}{2\xi^2}\right]^{\alpha} \sum_{j=0}^{\alpha} \begin{bmatrix}\alpha \\ j\end{bmatrix}$$

$$\Gamma(\frac{N}{2} + \alpha - j - 1)\ \Gamma\ (\frac{N}{2} + j)$$

$$\left[\sigma(\hat{B}-\rho) + \frac{\xi^2}{\sigma^2(B-\rho)}\right]^{2j} \left[\frac{1}{\sigma^2(\hat{B}-\rho)^2 + \xi^2}\right]^{(N/2)+j}$$

$$-\infty \leq \hat{B} \leq \infty \qquad 2.6.3$$

where $$C_2 = \frac{\sigma\xi^{N-1}}{\sqrt{\pi}\ \Gamma(\frac{N}{2} - 1)} \exp\left[- \frac{\tau^2}{2\sigma^2}\{1 + \frac{\sigma^2(B-\rho)^2}{\xi^2}\}\right] \qquad 2.6.4$$

$$\xi^2 = \sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}} \qquad 2.6.5$$

$$\rho = \frac{\sigma_{12}}{\sigma_{11}} \qquad 2.6.6$$

$$\tau^2 = N \sum_{j=1}^{k} \pi_{1j}^2 \qquad 2.6.7$$

and $$\Sigma = \sigma_{ij} = E(v_{it}v_{jt}) , \sigma_{11} = \sigma^2 \qquad 2.6.8$$

Alternative forms were given for this density function in Sawa's article (pp. 929-30, eq. 3.18 and 3.24, 1969). When $B = \rho$, this density function was

$$f(\hat{B}) = \frac{\sigma e^{-\tau^2/2\sigma^2}}{\sqrt{\pi}\ \xi} \sum_{j=0}^{\infty} \left[\frac{\Gamma(\frac{N}{2} + j)}{j!(\frac{N-1}{2} + j)}\right] \left[\frac{\tau^2}{2\sigma^2}\right]^j \left[\frac{\xi^2}{\sigma^2(B-\hat{B})^2 + \xi^2}\right]^{\frac{N}{2}+j}$$

$$-\infty \leq \hat{B} \leq \infty \qquad 2.6.9$$

The ordinary least squares estimator possesses moments to the order N-2 and those of higher order do not exist. The distribution of the two stage least squares estimator can be derived directly from the distribution of ordinary least squares estimator by substituting K+1 for N in all distributions mentioned above. Moments existed to the order of K-1 for two stage least squares.[7]

By modifying equation 2.6.2 to

$$y_{2t} = \alpha + B\, y_{1t} + \sum_{j=1}^{K_1} \gamma_j\, z_{jt} + u_t \quad t=1,\ldots,N \qquad 2.6.10$$

Sawa studied a general equation which was identical to Richardson's equation 2.5.1. The density function for ordinary least squares and two stage least squares could have been derived from 2.6.3 by replacing N with $K_2 = K-K_1$ and letting

$$\tau^2 = \sum_{j=k_1+1}^{k} \pi^2_{ij}.$$

The two estimators' mathematical structures were too complex to deduce any definite conclusion involving small sample properties. Therefore, Sawa evaluated the density function numerically. The density function contained the following parameters: B, σ_{11}, σ_{22}, τ^2, N, K and σ_{12}. The first three parameters do not play an important role in determining the behavior of the estimator and therefore were given constant values of 0.6, 1.0 and 1.0, respectively. Based on past econometric models, $\frac{\tau^2}{N}$ was fixed at 4.0. K and N were varied by two's from two to ten and four to twenty, respectively. σ_{12} equaled 0.0, 0.2, 0.4, 0.8, and 0.5. The results were summarized as follows:

---

[7]Whenever the name of an estimator is made reference to without clarification, the reference refers to the method of estimation.

1. The distributions of both estimators were highly sensitive to changes in $\sigma_{12} = \rho$. As $(B-\rho)$ increased in magnitude, the bias of ordinary least squares increased as compared to two stage least squares.

2. As the sample size, (N), increased, both estimators' dispersions decreased. The estimates of two stage least squares were concentrated around the parameter while the estimates of ordinary least squares were concentrated around a different point, illustrating the well-documented property that two stage least squares is a consistent estimator.

3. As K decreased for a given sample size, the bias and dispersion of two stage least squares estimates decreased. If K was less than N, the estimates of ordinary least squares and two stage least squares were similar. In conclusion, the estimates of the two stage least squares estimator were at least as "good" as the estimates of the ordinary least squares estimator in all cases studied in this paper.

K. Takeuchi (1970) also studied the exact sampling moments of ordinary least squares, two stage least squares and instrumental variables estimators in the same model as Sawa (1969) and similar to Richardson's model (1968) under the same assumptions. The lower moments of ordinary least squares for equation 2.6.2 were

$$E(\hat{B}-B) = (\rho-B) \quad \{1-\eta \quad g_n \left(\frac{\eta}{2}\right)\} \qquad 2.7.1$$

and $E(\hat{B}-B)^2 = (\rho-B)^2\{1+ (\frac{\eta^2}{2} - \frac{5}{2}\eta)\ g_n(\frac{\eta}{2}) + (\frac{\eta}{2}) g_{n-2}(\frac{\eta}{2})$

$$-\frac{\eta}{2}\ g_{n+2}\ (\frac{\eta}{2})\} + \frac{\tau^2}{\sigma_1^2}\ g_{n-2}\ (\frac{\eta}{2}) \qquad 2.7.2$$

where the reduced form disturbance terms variance-covariance matrix was

$$\begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} \qquad 2.7.3$$

$$\rho = \frac{\sigma_{12}}{\sigma_1^2} \qquad 2.7.4$$

$$\tau^2 = \sigma_2^2 - \sigma_{12}^2\ /\sigma_1^2 \qquad 2.7.5$$

$$g_n(X) = e^{-X} \sum_{k=0}^{\infty} \frac{1}{(n+2k)k!}\ X^k \qquad 2.7.6$$

$$\eta = \sum_{i=1}^{n} \xi_i^2 / \sigma_1^2 \qquad 2.7.7$$

and ξ_i was a linear combination of the exogenous variables.

The sampling moments of the instrumental variables

$$X_{1i}, \ldots, X_{pi}, \quad \sum_{i=1}^{n} X_{ji}\ X_{Ki} = \begin{matrix} 1 & j = K \\ 0 & j \neq K \end{matrix}, \qquad 2.7.8$$

were the same as equations 2.7.1 and 2.7.2 where n is replaced by p and $\sum_{j=1}^{p} \zeta_j^2$ for $\sum_{i=1}^{n} \xi_i^2$

$$\zeta_j = \frac{E(Z_{2j})}{B} \qquad j=1,\ldots,p \qquad 2.7.9$$

$$Z_{2j} = \sum_{i=1}^{n} X_{ji}\, y_{2i} \qquad j=1,\ldots,p. \qquad 2.7.10$$

By making certain adjustments, it was shown that the instrumental variables estimates were identical to the ordinary least squares estimates for the regression coefficient of y_{2i} on y_{1i} for equation 2.6.2. When the instrumental variables were the same as the exogenous variables within the model, the distribution of the estimates of the instrumental variables estimator was identical to the distribution of the two stage least squares estimates. When the number of instrumental variables was greater than the number of exogenous variables but less than n and the instrumental variables were orthogonal to the exogenous variables, the estimator was $\hat{B}^*$ with $\Sigma\zeta^2 = \Sigma\xi^2 = \Phi^2$. By holding Φ^2 constant, varying p and observing equation 2.7.1 and 2.7.7, the bias of two stage least squares was established to be smaller than the bias of the ordinary least squares estimator, when the number of exogenous variables within the system was $K < p < n$.

Takeuchi derived an improved estimator due to his knowledge of the theoretical small sample bias of ordinary least squares and two stage least squares estimators.

$$\hat{\hat{B}} = \hat{B} + (n-2)\,\frac{\hat{\sigma}_1^2\,\hat{B} - \hat{\sigma}_{12}}{\Sigma y_{1i}^2} \qquad 2.7.11$$

$$\hat{\hat{B}}^* = \hat{B}^* + (p-2)\,\frac{\hat{\sigma}_1^2\,\hat{B}^* - \hat{\sigma}_{12}}{\Sigma \hat{y}_{1i}^2} \qquad 2.7.12$$

where $\hat{B}$ and $\hat{B}^*$ are ordinary least squares and two stage least squares estimators, respectively.[8]

When exogenous variables appeared in equation 2.6.2 and the instrumental variables (r) were orthogonal to themselves and the exogenous variables, the moments of the generalized instrumental variable were given by 2.7.1 and 2.7.2, where r replaced n and $\Sigma\eta_j^2$ replaced $\Sigma\xi_i^2$.

$$\sum_j^r \eta_j^2 = \sum_j^r \left[\sum_i^n X_{ji}\, E(y_{1i})\right]^2. \qquad 2.7.13$$

---

[8]Professor Takeuchi noted that if K is large, two stage least squares estimates may have a large bias and mean square error due to the possibility of correlation between the K exogenous variables. In this situation, Takeuchi suggested taking part of the exogenous variables as instrumental variables, and obtaining a better estimator than the two stage least square estimator.

$$\hat{B}^* = \sum_{j=1}^{r} \left\{ \sum_{i=1}^{n} (X_{ji} y_{1i}) \sum_{i=1}^{n} (X_{ji} y_{2i}) \right\} \Big/ \sum_{j=1}^{r} \left(\sum_{i=1}^{n} X_{ji} y_{1i} \right)^2$$

Following a similar procedure as outlined in 2.7.8 - 2.7.12 a generalized two stage least squares estimator was derived. The results of this paper were similar to those of Richardson's paper, except Richardson used a different approach.

Still another approach was taken by Koteswara Rao Kadiyala (1970), Walter Oi (1969), and Asatoshi Maeshiro (1966) in determining theoretical criteria for selecting the "best" estimator in simultaneous equations with finite samples. They developed similar mathematical relationships among k-class estimators. These relationships were helpful in choosing an appropriate estimator under given assumptions and in evaluating Monte Carlo studies.

Asatoshi Maeshiro (1966), studied relationships for the following equation

$$y_{0t} = B_1 y_{1t} + \gamma_1 X_{1t} + \ldots + \gamma_\lambda X_{\lambda t} + u_t$$

$$t=1,\ldots,n. \qquad 2.9.1$$

Given n observations, the calculation of $k^*>1$, b_1^* , c_i^*, h_{b_1} , and h_{c_1} $(i=1,\ldots,\lambda)$ were made such that b_1 and c_i were related to the value of k by the rectangular hyperbola with the following form:

$$(b_1 - b_1^*)\ (k - k^*) = h_{b_1} \qquad 2.9.2$$

and $$(c_i - c_i^*)\ (k - k^*) = h_{c_i} \qquad 2.9.3$$

where the coefficients of 2.9.2 and 2.9.3 were defined by Maeshiro (p. 371, eq. 12-16, 1966). The results were as follows:

1. Given 2.9.2 and 2.9.3, a set of estimates for any k could have been directly derived.

2. Except for k^*, each estimated coefficient was a monotonic increasing or decreasing function of k depending on whether h was negative or positive, respectively.

3. If k was near k^*, the estimated coefficients were highly unstable.

4. Estimates for two stage least squares estimators ($k=1$) were between ordinary least square ($k=0$) and limited information maximum likelihood estimates ($k>1$).

5. Estimates from limited information maximum likelihood estimators were less stable than estimates from ordinary least squares and two stage least squares estimators.

6. It was useful to calculate k^* before choosing an estimator because whenever k was close to one, two stage least squares estimates were unstable.

Walter Oi (1969) generalized equation 2.9.1 by introducing n dependent endogenous variables in the equation observed.

$$y_1 = X^* \alpha + Y_2 B + u \qquad 2.10.1$$

Oi showed that the two stage least squares estimates, denoted by subscripts of 1, were related to all k-class estimates by the following identities:

$$\alpha_k = \alpha_1 - \lambda_2 \, q_k(\hat{G} - B_1) \qquad 2.10.2$$

$$B_k = B_1 + q_k(\hat{G} - B_1) \qquad 2.10.3$$

and $$\alpha_k - \alpha_1 = -\lambda_2 (B_k - B_1) \qquad 2.10.4$$

where $$q_k = (1-k) \quad \{\hat{e}_2' \, \hat{e}_2 - k \, \hat{V}_2' \, \hat{V}_2\}^{-1} \, \hat{V}_2' \, \hat{V}_2 \qquad 2.10.5$$

$$\lambda_2 = (X^{*\prime}X^*)^{-1} X^{*\prime} \, Y_2 \qquad 2.10.6$$

$$\hat{e}_2' \, \hat{e}_2 = Y_2' \, Y_2 - Y_2' \, X^*\lambda_2 \qquad 2.10.7$$

$$\hat{V}_2 = Y_2 - X(X' \, X)^{-1} \, X'Y_2 \qquad 2.10.8$$

and $$\hat{G} = (\hat{V}_2' \, \hat{V}_2)^{-1} \, \hat{V}_2' \, (y_1 - X(X'X)^{-1} \, X'y_1). \qquad 2.10.9$$

The discrepancies between k-class estimators and two stage least squares estimator were determined by three parameters, G, λ_2, and q_k. Oi studied three conditions: (1) $q_k=0$, (2) $\hat{G}-B_1=0$ and (3) $\lambda_2=0$. If (1) and (2) hold, all estimates of the k-class estimators gave identical results. However, if only (3) holds, all k-class estimators gave identical estimates for the α parameters but not for the B parameters.

With no co-linearity between Y_2 and X^*, $\alpha_k=(X^{*\prime}X^*)^{-1} X^{*\prime}y_1$. When y_2 and X^* were correlated, $(X^{*\prime}X^*)^{-1} X^{*\prime}y_1=\alpha_k+\lambda_2 B_k$. When the difference between $\hat{G}$ and B_1 decreased, the dispersion between estimates from different k-class estimators diminished. The smaller dispersion implied that the linear association between systematic jointly dependent variables was approaching the linear association between the measured reduced form disturbance terms. Finally, q_k was dependent on k, $\hat{V}_2' \hat{V}_2$ and $\hat{e}_2' \hat{e}_2$ which implied that q_0 was a generalized variance ratio:

$$q_0 = (\hat{e}_2' \hat{e}_2)^{-1} \hat{V}_2' \hat{V}_2 \qquad 2.10.10$$

If $q_k=0$, then k=1. Furthermore, if q_0 was small and between zero and one, the different k-class estimators yielded parameter estimates within a small range. For $k > 1$, q_k was a negative definite matrix. Therefore $\{I-kq_0\}$ was a positive definite matrix.

Given the above information, ordinary least squares estimates and two stage least squares estimates bracketed the estimates from the k-class estimators for k between zero and one. When k was greater than one, ordinary least squares and k-class estimates bracketed the two stage least squares estimates.

The last study of this type was by Koteswara Kadiyala (1970). Kadiyala used an equation with assumptions similar to equation 2.10.1. He showed that the residual sum of squares was a monotonic increasing function of k for $0 \leq k < k^*$. Kadiyala's results were similar to Maeshiro's (1966) and Oi's (1969).

In conclusion, the theoretical contributions of Basmann (1961; 1963) and Nagar (1959; 1962) constituted important first steps in theorizing on the small sample properties of different estimators. Basmann's major contributions were: (1) solving for the exact finite sample frequency function of the GCL estimators; (2) questioning the existence of the lower moments of the GCL estimators; and (3) showing that the density function of the GCL estimators had a general form in two and three equation models. Nagar's major contributions were: (1) solving for the approximate distribution of the k-class and double k-class estimators; (2) showing that the optimal value of k of the k-class estimators was below one; and (3) showing the relationship among a family of estimators.

After Basmann's and Nagar's contributions, econometricians tried to clarify and continue the search of these two men. Gerhard Kabe (1963; 1964) expanded on Basmann's contributions by using the non-central Wishart distribution. His results concurred with Basmann's results. David Richardson (1968) carried on the work of Basmann and Kabe using the non-central Wishart distribution. Richardson's major contributions were: (1) showing that the sign of the bias of the GCL estimates was opposite the sign of the parameters and that the size of the relative bias was between zero and minus one, and (2) showing that the estimates of the GCL estimators could converge to the parameters when the sample size was held constant.

Sawa (1969) and Takeuchi (1970) made contributions similar to Basmann, Kabe, and Richardson. Sawa showed that the distribution of the two stage least squares estimator could be derived as a corollary to the distribution of the ordinary least squares estimator. Sawa also showed that ordinary and two stage least squares estimates were sensitive to changes in the parameters of their respective density functions. Takeuchi's results were similar to Sawa's results. Takeuchi's major contribution was to devise improved estimators from his theoretical findings (<u>supra</u>, pp. 28-29).

A. R. Bergstrom (1962) contribution was in solving for the exact sampling distributions of ordinary least squares and maximum likelihood estimators in a basic Keynesian model. Maeshiro (1966), Oi (1969), and Kadiyala (1970) contributed to the theoretical knowledge of the k-class estimators by developing mathematical relationships between the different estimators belonging to the k-class estimators (<u>supra</u>, pp. 30-34).

The authors of the preceding theoretical works conjectured about choosing the "best" estimator when the sample size was finite for simultaneous equations. These conjectures were usually limited to single equation estimators, two or three equation models and models with no specification errors under restricted assumptions. Due to these limitations, it seems important to review some of the major Monte Carlo studies in Chapter Three.

CHAPTER III

MAJOR MONTE CARLO CONTRIBUTIONS TO THE STUDY OF ESTIMATES FROM DIFFERENT ESTIMATORS IN SIMULTANEOUS EQUATIONS OF FINITE SAMPLE SIZES

Monte Carlo methods approach the study of the distribution of the estimates of different estimators by empirically generating a sufficient number of artificial finite samples to make conjectures about the distributions of these estimates. The theoretical methods discussed in Chapter II are more powerful but generally do not result in the determination of the distribution of the estimates of different estimators due to extremely complex mathematical expressions. Therefore, econometricians have turned to Monte Carlo methods for supplementary knowledge about the "best" estimator for simultaneous equations with finite samples. These studies are always suspect due to lack of generality. But as more Monte Carlo studies are conducted, one can hope that a pattern will be generated to give general results.

Guy Orcutt and Donald Cochrane were pioneers in Monte Carlo studies (1949A; 1949B). In their first study (1949A), they examined autocorrelation of the error terms. The general equation studied was

$$X_1 = a + b_{12.3t}X_2 + b_{13.2t}X_3 + b_{1t.23}t + u \qquad 3.1.1$$

where X_2, X_3 and u were independently constructed series possessing the same autoregressive scheme, t represented a linear time trend, a=0, $b_{12.3t} = 2$, $b_{12.2t} = 1$ and $b_{1t.23} = 0$. Therefore the actual generation of X could be denoted by equation 3.1.2.

$$X_1 = 2X_2 + X_3 + u. \qquad 3.1.2$$

Five series were used to generate the data for X_2 and X_3.

A. $X_{t+1} = X_t + 0.3(X_t - X_{t-1}) + e_{t+1}$ 3.1.3

B. $X_{t+1} = X_t + e_{t+1}$ 3.1.4

C. $X_{t+1} = 0.3X_t + e_{t+1}$ 3.1.5

D. $X_{t+1} = e_{t+1}$ 3.1.6

E. $X_{t+1} = e_{t+1} - e_t$ 3.1.7

where the e's were random disturbances with zero mean and the number of observations was usually twenty. Series D was random; C and B were first order autoregressive schemes; A was a second order autoregressive scheme; and E was the first difference of random series.

The results of this experiment when applying least squares showed that there was strong evidence of positive autocorrelation in the error terms in the formation of economic models. This autocorrelation in the error terms caused an underestimation of the variance of the disturbance terms. In addition, the variances of the regression coefficients were overestimated. And finally, inefficient predictions would be obtained if the error terms were autocorrelated.

A suggested treatment was to add an independent variable to bias the residuals toward randomness. A second treatment suggested was to bias the series of residuals toward randomness by an autoregressive transformation. Suggested techniques to bias the series of residuals toward randomness were presented within this study (p. 53-55, Orcutt and Cochrane, 1949A).

In their second study, Orcutt and Cochrane (1949B) added lagged variables to autocorrelated residual terms in a simultaneous equation model of finite sample size. The data was generated by the following recursive system of equations:

$$X_t = aY_t + v_{1t}$$

$$Y_t = bX_{t-1} + v_{2t} \qquad 3.1.8$$

where v_{1t} and v_{2t} were series of random errors.

Assuming $E(v_{2t}^2) = 0$, model 3.1.8 became

1. $X_t = aY_t + v_{1t}$

2. $Y_t = bX_{t-1}$

3. $X_t = \rho X_{t-1} + v_{1t}$. 3.1.9

Assuming $E(v_{1t}^2) = 0$, the system became

4. $X_t = aY_t$

5. $Y_t = bX_{t-1} + v_{2t}$

6. $X_t = \rho X_{t-1} + av_{2t}$. 3.1.10

Equations 1 and 4 were indicative of structural equations; equations 2 and 5 were indicative of reduced form equations; and equations 3 and 6 were the reduced form equations of 1 and 4, respectively. Model 3.1.9 had one specification error: lagged exogenous variables. Model 3.1.10 had two specification errors: lagged exogenous variables and autocorrelation of the disturbance terms which caused correlation between the independent variables and the error terms.

This research resulted in skepticism in using the method of least squares for estimating structural parameters when autocorrelation was present. When autocorrelation was present and the correct intercorrelation of the error terms was known, the skepticism associated with the use of the method of least squares was reduced due to an appropriate autoregressive transformation.

One can conclude about Orcutt and Cochrane's study that autocorrelation alone would not have led one to expect biased estimates. Lagged variables would result in negative biases for least squares estimates. In this study, when combining these two specification errors, the least squares estimates showed a substantial positive bias. "This result highlights once again a striking weakness of the current state of econometrics, in that the joint result of several complications cannot be inferred as the sum of their separate result." (p. 216, Johnston, 1963)

George W. Ladd (1956) studied a complete, overidentified, non-dynamic model.

$$1.\quad Y_1(t) = b_{12}Y_2(t) + C_{11}Z_1(t) + C_{12}Z_2(t) + C_{10} + u_1(t)$$

$$2.\quad Y_1(t) = b_{22}Y_2(t) + C_{23}Z_3(t) + C_{24}Z_4(t) + C_{20} + u_2(t). \qquad 3.2.1$$

The population parameters of the exogenous variables and the disturbance terms were specified a priori and assumed to be normally distributed. Given the structural parameters, the means and variance-covariance of the endogenous variables can be computed from the reduced form equations. From the above information, samples of size thirty were generated for the endogenous variables from the reduced form equations.[9] A

---

[9]Call this sample of size thirty data set one.

second data set was generated from the first data set. This second data set contained errors in observations for all exogenous and endogenous variables by adding to each observation in data set one a normally distributed error term with mean zero, given variance, which was serially independent and independent of each other. Thirty samples of this type were constructed and called data set two. In addition, two infinite samples were constructed corresponding to data set one and data set two, respectively.

The question asked by Ladd was "whether the presence of errors of observation biases L.I.S.E. (Limited Information Single Equation) estimates, or simply affects their random errors of sampling" (p. 488, 1956). Comparisons were made between the computed coefficients from the no errors in observations data to the means of the computed estimates in the errors in observations data. The results from the finite sampling experiments suggested that errors in observation affect the limited information single equation estimates' random errors of sampling. In two cases (b_{12} and c_{23}), the absolute difference of the means of the estimates of data set two from the structural parameters was larger than the difference between the coefficients of data set one and the structural parameters, while in four cases the opposite was true. A closer analysis showed that in the four cases cited above (c_{11}, c_{12}, c_{24} and b_{22}), forty-five of the one hundred and twenty estimates from data set two were closer to the parameters than the estimates from data set one. Only in thirteen of the sixty estimates of the initial two cases mentioned above were the limited information single equation estimates closer to the values of the parameters for data set two. In every case, the difference between estimates from data set one and the mean estimates from data set two were smaller than one standard error.

When the method of least squares was applied to equation one for both data sets, the mean estimates of data set two were consistently closer to the parameters than the least squares estimates from data set one. This concealed the fact that only forty of the ninety estimates from data set two for equation one were closer to the parameters as compared to the estimated coefficients from data set one. For equation two, only twenty-eight out of ninety estimates using data set two were closer to the parameter for ordinary least squares estimates. Other important observations in this Monte Carlo experiment were:

1. The limited information single equation estimates approached a normal distribution quite rapidly.

2. The estimated variances of the coefficients when using the limited information single equation estimator were overestimated and therefore understated the reliability of the estimated coefficients in this experiment when using this estimator.

3. The presence of errors in variables did not increase

the bias of least squares or limited information single equation estimates significantly, but did increase the standard error of the estimated coefficients.

4. The ordinary least squares estimator can safely be applied to structural equations when the correlations between the exogenous variables and the disturbance terms are small.

5. When the disturbance terms and the exogenous variables were highly correlated, the least squares estimates of the standard errors overestimated the true value of the standard errors.

Harvey M. Wagner (1958), studied the following three-equation model, with equation one being the only overidentified equation within the model.

$$
\begin{aligned}
&1.\quad Y_1 - B_1Y_2 && -\gamma_1 = u_1 \\
&2.\quad -B_2Y_2 + Y_3 + \gamma_2Z_1 && -\gamma_3 = u_2 \\
&3.\quad Y_1 - Y_2 + Y_3 + Z_2 && = 0. \qquad 3.3.1
\end{aligned}
$$

The disturbance terms in the above model were independently and identically normally distributed with zero mean and finite variance. The exogenous variable Z_1 was the lagged endogenous variable Y_2 such that $Z_1(t) = Y_2(t-1)$. Z_2 was a trend variable. Given a priori information about the specification of the parameters, a hundred samples of size twenty were generated for model I and model II. The difference between models was that the parameters of the variance-covariance matrix of model II were greater than those of model I. In addition, the covariance terms of model II were negative. The parameters of the variance-covariance matrix for both models were relatively small.

Wagner estimated only the first equation, which can be interpreted as an overidentified consumption function of the following form:

$$C = 0.25 + 0.5Y(t) + u_1(t) \qquad 3.3.2$$

The estimators used for B_1 and γ_1 were ordinary least squares, limited information single equation, and instrumental variables. The results showed that ordinary least squares and limited information single equation estimates have approximately the same mean square error for both model I and model II.[10] One exception to the above results occurred in model I where the mean square error of the ordinary least squares estimate for γ_1 was less than the mean square error of the limited information single equation estimate. This study showed, as expected, that the ordinary least squares estimates

---

[10]Note that the "good" performance of the mean square error of ordinary least squares may be due to the small values of the parameters of the variance-covariance matrix of model I and model II. Johnston indicated that the smaller the values of the variance-covariance matrix of the disturbance terms, the smaller the difference in estimates by different estimators (p. 287, 1963).

have a larger bias but a smaller variance than the limited information single equation estimates.

The performance of the estimates of the instrumental variables could not be generalized. In model II, the mean square errors of each estimator were approximately equal. In model I, the mean square errors of the estimates of the instrumental variables were smaller when using the variable Z_2 but larger when using the variable Z_1. For Y_1, in model I, the mean square error of the instrumental variables estimates was larger than the ordinary least squares estimates but smaller than the limited information single equation estimates.

In summary, this experiment showed that the estimates from all three estimators were relatively equal when judged by the mean square error loss function. The time trend variable was recommended to achieve the above results for the estimates of the instrumental variables.

William Neiswanger and Thomas Yancey (1959) studied a model which contained autonomous growth.[11] Autonomous growth of this type tends to cause correlation of the error terms with the exogenous variables. This type of autonomous growth is a specification error frequently observed in econometric models. A second experiment was conducted without including autonomous growth. This second study involved a model which

---

[11]Autonomous growth is a secular change in the endogenous variables which is unexplained by the parameters or exogenous variables of the structural equations.

contained stochastic exogenous variables.

Model I was identical to Ladd's (1956) model 3.2.1. A second model was constructed identical to 3.2.1. except an exogenous variable $Z_5(t)$ was added to each equation. This new variable represented a time trend. Data set I was generated from a priori information for a hundred and twenty samples of size twenty-five each. Data set II was generated by adding a growth trend to each error term and exogenous variable $u_i(t) + \delta_i(t-13)$, $Z_i(t) + \delta_i(t-13)$. The magnitude of each growth trend was as follows: $Z_1(t) = 4.00$, $Z_2(t) = 10.00$, $Z_3(t) = 4.00$, $Z_4(t) = 0.50$, $U_1(t) = 4.40$ and $U_2(t) = 3.40$. The trends in the error terms were characteristic of omitted variables in econometric models. This specification error increased the correlation between all variables within the model. Data set I contained multicollinearity. Data set II contained multicollinearity plus autonomous growth trends which caused correlation between the exogenous variables and the disturbance terms. These specification errors violated the assumptions underlying the application of the ordinary least squares and limited information single equation estimators.

This Monte Carlo study showed that when applying data set I to model I, limited information single equation estimates performed better than ordinary least squares estimates. In this case, the ordinary least squares estimates exhibited a larger bias but a smaller variance around the sample mean of the least squares estimates. The mean square error of the limited information single equation estimates was smaller than the ordinary least squares mean square error, but the difference was marginal. An interesting test conducted by the authors showed that the bias of the ordinary least squares estimates involved a high risk of incorrect inferences based on the estimated coefficients with respect to the parameters. When data set II was applied to model I, both estimators performed poorly. For this experiment, there was a high risk that the estimates of both estimators would yield incorrect inferences about the parameters, especially for the coefficients of the endogenous variables.

In a second experiment both data sets were applied to Model II. When data set II was applied to Model II, the ranking of the estimates of the estimators was similar to that of Model I, data set I. The estimates of both estimators improved as compared to applying data set II to Model I which did not contain a trend variable. Once again limited information single equation estimates had a smaller bias and a larger variance than ordinary least squares estimates. The mean square errors of the limited information single equation estimates were smaller than the ordinary least squares estimates of the mean square errors for Model II. When data set I was applied to Model II, there were no noticeable effects upon the performance of the estimates of either estimator.

The authors then applied trend free data to a model in which the exogenous variables were stochastic from sample to sample. The results were similar to the results of applying data set II to Model II.

The conclusions of the Monte Carlo study were:

1. Ordinary least squares estimates of the standard errors were always smaller on the average than the limited information single equation estimates of the standard errors, but the limited information single equation estimates had a smaller bias.

2. Ordinary least squares estimates performed poorly when the disturbance terms and the exogenous variables were correlated.

3. Multicollinearity did not seem to increase on the average the standard errors of the estimates of either estimator.

4. Other disparate rates of growth besides autonomous growth made little difference in the estimates of the coefficients and their standard errors on the average.

5. An indication that $Z_5(t)$ should be omitted was when the estimates of C_{15} and C_{25} were small relative to their respective standard errors. The authors recommended the use of the variable time in economic models in which time series are used. The addition of the variable time seemed to improve the estimates of each estimator, and on the average the estimates were not made worse.

A. L. Nagar (1960) used Wagner's (1958) model I and II, 3.3.1, with the omission of the constant terms. He applied four different k-class estimators to equation I and equation II using a hundred samples of size twenty each. The estimators were ordinary least squares, two stage least squares, Nagar's unbiased k-class and minimum second moment estimators.[12] The first equation was similar to 3.3.2.

$$Y_1(t) - \gamma_1 Y_2(t) = u_1(t). \qquad 3.4.1$$

Equation II was a simple investment function depicted in 3.4.2.

$$-Y_1(t) + \gamma_2 Y_2(t) + B_2 Z_1(t) - Z_2(t) = u_2(t). \qquad 3.4.2$$

Y_1 was current consumption; Y_2 was current investment; Y_3 was current income; $Z_1(t) = Y_2(t-1)$; and Z_2 was a trend variable. Equation I was overidentified, and equation II was just identified.

General conclusions were that ordinary least squares estimates usually had the largest bias, but the smallest variance around the sample mean of the least squares estimates. Two stage least squares estimates usually exhibited the smallest bias and best ranking among estimators for this experiment (except in Model II, equation II). This ranking was made

---

[12]Note that the last two estimators are unbiased when there are no lagged endogenous variables (Nagar, 1959). The model here has lagged endogenous variables.

according to increasing distance from the parameter value. Ordinary least squares (except for Model II, equation II) exhibited the largest second moment about the parameters. The last conclusion was that as exogenous variables were added to the equations, minimum second moment estimates received the best ranking.

A conjecture which may be of interest in this Monte Carlo experiment was that if either the size of the parameters in the variance-covariance matrix of the disturbance terms increased or the covariance parameters were negative or both, there was a significant improvement in ordinary least squares estimates as compared to two stage least squares estimates.

A classic study was accomplished by Robert Summers (1965) which entailed the study of alternative estimators with respect to forecasting precision, estimating parameters of the reduced form equations, and estimating parameters of the structural equations. The model under consideration was an overidentified two equation model.

$$1.\quad Y_{1t} + B_{12}Y_{2t} + \gamma_{11}Z_{1t} + \gamma_{10} = u_{1t}$$

$$2.\quad Y_{1t} + B_{22}Y_{2t} + \gamma_{23}Z_{3t} + \gamma_{24}Z_{4t} + \gamma_{20} = u_{2t}. \tag{3.5.1}$$

The disturbance terms were distributed as a bivariate normal distribution with zero mean and variance-covariance $= \begin{bmatrix} 400 & 200 \\ 200 & 400 \end{bmatrix}$.

The four estimators used in estimating the structural coefficients were full information maximum likelihood, limited information single equation, two stage least squares and ordinary least squares. Five different sets of parameters were used with two different data sets. Data set A had very little multicollinearity relative to data set B. This experiment was conducted for fifty samples of size twenty. An additional experiment was conducted for a sample of size forty. In addition, for two experiments, the data was generated by using equation I in 3.5.1 and equation 2A in 3.5.2.

$$\text{2A.}\quad Y_{1t} + B_{22}Y_{2t} + \gamma_{21}Z_{1t} + \gamma_{23}Z_{2t} + \gamma_{24}Z_{4t} + \gamma_{20} = u_{2t}. \tag{3.5.2}$$

The data was then applied to model 3.5.1. The loss function used to rank the different estimators was the root mean square error.

The results were as follows:

1. The minimum variance property of ordinary least squares estimates of structural coefficients was reinforced, the bias of ordinary least squares estimates was usually the largest among the estimates obtained from the four estimators.

2. A conjecture was that the root mean square error of consistent estimators was inversely proportional to the square root of the sample size.

3. With a low degree of multicollinearity, the full

information maximum likelihood method clearly gave the best estimates. Two stage least squares estimates were ranked second and ordinary least squares estimates were ranked last.

4. With a high degree of multicollinearity, two stage least squares estimates received the highest ranking, full information maximum likelihood estimates were ranked second, and ordinary least squares estimates were ranked third.

5. With specification errors that Summers introduced by equation 3.5.2, full information maximum likelihood estimates performed poorly. This would be expected because this was the only estimator in this study which considers all *a priori* restrictions. Two stage least squares estimates received the highest ranking and were judged less sensitive to specification errors. A surprising result was that the ordinary least squares estimates received the second best ranking.

6. The results of applying the Kolmogoroff-Smirnov test to the sample distribution of Studentized structural coefficient estimates to test for normality showed that limited information single equation and two stage least squares estimates fared well with regard to accepting the null hypotheses of normality.

From the above results, the recommendation of using two stage least squares was justified for estimating the coefficients of structural equations in this Monte Carlo study.

Richard E. Quandt (1965) presented the results of a sampling experiment on a four equation model. This study involved the comparison of the k-class estimators for alternative values of k with special emphasis given to ordinary least squares and two stage least squares estimates.

The two models used were from the following basic model:

1. $B_{11}Y_1 + B_{12}Y_2 + B_{13}Y_3 + B_{14}Y_4 + \gamma_{11}Z_1 + \gamma_{12}Z_2 + \gamma_{13}Z_3 = u_1$

2. $B_{21}Y_1 + B_{22}Y_2 + B_{23}Y_3 + B_{24}Y_4 + \gamma_{22}Z_2 + \gamma_{25}Z_5 = u_2$

3. $B_{31}Y_1 + B_{32}Y_2 + B_{33}Y_3 + B_{34}Y_4 + \gamma_{33}Z_3 + \gamma_{34}Z_4 + \gamma_{35}Z_5 = u_3$

4. $B_{41}Y_1 + B_{42}Y_2 + B_{43}Y_3 + B_{44}Y_4 + \gamma_{43}Z_3 + \gamma_{45}Z_5 + \gamma_{46}Z_6 = u_4$. 3.6.1

The above model was denoted as Model I. Model II was identical to model I except that a new exogenous variable, Z_7, was introduced into the second equation. Equation I was the only equation estimated. Equation I for model I was just identified and equation I for model II was overidentified. The exogenous variables were fixed for repeated sampling and did not include lagged values of the endogenous variables. The structural disturbance terms were jointly normally distributed with zero mean and covariance matrix Σ. There were one hundred samples of size twenty.

The experiment studied the effects of sparseness of the B matrix, sparseness of the Σ matrix and multicollinearity.[13] To accomplish this, the data sets were generated as follows:

1. Five alternative B matrices were used. These five matrices ranged from low sparseness to high sparseness with B_5 being a lower triangular matrix.

2. Two Σ matrices were used, one of which was perfectly sparse (Σ_2).

3. Two data sets were used. Data set II contained relatively less multicollinearity between exogenous variables than data set I. Six logical sets were constructed from the above generations: (B_5,Σ_1), (B_4,Σ_1), (B_3,Σ_1), (B_2,Σ_1), (B_1,Σ_1), and (B_1,Σ_2). Each logical set was run four times; each run consisted of a different data set and different model.

The results were as follows:[14]

1. For covariance Σ_1, the ordinary least squares standard deviations of the coefficients were smaller than two stage least squares standard deviations of the coefficients. However, the ordinary least squares bias of the estimates was

---

[13]Perfectly sparse refers to a matrix which has only diagonal elements.

[14]The ranking of estimators was based upon the following combination of loss functions: bias, sample standard deviation, coefficient of concentration, coefficient of decentralization (estimates which have the wrong sign), largest deviation of an estimate from the parameter, and the root mean square error.

greater than was the two stage least squares bias of the estimates. This tended to confirm the hypothesis that the tails of the distribution of two stage least squares estimates are substantially thicker than ordinary least squares distribution of the estimates. In addition two stage least squares estimates exhibit a greater density around the parameter than ordinary least squares estimates. Given the above results, two stage least squares performed better in this experiment.

2. When Σ_2 was used, the question of superiority of ordinary least squares and two stage least squares was unsettled in this experiment.

3. When multicollinearity was low, the estimates of both estimators were generally good, with two stage least squares estimates performing best. The performance of ordinary least squares estimates improved as compared to two stage least squares estimates when data set I was applied.

4. Sparseness and triangularity of the B matrix improved the estimates of both estimators.

5. Sparseness of the Σ matrices improved the estimates of both estimators. This was especially true for ordinary least squares estimates.

6. Generally, the estimates of different estimators performed better for the overidentified equation.

Quandt has shown in this Monte Carlo study that ordinary least squares estimates were not necessarily poorer than two stage least squares estimates, given all the results of the above models. He suggested that both estimators merit consideration when choosing the "best" estimator for a particular model. A guideline suggested was to calculate estimates from both estimators and if the difference of the two estimates is large, use ordinary least squares as the estimator. If the difference is small, use two stage least squares.[15]

John Cragg has done extensive work on Monte Carlo studies (1966, 1967 and 1968). In his first study, Cragg (1966) investigated the sensitivity of ordinary least squares, two stage least squares, Nagar's unbiased k-class estimator ($k = 1 + \frac{L-1}{T}$), (<u>supra</u>, p. 12, eq. 2.2.3), three stage least squares, limited information and full information maximum likelihood estimates, when errors of measurement in the exogenous variables, stochastic coefficients, heteroskedastic disturbances, or autocorrelation of the disturbance terms was present in an econometric model. Each equation in his three equation model was overidentified. The model was as follows in matrix notation:

$$BY_t = \Gamma Z_t + U_t \qquad 3.7.1$$

where U_t had a multivariate normal distribution with mean zero,

---

[15]The author did not give a criterion for distinguishing between large and small differences between estimates from different estimators.

and covariance

$$E(U_{it}\ U_{jm}) = \Sigma,\ t = m$$

$$E(U_{it}\ U_{jm}) = 0,\quad t \neq m;\ t,m = 1,\ldots,T \quad i,j = 1,2,3$$

$$|\Sigma| \neq 0.$$

Fifty samples of size twenty were generated for each experiment. Each experiment contained a specification error, except the initial experiment.

The results were as follows:

1. In the initial experiment without a specification error, the results suggested that full information maximum likelihood and three stage least squares estimates were superior to the second group which included Nagar's unbiased k-class estimates, limited information single equation and two stage least squares estimates. Ordinary least squares estimates exhibited the poorest performance in this experiment due to a large bias. This ranking was based upon the absolute deviation of the estimated coefficients from the true coefficients. The standard errors of the coefficients of the consistent estimators were reliable for making inferences about the true value of the structural coefficient. A second experiment which altered the parameters of the variance-covariance matrix of the error terms gave different results. When these para-

meters were reduced in magnitude, all the estimates except the ordinary least squares estimates were negligible (p. 141, Cragg, 1966). The inferiority of ordinary least squares estimates was not as pronounced in this second experiment.[16]

2. Errors of measurement in the exogenous variables were introduced into the initial experiment after the generation of the data. This was done four different times with the variance of the errors of measurement being one, four, sixteen, and sixty-four. The ranking of the estimates of different estimators resulted in identical ranking as depicted in the results of the initial experiment by Cragg. An interesting result was that the standard errors of the structural coefficients increased approximately in proportion to increases in the size of the standard deviation of the measurement errors. Structure II was used when the variance of the errors of measurement was four. The ranking of the estimator was the same as depicted in the initial experiment.

3. The stochastic coefficients were introduced by adding independent normal deviates to each coefficient before an observation was generated. For three different replications of this experiment, the stochastic additions were multiplied by various scale factors so that the variances of the additions were equal to four, sixteen, and sixty-four percent of the true coefficients. The ranking of the estimates from

---

[16]Call this second experiment, experiment on structure II.

different estimators changed drastically. Ordinary least squares estimates were ranked best with respect to absolute deviation of estimates from the true coefficient; limited information and full information maximum likelihood estimates received the lowest ranking. Two stage least squares and three stage least squares estimates were ranked second behind the ordinary least squares estimates. Stochastic coefficients affected the central tendency of the estimates of all estimators. There was a marked increase in the dispersion of the estimates. With respect to increases in dispersion, there seemed to be an approximate proportional relationship to increases in the variance of the addition to the standard deviations of the coefficients.

4. Heteroskedastic disturbance terms were introduced as a specification error in the following forms:

$$Y_{it} = r_{it} \left[1 + \left[\frac{\sum_{t=1}^{20} r_{it}^2}{t} \right]^{-\frac{1}{2}} w_{it} \right] \qquad 3.7.2$$

$$Y_{it} = r_{it} + \left[\frac{20 \sum_{K=1}^{7} Z_{Kt}}{\sum_{t=1}^{20} \sum_{K=1}^{7} Z_{Kt}} \right] w_{it} \qquad 3.7.3$$

where r_{it} was the value of Y_{it} when there were no disturbances and w_{it} was the value of the reduced form disturbance term when there was heteroskedasticity. Autocorrelation of the disturbance terms was introduced in the following forms:

1. $U_{it} = 0.6\ v_{it} + 0.4\ U_{i(t-1)}$ $\quad$ $i=1,\ t\neq 1$ $\quad$ 3.7.4

$U_{it} = v_{it}$ $\quad$ $i=1,\ t=1$

. $\quad$ $i=1$, all t

2. $U_{it} = 0.5\ v_{it} + 0.5\ U_{i(t-1)}$ $\quad$ $t\neq 1$, all i $\quad$ 3.7.5

$U_{it} = v_{it}$ $\quad$ $t=1$, all i

3. $U_{it} = 0.2\ v_{it} + 0.8\ U_{i(t-1)}$ $\quad$ $t\neq 1$, all i $\quad$ 3.7.6

$U_{it} = v_{it}$ $\quad$ $t=1$, all i

where v_{it} represented the structural disturbance term usually used. The results of these two specification errors on structure I did not alter the ranking given in the initial experiment. The standard errors of the estimates were relatively unaffected.

Overall, the only specification error that had any pronounced effect on the ranking of the estimates as depicted in the initial experiment was that of the stochastic coefficients.

In Cragg's second study (1967), the basic model was identical to 3.7.1. Five different B matrices were used for a given Σ matrix denoted as structure I, II,..., and structure V, respectively. Structures VI through VIII used the B matrix of structure I in conjunction with different Σ matrices. There were three different data sets available for all experiments. Theoretically each data set had no multicollinearity attached to the exogenous variables. When the experiment involving multicollinearity was investigated, data sets ranging from low to high multicollinearity were used. The results were as follows:

1. With no specification errors, the results were identical to Cragg's earlier work (1966).[17]

2. When different sets of disturbance terms were used, the ranking was the same as in the initial experiment.

3. When different exogenous data sets were introduced into the initial experiment which were similar but generated independently of the original data set of the exogenous variables, the ranking was identical to the initial experiment. An important conjecture about this sampling experiment was that "the precise results of sampling experiments depend on the exact set of exogenous data used" (Cragg, p. 103, 1967).

4. When observing the effects of different structural coefficients upon the estimates of different estimators, no conclusions were drawn about the performances relative to

---

[17]Denoted in 1, page 58 of this paper and referred to as the results of the initial experiment.

ranking. The results show that the estimates of each estimator were sensitive to changes in structural coefficients.

5. With different Σ matrices, the results showed great improvement in ordinary least squares estimates. The estimates of k-class estimators performed better than those of three stage least squares and full information maximum likelihood estimates. These results were probably due to the fact that the changes in the Σ matrices reduced the elements of the off-diagonal terms, i.e., reduced contemporaneous correlation.

6. The size of the disturbance terms of the reduced form equations changed the results of the ranking of estimates from different estimators as given in the initial experiment. As the size of the disturbance terms increased, the dispersion of the estimates of each method increased. The standard errors of the estimates were approximately proportional to the difference in the size of the disturbance terms. With respect to the above results, limited information and full information maximum likelihood estimates deteriorated the most. The bias of the estimates from Nagar's unbiased estimator, two stage least squares and three stage least squares estimators increased as the size of the disturbance terms increased. At first, the bias of the limited information and full information maximum likelihood estimates decreased but the bias eventually increased as the size of the disturbance terms increased.

7. When the sample size was increased, the dispersion

of the estimates from each estimator decreased about the parameter except for ordinary least squares.

A general ranking of estimates from different estimators gave a marginal preference for three stage least squares and full information maximum likelihood estimates. The remaining estimates, except for ordinary least squares estimates, were ranked second. Ordinary least squares estimates performed poorly. Generally, the central tendency of the estimates from the consistent estimators behaved well, with full information and limited information maximum likelihood estimates giving superior results. With respect to the above results concerning central tendency, the presence of large disturbance terms or multicollinearity changed the conclusions. Standard errors of the estimates from consistent estimators usually gave reliable inferences. Standard errors of ordinary least squares estimates were not useful in making inferences about the parameters. Given the results of this Monte Carlo study, the two stage least squares estimator was chosen from among the consistent estimators because of its low cost and ease of application.

Cragg's final article (1968) studied the specification error of incorrect exclusion of variables from certain equations or the exclusion of entire equations from a model, the estimators and models examined being identical to those of his earlier study (p. 137-38, 1966). The summary of the general results of structure I was given in Cragg's earlier article (<u>supra</u>, footnote 17, p. 62).

The specification that coefficients were zero when in reality they were almost zero had little effect upon the ranking of the estimates of different estimators as given in his initial experiment (supra, p. 49). Ignoring an unimportant exogenous variable or omitting a structural equation affected the estimates of different estimators only slightly and did not alter their previous ranking. The above results suggested that econometric models built on approximate or partial knowledge give useful information.

Contrary results were presented when the specification error designated a non-zero coefficient to be zero when the coefficient had important economic significance. Full information maximum likelihood and three stage least squares estimates deteriorated in performance. The estimates of the single equation consistent estimators ranked as the best estimates under this specification error with respect to central tendency. The standard errors of the estimates of all estimators were useless in making inferences about the parameter with this type of specification error present. Another specification error studied was that of estimating the coefficients when the true values of these coefficients were zero. This specification error seriously affected the central tendency and dispersion of the estimates of different estimators. Many times, a specification error of the above type is introduced in econometric models so that identification can be achieved.

The results of this Monte Carlo experiment emphasized the importance of a priori knowledge about coefficients in structural equations which are close to or equal to zero and the importance of these variables in economic theory. This illustrates the econometrician's problem of misspecification versus underspecification and the econometric consequences.

Jan Kmenta and Roy Gilbert (1968) studied a model which was already in the reduced form. Though this is not the subject being discussed here, it is interesting to see the effects of specification errors on the reduced form coefficients. It must be recalled that two stage and three stage least squares estimates use information from the reduced form equations in the initial stages.

Four experiments were conducted. Experiments 1 and 2 were conducted on the following model:

$$Y_1(t) = 10 + 2X_{11}(t) - 5X_{12}(t) + u_1(t)$$

$$Y_2(t) = -10 + 6X_{21}(t) + 3X_{22}(t) + u_2(t). \qquad 3.8.1$$

The difference between experiments 1 and 2 was that in experiment 2 there was an increase in the pairwise correlation between explanatory variables. Experiment 3 added the following two equations to experiment 1 and 2.

$$Y_3(t) = 10 + 2X_{31}(t) - 5X_{32}(t) + u_3(t)$$

$$Y_4(t) = -10 + 6X_{41}(t) + 3X_{42}(t) + u_4(t). \qquad 3.8.2$$

For experiment 3, the values of the explanatory variables for equations 3.8.1 were the same as experiment 1 and the explanatory values for 3.8.2. were the same as experiment 2. The purpose of experiment 3 was to examine the effects of increasing the number of equations in a system of equations which contained a high pairwise correlation between explanatory variables. Experiment 4 studied the effects on a two equation model which had lagged values of the dependent variable as one of the independent variables.

$$Y_1(t) = -10 + 6X_{11}(t) + 0.25Y_1(t-1) + u_1(t)$$

$$Y_2(t) = 10 + 2X_{21}(t) + 0.75Y_2(t-1) + u_2(t). \quad 3.8.3$$

Nine different data-models were generated. Each data model consisted of a hundred samples of size ten, twenty, and one hundred. Each data-model set was applied to each experiment. Data-models A, B, and D were identical except for the degree of correlation of the disturbance terms across equations. Data-model A had high correlation, data-model B had moderate correlation, and data-model D had no correlation of disturbance terms across equations. Data-model C was the same as data-model B except that the variance of $u_2(t)$ had been increased greatly. Data-models E and F had heteroskedasticity built into the error terms in the following form:

$$\text{Var}\ \{u_m(t)\} = \{\frac{1}{10}\ E\ \{Y_m(t)\}\}^2. \quad 3.8.4$$

Data-model E had moderate correlation of disturbance terms across equations while data-model F had none. Data-models G, F, and I had an autocorrelated scheme of the following form:

$$u_1(t) = 0.8\ u_1(t-1) + 0.6\ v_1(t)$$

$$u_2(t) = 0.8\ u_2(t-1) + \lambda v_1(t) + \mu v_2(t) \qquad 3.8.5$$

where $v_1(t) \sim NID(0,1),\ v_2(t) \sim NID(0,1)$,

$$E\{v_1(t)\ v_2(t)\} = 0,\ Var\ \{u_m(t)\} = 1. \qquad 3.8.6$$

$$\text{Model G: } \lambda = 0.555,\quad \mu = 0.228\ . \qquad 3.8.7$$

$$\text{Model H: } \lambda = 0.360,\quad \mu = 0.480\ . \qquad 3.8.8$$

$$\text{Model I: } \lambda = 0\quad ,\quad \mu = 0.600\ . \qquad 3.8.9$$

All data-models had the same autoregressive coefficients of 0.8 but were different in the correlation of the disturbance terms across equations. Data-model G contained high, data-model H contained moderate, and data-model I contained no contemporaneous correlation.

The estimators considered were ordinary least squares, (OLS), Zellner's two stage Aitken's (ZEF), Zellner's iterative Aitken's (IZEF), Telser's iterative (TIE), and maximum likelihood (ML). In this sampling experiment ML, IZEF, and TIE estimates gave identical results; therefore, ML estimates were used to measure the performance of this group. The results were as follows:

1. Theoretical studies have been completed using large sample sizes that coincided with the results of this sampling experiment, when the size of the sample was one hundred. The unbiasness of ordinary least squares estimates for Experiments 1, 2, 3, and 4 with data-models A and B coincided with the theoretical results. The ZEF and ML estimates theoretically should have given the author identical results and they did.

2. ZEF estimates were superior to OLS estimates except when there was no correlation of the disturbance terms across equations. In the latter case, OLS estimates were only marginally better than ZEF estimates for all experiments and models.

3. ZEF and ML estimates had equal efficiency for large samples for all experiments and data-models. In experiment I, data-model A, maximum likelihood estimates were more efficient than ZEF estimates. This result also held true for all experiments when data-models H and G were used. In all other cases, ZEF estimates showed greater efficiency than ML estimates.

4. When data-model C was used the relative performance of the estimates from all estimators was not affected.

Overall results for these experiments suggested that ZEF estimates were preferable in most situations. This conclusion was drawn due to the overall superiority of the efficiency and ease of application of the ZEF estimator.

Recently, Potluri Rao and Zvi Griliches (1969) studied the effects of autocorrelation on the following model:

$$y_y = Bx_t + u_t$$

$$x_t = \lambda x_{t-1} + v_t \qquad t = 1,\dots,T$$

$$u_t = \rho u_{t-1} + w_t. \qquad 3.9.1$$

Where

$$E(v_t) = E(w_t) = E(v_t w_t) = E(w_t w_{t-1}) = E(v_t v_{t-1}) = 0 \qquad 3.9.2$$

$$E(v_t^2) = \sigma_v^2,\ E(w_t^2) = \sigma_w^2,\quad |\lambda| < 1,\ |\rho| < 1. \qquad 3.9.3$$

The initial values of u and x were derived from a normal population with mean zero and variance $\sigma_v^2 = \sigma_w^2/(1-\rho^2)$ and $\sigma_x^2 = \sigma_v^2/(1-\lambda^2)$, respectively. Six methods were used to estimate B (3.9.1). They were generalized least squares (GLS) with ρ known and ordinary least squares (OLS) with ρ unknown. Four other estimators were used in an attempt to improve on the ordinary least squares estimates. They were the Cochrane and Orcutt estimates (p. 53-55, 1949A) where

$$\hat{\rho} = \sum_{t=2}^{T} e_t e_{t-1} \Big/ \sum_{t=1}^{T} e_t\ ,$$

the Durbin estimates (p. 256-57, Roa and Griliches, 1969) of

$\hat{\rho}$ from the following equation:

$$y_t = \rho y_{t-1} + Bx_t - B\rho x_{t-1} + w_t \qquad t=2,\ldots,T \qquad 3.9.5$$

$$\text{where} \quad \hat{\rho} = \sum_{t=2}^{T} y_t y_{t-1} \Big/ \sum_{2}^{T} y_t^2 \,, \qquad 3.9.6$$

the Prais-Winsten (p. 256-57, Roa and Griliches, 1969) estimates which solve for $\hat{\rho}$ the same as the Cochran and Orcutt method but transforms the initial value of the original data by $\sqrt{1-\rho^2}$, and a nonlinear estimate identical to 3.9.5 except that $\widehat{B\rho} = \hat{B}\,\hat{\rho}$. For this last estimator, the criterion was to find the absolute minimum of the residual sum of squares with respect to B and ρ.

Except for the nonlinear method, the last four estimators required two stages to estimate B. In the first stage, ρ was estimated, and from the estimate of ρ the estimate of B was made. For the estimate of ρ, when ρ was positive, the Durbin estimates showed significantly smaller bias than the Cochrane-Orcutt or nonlinear estimates of ρ. For small negative values of ρ, the Cochrane-Orcutt estimates showed the least bias while for large negative values of ρ, the nonlinear method showed the least bias. It is noteworthy that the superiority of the estimates of the above two estimators as compared with the Durbin estimates of ρ, when ρ was negative, was slight. In addition, most economic data generally does have a positive autocorrelated scheme. With respect to the mean square error of the above estimates of ρ, the results were similar to those given above.

In estimating B, ordinary least squares estimates were inefficient as compared to all other estimates produced by the other estimators. The estimates of the nonlinear estimator were somewhat inferior to two stage estimates for ρ greater than zero. The Cochrane-Orcutt method was slightly inferior to the other two stage methods of estimating B. Therefore, when ρ is thought to be larger than 0.3 in absolute value, either the Durbin or Prais-Winsten method should be applied. Even when ρ is less than 0.3 in absolute value, little is lost by using the estimates of these two stage estimators. The authors suggested that the best method was to use the Prais-Winsten method in the first stage, estimating ρ, and the Durbin estimates in the second stage in calculating B.

Recently, there has been a trend toward investigating alternative methods of estimating small samples in simultaneous equations. Fred Glahe and Jerry Hunt (1970) studied the performance of least absolute ordinary least squares and least absolute two stage least squares estimates versus ordinary least squares and two stage least squares estimates.[18] The model selected was:

---

[18]Least absolute estimators minimize the absolute values of the residuals where ordinary least squares and two stage least squares minimize the squares of the residuals.

$$1.\quad Y_1 + B_{12}Y_2 + \gamma_{11}Z_1 + \gamma_{12}Z_2 + \gamma_{10} = u_1$$

$$2.\quad Y_1 + B_{22}Y_2 + \gamma_{23}Z_3 + \gamma_{24}Z_4 + \gamma_{20} = u_2 \qquad 3.10.1$$

where $u_i \sim N(0,100)$, $i=1,2$.

Sample sizes of ten and twenty were used for four experiments. Experiment I contained no specification errors. Experiment 2 contained a degree of multicollinearity common to most economic data. Experiment 3 introduced the following heteroskedasticity scheme:

$$\sigma_u = (\sigma_0^2 + i)^2 \qquad 3.10.2$$

where $\sigma_0^2 = 5$ and $i = 0,\ldots,N-1$. Experiment 4 investigated the specification error of a variable added to equation 2, 3.10.1, when the true value of the parameter of this new coefficient was zero. The results showed that for structural estimations, ordinary least squares and two stage least squares estimates performed better than the estimates from either of the two least absolute estimators for samples of size ten and twenty for all four experiments. The only exception to the above statement occurred when the loss function, mean absolute error, was used for samples of size ten for all four experiments. In this case, two stage least absolute estimates were better than the estimates from the other estimators. Kendall's W in this experiment was usually not significant at the .01 level; therefore, the differences of the two stage least absolute estimates were slight as compared to the estimates from other estimators for samples of size ten given this loss function. This study showed that continued use of the squared deviations of the residuals as a criterion in developing estimators for simultaneous equations with small sample sizes was a valid choice.

In summary, each Monte Carlo study has made contributions to the understanding of the estimators used in estimating coefficients in structural equations when the sample size is finite. Though each Monte Carlo experiment was unique in that the models used were never identical, a few generalizations and conclusions seem to follow from one Monte Carlo study to other Monte Carlo studies. They are as follows:

1. When autocorrelation was present, the ordinary least squares estimates (OLS) performed poorly (Cochrane and Orcutt, 1949A, 1949B; Cragg, 1966; Kmenta and Gilbert, 1968; Roa and Griliches, 1969).

2. The addition of the exogenous variable time (t) improved the performance of different estimators (Neiswanger and Yancey, 1959; Wagner, 1958).

3. The estimates of the OLS estimator exhibited a larger bias but a smaller variance than other estimates from different estimators in different Monte Carlo studies (Wagner, 1958; Nagar, 1960; Summers, 1965; Quandt, 1965; Cragg, 1966; Kmenta and Gilbert, 1967; Roa and Griliches, 1969). The major exception to the above property was found in Quandt (1965)

when multicollinearity was present. Summers (1965), Cragg (1966), and Kmenta and Gilbert (1968) contradicted Quandt's findings and reinforced the above property. The results indicated that multicollinearity has an adverse effect on the bias and standard error of the estimates from different estimators but this adverse effect was less for consistent estimators.

4. When errors in observation were analyzed in Monte Carlo experiments, the results were that the bias of OLS and limited information single equation (LISE) estimates did not increase but the standard errors of the estimated coefficients did (Ladd, 1956). Cragg's (1966) results were similar to Ladd's results. Cragg also ranked the estimators with three stage least squares (3SLS) and full information maximum likelihood (FIML) receiving the highest rank and the consistent single equation estimators being ranked second.

5. When lagged exogenous variables were present in a simultaneous equation model (Wagner, 1958), the results showed that OLS, instrumental variables (IV), and LISE estimates were similar. This Monte Carlo experiment recommended two stage least squares estimates (TSLS) with lagged exogenous variables.

6. Ladd's study (1956) showed that the OLS estimator can be used when the disturbance terms and the exogenous variables were slightly correlated. Ladd also showed that when the disturbance terms and exogenous variables were highly

correlated, OLS estimates of the standard errors overestimated the true values of the standard errors. Neiswanger and Yancey (1959) substantiated Ladd's findings.

7. Neiswanger and Yancey (1959) observed that the LISE estimates performed better than the OLS estimates when the exogenous variables were stochastic. Cragg (1967) substantiated the above observation but ranked FI and 3SLS estimates ahead of the estimates of the consistent single equation estimators.

8. Cragg (1967) ranked FI and 3SLS estimators first when the specification error was heteroskedasticity. Kmenta and Gilbert (1968) verified Cragg's findings.

9. Summers (1965) concluded that the 2SLS estimator performed best with an incorrect specification of a model. Cragg's (1968) Monte Carlo study verified Summer's findings.

10. Many econometricians have analyzed the effects of different B matrices on different estimators. Quandt (1965) showed that OLS and 2SLS estimates improved when the B matrix was sparse. Cragg's (1966) Monte Carlo study showed that stochastic coefficients caused the OLS estimates to be superior to the estimates from the consistent estimators. Cragg (1967) showed that the estimates of different estimators were sensitive to changes in the structural coefficients.

11. As the size of the parameters in the variance-covariance matrix (Σ) of the disturbance terms increased, became negative or both, OLS estimates improved (Nagar, 1960).

Quandt (1965) showed that as the Σ matrix became sparse, OLS estimates improved. Cragg (1967) and Kmenta and Gilbert (1968) reinforced the conclusions of Nagar and Quandt. Cragg (1967) also showed that as the size of the disturbance terms increased, the performance of all the estimators in his Monte Carlo study deteriorated in performance with respect to bias and efficiency.

CHAPTER IV

A MONTE CARLO EXPERIMENT WITH TWO SPECIFICATION ERRORS IN A TWO EQUATION SIMULTANEOUS MODEL

The two preceding chapters reviewed the major theoretical and Monte Carlo studies in estimating structural parameters in a system of simultaneous equations when the samples were of finite size. Numerous Monte Carlo studies have been attempted observing only one specification error within the model. Even when more than one specification error has been observed, the purpose of the experiment has not generally been to isolate the effects of each specification error upon the accuracy of each estimate. The purpose of this chapter is to attempt to ascertain the effects of each specification error on the performance of the estimates from different estimators. The specification errors are autocorrelation of the disturbance terms and multicollinearity of the exogenous variables.[19] These two specification errors were chosen because they are inherent in many economic models.

---

[19]The assummption of the autocorrelated disturbance terms considers both serial and contemporaneous correlation.

The model chosen is the following:[20]

$$Y_1(t) = B_{12}Y_2(t) + C_{11}Z_1(t) + C_{12}Z_2(t) + C_{10} + U_1(t)$$

$$Y_1(t) = B_{22}Y_2(t) + C_{23}Z_3(t) + C_{24}Z_4(t) + C_{20} + U_2(t). \qquad 3.1$$

In model 3.1, the first equation is a demand equation and the second equation is a supply equation in partial equilibrium. Both equations are overidentified. The Y variables are endogenous variables, the Z variables are exogenous variables and the U variables are error terms. The parameter values are specified a priori. The independent observations of the Z and U variables are generated from random normal deviates. The exogenous variables are stochastic because they are, in reality, endogenous variables or some larger system (p. 392, Neiswanger and Yancey, 1959). Fifty samples of size twenty each are generated for each case.[21] Corresponding to each sample, the disturbance terms are generated consistent with a positive first order autocorrelated scheme. The exogenous variables have varying degrees of multicollinearity.

---

[20]The above model was chosen because it is indicative of economic models but small enough to be handled within time and cost constraints (Summers, 1965; Neiswanger and Yancey, 1959)

[21]The size of the samples, twenty, and the number of samples in one experiment, fifty, were chosen arbitrarily. Both choices are deemed adequate in conducting Monte Carlo studies (Summers, 1965; Roa and Griliches, 1969; Glahe and Hunt, 1970).

The exogenous variables have the following means and the following variance covariance matrices:

$$Z_1 = 478 \qquad Z_2 = 381 \qquad Z_3 = 478 \qquad Z_4 = 43 \tag{3.2}$$

Low Multicollinearity Correlation Matrix of the Exogenous Variables

High Multicollinearity Correlation Matrix of the Exogenous Variables

$$\begin{array}{c} \\ Z_1 \\ Z \\ Z \\ Z \end{array} \begin{array}{c} \begin{array}{cccc} Z_1 & Z_2 & Z_3 & Z_4 \end{array} \\ \left[\begin{array}{cccc} 1 & .078 & .016 & .38 \\ & 1 & .017 & -.057 \\ & & 1 & .31 \\ & & & 1 \end{array}\right] \end{array} \qquad \begin{array}{c} \\ Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \end{array} \begin{array}{c} \begin{array}{cccc} Z_1 & Z_2 & Z_3 & Z_4 \end{array} \\ \left[\begin{array}{cccc} 1 & -.37 & .76 & .80 \\ & 1 & -.52 & -.43 \\ & & 1 & .83 \\ & & & 1 \end{array}\right] \end{array} . \tag{3.3}$$

The structural coefficients are as follows:[22]

$$\begin{aligned} B_{12} &= -0.4 \\ B_{22} &= 0.6 \\ C_{11} &= 0.1 \\ C_{12} &= 0.45 \\ C_{23} &= 0.25 \\ C_{24} &= 0.8 \\ C_{10} &= 100.0 \\ C_{20} &= 50.0 \, . \end{aligned} \tag{3.4}$$

[22]The structural coefficients are identical to the coefficients used by Neiswanger and Yancey (p. 391, 1959).

The disturbance terms and the exogenous variables are independently generated; therefore, the study of the effects of each specification error on estimates from each estimator is achieved by using different data sets with different degrees of multicollinearity and autocorrelation.[23]

For example, data set one has a given autocorrelation scheme of P_1 and varying degrees of multicollinearity--low and high.[24] Data set two has an autocorrelation scheme P_2, where P_2 is greater than P_1, and maintains the same low and high multicollinearity as each respective sub data set of data set one. The autocorrelation scheme is varied to include the autoregressive parameters 0.2, 0.6, 0.8, and 0.9. Four data sets are generated with two sub-sets each and are denoted as S_{1L}, and S_{1H}, S_{2L} and S_{2H}, ... , and S_{4H}, in which the first subscript is for the autoregressive parameter and the second subscript is for the degree of multicollinearity.[25]

---

[23]This is true if the autocorrelated disturbance terms are due to omitted exogenous variables. If autocorrelation is present and lagged endogenous variables are used as exogenous variables in the same equation, the assumption of independence does not hold true.

[24]Exogenous data sets having low and high multicollinearity will be denoted by L and H, respectively.

[25]Hobson Thornber (1967) criticized past Monte Carlo studies for concentrating attention to one or two points in a parameter space when the risk function should be evaluated over the whole parameter space. In this Monte Carlo study, eight parameter points are observed because an arbitrary choice had to be made with respect to the constraints of time and cost (Thornber, p. 809, 1967).

The variance-covariance matrix and means of the exogenous variables are shown in the following form for generating the data with the desired level of multicollinearity.

$$\begin{bmatrix} Z_1(1) & \cdots & Z_1(20) \\ Z_2(1) & \cdots & Z_2(20) \\ Z_3(1) & \cdots & Z_3(20) \\ Z_4(1) & \cdots & Z_4(20) \end{bmatrix} = \begin{bmatrix} M_{11} & 0 & 0 & 0 \\ M_{21} & M_{22} & 0 & 0 \\ M_{31} & M_{32} & M_{33} & 0 \\ M_{41} & M_{41} & M_{43} & M_{44} \end{bmatrix} \begin{bmatrix} S_1(1) & \cdots & S_1(20) \\ S_2(1) & \cdots & S_2(20) \\ S_3(1) & \cdots & S_3(20) \\ S_4(1) & \cdots & S_4(20) \end{bmatrix} + \begin{bmatrix} M_1 \\ M_2 \\ M_3 \\ M_4 \end{bmatrix}. \quad 3.5$$

The M matrix is a lower triangular matrix which is related to the population variance-covariance matrix of the exogenous variables such that MM'= population variance-covariance matrix. The S variables are standardized normal deviates. M_1, M_2, M_3 and M_4 are the means of Z_1, Z_2, Z_3 and Z_4, respectively.

To derive the M matrix, let Σ be a symmetric, positive definite matrix of rank four. Then there exists a lower triangular matrix M such that $\Sigma = M M'$ (p. 3, Graybill, 1961).

$$\begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} & \Sigma_{14} \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{23} & \Sigma_{24} \\ \Sigma_{31} & \Sigma_{32} & \Sigma_{33} & \Sigma_{34} \\ \Sigma_{41} & \Sigma_{42} & \Sigma_{43} & \Sigma_{44} \end{bmatrix} = \begin{bmatrix} M_{11} & 0 & 0 & 0 \\ M_{21} & M_{22} & 0 & 0 \\ M_{31} & M_{32} & M_{33} & 0 \\ M_{41} & M_{42} & M_{43} & M_{44} \end{bmatrix} \begin{bmatrix} M_{11} & M_{21} & M_{31} & M_{41} \\ 0 & M_{22} & M_{32} & M_{42} \\ 0 & 0 & M_{33} & M_{43} \\ 0 & 0 & 0 & M_{44} \end{bmatrix}. \quad 3.6$$

$$\Sigma_{11} = M_{11}^2 \qquad \Sigma_{22} = M_{21}^2 + M_{22}^2$$

$$\Sigma_{33} = M_{31}^2 + M_{32}^2 + M_{33}^2 \qquad \Sigma_{12} = M_{11}M_{21}$$

$$\Sigma_{23} = M_{21}M_{31} + M_{22}M_{32} \qquad \Sigma_{34} = M_{31}M_{41} + M_{32}M_{42} + M_{33}M_{43}$$

$$\Sigma_{13} = M_{11}M_{31} \qquad \Sigma_{24} = M_{21}M_{41} + M_{22}M_{42}$$

$$\Sigma_{44} = M_{41}^2 + M_{42}^2 + M_{43}^2 + M_{44}^2 \qquad \Sigma_{14} = M_{11}M_{41}. \tag{3.7}$$

$$M_{11} = \sqrt{\Sigma_{11}} \qquad M_{22} = \sqrt{\Sigma_{22} + M_{21}^2}$$

$$M_{33} = \sqrt{\Sigma_{33} - M_{31}^2 - M_{32}^2} \qquad M_{21} = \Sigma_{12}/M_{11}$$

$$M_{32} = (\Sigma_{23} - M_{21} - M_{31})/M_{22} \qquad M_{43} = (\Sigma_{34} - M_{31}M_{41} - M_{32}M_{42})/M_{33}$$

$$M_{31} = \Sigma_{13}/M_{11} \qquad M_{42} = (\Sigma_{24} - M_{21}M_{41})/M_{22}$$

$$M_{44} = \sqrt{\Sigma_{44} - M_{41}^2 - M_{42}^2 - M_{43}^2} \qquad M_{41} = \Sigma_{14}/M_{11}. \tag{3.8}$$

The values of the M matrix are all positive on the diagonal so a unique matrix M is determined.

The true values of the disturbance terms are calculated from a first order autocorrelated scheme

$$U_{it} = P_i U_{it-1} + e_{it} \; . \tag{3.9}$$

Assume (Parks, 1968)

$$e_{it} \sim N(0,1) \qquad i = 1,2 \qquad t = 1,\ldots,20. \tag{3.10}$$

Let[26]

$$E(e_{it}\, e_{jt'}) = \sigma_{ij} \qquad i,j = 1,2 \qquad t = t'$$

$$E(e_{it}\, e_{jt'}) = 0 \qquad t \neq t'. \tag{3.11}$$

The initial value of the autocorrelated scheme is

$$U_{11} = (1 - P_i^2)^{-\frac{1}{2}} \sqrt{\sigma_{11}}\, e_{11}\,. \tag{3.12}$$

The value of the first element in the covariance matrix is[27]

$$E(e_{i1}e_{j1}) = \frac{\sigma_{ij}\ (1 - P_i^2)^{-\frac{1}{2}}\ (1 - P_j^2)^{-\frac{1}{2}}}{1 - P_iP_j}\,. \tag{3.13}$$

Now note that $$U_{it} = \sum_{\tau=0}^{\infty} P_i^{\tau}\, e_{it-\tau}\,. \tag{3.14}$$

The variance-covariance matrix of the structural disturbance terms contain the following elements:

---

26 $\sigma_{12} = \sigma_{21}$

27 Note that the assumption is that P_i=P_j in all data sets. Therefore $E(e_{i1}e_{j1}) = \sigma_{ij}$.

$$E(U_{it}\ U_{it+\tau}) = \frac{\sigma_{ij}\ P_j^{\tau}}{1 - P_i^2} \qquad i = 1,2$$

$$E(U_{it}\ U_{jt+\tau}) = \frac{\sigma_{ij}\ P_j^{\tau}}{1 - P_iP_j} \qquad i = 1,2 \quad \tau \geq 0$$

$$E(U_{it}\ U_{jt+\tau}) = \frac{\sigma_{ij}\ P_i^{\tau}}{1 - P_iP_j} \qquad i = 1,2 \quad \tau \leq 0. \qquad 3.15$$

Given the above assumptions, the generation of the disturbance terms U_{it}, $i = 1,2$, $t = 1,2, \ldots, 20$ is as follows:

Let $E(e_{it}e_{jt}) = \sigma_{ij}$, $i,j = 1,2$. Therefore $QQ' = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$

where Q is a lower triangular matrix (p. 3, Graybill, 1961). The QQ´ matrix of the disturbance terms for this study equals

$$\begin{bmatrix} 400 & 200 \\ 200 & 400 \end{bmatrix}.$$

The generation is

$$
\begin{bmatrix} U_{11} \\ U_{12} \\ \vdots \\ U_{120} \\ U_{21} \\ U_{22} \\ \vdots \\ U_{220} \end{bmatrix}
=
\begin{bmatrix}
(1-P_i^2)^{-\frac{1}{2}} & \cdots & 0 & 0 & \cdots & 0 \\
P_i(1-P_i^2)^{-\frac{1}{2}} & 1 \cdots & 0 & & & \\
\vdots & & \vdots & \vdots & \ddots & \vdots \\
P_i^{19}(1-P_i^2)^{-\frac{1}{2}} & P_i^{18} \cdots & 1 & 0 & \cdots & 0 \\
0 & \cdots & 0 & (1-P_i^2)^{-\frac{1}{2}} & \cdots & 0 \\
 & & & P_i(1-P_i^2)^{-\frac{1}{2}} & 1 \cdots & 0 \\
\vdots & & \vdots & \vdots & & \vdots \\
0 & \cdots & 0 & P_i^{19}(1-P_i^2)^{-\frac{1}{2}} & P_i^{18} \cdots & 1
\end{bmatrix} X
$$

$$
\begin{bmatrix}
\sqrt{\sigma_{11}} & \cdots & 0 & 0 & \cdots & 0 \\
\vdots & & \vdots & \vdots & & \vdots \\
0 & \cdots & \sqrt{\sigma_{11}} & 0 & \cdots & 0 \\
\frac{\sigma_{12}}{\sqrt{\sigma_{11}}} & \cdots & 0 & \sqrt{\sigma_{22}-\frac{\sigma_{12}^2}{\sigma_{11}}} & \cdots & 0 \\
\vdots & & \vdots & & & \vdots \\
0 & \cdots & \frac{\sigma_{12}}{\sqrt{\sigma_{11}}} & 0 & \cdots & \sqrt{\sigma_{22}-\frac{\sigma_{12}^2}{\sigma_{11}}}
\end{bmatrix}
\begin{bmatrix} e_{11} \\ \vdots \\ e_{120} \\ e_{21} \\ \vdots \\ e_{220} \end{bmatrix}. \tag{3.16}
$$

The above generation gives the disturbance terms both serial and contemporaneous correlation as given in Park's article (1968) when $P_i = P_j$.[28]

The proof is as follows:

$$P = \begin{bmatrix} (1-P_i^2)^{-\frac{1}{2}} & 0 & \cdots & 0 \\ P_i(1-P_i^2)^{-\frac{1}{2}} & 1 & \cdots & 0 \\ \vdots & & & \vdots \\ P_i^{19}(1-P_i^2)^{-\frac{1}{2}} & P_i^{18} & \cdots & P_i 1 \end{bmatrix} \qquad Q_1 = \begin{bmatrix} \sqrt{\sigma_{11}} & 0 & \cdots & 0 \\ 0 & \sqrt{\sigma_{11}} & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & & \sqrt{\sigma_{11}} \end{bmatrix}$$

$$Q_2 = \begin{bmatrix} \frac{\sigma_{12}}{\sqrt{\sigma_{11}}} & 0 & \cdots & 0 \\ 0 & & & \vdots \\ \vdots & & & \\ 0 & & \cdots & \frac{\sigma_{12}}{\sqrt{\sigma_{11}}} \end{bmatrix} \qquad Q_3 = \begin{bmatrix} \sqrt{\sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}}} & 0 & \cdots & 0 \\ 0 & & & \vdots \\ \vdots & & & \vdots \\ 0 & \cdots & & \sqrt{\sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}}} \end{bmatrix}$$

$$U = \begin{bmatrix} U_{11} \\ \vdots \\ U_{120} \\ U_{21} \\ \vdots \\ U_{220} \end{bmatrix} \qquad e = \begin{bmatrix} e_{11} \\ \vdots \\ e_{120} \\ e_{21} \\ \vdots \\ e_{220} \end{bmatrix}$$

---

[28]Note that this scheme also generates data when $P_i \neq P_j$.

Then

$$UU' = \left[\begin{array}{c|c} P & 0 \\ \hline 0 & P \end{array}\right] \left[\begin{array}{c|c} Q_1 & 0 \\ \hline Q_2 & Q_3 \end{array}\right] ee' \left[\begin{array}{c|c} Q_1 & Q_2 \\ \hline 0 & Q_3 \end{array}\right] \left[\begin{array}{c|c} P' & 0 \\ \hline 0 & P' \end{array}\right]$$

$$E(UU') = \left[\begin{array}{c|c} P & 0 \\ \hline 0 & P \end{array}\right] \left[\begin{array}{c|c} Q_1 & 0 \\ \hline Q_2 & Q_3 \end{array}\right] \left[\begin{array}{c|c} Q_1 & Q_2 \\ \hline 0 & Q_3 \end{array}\right] \left[\begin{array}{c|c} P' & 0 \\ \hline 0 & P' \end{array}\right] E(ee')$$

$$E(UU') = \left[\begin{array}{c|c} P\,Q_1Q_1P' & PQ_1Q_2P' \\ \hline PQ_1Q_2P' & PQ_2Q_2P'+PQ_3Q_3P' \end{array}\right] \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & 0 \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

$$E(UU') = \left[\begin{array}{c|c} \sigma_{11}PP' & \sigma_{12}PP' \\ \hline \sigma_{21}PP' & \sigma_{22}PP' \end{array}\right]. \qquad 3.18$$

The reduced form of the structural equations is

1. $$Y_1(t) = \frac{1}{\left[B_{12}-B_{22}\right]} \Big[B_{12}C_{20} - B_{22}C_{10} - B_{22}C_{11}Z_1(t)$$

$$- B_{22}C_{12}Z_2(t) + B_{12}C_{23}Z_3(t) + B_{12}C_{24}Z_4(t)\Big]$$

$$+ \frac{1}{\left[B_{22}-B_{12}\right]} \left[B_{22}U_1(t) - B_{12}U_2(t)\right]$$

$$2. \quad Y_2(t) = \frac{1}{\left[B_{12} - B_{22}\right]} \Big[C_{20} - C_{10} - C_{11}Z_1(t) - C_{12}Z_2(t) + C_{23}Z_3(t) + C_{24}Z_4(t)\Big] + \frac{1}{\left[B_{22} - B_{12}\right]} \Big[U_1(t) - U_2(t)\Big]. \qquad 3.19$$

Having specified the values of the parameters a priori and having generated data sets for the exogenous variables and disturbance terms, the endogenous variables are generated from the reduced form equations. After generating the data sets, the data sets are then applied to the structural equations. The estimates of the structural coefficients are determined by applying different estimates with different data sets to the structural equations disregarding all a priori information. The estimators are ordinary least squares, two stage least squares, and three stage least squares.[29] The

---

[29]The properties of the estimators chosen are well documented in (Dhrymes, 1970), (Johnston, 1963), (Theil, 1971), (Christ, 1966) and the studies which are referenced in chapters II and III. In simultaneous linear models, the explanatory variables are many times not independent of the error terms. An example of such a model is as follows:

$$Y_{ti} = \sum_{j=1}^{m} B_{ji} Y_{tj} + \sum_{s=1}^{G} \gamma_{si} X_{ts} + U_{ti} \qquad i = 1,2,\ldots,m \quad t = 1,2,\ldots t$$

where, for purposes of identification, it is understood that some of the B_{ji} and γ_{si} may be zero; $E(U_{ti}) = 0$; $E(U_{ti}U_{tj}) = \delta_{tt'}\sigma_{ij}$ for $t = t'$; $E(U_{ti}U_{t'j}) = 0$ for $t \neq t'$; and $E(X_{ts}U_{ti}) = 0$. In the above system, ordinary least squares gives inconsistent and generally biased estimates. Two stage least squares gives consistent but generally biased estimates in the above model. Under the assumptions of the above model, choice of these estimates is due to their popularity over a wide range of individuals who are involved in estimating parameters in economic models.

The estimates of each estimator have a sampling distribution of size fifty for each parameter. The results of these sampling distributions plus our *a priori* knowledge of the parameters are analyzed to compare these estimators under the specification errors involved.

Certain measures of performance are compared throughout this experiment to observe if any estimator shows superiority over other estimators under given specification errors. The ideal result would be that one estimator performs better for all data sets, for all data sets given a level of autocorrelation, or for all data sets given a level of multicollinearity. The measures of performance are the absolute values of the deviations of the median of the sampling distribution

---

the asymptotic distribution of two stage least squares estimates approaches the normal distribution with mean zero and constant covariance. Two stage least squares estimates are efficient when $\sigma_{ij} = 0$ for $i \neq j$.

Three stage least squares estimates are consistent under the assumptions of the above model. The asymptotic distribution of three stage least squares estimates is distributed with zero mean and constant covariance. Assuming $E(U_{ti}U_{t'j}) = \delta_{tt'}\,\sigma_{ij}$, three stage least squares estimates are efficient estimates.

Note that the above statements hold true for stochastic exogenous variables but not for finite samples or autocorrelated disturbance terms. Multicollinearity can cause problems with respect to matrix inversion.

of the estimators from the parameter, a coefficient of concentration C,[30] and mean square error.[31]

For each measure of performance, the three estimators are ranked. The criteria for ranking are the smallest absolute deviation of the median from the parameter, largest coefficient of concentration, and the smallest mean square error.[32]

The Kendall Coefficient of Concordance, W, and the Friedman Two Way Analysis of Variance non-parametric tests are used to analyze the loss functions with respect to each estimator (Siegel, 1965).[33] Because of the quantity of data generated by these estimators, general observations of the data are also made to shed light on the characteristics of the estimators in this experiment.

---

[30]C is defined as the number of estimates within thirty percent of the parameter value.

[31]The median and coefficient of concentration always exist but the mean square error may not exist because the finite sampling distribution of the estimators may not have first or second moments. Professor Johnston states (1963) that even if appropriate moments do not exist for one or more of the estimators, comparisons between sample mean square errors still provide useful summary information on the dispersion of the estimates of different estimators. In this experiment, the mean square error is observed because past Monte Carlo experiments have used this loss function. Conclusions based upon the mean square error as a loss function will be judged less significant than those based on other loss functions used in this study.

[32]Data from the Monte Carlo experiment is summarized in the Appendix I, tables 1-6, to facilitate the understanding of the rankings of the different estimators.

[33]The description of the non-parametric tests are given in Appendix II.

In observing each level of autocorrelation when the level of multicollinearity is given, the results show, using the Kendall Coefficient of Concordance, W, that the medians of the two stage least squares estimates are ranked best. Of the other two estimators, the medians of ordinary least squares estimates are ranked last. This ranking held true for all data sets with low multicollinearity except data set S_{2L} where the medians of two stage and three stage least squares estimates are tied for the best ranking. When combining all the data sets, the Friedman non-parametric test for S_{1L}, S_{2L}, S_{3L}, and S_{4L} ranked the estimates of two stage least squares first, three stage least squares second, and ordinary least squares last if the loss function is the median. When the level of multicollinearity is increased to high, the Friedman non-parametric test for related samples, (S_{1H}, S_{2H}, S_{3H}, and S_{4H}), gave the estimates of two stage least squares the highest ranking and the estimates of three stage least squares the second highest ranking when the loss function is the median.[34] This ranking is not as conclusive as when the level of multicollinearity is low. Kendall's W for S_{2H} ranks three stage least squares estimates ahead of two stage least squares estimates. From a closer analysis of data set S_{2H}, the ranking of two stage least squares estimates behind three stage least squares

[34]All the rankings are achieved at the .05 level of significance. The ranking of the median given above is achieved at the .06 level of significance.

estimates is negligible. In S_{3H}, these two estimators tied for the highest ranking. These results show that two stage least squares estimates generally give the best estimates of the central location at different levels of autocorrelation given the level of multicollinearity.

When autocorrelation is held constant and multicollinearity is varied, the results are as follows: For S_{1L} and S_{1H}, S_{3L},and S_{4L} and S_{4H}, the results are that the medians of two stage least squares estimates perform better than the medians of the three stage least squares estimates. In data sets S_{2L} and S_{3H}, two stage and three stage least squares estimates of the medians are tied for the best ranking. In data set S_{2H}, the three stage least squares estimates are ranked ahead of the two stage least squares estimates. In all cases, the medians of the ordinary least squares estimates perform poorly.

When the coefficient of concentration is used as the loss function, the results of the Friedman test for related samples are inconclusive. The Kendall coefficient of concordance, W, shows in four out of eight data sets (S_{1L}, S_{3L}, S_{2H} and S_{3H}) that the coefficients of concentration of the three stage least squares are ranked first. Of the other two estimators for these data sets, the coefficients of concentration of the two stage least squares estimates are ranked second. In the other four data sets, Kendall's test does not establish a ranking of estimators at the .05 level of significance. In analyzing all the data sets, the coefficients of concentration of the ordinary least squares estimates perform very poorly as compared to the coefficients of concentration of the two and three stage least squares estimates.

When the mean square error loss function is used, no conclusive statistical statements can be made about the superiority of one estimator's performance as compared to the performance of other estimators as the level of the specification error changes. The Friedman test shows that the mean square errors of the ordinary least squares estimates perform as well as the mean square errors of the other two estimators for related data sets. For data sets S_{1L}, S_{3L}, S_{1H}, and S_{2H}, the Kendall test gives the estimators identical rankings. For S_{2L} ans S_{4H}, the Kendall test ranks the ordinary least squares estimates best, the three stage least squares estimates second, and two stage least squares estimates last. When applying the Kendall test to data set S_{3H}, the mean square errors of the three stage least squares estimates perform better than the mean square errors of the ordinary least squares estimates. The two stage least squares estimates are ranked last for data set S_{3H}. The ordinary least squares and three stage least squares estimates are tied for highest ranking for data set S_{4L}. In observing these results, a general conclusion is that the estimates of ordinary and three stage least squares are better than the estimates of two stage least squares as autocorrelation increases when the loss function is the mean square error.

In observing the estimates of different estimators in samples from different data sets the following observations are made:

1. When the autocorrelation parameter is increased for a given level of multicollinearity, the absolute distances between the medians and the parameters increase, the coefficients of concentration decrease, and the mean square errors increase. In the few exceptions to the above observations, the behavior of the loss functions is very similar to the above pattern; even if some deviation occurs this deviation usually returns to the above pattern as the specification error changes. The above pattern holds true when the level of autocorrelation is held constant and multicollinearity is varied. These changes are important in that they illustrate the importance of analyzing Monte Carlo studies over a whole parameter space (pp.806-810, Thornber, 1968).

2. With the exception of the intercept parameters, the medians for all parameters underestimate the parameters if the parameter is positive or overestimate the parameters if the parameter is negative. The only time this observation does not hold true is for ordinary least squares estimates of C_{11} in all data sets, two stage least squares estimates of C_{11} in S_{1L}, S_{2L}, and S_{1H}, and three stage least squares estimates of C_{11} in S_{1H}. This observation is also violated for two stage least squares and three stage least squares estimates of C_{24}

in data set S_{4H}.[35] This may be a clue that the estimators in question are all biased in the same direction for small sample size with the specification errors observed in this Monte Carlo experiment.

3. When the level of autocorrelation or multicollinearity is increased, the standard errors of the regression coefficients of all estimators increase in magnitude.

A conclusion that can be drawn from this experiment is that the overall performance of the ordinary least squares estimates is poor. The improvement in the ordinary least squares estimates when the ranking is judged by the mean square error loss function only illustrates the conjecture that the ordinary least squares estimates cluster around a point other than the parameter of the sampling distribution of the estimates. Furthermore, using the mean square error as a loss function can lead one to make incorrect inferences about the performance of different estimators. With specification errors involved, the choice of two stage least squares as an estimator gives the best estimates when the median is the loss function. If the loss function is the coefficient of concentration, then the three stage least squares estimator is marginally superior to the two stage least squares estimator. In ranking the estimators over all data sets and all loss

---

[35]Note that the exceptions are 8.5 percent of the total possible chances. Of this 8.5 percent, 85.7 percent occurred when calculating estimates for C_{11}.

and then applying the Friedman test, three stage least squares estimates rank first, two stage least squares estimates rank second, and ordinary least squares estimates ranks last. The overall conclusion of this experiment is that the three stage least squares estimates is preferred wherever applicable. If difficulty arises due to complexities of a model, the recommended estimator is two stage least squares. All these conclusions are based upon the model and specification errors given in Chapter IV.

CHAPTER V

AN EMPIRICAL INVESTIGATION OF A REAL WORLD DEMAND AND SUPPLY MODEL AND CONCLUDING REMARKS

The results of Chapter IV, though not entirely conclusive, should provide some insight into the analysis of other economic models with similar specification errors.[36] A model chosen to study is similar to Karl Fox's model (p. 403, 1968) for the consumption of pork and is given as follows:

$$\text{Demand} \qquad Q = B_{10} + B_{11}P + B_{12}Y + B_{13}W$$

$$\text{Supply} \qquad Q = B_{20} + B_{21}P + B_{24}R_{-1} + B_{25}Z\,. \tag{5.1}$$

The variable P is the average retail price of pork in cents per pound, Q is per capita consumption of pork, Y is per capita

[36]The Durbin-Watson statistics for the demand and supply equations in 5.1 are 0.385 and 0.802, respectively. The correlation matrix of the exogenous variables is as follows:

$$\begin{array}{c c} & \begin{array}{cccc} Y & W & R_{-1} & Z \end{array} \\ \begin{array}{c} Y \\ W \\ R_{-1} \\ Z \end{array} & \begin{bmatrix} 1 & .3445 & -.1146 & .2127 \\ & 1 & -.0706 & .0464 \\ & & 1 & .1748 \\ & & & 1 \end{bmatrix}. \end{array}$$

disposable consumer income, W is the per capita consumption of beef, Z is an estimate of per capita pork production based upon exogenous and predetermined variables, and R_{-1} is the ratio of the average Chicago wholesale price of hogs per hundred pounds to the average market price of a bushel of corn, lagged one year. The data observed is for the years 1922 to 1941.

The model described above is similar to the model studied in Chapter IV, equations 3.1. There are two equations, each equation is overidentified to the following degree, $K^{**} - G^{\Delta} - 1 = 1$.[37] The exogenous variable Z is determined from some larger model and therefore is stochastic. The Durbin-Watson statistics show that both equations have autocorrelation, and the exogenous variables have a moderate degree of multicollinearity. The Monte Carlo model (3.1) and the "real world" model (4.1) studied exhibit a high degree of similarity, but one must proceed with caution in comparing them because the real world model may have other characteristics not included in the Monte Carlo model.

The results from the application of ordinary, two stage, and three stage least squares to the pork consumption model are:

---

[37] K^{**} is the number of excluded exogenous variables and G^{Δ} is the number of included endogenous variables in the equation being analyzed.

Ordinary Least Squares

$$\text{Demand} \quad Q = \underset{(13.0235)}{70.5756} - \underset{(0.312185)}{1.30576}P + \underset{(0.0172713)}{0.0843832}Y - \underset{(0.250361)}{0.297492}W$$

$$\text{Supply} \quad Q = \underset{(6.1854)}{15.8244} - \underset{(0.114252)}{.0395168}P - \underset{(0.120233)}{0.0685225}R_{-1} + \underset{(0.079457)}{0.750699}Z. \tag{5.2}$$

Two Stage Least Squares

$$\text{Demand} \quad Q = \underset{(19.8886)}{73.3939} - \underset{(0.780463)}{2.84232}P + \underset{(0.0365858)}{0.143381}Y - \underset{(0.381654)}{0.285731}W$$

$$\text{Supply} \quad Q = \underset{(7.26034)}{9.10129} + \underset{(0.145958)}{0.166267}P - \underset{(0.130958)}{0.0649537}R_{-1} + \underset{(0.0778449)}{0.775114}Z. \tag{5.3}$$

Three Stage Least Squares

$$\text{Demand} \quad Q = \underset{(16.5251)}{92.7631} - \underset{(0.780397)}{2.82456}P + \underset{(0.0364261)}{0.149357}Y - \underset{(0.295163)}{0.709185}W$$

$$\text{Supply} \quad Q = \underset{(7.01589)}{7.52553} + \underset{(0.145723)}{0.173896}P + \underset{(0.104180)}{0.00816743}R_{-1} + \underset{(0.0775719)}{0.781116}Z. \tag{5.4}$$

The results yielded by applying these three estimators to the real world model for the consumption of pork are not conclusive but some evidence is generated which seems to substantiate the conclusions of Chapter IV. For example, the ordinary least squares estimate of the coefficient of the price variable in the supply equation has a negative sign while the two stage and three stage least squares estimates of the same variable have a positive sign. Based upon economic theory, one would expect a positive sign for the coefficient of the price variable in a supply function for a non-durable commodity. A second result that one would expect is that the sign of the coefficient of the variable R_{-1} should be positive, and only the three stage least squares estimate gives this sign. Based upon these results, the choice of an estimator should be three stage least squares first, two stage least squares second, and ordinary least squares third.

The evidence yielded by the demand equation supporting the two and three stage least squares estimators is not conclusive. The signs of the estimated coefficients produced by each estimator appear as might be expected. For the estimated coefficients of the demand and supply equations, there seems to be a directional pattern of the estimated coefficients of all variables except W. This pattern seems similar to the conjectures made in Chapter IV (<u>supra</u>, para. 2, p. 95 and pp. 92-93). The conjectures can best be summarized by stating that the bias of the ordinary least squares estimates is larger than the bias of the two and three stage least squares estimates, and that the bias of all estimators is opposite the sign of the parameter. An example of this pattern is illustrated for the estimated coefficients of the P and Y variables in the demand equation. From the above pattern and conjectures, one would conclude that the parameters of P and Y are less than -2.82456 and greater than .149357 respectively.

In general, the Monte Carlo study presented in Chapter

IV and the analysis of the pork consumption model substantiates the conclusions of past theoretical works and Monte Carlo studies dealing with simultaneous equation models with finite sample sizes. Basmann's and Takeuchi's theoretical findings under restrictive assumptions stated that the bias of ordinary least squares is larger than the bias of two stage least squares estimates and that the standard error of the regression coefficients are smaller for ordinary least squares estimates than two stage least squares estimates. The results derived from pure theory were confirmed by our Monte Carlo studies, when specification errors were involved. Richardson's theoretical conclusion also applies to this Monte Carlo study. His conclusions, under restrictive assumptions, were that the signs of the biases of the two stage least squares estimates are opposite to the sign of the parameter and that the size of the relative bias lies between zero and minus one. Not only were these conclusions warranted for the two stage least square estimates, but also for the estimates produced by the other two estimators in most cases. Sawa showed that two stage least squares estimates are concentrated around the parameter while ordinary least squares estimates are concentrated around some other value. In our Monte Carlo experiment with specification errors, Sawa's results held and three stage least squares estimates exhibited the same pattern as the two stage least squares estimates.

A general conclusion reached in the Monte Carlo studies discussed in Chapter III was that the use of ordinary least squares gives estimates that are inferior to other estimates for different estimators. This coincides with the findings of the Monte Carlo study performed in Chapter IV. A basic conclusion of Monte Carlo studies was that the ordinary least squares estimates have a large bias but a small variance around the mean of the least squares estimates, and the Monte Carlo study of Chapter IV verified this. The two stage least squares estimates produced in most of the Monte Carlo studies were found to be less sensitive to specification errors as compared to the estimates generated by ordinary least squares. This was also substantiated in Chapter IV.

In Monte Carlo studies where all three estimators were used, the basic results of the ranking are identical to those given in Chapter IV. When the ranking was changed due to specification errors, two stage least squares became the recommended estimator over three stage least squares and ordinary least squares was always considered an inferior estimator. In Chapter IV, the three stage least squares estimator was less sensitive to the specification errors that were introduced when compared to other specification errors in other Monte Carlo studies.

Therefore in general, given the three estimators which were analyzed in Chapter IV and given the results of the study presented and studies in the past, the present investigation seems to provide rational evidence for the choice of three stage least squares as the appropriate estimator in simultaneous models with finite samples. If it is not feasible to apply three stage least squares, then two stage least squares would be the next best choice. Ordinary least squares is not recommended.

LIST OF REFERENCES

Basmann, Robert L. "A Note on the Exact Finite Sample Frequency Functions of Generalized Classical Linear Estimates in Two Leading Overidentified Cases," Journal of American Statistical Association, LVI, No. 56 (September, 1961), pp. 619-636.

Basmann, Robert L. "A Note on the Exact Finite Sample Frequency Functions of Generalized Classical Linear Estimates in a Leading Three Equation Case," Journal of American Statistical Association, LVIII, No. 301 (March, 1963), pp. 161-171.

Bergstrom, A. R. "The Exact Sampling Distributions of Least Squares and Maximum Likelihood Estimators of the Marginal Propensity to Consume," Econometrica, XXX, No. 3 (July, 1962), pp. 480-490.

Christ, Carl F. Econometric Models and Methods. New York: John Wiley and Sons, Inc., 1966.

Cochrane, Donald and Orcutt, Guy H. "Application of Least Squares Regression to Relationships Containing Autocorrelated Error Terms," Journal of the American Statistical Association, XLIV, No. 245 (March, 1949), pp. 32-61.

Cochrane, Donald and Orcutt, Guy H. "A Sampling Study of the Merits of Autoregressive and Reduced Form Transformations in Regression Analysis," Journal of the American Statistical Association, XLIV, (September, 1949), pp. 356-372.

Cragg, John G. "On the Sensitivity of Simultaneous Equation Estimators to Stochastic Assumptions of the Models," Journal of the American Statistical Association, LXI No. 313 (March, 1966), pp. 136-151.

Cragg, John G. "On the Relative Small Sample Properties of Several Structural Equation Estimators," Econometrica, XXXV, No. 1 (January, 1967), pp. 89-110.

Cragg, John G. "Some Effects of Incorrect Specification on the Small Sample Properties of Several Simultaneous Equation Estimators," International Economic Review, IX, No. 1 (February, 1968), pp. 63-86.

Dhrymes, Phoebus J. Econometrics: Statistical Foundations and Applications. New York: Harper and Row, 1970.

Fox, Karl. Intermediate Economic Statistics. New York: John Wiley and Sons, Inc., 1968.

Glahe, Fred R. and Hunt, Jerry G. "Small Sample Properties of Simultaneous Equation Least Absolute Estimators Vis-A-Vis Least Squares Estimators," Econometrica, XXXVIII, No. 5 (September, 1970), pp. 742-753.

Graybill, Franklin A. An Introduction to Linear Statistical Models. New York: McGraw-Hill Book Co., 1961.

Johnston, John. Econometric Methods. New York: McGraw-Hill Book Co., Inc., 1963.

Kabe, D. Gerhard. "A Note on the Exact Distribution of the GCL Estimators in Two Leading Overidentified Cases," Journal of the American Statistical Association, LVIII, No. 302 (June, 1963), pp. 535-537.

Kabe, D. Gerhard. "A Note on the Exact Distributions of the GCL Estimators in a Leading Three-Equation Case," Journal of the American Statistical Association, LIX, No. 307 (September, 1964), pp. 881-894.

Kadiyala, Koteswara Rao. "An Exact Small Sample Property of the K-Class Estimators," Econometrica, XXXVIII, No. 6 (November, 1970), pp. 930-932.

Kmeta, Jan and Gilbert, Roy F. "Small Sample Properties of Alternative Estimators of Seemingly Unrelated Regressions," Journal of the American Statistical Association, LXIII, No. 324 (December, 1968), pp. 1180-1203.

Ladd, George W. "Effects of Shocks and Errors in Estimations: An Empirical Comparison," Journal of Farm Economics, XXXVIII, No. 2 (May, 1956), pp. 485-495.

Maeshiro, Asatoshi. "A Simple Mathematical Relationship Among K-Class Estimators," Journal of the American Statistical Association, LXI, No. 314 (June, 1966), pp. 368-374.

Nagar, A. L. "The Bias and Moment Matrix of the General K-Class Estimators of the Parameters in Simultaneous Equations," Econometrica, XXVII, No. 4 (October, 1959), pp. 575-595.

Nagar, A. L. "A Monte Carlo Study of Alternative Simultaneous Equation Estimators," Econometrica, XXVIII, No. 3 (July, 1960), pp. 573-590.

Nagar, A. L. "Double K-Class Estimators of Parameters in Simultaneous Equation and Their Small Sample Properties," International Economic Review, III, No. 2 (May, 1962), pp. 168-188.

Neiswanger, William A. and Yancey, Thomas A. "Parameter Estimates and Autonomous Growth," Journal of the American Statistical Association, LIV, No. 286 (June, 1969), pp. 389-402.

Oi, Walter Y. "On the Relationship Among Different Members of the K-Class," International Economic Review, X, No. 1 (February, 1969), pp. 36-45.

Orcutt, Guy H. and Winokin, Herbert S. "First Order Autoregression: Inference, Estimation, and Prediction," Econometrica, XXXVII, No. 1 (January, 1969), pp. 1-14.

Parks, Richard W. "Efficient Estimation of a System of Regression Equation When Disturbances are Both Serially and Contemporaneously Correlated," Journal of the American Statistical Association, LXII, No. 318 (June, 1968), pp. 500-509.

Quandt, Richard E. "On Certain Sample Properties of K-Class Estimates," International Economic Review, VI, No. 1 (January, 1965), pp. 92-104.

Rao, Potluri and Griliches, Zvi. "Small Sample Properties of Several Two Stage Regression Methods in the Context of Auto-correlated Errors," Journal of the American Statistical Association, LXIV, No. 325 (March, 1969), pp. 253-273.

Richardson, David H. "The Exact Distribution of a Structural Coefficient Estimator," Journal of the American Statistical Association, LXII, No. 324 (December, 1968), pp. 1214-1226.

Sawa, Takamitsu. "The Exact Sampling Distribution of Ordinary Least Squares and Two Stage Least Squares Estimators," Journal of the American Statistical Association, LXVI, No. 316 (September, 1969), pp. 923-937.

Siegel, Sidney. Nonparametric Statistics. New York: McGraw-Hill Book Co., Inc., 1956.

Summers, Robert M. "A Capital Intensive Approach to the Small Sample Properties of Various Simultaneous Equation Estimators," Econometrica, XXXIII, No. 1 (January, 1965), pp. 1-41.

Takeuchi, K. "Exact Sampling Moments of the Ordinary Least Squares, Instrumental Variables, and Two Stage Least Squares Estimators," International Economic Review, XI, No. 1 (February, 1970), pp. 1-12.

Theil, Henry. Principles of Econometrics. New York: John Wiley and Sons, Inc., 1971.

Thornber, Hodson. "Finite Sample Monte Carlo Studies: An Autoregressive Illustration," Journal of the American Statistical Association, LXII, No. 319 (September, 1967), pp. 801-818.

Wagner, Harvey M. "A Monte Carlo Study of Estimates of Simultaneous Linear Structural Equations," Econometrica, XXVI, No. 1 (January, 1958), pp. 117-133.

Zellner, Arnold. "Estimators for Seemingly Unrelated Regression Equations: Some Exact Finite Sample Results," Journal of the American Statistical Association, LVIII, No. 304 (December, 1963), pp. 977-992.

APPENDIX I

TABLE 1

THE VALUES OF THE MEDIANS OF THE SAMPLING DISTRIBUTION OF THE ESTIMATORS WHEN THE LEVEL OF MULTICOLLINEARITY IS LOW

	P =0.2	P =0.6	P =0.8	P =0.9
C_{10} = 100.0				
OLS	109.8720	109.8025	101.3255	109.2455
2SLS........	104.9565	100.5674	98.8806	102.4060
3SLS........	110.7380	105.5630	108.1920	107.6220
B_{12} = -0.4				
OLS	-.2127	-.1765	-.1320	-.1140
2SLS........	-.3873	-.3646	-.3545	-.3276
3SLS........	-.3715	-.3659	-.3369	-.3422
C_{11} = 0.1				
OLS	.1055	.1438	.1519	.1425
2SLS........	.1067	.1068	.0936	.0904
3SLS........	.1130	.0889	.0791	.0626
C_{12} = 0.45				
OLS	.3585	.3362	.3063	.2925
2SLS........	.4380	.4022	.4014	.4081
3SLS........	.4283	.4129	.3999	.3920
C_{20} = 50.0				
OLS	104.9230	119.5995	124.0705	138.1465
2SLS........	56.6420	72.2471	82.8171	74.4665
3SLS........	60.4998	67.4109	62.3273	56.4067
B_{22} = 0.6				
OLS	.4045	.4030	.3410	.2718
2SLS........	.5804	.4680	.5490	.5175
3SLS........	.5784	.5559	.5281	.5150
C_{23} = 0.25				
OLS	.1863	.1535	.1464	.1690
2SLS........	.2217	.2117	.2055	.2109
3SLS........	.2272	.2189	.2107	.1979
C_{24} = 0.8				
OLS	.6731	.6077	.5843	.5871
2SLS........	.7883	.7604	.7731	.7269
3SLS........	.7608	.7371	.7512	.7120

TABLE 2

THE COEFFICIENTS OF CONCENTRATION C WHEN THE LEVEL OF MULTICOLLINEARITY IS LOW

	P = 0.2	P = 0.6	P = 0.8	P = 0.9
C_{10} = 100.0				
OLS	19	18	14	16
2SLS........	21	20	18	13
3SLS........	22	18	19	18
B_{12} = -0.4				
OLS	14	9	11	9
2SLS........	24	23	28	16
3SLS........	24	23	18	15
C_{11} = 0.1				
OLS	29	29	25	27
2SLS........	27	25	22	22
3SLS........	35	30	28	21
C_{12} = 0.45				
OLS	27	25	21	15
2SLS........	30	27	24	19
3SLS........	30	27	26	22
C_{20} = 50.0				
OLS	5	9	3	3
2SLS........	12	4	4	7
3SLS........	17	12	10	9
B_{22} = 0.6				
OLS	21	21	15	13
2SLS........	30	29	27	24
3SLS........	30	29	27	24
C_{23} = 0.25				
OLS	22	21	18	19
2SLS........	26	19	18	16
3SLS........	29	24	25	18
C_{24} = 0.8				
OLS	33	24	25	23
2SLS........	35	33	30	26
3SLS........	36	29	31	25

TABLE 3

THE MEAN SQUARE ERRORS OF THE ESTIMATES OF EACH ESTIMATOR WHEN THE LEVEL OF MULTICOLLINEARITY IS LOW

	P = 0.2	P = 0.6	P = 0.8	P = 0.9
C_{10} = 100.0				
OLS	4656.7258	5899.6293	7388.8680	10050.0693
2SLS........	5140.5873	7691.2040	11606.3994	15548.6741
3SLS........	3854.0662	7658.0810	12560.8300	15284.1413
B_{12} = −0.4				
OLS	.0484	.0632	.0844	.1215
2SLS........	.0337	.0444	.0716	.1239
3SLS........	.0335	.0441	.0733	.1179
C_{11} = 0.1				
OLS	.0159	.0178	.0212	.0307
2SLS........	.0163	.0188	.0281	.0435
3SLS........	.0099	.0143	.0234	.0316
C_{12} = 0.45				
OLS	.0248	.0344	.0448	.0548
2SLS........	.0236	.0349	.0471	.0681
3SLS........	.0237	.0372	.0514	.0717
C_{20} = 50.0				
OLS	5678.1693	7280.3035	10221.2850	13925.7928
2SLS........	12232.0317	9291.0161	12207.5704	16821.1699
3SLS........	11454.3593	7992.1340	10007.3940	13546.1595
B_{22} = 0.6				
OLS	.0151	.0665	.0979	.1418
2SLS........	.1021	.0814	.0992	.1310
3SLS........	.0997	.0767	.0881	.1099
C_{23} = 0.25				
OLS	.0114	.0143	.0202	.0277
2SLS........	.0193	.0162	.0228	.0309
3SLS........	.0169	.0120	.0255	.0203
C_{24} = 0.8				
OLS	.0580	.0810	.1248	.1611
2SLS........	.1363	.1211	.1567	.2264
3SLS........	.1343	.1132	.1357	.1911

TABLE 4

THE VALUES OF THE MEDIANS OF THE SAMPLING DISTRIBUTION OF THE ESTIMATORS WHEN THE LEVEL OF MULTICOLLINEARITY IS HIGH

	P = 0.2	P = 0.6	P = 0.8	P = 0.9
C_{10} = 100.0				
OLS	83.4352	71.5492	57.9859	73.5192
2SLS........	99.3240	83.4078	102.0270	105.9775
3SLS........	94.0192	100.8200	110.9510	116.0780
B_{12} = -0.4				
OLS	-.1135	-.0783	-.0158	.0488
2SLS........	-.3781	-.3294	-.3268	-.3024
3SLS........	-.3537	-.3571	-.3293	-.2992
C_{11} = 0.1				
OLS	.1915	.2246	.2410	.2467
2SLS........	.1027	.0973	.0816	.1088
3SLS........	.1066	.0991	.0995	.0998
C_{12} = 0.45				
OLS	.2645	.2433	.2214	.2151
2SLS........	.4179	.4007	.3639	.3667
3SLS........	.4159	.3963	.3573	.3560
C_{20} = 50.0				
OLS	142.3880	166.8395	172.2160	185.3105
2SLS........	67.6479	83.6484	88.9468	79.2961
3SLS........	74.8051	79.1524	85.3350	98.6380
B_{22} = 0.6				
OLS	.3352	.3263	.2707	.2303
2SLS........	.5727	.5115	.5086	.4512
3SLS........	.5612	.5262	.5133	.4614
C_{23} = 0.25				
OLS	.1258	.0874	.0453	.0429
2SLS........	.2368	.1961	.2050	.1932
3SLS........	.2330	.2008	.1915	.1811
C_{24} = 0.8				
OLS	.6778	.5961	.5738	.5696
2SLS........	.8260	.7664	.7820	.6953
3SLS........	.8058	.7784	.7330	.6265

TABLE 5

THE COEFFICIENTS OF CONCENTRATION C WHEN THE LEVEL OF MULTICOLLINEARITY IS HIGH

	P = 0.2	P = 0.6	P = 0.8	P = 0.9
C_{10} = 100.0				
OLS	21	15	11	11
2SLS........	17	10	8	8
3SLS........	19	15	17	16
B_{12} = −0.4				
OLS	8	4	4	6
2SLS........	20	20	15	12
3SLS........	18	18	21	12
C_{11} = 0.1				
OLS	23	17	15	14
2SLS........	25	24	22	20
3SLS........	27	30	28	28
C_{12} = 0.45				
OLS	17	11	7	9
2SLS........	25	26	20	19
3SLS........	27	29	23	17
C_{20} = 50.0				
OLS	5	3	3	3
2SLS........	5	3	2	5
3SLS........	7	7	5	5
B_{22} = 0.6				
OLS	16	12	14	12
2SLS........	29	21	15	15
3SLS........	32	21	17	15
C_{23} = 0.25				
OLS	17	13	12	10
2SLS........	16	16	14	15
3SLS........	15	19	15	13
C_{24} = 0.8				
OLS	29	21	17	20
2SLS........	25	22	23	20
3SLS........	21	24	21	18

TABLE 6

THE MEAN SQUARE ERRORS OF THE ESTIMATES OF EACH ESTIMATOR WHEN THE LEVEL OF MULTICOLLINEARITY IS HIGH

	P = 0.2	P = 0.6	P = 0.8	P = 0.9
$C_{10} = 100.0$				
OLS	7074.8906	9227.9728	10574.5503	13350.6744
2SLS........	8622.7975	13745.0508	21882.6208	31298.9262
3SLS........	6388.3628	10336.1283	17819.7372	25936.7775
$B_{12} = -0.4$				
OLS	.1032	.1296	.1587	.2022
2SLS........	.0836	.1008	.1406	.2399
3SLS........	.0873	.1036	.1265	.1945
$C_{11} = 0.1$				
OLS	.0342	.0418	.0442	.0584
2SLS........	.0294	.0338	.0593	.1211
3SLS........	.0197	.0202	.0381	.0730
$C_{12} = 0.45$				
OLS	.0505	.0671	.0862	.1042
2SLS........	.0540	.0748	.0935	.1366
3SLS........	.0570	.0799	.0932	.1315
$C_{20} = 50.0$				
OLS	16400.3267	18969.2250	24874.9588	38195.9810
2SLS........	33644.8904	32500.4010	28250.1966	41025.9509
3SLS........	31596.4031	28772.8203	22033.9556	30573.8113
$B_{22} = 0.6$				
OLS	.0741	.0983	.1386	.1900
2SLS........	.1904	.2472	.1965	.2649
3SLS........	.1841	.2319	.1726	.2307
$C_{23} = 0.25$				
OLS	.0462	.0497	.0657	.1072
2SLS........	.0669	.0515	.0586	.0961
3SLS........	.0631	.0461	.0389	.0594
$C_{24} = 0.8$				
OLS	.0945	.1355	.2256	.2997
2SLS........	.2887	.4195	.4352	.6308
3SLS........	.2809	.4931	.3936	.5757

APPENDIX II

The parameters for each measure of performance are ranked in the following form for each data set.

Estimators

	OLS	2SLS	3SLS
C_{10}	r	r	r
B_{11}	r	r	r
⋮	⋮	⋮	⋮
C_{24}	r	r	r
R_j	Σr	Σr	Σr

r is the rank of the estimators--1,2, or 3.

R_j is the sum of the column ranks for each estimator.

The Kendall Coefficient of Concordance: W (Siegel, 1956) is used to test the null hypothesis that there is no difference in relative performances between estimators for a given data set. The alternative hypothesis would be that there is a difference in relative performances between estimators for a given data set. The level of significance used is five per cent. If the null hypothesis is rejected, the estimators will be ranked according to the following procedure. The estimator with the smallest R_j receiving the rank 1, the estimator with the largest R_j receiving the rank 3, and the other estimator receiving the rank 2.

The Friedman Two-Way Analysis of Variance by Rank test for matched groups (Siegel, 1956) is applied to within data sets and among data sets. The null hypothesis will be that there is no difference in relative performance within (among) data sets. The alternative hypothesis is that there is a difference in relative performance within (among) data sets. The significance level is five percent. The form is as follows:

Within Data Sets

Estimators

Data Sets	OLS	2SLS	3SLS
S_{iL}	r	r	r
S_{iH}	r	r	r
R_j	Σr	Σr	Σr

$i = 1,2,3,4$

Among Data Sets

Estimators

Data Sets	OLS	2SLS	3SLS
S_{1i}	r	r	r
S_{2i}	r	r	r
S_{3i}	r	r	r
S_{4i}	r	r	r
R_j	Σr	Σr	Σr

$i = L, H$

r now is the rank of the estimator received by the Kendall coefficient of concordance: W non-parametric test.

An example of this procedure is given as follows:

Loss Function - Median

Data Set - S_{1H}

Kendall Coefficient of Concordance: W

	OLS	2SLS	3SLS
C_{10}	3	1	2
B_{12}	3	1	2
C_{11}	3	1	2
C_{12}	3	1	2
C_{20}	3	1	2
B_{22}	3	1	2
C_{23}	3	1	2
C_{24}	3	2	1
R_j	24	9	15
Rank	3	1	2

$$W = \frac{\Sigma\left[R_j - \frac{\Sigma R_j}{N}\right]^2}{\frac{1}{12}k^2(N^3 - N)}$$

Critical value of W is .37

W = .82 > .37

Therefore reject H_0 and accept H_A

Friedman Two-Way Analysis
of Variance by Rank

	OLS	2SLS	3SLS
S_{1H}	3	1	2
S_{2H}	3	1.5	1.5
S_{3H}	3	1	2
S_{4H}	3	1	2
R_j	12	4.5	7.5
Rank	3	1	2

$$\chi_r^2 = \frac{12}{Nk(k-1)} \sum_{j=1}^{k} (R_j)^2 - 3N(k+1$$

Critical value of χ_r^2 is 6.0

$\chi_r^2 = 43.1 > 6.0$

Therefore reject H_0 and accept H_A