ACKNOWLEDGEMENTS

I would like to thank Dr. T.D. Dwivedi for his constant counselling and encouragement throughout this undertaking.

I am also grateful to Stephanie Roberts for the speed and efficiency with which she typed this thesis.

TABLE OF CONTENTS

INTRODUCTION

There is a constant search by Economists and Statisticians for appropriate statistical tools to provide meaningful inferences about the structure of the economic system. In this on-going process, they have to take account of the nonreproducible nature of economic observations, the simultaneity of the structural relations making up the economic system, and the dynamic characteristics of the economic process.

Indeed, it was necessary to develop fundamental concepts such as completeness of model, identification, exogenity, endogenity and dynamic multipliers, to name a few, that play an important role in linking statistical analysis and economic theory. Many statistical estimation, testing and prediction procedures have been developed to handle the various kinds of problems in econometric models, including linear and non-linear interdependent structural models, models with time-series implications, models for combined time-series and cross-section data, and models with random parameters.

In this thesis we shall be focusing our attention on statistical estimation procedures for simultaneous linear econometric models. These procedures will include various variants of least squares, maximum likelihood etc. and some other alternate asymptotically justified procedures. In addition, we shall also discuss the asymptotic and finite-sample properties of these estimators.

CHAPTER I

THE NATURE OF SIMULTANEOUS EQUATION MODELS

1. Research since the 1940's has led to the development of sophisticated estimating techniques to handle the problem of simultaneity of economic relationships. The pioneers in this area postulated that economic activity could be analysed as a system of simultaneous stochastic equations. Instead of determining a single dependent variable in a single equation, they obtained a joint distribution of dependent variables from the simultaneous structure.

Such a formulation led to the re-examination of the Ordinary Least Squares (OLS) method of estimation. It was discovered that if more than one jointly dependent variable appeared in a particular equation, least squares estimation yielded biased and inconsistent estimates of the population parameters. This is a direct result of the intercorrelation between jointly dependent variables and the disturbances. In the classical regression model this difficulty was obviated by the assumed fixed nature of all the independent variables. When the simultaneous nature of economic activity is admitted into the model, the foregoing assumption becomes invalid.

A simultaneous equation model differs from the classical regression models in the sense that all of the relationships involved are needed for determining the value of even one of the endogenous variables included in the model. When an economic model has been specifically formulated as a set of well-defined stochastic relationships, it is termed an econometric model. The variables involved in these equations may be classified into two groups, viz., endogenous and exogenous. The endogenous variables are those which are required to be explained by the equation system and the exogenous variables are considered as given for the purposes of explaining the endogenous variables.

For statistical purposes we distinguish between jointly dependent and predetermined variables. The current endogenous variables are called jointly dependent, which are to be explained by the model, and the lagged variables (both endogenous and exogenous) along with the current exogenous variables may be grouped in a class called predetermined variables. The values of the exogenous variables are completely determined outside the system whereas the values of lagged endogenous variables are represented by the past values of the endogenous variables of the model. We should observe that in a particular equation there shall always be one jointly dependent variable which is "to be explained" and the others (both jointly dependent and predetermined) occurring in the equation may be called "explanatory."

In an attempt to effect a statistical estimation of Simultaneous Equation Models (SEM), we must first make specific assumptions about the structure of various equations in the particular model. Some of these assumptions are imposed by the very nature of economic theory while others are a direct result of empirical evidence. If the number of equations in the model is equal to the number of variables to be explained, i.e., the jointly dependent variables, the system of equations is said to be complete.

As an illustration consider the simple dynamic Keynesian Model:

$$C_t = \beta_0 + \beta Y_{t-1} \tag{1.1.1}$$

$$Y_t = C_t + I_t \tag{1.1.2}$$

Here I is exogenous and C and Y are endogenous. C [consumption], Y [income] and I [investment expenditure]. I_t and Y_{t-1} are predetermined and C_t and Y_t are jointly dependent. Also Y_{t-1} is a lagged endogenous variable. There are two jointly dependent variables and two equations. Thus, the model is complete. Equations (1.1.1) and (1.1.2) are called structural equations.

In general, the structural form of a simultaneous equation system can be written as follows:

$$\begin{aligned}
&\beta_{11}y_{1t} + \beta_{12}y_{2t} + \ldots\beta_{1M}y_{Mt} + \gamma_{11}x_{1t} + \gamma_{12}x_{2t} + \ldots\gamma_{1K}x_{Kt} = u_{1t} \\
&\beta_{21}y_{2t} + \beta_{22}y_{2t} + \ldots\beta_{2M}y_{Mt} + \gamma_{21}x_{1t} + \gamma_{22}x_{2t} + \ldots\gamma_{2K}x_{Kt} = u_{2t} \\
&\quad\vdots \\
&\beta_{M1}y_{1t} + \beta_{M2}y_{2t} + \ldots\beta_{MM}y_{Mt} + \gamma_{M1}x_{1t} + \gamma_{M2}x_{2t} + \ldots\gamma_{MK}x_{Kt} = u_{Mt}
\end{aligned} \tag{1.1.3}$$

for $t = 1,2,\ldots,T$,

where the y's denote the endogenous variables, the x's denote the predetermined variables, the u's represent the unobserved stochastic disturbances. The β's and the γ's are called structural coefficients. It is assumed that theory typically specifies some of the β's and the γ's to be zero

The entire system outlined in (1.1.3) can have the following matrix representation:

$$(1.1.4) \qquad BY_t + \Gamma X_t = U_t \qquad t = 1,2,\ldots.T$$

where

$$T_t = \begin{bmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{Mt} \end{bmatrix}, \quad X_t = \begin{bmatrix} x_{1t} \\ x_{2t} \\ \vdots \\ x_{Kt} \end{bmatrix}, \quad U_t = \begin{bmatrix} u_{1t} \\ u_{2t} \\ \vdots \\ u_{Mt} \end{bmatrix}$$

$$(M \times 1) \qquad (K \times 1) \qquad (M \times 1)$$

$$B = \begin{bmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1M} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2M} \\ \vdots & \vdots & & \vdots \\ \beta_{M1} & \beta_{M2} & \cdots & \beta_{MM} \end{bmatrix}, \quad \Gamma = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1K} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2K} \\ \vdots & \vdots & & \vdots \\ \gamma_{M1} & \gamma_{M2} & \cdots & \gamma_{MK} \end{bmatrix}$$

$$(M \times M) \qquad (M \times K)$$

We assume that B is a nonsingular matrix. Hence we can solve (1.1.4) for Y_t to obtain the reduced form given by

$$(1.1.5) \qquad \begin{aligned} Y_t &= -B^{-1}\Gamma X_t + B^{-1}U_t \\ &= \Pi X_t + V_t \end{aligned}$$

Here $\Pi = -B^{-1}\Gamma$ and $V_t = B^{-1}U_t$

In (1.1.5)

$$\Pi = \begin{bmatrix} \pi_{11} & \pi_{12} \cdots\cdots & \pi_{1K} \\ \pi_{21} & \pi_{22} \cdots\cdots & \pi_{2K} \\ \vdots & \vdots & \vdots \\ \pi_{M1} & \pi_{M2} \cdots\cdots & \pi_{MK} \end{bmatrix}, \qquad V_t = \begin{bmatrix} v_{1t} \\ v_{2t} \\ \vdots \\ v_{Mt} \end{bmatrix}$$

$$(M \times K) \qquad\qquad (M \times 1)$$

The distinguishing feature of the reduced form [i.e. (1.1.5)] is that in each of its equations only one jointly dependent variable appears. This can be shown explicitly as follows:

$$
\begin{aligned}
y_{1t} &= \pi_{11}x_{1t} + \pi_{12}x_{2t} + \cdots + \pi_{1K}x_{Kt} + v_{1t} \\
y_{2t} &= \pi_{21}x_{1t} + \pi_{22}x_{2t} + \cdots + \pi_{2K}x_{Kt} + v_{2t} \\
&\vdots \\
y_{Mt} &= \pi_{M1}x_{1t} + \pi_{M2}x_{2t} + \cdots + \pi_{MK}x_{Kt} + v_{Mt}
\end{aligned} \tag{1.1.6}
$$

$$(t = 1,2,\ldots.T)$$

The π's represent the reduced form coefficients and the v's the reduced form disturbances. It may be observed that, in general, each reduced form disturbance is a linear function of all structural disturbances.

We assume that:

(1.1.7) $E(u_t)$ is a null vector and

(1.1.8) $E(U_t U_t') = \Sigma$ is a nonsingular matrix whose elements are finite and constant but unknown.

(1.1.9) the u's are independent between different observations in the sample, i.e. $E(u_s u_t')$ is a null matrix so long as s is not equal to t. $(s, t = 1, 2, \ldots. T)$

(1.1.10) the predetermined variables are generated by a stationary multivariate process which is independent of the stochastic process generating structural disturbances.

(1.1.11) $\text{plim} \frac{X'X}{T}$ is assumed to be positive definite, where

$$X = \begin{bmatrix} x_1' \\ x_2' \\ \vdots \\ x_T' \end{bmatrix} \quad \text{is the } T \times K \text{ matrix of}$$

observations on the predetermined variables.

Since the reduced form disturbances are linear combinations of the structural disturbances, it follows that $E(v_t) = 0$, and that the variance-covariance matrix of the reduced form disturbances, Ω is given by

(1.1.12)
$$\Omega = E(V_t V_t') = E[B^{-1} U_t U_t' (B^{-1})']$$
$$= B^{-1} \Sigma (B^{-1})'$$

2. IDENTIFICATION

Before attempting to discuss methods of estimating the unknown parameters in the model, $BY_t + \Gamma X_t + V_t$, it is necessary to consider the conditions that must be satisfied for any solution of these equations. This is generally termed as the problem of identification in SEM. As soon as our model is expanded to make for more than one structural relationship, the problem of identification emerges.

In the case of the Simultaneous Equation Model (SEM), it does not suffice to know the precise list of variables contained in the equation to be estimated but it is also necessary to know what variables are contained in the other simultaneously holding equations or even to have additional information about the equation in question. Empirical observations alone, no matter how extensive or complete, cannot give all the information about the parameters in question.

In developing a model, more than one structure or point in the parameter space may be consistent with the available data. For example, the structural model might imply that we have linear supply and demand functions, both of which involve quantity and price. In this case, no amount of observations, however large, on price and quantity can estimate the parameters of the demand equation or the supply equation uniquely. In other words, neither function is identifiable. If, however, our specification of the model includes a sufficient number of restrictions on the admissible parameter space, then a series of observations on the set of variables will permit identification of both demand and supply equations.

It is therefore imperative that the structural model must possess a priori identification properties which are determined independently of the parameters. An analysis of the identification properties of the model may help in the selection of appropriate estimation procedures. In general, economic theory usually leads to the identification of most structural relations. This is mainly due to the fact that the theory often stipulates that several independent variables are to be excluded from each structural relation so that overidentification is usually quite common.

The first systematic work in the area of identification was done by Koopmans (1949). In his article entitled, "Identification Problems In Economic Model Construction", he made the following opening remark: "Statistical inference, from observations to economic parameters, can be made in two steps: inference from the observations to the parameters of the assumed joint distribution of observations, and inference from that distribution to the parameters of the structural equations describing economic behaviour. The latter problem of inference is described by the term, "identification problem".

Other important early references on the identification problem include Koopmans and Hood (1953), Koopmans and Rubins and Leipnik (1950) and Basmann (1960). Fisher in his book entitled, "The Identification Problem", McGraw Hill Book Company, New York, (1966), provides a generalized treatment of the conditions for identifiability. Fisher (1966) looked in detail at identification through "exclusion restriction." i.e., some variables are included and some excluded. He also discussed identification through restrictions on the disturbance variance-covariance matrix, identification under homogenous linear restrictions and the complications of nonhomogenous restrictions, nonlinearities, and cross-equation restrictions.

Let

$$(1.2.1) \qquad Q_t = \alpha_o + \alpha_1 p_t + u_{1t} \qquad \text{(DEMAND)}$$

$$(1.2.2) \qquad Q_t = \beta_o + \beta_1 p_t + u_{2t} \qquad \text{(SUPPLY)}$$

where Q is the quantity demanded and supplied, p is the price, the u's are structural disturbances. As both of these equations involve the same variables, any linear combination of then, viz.,

$$(\lambda_1 + \lambda_2)Q_t = (\lambda_1\alpha_o + \lambda_2\beta_o) + (\lambda_1\alpha_1 + \lambda_2\beta_1)p_t + (\lambda_1 u_{1t} + \lambda_2 u_{2t})$$

involves the same variables and

$$\lambda_1 = \lambda \ , \quad \lambda_2 = 1 - \lambda \qquad \text{for } (o < \lambda \leq 1).$$

Now it is impossible to identify an empirical linear relation between p and Q as demand function or supply function or the above linear combination of them. Even the specification that $\alpha_1 < o$ and $\beta_1 > o$ does not help because the λ's are arbitrary constants and can be so appropriately chosen as to keep intact the sign of α and β.

Therefore, while comparing any two equations of the complete system we must observe that the equations have at least one variable which is not common to the other. It means that equations in the system must exclude some variables. The equations (1.2.1) and (1.2.2) become identifiable if we modify them as follows:

$$(1.2.3) \qquad Q_t = \alpha_0 + \alpha_1 P_t + \alpha_2 I_t + u_{1t} \qquad \text{(DEMAND)}$$

$$(1.2.4) \qquad Q_t = \beta_0 + \beta_1 P_t + \beta_2 P_{t-1} + u_{2t} \qquad \text{(SUPPLY)}$$

where Q_t and P_t are endogenous and I_t and P_{t-1} are predetermined.

Since these are endogenous variables among the explanatory variables in the Simultaneous Equation Model (SEM), (OLS) estimators of the structural coefficients are not consistent, at least in general. We may try to estimate the structural coefficients by way of the reduced form. Since in the reduced form equations, the explanatory variables are represented by the predetermined variables of the system, then (OLS) estimators of the reduced form coefficients are consistent. The question then is whether we can derive estimates of structural coefficients from the consistent estimates of the reduced form coefficients. In other words, a structural parameter is identified if and only if it can be deduced uniquely from the reduced form parameters.

The reduced form equation given by

$$Y_t = \Pi X_t + V_t \, ,$$

represents the unrestricted version of these equations, while the form given by

$$Y_t = -B^{-1}\Gamma X_t + B^{-1}U_t$$

represents the restricted version. Whenever there is one-to-one correspondence between the restricted and unrestricted parameters we have exact identification. On the other hand, when the number of the unrestricted coefficients exceeds the number of the restricted parameters and there is no unique solution, we have overidentification. If the number of unrestricted coefficients is insufficient for the solution, we have underidentification. We say that an equation is identified if it is either exactly identified or overidentified.

In his articles "A Simple Forecasting Model For The U.S. Economy" (1955) and "Underidentification, Structural Estimation, and Forecasting" (1960), T.C. Liu takes issue with the emphasis placed on overidentification in the estimation of simultaneous equation models. He argued that economic theory supports the fact that underidentification is the usual one. He further claimed that apparently "reasonable" overidentified structures have been obtained probably only because the specification errors have cancelled one another and as a consequence of the prevalence of underidentified structures, the least squares reduced form equations are likely to be the best forecasting results.

Therefore, says Lui, the current emphasis on techniques of estimation in overidentified systems is entirely misplaced; structural estimation is generally not possible in simultaneous systems, and only reduced forms can be obtained. Furthermore, unrestricted least squares estimation of the reduced form is the appropriate method, for restricting the reduced form by the a priori restrictions not only adds no more information, it is positively harmful, as it adds misinformation. Forecasting should therefore always be done with the unrestricted reduced form which, as is well known, has smallest error variance for the sample observations, as there is no reason to expect mistakenly oversimplified and restricted models to do as well or better.

Fisher in his article entitled "On the Cost of Approximate Specification in Simultaneous Equation Estimation" (1961) examined Liu's objections to simultaneous equation methods. Fisher claimed that the problem is not the discontinuous one of exact or overidentification if the restrictions hold exactly and underidentification if they do not, but rather one of diminishing estimation inconsistency as the restrictions are better and better approximations. However, we shall not indulge in this kind of controversy and shall assume that structural equations under study are not underidentified.

The following model serves to illustrate the notation of identification. Consider the simplified Demand and Supply model for a given commodity:

$$Q_t = \alpha_o + \alpha_1 P_t + \alpha_2 I_t + u_{1t} \qquad \text{(DEMAND)} \tag{1.2.5}$$

$$Q_t = \beta_o + \beta_1 P_t + u_{2t} \qquad \text{(SUPPLY)} \tag{1.2.6}$$

$$\alpha_1 \leq o, \quad \alpha_2 \geq o, \quad \beta_1 \geq o \qquad \text{where}$$

Q is the equilibrium quantity exchanged on the market.

P is the equilibrium price.

I is the income of the consumer.

The variables Q and P are endogenous and I is exogenous. The α's and β's are the parameters to be estimated. The u's are random disturbances, and t represents a specific time period.

Now equations (1.2.5) and (1.2.6) can be written in the pattern of $BY_t + \Gamma X_t = U_t$ as follows:

$$\begin{bmatrix} 1 & -\alpha_1 \\ 1 & -\beta_1 \end{bmatrix} \begin{bmatrix} Q_t \\ P_t \end{bmatrix} + \begin{bmatrix} -\alpha_o & -\alpha_2 \\ -\beta_o & o \end{bmatrix} \begin{bmatrix} 1 \\ I_t \end{bmatrix} = \begin{bmatrix} U_{1t} \\ U_{2t} \end{bmatrix} \tag{1.2.7}$$

The reduced form of (1.2.7) is given by

$$(1.2.8) \qquad \begin{bmatrix} Q_t \\ P_t \end{bmatrix} = \begin{bmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{bmatrix} \begin{bmatrix} 1 \\ I_t \end{bmatrix} + \begin{bmatrix} V_{1t} \\ V_{2t} \end{bmatrix}$$

where

$$\begin{bmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{bmatrix} = -B^{-1}\Gamma = \begin{bmatrix} 1 & -\alpha_1 \\ 1 & -\beta_1 \end{bmatrix}^{-1} \begin{bmatrix} -\alpha_0 & -\alpha \\ -\beta_0 & 0 \end{bmatrix}$$

$$= -\left(\frac{1}{\alpha_1-\beta_1}\right) \begin{bmatrix} -\beta_1 & \alpha_1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} \alpha_0 & -\alpha_2 \\ -\beta_0 & 0 \end{bmatrix}$$

$$= -\left(\frac{1}{\alpha_1-\beta_1}\right) \begin{bmatrix} (\alpha_0\beta_1 - \alpha_1\beta_0) & \alpha_2\beta_1 \\ (\alpha_0 - \beta_0 & \alpha_2 \end{bmatrix}$$

and

$$\begin{bmatrix} v_{1t} \\ v_{2t} \end{bmatrix} = B^{-1} \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} = \left(\frac{1}{\alpha_1-\beta_1}\right) \begin{bmatrix} -\beta_1 & \alpha_1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

$$= \left(\frac{1}{\alpha_1-\beta_1}\right) \begin{bmatrix} -\beta_1 u_{1t} + \alpha_1 u_{2t} \\ -u_{1t} + u_{2t} \end{bmatrix}$$

From (1.2.8) we can rewrite the reduced form as

$$(1.2.9) \qquad Q_t = \pi_{11} + \pi_{12} I_t + v_{1t}$$

$$(1.2.10) \qquad P_t = \pi_{21} + \pi_{22} I_t + v_{2t}$$

To determine whether the α's and the β's can be expressed in terms of the π's, we substitute for Q_t and P_t from the reduced form

(1.2.9) and (1.2.10). Hence we have

$$\text{(DEMAND)} \quad (\pi_{11} + \pi_{12}\, I_t + v_{1t}) = \alpha_0 + \alpha_1(\pi_{21} + \pi_{22}\, I_t + v_{2t}) + \alpha_2 I_t + u_{1t}$$

$$\text{(SUPPLY)} \quad (\pi_{11} + \pi_{12}\, I_t + v_{1t}) = \beta_0 + \beta_L(\pi_{21} + \pi_{22}\, I_t + v_{2t}) + u_{2t}$$

Therefore,

$$\pi_{11} + \pi_{12}\, I_t = (\alpha_0 + \alpha_1\, \pi_{21}) + (\alpha_1\, \pi_{22} + \alpha_2) I_t \qquad \text{(DEMAND)}$$

$$\pi_{11} + \pi_{12}\, I_t = (\beta_0 + \beta_1\, \pi_{21}) + \beta_1\, \pi_{22}\, I_t \qquad \text{(SUPPLY)}$$

(By virture of the definition of v_{1t} and v_{2t} given previously, the stochastic disturvances in each equation cancel out each other).

The qualities from the Demand equation are

$$\pi_{11} = \alpha_0 + \alpha_1\, \pi_{21} \;; \quad \pi_{12} = \alpha_1\, \pi_{22} + \alpha_2$$

Since these are two equations, we cannot solve for the three unknowns α_0, α_1 and α_2.

From the supply equation,

$$\pi_{11} = \beta_0 + \beta_1\, \pi_{21} \quad ; \quad \pi_{12} = \beta_1\, \pi_{22} \;,$$

hence, $\quad \beta_0 = \pi_{11} - \dfrac{\pi_{12}}{\pi_{22}}\, \pi_{21} \quad$ and $\quad \beta_1 = \dfrac{\pi_{12}}{\pi_{22}}$

Therefore, the Demand equation is underidentified and the supply equation is exactly identified.

Let us consider once again the general representation of the M structural equations given by $BY_t + \Gamma X_t = U_t$ and the reduced form given by -

$$Y_t = \Pi X_t + V_t .$$

Substituting for Y_t from the reduced form into the structural form to obtain

$$B\Pi X_t + BV_t + \Gamma X_t = U_t$$

$$\therefore \quad B\Pi X_t = -\Gamma X_t$$

(1.2.11) $\qquad \therefore \quad B\Pi = -\Gamma$

We shall use the relation (1.2.11) to derive a general identification rule for each structural equation. Expressing $B\Pi = -\Gamma$ in expanded form gives

(1.2.12)
$$\begin{bmatrix} \beta_{11} & \beta_{12} \cdots \beta_{1M} \\ \beta_{21} & \beta_{22} \cdots \beta_{2M} \\ \vdots & \vdots \quad\;\; \vdots \\ \beta_{M1} & \beta_{M2} \cdots \beta_{MM} \end{bmatrix} \begin{bmatrix} \Pi_{11} & \Pi_{12} \cdots \Pi_{1K} \\ \Pi_{21} & \Pi_{22} \cdots \Pi_{2K} \\ \vdots & \vdots \quad\;\; \vdots \\ \Pi_{M1} & \Pi_{M2} \cdots \Pi_{MK} \end{bmatrix} = - \begin{bmatrix} \gamma_{11} & \gamma_{12} \cdots \gamma_{1K} \\ \vdots & \vdots \quad\;\; \vdots \\ \gamma_{M1} & \gamma_{M2} \cdots \gamma_{MK} \end{bmatrix}$$

If we consider the m^{th} equation, we have from (1.2.12)

(1.2.13)
$$\begin{bmatrix} \beta_{m1} & \beta_{m2} \cdots \beta_{mM} \end{bmatrix} \begin{bmatrix} \Pi_{12} & \Pi_{12} \cdots \Pi_{1K} \\ \Pi_{21} \cdots\cdots & \Pi_{2K} \\ \Pi_{M1} \cdots\cdots & \Pi_{MK} \end{bmatrix} = - \begin{bmatrix} \gamma_{m1} & \gamma_{m2} \cdots \gamma_{mK} \end{bmatrix}$$

or

(1.2.14) $\qquad \beta_m \Pi = - \gamma_m$

If all of the endogenous and predetermined variables of the system do not appear in the m^{th} equation, some of the β's and some of the γ's will be zero.

(1.2.15). Let M^{Δ} = number of included endogenous variables in the m^{th} equation where

$M^{\Delta\Delta} = M - M^{\Delta}$ be number of excluded endogenous variables.

K^* = number of predetermined variables which appear in the m^{th} equation and

$K^{**} = K - K^*$ which are not included.

Let us assume without any loss of generality that it is possible to partition β_m and γ_m as follows:

(1.2.16) $$\beta_m = \left[\beta_{\Delta} \quad 0_{\Delta\Delta}\right]$$

(1.2.17) $$\gamma_m = \left[\gamma_{*} \quad 0_{**}\right]$$

where

$$\beta_{\Delta} = \left[\beta_{m1}, \beta_{m2} \ldots, \beta_{mM^{\Delta}}\right] \quad \text{is } 1 \times M^{\Delta}$$

$$0_{\Delta\Delta} = \left[0, \quad 0 \ldots\ldots, 0\right] \quad \text{is } 1 \times M^{\Delta\Delta}$$

$$\gamma_{*} = \left[\gamma_{m1}, \gamma_{m2} \ldots, \gamma_{mK^*}\right] \quad \text{is } 1 \times K^*$$

$$0_{**} = \left[0, \quad 0 \ldots\ldots, 0\right] \quad \text{is } 1 \times K^{**}$$

The matrix Π can be portioned conformably into

$$(1.2.18) \qquad \Pi = \begin{bmatrix} \Pi_{\Delta,*} & \Pi_{\Delta,**} \\ \Pi_{\Delta\Delta,*} & \Pi_{\Delta\Delta,**} \end{bmatrix}$$

where

	K^* column	K^{**} column
M^{Δ} rows	$\Pi_{\Delta,*}$	$\Pi_{\Delta,**}$
$M^{\Delta\Delta}$ rows	$\Pi_{\Delta\Delta,*}$	$\Pi_{\Delta\Delta,**}$

From (1.2.16), (1.2.17) and (1.2.18) we can now write (1.2.14) as

$$(1.2.19) \qquad \left[\beta_{\Delta}\ ,\ 0_{\Delta\Delta}\right] = \begin{bmatrix} \Pi_{\Delta,*} & \Pi_{\Delta,**} \\ \Pi_{\Delta\Delta,*} & \Pi_{\Delta\Delta,**} \end{bmatrix} = \left[\gamma_{*}\ ,\ 0_{**}\right]$$

which gives

$$(1.2.20) \qquad \beta_{\Delta}\,\Pi_{\Delta,*} = -\gamma_{*}$$

$$(1.2.21) \qquad \beta_{\Delta}\,\Pi_{\Delta,**} = 0_{**}$$

Now since one of the β's in each structural equation equals unity, (1.2.20) and (1.2.21) involve $M^{\Delta}-1$ unknown β's and K^* unknown γ's. In particular (1.2.21) contains altogether K^{**} equations, one for each element of the $1 \times K^{**}$ vector. It is obvious that if we want a solution for the $M^{\Delta}-1$ unknown elements of β_{Δ}, we need at least $M^{\Delta}-1$ equations. This means that we require that

(1.2.22) $\qquad K^{**} \geq M^{\Delta} - 1.$ The relation (1.2.22) is known as the order condition for identifiability.

The order condition for identification given by (1.2.22) is only a necessary condition; it is not a sufficient condition, since the K^{**} equations in (1.2.21) may not be independent, i.e., it may happen that the equations (1.2.21) contain fewer than $M^{\Delta} - 1$ different pieces of information about the relation between the β's and the Π's. Consequently, a necessary and sufficient condition for identification is that the number of independent equations in (1.2.21) is $M^{\Delta} - 1$, i.e., if and only if

(1.2.23) $$\operatorname{rank}(\Pi_{\Delta,**}) = M^{\Delta} - 1.$$

The relation (1.2.24) is known as the rank condition for identifiability.

The rank and order condition for identification derived above utilize only "exclusion restrictions". For a discussion of other types of restrictions and derivation of the corresponding appropriate conditions, see Fisher (1966).

Investigating the rank of $\Pi_{\Delta,**}$ is difficult. However, an alternative rank condition is the following. Partition the matrices of structural coefficients conformably to the partitioning of β_m and γ_m as follows:

(1.2.24) $$B = \begin{bmatrix} \beta_{\Delta} & 0_{\Delta\Delta} \\ B_{\Delta} & B_{\Delta\Delta} \end{bmatrix} ; \quad \Gamma = \begin{bmatrix} \gamma_{*} & 0_{**} \\ \Gamma_{*} & \Gamma_{**} \end{bmatrix}$$

where the dimensions of β_{Δ}, $0_{\Delta\Delta}$, γ_{*} and 0_{**} have been specified previously and

$$(1.2.25) \qquad \begin{aligned} B_{\Delta} &\text{ is } (M-1) \times M^{\Delta} \\ B_{\Delta\Delta} &\text{ is } (M-1) \times M^{\Delta\Delta} \\ \Gamma_{*} &\text{ is } (M-1) \times K^{*} \\ \Gamma_{*} &\text{ is } (M-1) \times K^{**} \end{aligned}$$

It should be noted that $B_{\Delta\Delta}$ and Γ_{**} represent matrices of the structural coefficients for variables omitted from the m^{th} equation but included in the other structural equations

If we define a new matrix A as follows

$$(1.2.26) \qquad A = \left[B_{\Delta\Delta} \, , \; \Gamma_{**}\right] ,$$ then it can be shown that

$$(1.2.27) \qquad \text{Rank } (\Pi_{\Delta,**}) = \text{rank } (A) - M^{\Delta\Delta} .$$

We show this by defining

$$(1.2.28) \qquad A_{*} = \begin{bmatrix} 0_{**} & 0_{\Delta\Delta} \\ \Gamma_{**} & B_{\Delta\Delta} \end{bmatrix}$$

It is clear that rank A_{*} and rank A is the same since the rank of a matrix is unaltered by switching columns or by enlarging it by a row of zeros. Now we can write A_{*} as

$$(1.2.29) \qquad A_{*} = \begin{bmatrix} B_{\Delta} & 0_{\Delta\Delta} \\ B_{\Delta} & B_{\Delta\Delta} \end{bmatrix} \begin{bmatrix} -\Pi_{\Delta,**} & 0_{\Delta.\Delta\Delta} \\ -\Pi_{\Delta\Delta,**} & I_{\Delta\Delta} \end{bmatrix}$$

where $0_{\Delta,\Delta\Delta}$ is a $M^{\Delta} \times M^{\Delta\Delta}$ matrix of zeros and $I_{\Delta\Delta}$ is an $M^{\Delta\Delta} \times M^{\Delta\Delta}$ identity matrix.

From (1.2.29)

$$A_* = \begin{bmatrix} -\beta_\Delta \Pi_{\Delta,**} & 0_{\Delta\Delta} \\ -(-B_\Delta \Pi_{\Delta**} - B_{\Delta\Delta} \Pi_{\Delta\Delta**} & B_{\Delta\Delta} \end{bmatrix}$$

But from (1.2.21)

$$-\beta_\Delta \Pi_{\Delta,**} = 0_{**}$$ and from the relation

$$B\Pi = -\Gamma$$ it follows that:

$$-B_\Delta \Pi_{\Delta**} - B_{\Delta\Delta} \Pi_{\Delta\Delta**} = \Gamma_{**}$$

Since the rank of a matrix is unaltered by premultiplication by a nonsingular matrix, then

$$\text{rank}(A_*) = \text{rank}(B^{-1}A_*) \quad \text{i.e.,}$$

$$\text{rank}(A_*) = \text{rank}\begin{bmatrix} -\Pi_{\Delta,**} & 0_{\Delta,\Delta\Delta} \\ -\Pi_{\Delta\Delta,**} & I_{\Delta\Delta} \end{bmatrix}$$

$$= \text{rank}\begin{bmatrix} -\Pi_{\Delta**} & 0_{\Delta,\Delta\Delta} \\ -\Pi_{\Delta\Delta,**} & I_{\Delta\Delta} \end{bmatrix}\begin{bmatrix} I_{**} & 0_{**\Delta\Delta} \\ \Pi_{\Delta\Delta,**} & I_{\Delta\Delta} \end{bmatrix}$$

$$= \text{rank}\begin{bmatrix} -\Pi_{\Delta**} & 0_{\Delta,\Delta\Delta} \\ 0_{\Delta\Delta,**} & I_{\Delta\Delta} \end{bmatrix}$$

$$= \text{rank}(\Pi_{\Delta**}) + M^{\Delta\Delta}$$

where $0_{**,\Delta\Delta}$ and $0_{\Delta\Delta,**}$ are zero matrices of order $K^{**} \times M^{\Delta\Delta}$ and $M^{\Delta\Delta} \times K^{**}$ respectively, and I_{**} is an identity matrix of order $K^{**} \times K^{**}$.

The order and rank conditions provide the following general rule for determining the identification of a structural equation:

1. If $K^{**} > M^{\Delta} - 1$ and rank $(\Pi_{\Delta **}) = M^{\Delta} - 1$ we have overidentification.

2. If $K^{**} = M^{\Delta} - 1$ and rank $(\Pi_{\Delta **}) = M^{\Delta} - 1$ we have exact identification.

In all other cases the structural equation is underidentified.

CHAPTER 2

METHODS OF STRUCTURAL ESTIMATION

1. There are chiefly two approaches to estimate the coefficients of the structural equations, namely (a) single-equation methods, also known as limited information methods, and (b) complete system methods, also known as full information methods. In the single-equation methods, we estimate each equation in the system individually, taking into account any restrictions placed on that equation to be estimated without worrying about the restrictions on the other equations of the system. In the complete system methods, we estimate all the equations in the model simultaneously, taking into account all restrictions on such equations.

Ideally, we should use the full-information methods in order to preserve the spirit of the simultaneity of the structural equations. In practice, however, such methods are rarely used for a variety of reasons. For instance, methods such as Full Information Maximun Likelihood (FIML) often lead to equations which are highly non-linear and hence difficult to solve even in this era of high-speed computers. The major criticism of the limited information methods is that they fail to utilize adequately the simultaneity of the equations in the model and, therefore they may not be generally as efficient, even asymptotically as full information methods.

It is well known that the application of ordinary least squares (OLS) to the structural equations of a simultaneous equation model yields inconsistent estimators except in the very special case of fully recursive models. The inconsistency of the (OLS) estimators is a direct consequence of the fact that some of the explanatory variables are typically correlated with the structural disturbances. A variety of estimation procedures that provide consistent estimators are available in the literature. A brief account of these estimation procedures is presented in this chapter.

2. INDIRECT LEAST SQUARES (ILS) ESTIMATION

In this method we start with estimation of the coefficients of the reduced form by least squares methods, and from these estimates derive desirable estimates of the coefficients of the structural equation.

We consider the model described in Chapter I, i.e.,

$$BY_t + \Gamma X_t = U_t \qquad (t = 1,2,\ldots.T) \tag{2.2.1}$$

with reduced form given by

$$Y_t = \Pi X_t + V_t \tag{2.2.2}$$

We consider a single structural equation of the model described in (2.2.1) and assume that the equation is identified.

Let this structural equation be denoted by

$$\beta y_t + \gamma x_t = u_{1t} \qquad (t = 1,2,\ldots T) \tag{2.2.3}$$

In keeping with the notation used in Chapter I, let

$$\beta = \left[\beta_{11},\beta_{12}\ldots.\beta_{1M^\Delta} \;\; 0_{1,M^\Delta+1},\ldots.0_{1M}\right] \qquad \text{and} \tag{2.2.4}$$

$$\gamma = \left[\gamma_{11}\ldots\ldots\ldots\gamma_{1K*} \;\; 0_{1,K*+1}\ldots.0_{1K}\right] \qquad \text{and} \tag{2.2.5}$$

$$\beta_{11} = -1 \tag{2.2.6}$$

(2.2.7)
$$\Pi = \begin{bmatrix} \Pi_{\Delta,*} & \Pi_{\Delta**} \\ \Pi_{\Delta\Delta,*} & \Pi_{\Delta\Delta,**} \end{bmatrix}$$

where the dimensions of the submatrices of Π are given in (1.2.18).

As in (1.2.16) and (1.2.17), partition β and γ as follows:

(2.2.8)
$$\beta = \begin{bmatrix} \beta_{\Delta} & 0_{\Delta\Delta} \end{bmatrix}$$

(2.2.9)
$$\gamma = \begin{bmatrix} \gamma_{*} & 0_{**} \end{bmatrix}$$

and from the relation $B\Pi = -\Gamma$, we obtain the relation between Π β and γ, i.e.

(2.2.10)
$$\beta_{\Delta}\,\Pi_{\Delta*} = -\gamma_{*}$$

(2.2.11)
$$\beta_{\Delta}\,\Pi_{\Delta,**} = 0_{**}$$

[see (1.2.19)] .

Since we specified that equation (2.2.3) is identified, then

(2.2.12)
$$\operatorname{rank}(\Pi_{\Delta**}) = M^{\Delta} - 1$$

Let the matrix P denote the unrestricted least-squares estimates of Π and arrange the partition of P to conform to the partition of Π in (2.2.7), i.e.

(2.2.13)
$$\begin{bmatrix} P_{\Delta,*} & P_{\Delta,**} \\ P_{\Delta\Delta,*} & P_{\Delta\Delta,**} \end{bmatrix},$$

After obtaining the least squares estimates P of Π, we shall show that it is possible to determine desirable estimates of the structural coefficients when

(2.2.14) $\quad K^{**} = M^{\Delta} - 1 \quad$ (i.e. the case of exact identification).

We shall show also that the determination from the matrix P of estimates of structural coefficients is not possible in the same manner when

(2.2.15) $\quad K^{**} > M^{\Delta} - 1 \quad$ (i.e. the case of exact overidentification)

Consider the estimators $\hat{\beta}_{\Delta}$ and $\hat{\gamma}_{*}$ defined by

(2.2.16) $\quad -\hat{\beta}_{\Delta} P_{\Delta,*} = \hat{\gamma}_{*}$

(2.2.17) $\quad \hat{\beta} P_{\Delta,**} = 0_{**}$

(2.2.18) $\quad \hat{\beta}_{11} = -1$

Although in the case of exact or overidentification, rank $(\Pi_{\Delta,**}) = M^{\Delta} - 1$, the rank of the estimating matrix $P_{\Delta,**}$ differs. In the case where $K^{**} = M^{\Delta} - 1$, the rank of the $M^{\Delta} \times K^{**}$ matrix $P_{\Delta,**}$ will be $M^{\Delta} - 1$. Having computed $P_{\Delta,**}$, we may use (2.2.17) to determine the elements of $\hat{\beta}_{\Delta}$ up to a factor of proportionality; and the normalization rule (2.2.18) to determine the elements of $\hat{\beta}_{\Delta}$ uniquely. Then insertion of $\hat{\beta}_{\Delta}$ into (2.2.16) determines the elements of $\hat{\gamma}_{*}$ uniquely. This method of estimation is known as Indirect Least Squares method - the structural estimates, are derived indirectly from the least-squares estimates of the reduced-form coefficients.

In the case where $K^{**} > M^{\Delta} - 1$, rank $(P_{\Delta,**}) > M^{\Delta} - 1$ and since $P_{\Delta,**}$ has M^{Δ} rows, rank $(P_{\Delta,**}) = M^{\Delta}$. Therefore, no nontrivial solution to $\hat{\beta} P_{\Delta,**} = 0_{**}$ exists and hence the indirect least-squares method of estimation cannot be applied.

Thus, it should be noted that the (ILS) procedure holds only when the structural equation is exactly identified. Since economic models usually have many restrictions, overidentification rather than exact identification is the prevailing condition. Hence, methods of estimation other than (ILS) have been developed to handle the case of overidentification.

3. TWO-STAGE LEAST SQUARES (2SLS) ESTIMATION

We wish to estimate the following structural equation

$$\beta_{11}y_{1t} + \beta_{12}y_{2t} + \dots \beta_{1M}y_{Mt} + \gamma_{11}x_{1t} + \gamma_{12}x_{2t} + \dots \gamma_{1K}X_{Kt} = u_{1t} \tag{2.3.1}$$

which is assumed to be identified and is the first equation of the model (2.2.1). Suppose β_{11} equals unity, and that the included jointly dependent and predetermined variables are

$$y_{1t}, y_{2t} \dots y_{M^{\Delta}t} \quad \text{and} \quad x_{1t}, x_{2t} \dots x_{K^*t}$$

Then we can rewrite (2.3.1) as

(2.3.2) $$y_{1t} = -\beta_{12}y_{2t} - \beta_{13}y_{3t} - \cdots - \beta_{1M^\Delta}y_{M^\Delta t} - \gamma_{11}x_{1t} \cdots \gamma_{1K^*}x_{K^*t} + u_{1t}.$$

In matrix form (2.3.2) can be written as

(2.3.3) $$y_1 = Y_1\beta_1 + X_1\gamma_1 + U_1$$

where y_1 is the $T \times 1$ vector of observations on the "dependent" endogenous variable, Y_1 is a $T \times (M^\Delta - 1)$ matrix of observations on the other included endogenous variables.

X_1 is a $T \times K^*$ matrix of observations on the included exogenous variables.

β_1 and γ_1 are $(M^\Delta - 1) \times 1$ and $K^* \times 1$ vectors of parameters, respectively.

U_1 is a $T \times 1$ vector of disturbances.

If we let $Z_1 = [Y_1 \; , X_1]$ and

$$\delta_1 = \begin{bmatrix} \beta_1 \\ \gamma_1 \end{bmatrix}$$ then

equation (2.3.3) can be written as (2.3.4) $y_1 = Z_1\delta_1 + U_1$

where δ_1 is an $(M^\Delta - 1 + K^*) \times 1$ vector.

Now the reduced form of the jointly dependent variables appearing as explanatory variables in (2.3.3) is

(2.3.5) $$Y_1 = X\Pi_1 + V_1 \; .$$

Inserting (2.3.5) in (2.3.3) and rearranging gives

(2.3.6) $$y_1 = X\Pi_1 \; \beta_1 + X_1\gamma_1 + (u_1 + V_1\beta_1)$$

where X is the $T \times K$ matrix of observations on all the predetermined variables in the complete system given in (2.2.1). Now if Π_1 were known, least squares applied to (2.3.6) would yield consistent estimates of β_1 and γ_1. Of course Π_1 is not known. Therefore, in the first stage of the (2SLS) methods we obtain the (OLS) estimates of Π_1 by regressing each variable in Y_1 on X. This yields

$$(2.3.7) \qquad \hat{\Pi}_1 = (X'X)^{-1} X'Y_1$$

and the calculated values in this regression are given by

$$(2.3.8) \qquad \hat{Y}_1 = X(X'X)^{-1} X'Y_1 .$$

In the second stage y_1 is regressed on $\hat{Y}_1$ and X_1. This yields the two-stage least squares normal equations:

$$(2.3.9) \qquad \begin{bmatrix} \hat{Y}_1'\hat{Y}_1 & \hat{Y}_1' X_1 \\ X_1'\hat{Y}_1 & X_1' X_1 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\gamma}_1 \end{bmatrix} = \begin{bmatrix} \hat{Y}_1' y_1 \\ X_1' y_1 \end{bmatrix}$$

which is a system of $M^{\Delta} - 1 + K^*$ equations in $M^{\Delta} - 1 + k^*$ unknowns and

$\begin{bmatrix} \hat{\beta}_1 \\ \hat{\gamma}_1 \end{bmatrix}$ denotes the (2SLS) estimates of $\begin{bmatrix} \beta_1 \\ \gamma_1 \end{bmatrix}$.

Now from (2.3.5), (2.3.7) and (2.3.8), let $\hat{V}_1$ be the matrix of estimated residuals from the reduced-form so that

$$(2.3.10) \qquad Y_1 = \hat{Y}_1 + \hat{V}_1$$

Then from the usual properties of least squares residuals, i.e. the (OLS) residuals are orthogonal to the fitted value of the dependent variable and to each explanatory variable, we have

$$(2.3.11) \qquad \hat{Y}_1 \hat{V}_1 = \hat{V}_1' \hat{Y}_1 = 0$$

$$X' \hat{V}_1 = \hat{V}_1' X = 0 \qquad \text{which implies } X_1' \hat{V}_1 = 0$$

Also

$$(2.3.11) \qquad \hat{Y}_1' \hat{Y}_1 = \hat{Y}'(Y_1 - \hat{V}_1) = \hat{Y}_1' Y_1 = Y_1' X(X'X)^{-1} X'Y_1$$

$$(2.3.13) \qquad \hat{Y}_1' \hat{Y}_1 = (Y_1 - \hat{V}_1)'(Y_1 - \hat{V}_1)$$

$$= Y_1' Y_1 - \hat{V}_1' \hat{V}_1$$

$$(2.3.14) \qquad \hat{Y}_1' X_1 = (Y_1 - \hat{V}_1)' X_1$$

$$= Y_1' X_1$$

and

$$(2.3.15) \qquad X_1' \hat{Y}_1 = X_1'(Y_1 - \hat{V}_1) = X_1' Y_1$$

$$(2.3.16) \qquad \hat{Y}_1' y_1 = (Y_1 - \hat{V}_1)' y_1 = Y_1' y_1 - \hat{V}_1' y_1$$

In view of the results (2.3.11) to (2.3.16), the normal equations of the (2SLS) estimator given by (2.3.9) can be written as

$$(2.3.19) \qquad \begin{bmatrix} Y_1'X(X'X)^{-1}X'Y_1 & Y_1' X_1 \\ X_1' Y_1 & X_1' X_1 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\gamma}_1 \end{bmatrix} = \begin{bmatrix} Y_1'X(X'X)^{-1}X'y_1 \\ X_1' y_1 \end{bmatrix}$$

or as

$$(2.3.18) \qquad \begin{bmatrix} Y_1' Y_1 - \hat{V}_1' \hat{V}_1 & Y_1' X_1 \\ X_1' Y_1 & X_1' X_1 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\gamma}_1 \end{bmatrix} = \begin{bmatrix} Y_1' y_1 - \hat{V}_1' y_1 \\ X_1' y_1 \end{bmatrix}$$

Using (2.3.10) we can write (2.3.3) as

$$(2.3.19) \qquad y_1 = \hat{Y}_1 \beta_1 + X_1 \gamma_1 + (U_1 + \hat{V}_1 \beta_1) \qquad \text{or,}$$

more compactly, as

$$(2.3.20) \qquad y_1 = \hat{Z}_1 \delta + (U_1 + \hat{V}_1 \beta_1)$$

where

$$(2.3.21) \qquad \hat{Z}_1 = \left[\hat{Y}_1 , X_1\right] \qquad \text{and} \qquad \delta = \begin{bmatrix} \beta_1 \\ \gamma_1 \end{bmatrix}$$

Now if we apply (OLS) to (2.3.21) we obtain the (2SLS) estimator in the form

$$(2.3.22) \qquad \hat{\delta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\gamma}_1 \end{bmatrix} = \left(\hat{Z}_1' Z_1\right)^{-1} \hat{Z}_1 y_1$$

From (2.3.20) and (2.3.22)

$$(2.3.23) \qquad \hat{\delta}_1 = \delta_1 + \left(\hat{Z}_1' Z_1\right)^{-1} \hat{Z}_1' U_1 \quad , \qquad \text{since}$$

$$\hat{Z}_1' \hat{V}_1 = 0$$

So,

$$(2.3.24) \qquad \text{plim } \hat{\delta}_1 = \delta_1 + \text{plim}\left(\frac{1}{T} \hat{Z}_1' \hat{Z}_1\right)^{-1} \text{plim}\left(\frac{1}{T} \hat{Z}_1' U_1\right)$$

since the variables in the system are stable, we assume that

$\text{plim}\left(\frac{1}{T}\hat{Z}_1'\hat{Z}_1\right)$ is finite.

$$\text{Now}\quad \text{plim}\left(\frac{1}{T}\hat{Z}_1' U_1\right) = \begin{bmatrix} \text{plim}\,\frac{1}{T}\hat{Y}_1' U_1 \\ \\ \text{plim}\,\frac{1}{T} X_1' U_1 \end{bmatrix}$$

$\text{plim}\left(\frac{1}{T}X_1' U\right) = 0$ since the X variables by assumption are uncorrelated with the disturbances in the limit.

$$\text{Now}\quad \text{plim}\left(\frac{1}{T}\hat{Y}_1' U_1\right) = \text{plim}\left(\frac{1}{T}\hat{\Pi}_1 X' U_1\right)$$

$$= 0$$

Therefore

(2.3.25) $\text{plim}\,\hat{\delta}_1 = \delta_1$ which establishes the consistency of the (2SLS) estimator

Also the (2SLS) estimator will in general be biased since

(2.3.26) $E(\hat{\delta}_1) = \delta_1 + E\left[(\hat{Z}_1'\hat{Z}_1)^{-1}\hat{Z}_1' U_1\right]$ from (2.3.23)

and $\hat{Z}_1$ is for finite samples correlated with U_1. Indeed, as pointed out by Dhrymes (1970), the expectation in (2.3.26) may not even exist

Alternately, we can obtain the (2SLS) estimator as follows:

Premultiply (2.3.4) by X' to obtain

(2.3.27) $X'y_1 = X'Z_1\,\delta_1 + X'U$

where X is the $T \times K$ matrix of observations on all the predetermined variables in the entire system with rank $(X) = K$.

The covariance matrix of the residuals of the transformed system (2.3.27) is given by

(2.3.28) $E(X'U_1U_1'X) = \sigma^2(X'X)$. The (G.L.S) estimator of δ from (2.3.27) is

(2.3.29) $\hat{\delta}_1 = \left[Z_1'X(X'X)^{-1}X'Z_1\right]^{-1}\left[Z_1'X(X'X)^{-1}X'y_1\right] = \left(\hat{Z}_1'Z_1\right)^{-1}\hat{Z}_1'y_1$

which is the (2SLS) estimator. Thus, the (2SLS) estimator of δ_1 is the (GLS) estimator of δ_1 from the transformed system (2.3.27).

We can actually make one more transformation of equation (2.3.27). We can find a nonsingular matrix R such that $(X'X) = RR'$. Now transform (2.3.27) by premultiplying throughout by R^{-1} to obtain

(2.3.30) $R^{-1}X'y_1 = R^{-1}X'Y_1\beta_1 + R^{-1}X'X_1\gamma_1 + R^{-1}X'U_1$.

The covariance matrix of residuals is now given by

(2.3.31) $E\left[R^{-1}X'UU'X(R^{-1})'\right] = \sigma^2 I.$

Equation (2.3.30) can be written compactly as

(2.3.32) $a_1 = A_1\delta_1 + q_1$ where

$a_1 \stackrel{\ell}{=} R^{-1}X'y_1$; $A_1 = R^{-1}X'Z_1 = \left[R^{-1}X'Y_1 , R^{-1}X'X_1\right]$

$q_1 = R^{-1}X'u_1$

The OLS) estimator applied to (2.3.32) is denoted by

(2.3.33) $d_1 = (A_1'A_1)^{-1}A_1'a_1$.

By some simple algebraic manipulations it can be shown that the (OLS) estimator of δ_1 in (2.3.30) is exactly the same as the (2SLS) estimator of δ_1 in (2.3.22).

ASYMPTOTIC PROPERTIES OF THE (2SLS) estimators

Now following Dhrymes (1970)

$$(2.3.34) \qquad \sqrt{T}\,(\hat{\delta}_1-\delta_1) = \left(\frac{A_1'A_1}{T}\right)^{-1} \frac{Z_1'X}{T}\ \frac{X'X}{T}\ \frac{X'U_1}{\sqrt{T}}$$

from the relations (2.3.32) and (2.3.33).

With the exception of $\frac{X'u_1}{\sqrt{T}}$, the probability limits of the quantities on the (R.H.S) of (2.3.34) are finite. Following Dhrymes (1970) once more, we establish the limiting distribution of $\frac{X'u_1}{\sqrt{T}}$ as follows.

Choose C_i a K-dimensional vector as the ith column of X'. The elements of C_i are by assumption either nonstochastic or, if stochastic, then independent of U_{i1} and

$$(2.3.35) \qquad \frac{X\ U_1}{\sqrt{T}} = \frac{1}{\sqrt{T}} \sum_{i=1}^{T} C_i\, u_{i1} \qquad \text{and}$$

$$(2.3.36) \qquad \mathrm{Cov}\,(C_i U_{u1}) = \sigma^2 C_i C_i'$$

Hence by an application of an appropriate Central Limit Thereom $\frac{X'U_1}{\sqrt{T}}$ is asymptotically normal with mean zero and covariance matrix

$$\sigma^2 \lim_{T\to\infty} \frac{1}{T} \Sigma\, C_i C_i' = \sigma^2 \lim_{T\to\infty} \left(\frac{X'X}{T}\right)$$

There

$$\sqrt{T}\,(\hat{\delta}_1-\delta_1) \sim N\left[0,\ \sigma^2 \operatorname{plim}\left(\frac{A_1 A_1}{T}\right)^{-1}\right]$$

4. k-CLASS ESTIMATORS

Thiel (1961) defined a family of estimators called the k-class extimator which are a generalization of the (2SLS) estimators defined in (2.3.18). The k-class estimators are given by.

$$(2.4.1)\quad \hat{\delta}_1(k) = \begin{bmatrix} \hat{\beta}_1(k) \\ \hat{\gamma}_1(k) \end{bmatrix} = \begin{bmatrix} Y_1'Y_1 - k\hat{V}_1'\hat{V}_1 & Y_1'X_1 \\ X_1'Y_1 & X_1'X_1 \end{bmatrix}^{-1} \begin{bmatrix} (Y_1-\hat{V}_1)'y_1 \\ X_1'y_1 \end{bmatrix}$$

where k is any scalar, stochastic or nonstochastic. For nonstochastic k, for $k = 0$ and $k = 1$ we have the (OLS) and the (2SLS) estimators respectively.

If k is stochastic, the k-class estimator will be consistent if and only if $\operatorname{plim}_{T\to\infty} k = 1$, i.e., if $\operatorname{plim}_{T\to\infty} k \neq 1$, the normal equations of the k-class estimator will converge to the normal equations of the (2SLS) estimators and in that case the k-class estimators will be consistent and will have the same asymptotic variance-covariance matrix as the (2SLS) estimators.

The k-class estimators described in (2.4.1) can be expressed compactly as

$$(2.4.2) \qquad \hat{\delta}_1(k) = \left[Z_1(I-kM)Z_1\right]^{-1} Z_1'(I-kM)y_1$$

where

$$(2.4.3) \qquad M = I - X(X'X)^{-1}X' \qquad \text{and}$$

$$Z_1 = [Y_1 \ , X_1] \qquad \text{as defined previously.}$$

Double k-class Estimator

To achieve more flexibility, Nagar (1962) proposed the double k-class, a family of limited information methods. The double k-class estimator of δ_1 is given by:

$$(2.4.4) \qquad \hat{\delta}_1(k_1,k_2) = \begin{bmatrix} Y_1'Y_1 - k_1\hat{V}_1'\hat{V}_1 & Y_1X_1 \\ X_1'Y_1 & X_1'X_1 \end{bmatrix}^{-1} \begin{bmatrix} Y_1' - k_2\hat{V}_1' \\ X_1' \end{bmatrix} y_1$$

where k_1 and k_2 are arbitrary scalars characterizing the estimator.

The Double k-class estimator of δ_1 can be expressed more compactly as

$$(2.4.5) \qquad \hat{\delta}_1(k_1,k_2) = \left[Z_1'(I-k_1M)Z\right]^{-1} Z_1'(I-k_2M)y_1$$

The Double k-class estimator will be consistent if

$$(2.4.6) \qquad \text{plim}(k_1-1) = \text{plim}(k_2-1) = 0$$

The h-class Estimator

Setting $k_1 = 1-h^2$ and $k_2 = 1-h$ we get the h-class estimator. Thus, we see that the h-class is a modification of the Double k-class estimators. The h-class estimator of δ_1 is given by

$$(2.4.7) \qquad \hat{\delta}_1(h) = \begin{bmatrix} Y_1'Y_1 - (1-h^2)\hat{V}_1'\hat{V}_1 & Y'X_1 \\ X_1'Y_1 & X_1'X_1 \end{bmatrix}^{-1} \begin{bmatrix} Y_1'-(1-h)\hat{V}_1' \\ X_1' \end{bmatrix} y_1$$

The k-class estimators of δ_1 can also be obtained from the Double k-class estimators by setting $k_1 = k_2 = k$. For $h = 1$ and $h = 0$ in (2.4.7) we obtain the (OLS) and the (2SLS) estimators of δ_1 respectively.

5. LIMITED-INFORMATION MAXIMUM-LIKELIHOOD METHOD (LIML)

The limited-information maximum-likelihood (LIML) estimator belongs to the k-class family and is consistent. This estimator is also called the least-variance-ratio (LVR) estimator, and Goldberger (1964) refers to it as the least generalized residual variance (LGRV) estimator.

In the (LIML) method due to Anderson and Rubin (1949, 1950), we consider the limited likelihood function of the M^{Δ} jointly dependent variables under the assumption of normality of the disturbances and maximize it subject to the restriction rank $(\Pi_{\Delta**}) = M^{\Delta} - 1$. The approach taken by Goldberger (1964) was to consider the likelihood function for the reduced form and maximize it subject to the implied restrictions on the reduced form. This approach reduces to maximizing the generalized residual variance $\left|T^{-1}V_{\Delta}'V_{\Delta}\right|$, where V_{Δ} is the matrix of reduced form residuals for the included endogenous variables, subject to the restriction $\beta_\Delta \Pi_{\Delta**} = 0_{**}$ (defined in Chapter I) on the reduced form.

Koopmans and Hood (1953) showed that we can obtain the results of Anderson and Rubin (1949, 1950) by the least-variance-ratio (LVR) principle and also demonstrated the equivalence of the two approaches. We obtain the (LIML) estimator by the (LVR) principle as follows:

Consider the first structural equations

(2.5.1) $$y_1 = Y_1\beta_1 + X_1\gamma_1 + U_1$$

defined in Chapter I and we can rewrite (2.5.1) as

(2.5.2) $$y_\Delta = Y_\Delta\beta_\Delta = X_1\gamma_1 + U_1$$ where

(2.5.3) $$Y_\Delta = [y_1 \quad Y_1] \quad ; \quad \beta_\Delta = \begin{bmatrix} 1 \\ -\beta_1 \end{bmatrix}$$

and

(2.5.4) $$y_\Delta = y_1 - Y_1\beta_1$$

which is a linear conbination of the jointly dependent variables appearing in equation (2.5.1), the coefficients of the combination being the unknown β_1 parameters. For observed values of y_Δ, we shall estimate γ_1 by the (OLS) method. This would lead to

(2.5.5) $$\hat{\gamma}_1 = (X_1'X_1)^{-1} X_1'y_\Delta$$ and the residual sum of squares would be

$$(2.5.6) \qquad y_\Delta' M_1 y_\Delta \qquad \text{where} \quad M_1 = I - X_1(X_1'X_1)^{-1}X_1'$$

$$= \beta_\Delta' Y_\Delta' Y_\Delta \beta_\Delta - \beta_\Delta' Y_\Delta' X_1 (X_1'X_1)^{-1} X_1' Y_\Delta \beta_\Delta$$

$$= \beta_\Delta' W_1 \beta_\Delta$$

where

$$(2.5.7) \qquad W_1 = Y_\Delta' Y_\Delta - Y_\Delta' X_1 (X_1' X_1)^{-1} X_1' Y_\Delta \, .$$

Similarly if y_Δ is regressed on all the predetermined variables $X = [X_1 , X_2]$ where X is a $T \times K$ matrix, the residual sum of squares would be

$$(2.5.8) \qquad y_\Delta' M y_\Delta \qquad \text{where} \quad M = I - X(X'X)^{-1} X' = \beta_\Delta' W \beta_\Delta$$

where

$$(2.5.9) \qquad W = Y_\Delta' Y_\Delta - Y_\Delta' X(X'X)^{-1} X' Y_\Delta$$

The addition of further explanatory variables will not increase the residual sum of squares, i.e., the second residual sum of squares will be no greater than the first since the second regression includes all the explanatory variables in the first regression, X_1, plus the set X_2. Hence the ratio $\lambda = \dfrac{\beta_\Delta' W_1 \beta_\Delta}{\beta_\Delta' W \beta_\Delta}$ can never be smaller than unity. The least-variance-ratio method suggests minimizing

$$\frac{\beta_\Delta' W_1 \beta_\Delta - \beta_\Delta' W \beta_\Delta}{\beta_\Delta' W \beta_\Delta} \, ,$$

or equivalently minimizing

$$\lambda = \frac{\beta_\Delta' W_1 \beta_\Delta}{\beta_\Delta' W \beta_\Delta} \quad . \tag{2.5.10}$$

In other words, the (LVR) method suggests that the estimate of β_Δ should be chosen to keep the reduction of the residual sum of squares as small as possible.

Differentiating λ with respect to β_Δ gives

$$\frac{\partial\lambda}{\partial\beta_\Delta} = \frac{\left(\beta_\Delta' W \beta_\Delta\right)\left(2W_1 \beta_\Delta\right) - \left(\beta_\Delta' W_1 \beta_\Delta\right)\left(2W \beta_\Delta\right)}{\left(\beta_\Delta' W \beta_\Delta\right)^2} \tag{2.5.11}$$

Setting $\frac{\partial\lambda}{\partial\beta_\Delta} = 0$ gives

$$\left(W_1 - \lambda W\right)\beta_\Delta = 0 \quad .$$

Hence the minimum value of the variance ratio (2.5.10) is given by the minimum root of the determinatal equation

$$\left|W_1 - \lambda W\right| = 0 \quad . \tag{2.5.12}$$

Since $W_1 - W \geq 0$, all the roots of equation (2.5.12) are ≥ 1. β_Δ obtained as the solution of the equations

$$\left(W_1 - \hat{\lambda} W\right) \beta_\Delta = 0 \tag{2.5.13}$$

where $\hat{\lambda}$ is the minimum root of (2.5.12) . Since equation (2.5.13) determines β_Δ only up to a multiplicative constant, the estimator $\hat{\beta}_\Delta$ of β_Δ is obtained from

$$\left(W_1 - \hat{\lambda} W\right) \hat{\beta}_\Delta = 0 \tag{2.5.14}$$

by setting the first element in $\hat{\beta}_\Delta$ equal to unity.

We now define

(2.5.15) $\hat{y}_\Delta = \hat{Y}_\Delta \hat{\beta}_\Delta$ and obtain γ_1 by regressing $\hat{y}_\Delta$ on X,

we get

(2.5.16) $$\hat{\gamma}_1 = (X_1' X_1)^{-1} X_1' Y_\Delta \hat{\beta}_\Delta \quad .$$

We can show that the (LIML) estimator is a member of the k-class family by choosing $k = \lambda$. We can write the k-class estimator of

$\delta_1(k) = \begin{bmatrix} \beta_1(k) \\ \gamma_1(k) \end{bmatrix}$ given earlier, i.e.

$$\hat{\delta}_1(k) = \begin{bmatrix} \hat{\beta}_1(k) \\ \hat{\gamma}_1(k) \end{bmatrix} = \begin{bmatrix} Y_1'Y_1 - k\hat{V}_1'\hat{V}_1 & Y_1'X_1 \\ X_1'Y_1 & X_1'X_1 \end{bmatrix}^{-1} \begin{bmatrix} (Y_1 - kV_1)^1 y_1 \\ X_1'y_1 \end{bmatrix} .$$

as

(2.5.17) $$\left(Y_1'Y_1 - k\hat{V}_1'\hat{V}_1\right)\hat{\beta}_1(k) + Y_1'X_1\hat{\gamma}_1(k) = \left(Y_1 - k\hat{V}_1\right)' y_1$$

and

(2.5.18) $$X_1' Y_1 \hat{\beta}_1(k) + (X_1'X_1)\hat{\gamma}_1(k) = X_1'y_1 \quad .$$

The solution of the equation (2.5.18) for $\hat{\gamma}_1(k)$ gives

(2.5.19) $\hat{\gamma}_1(k) = (X_1'X_1)^{-1} X_1' \left[y_1 - Y_1 \beta_1(k)\right]$ and

substituting (2.5.19) in equation (2.5.17), we get

(2.5.20) $$\left(Y_1'Y_1 - k\hat{V}_1'\hat{V}_1\right)\hat{\beta}_1(k) + Y_1'X_1\left(X_1'X_1\right)^{-1} X_1'Y_1\hat{\beta}_1(k)$$
$$= \left(Y_1 - k\hat{V}_1\right)^1 y_1 - Y_1'X_1\left(X_1'X_T\right)^{-1} X_1'y_1$$

or

$$\left(Y_1'M_1Y_1 - k\hat{V}_1'\hat{V}_1\right)\hat{\beta}_1(k) = \left(Y_1'M_1y_1 - k\hat{V}_1'\hat{y}_1\right)$$

where $\quad M_1 = I - X_1(X_1'X_1)^{-1}X_1'$

Also since $\quad \hat{V}_1 = Y_1 - \hat{Y}_1 \qquad$ (defined previously)

$$= Y_1 - X(X'X)^{-1}X'Y_1$$

$$= MY_1$$

where $\quad M = I - X(X'X)^{-1}X$, we can write (2.5.20) as

$$(2.5.21) \qquad \left(Y_1'MY_1 \;\; kY_1'MY_1\right)\hat{\beta}_1(k) = Y_1'M_1y_1 - kY_1'My_1 \;.$$

Since $Y_\Delta = [y_1 \;,\; Y_1]$ and $\beta_\Delta = \begin{bmatrix} 1 \\ -\beta_1 \end{bmatrix}$ it is easy to see that equations (2.5.19) and (2.5.16) are equivalent, and noting that $W_1 = Y_\Delta'M_1Y_\Delta$ and $W = Y_\Delta' M Y_\Delta$, it can be seen that (2.5.21) follows from (2.5.14).

Hence the (LIML) estimator is a k-class estimator with $k = \hat{\lambda}$. It should be observed that, in the case of LIML, k is a stochastic variable since it is obtained by minimizing (2.5.10).

Anderson and Rubin (1950) have shown that under general conditions the asymptotic distribution of $T(\hat{\lambda} - 1)$ is a $X^2(n)$ distribution with degrees of freedom equal to the degree of overidentification, $n = (K^{**} - M^\Delta + 1) =$ Since n and $2n$ are the expected value and variance of $X^2(n)$ respectively, it follows that

$$(2.5.22) \qquad A.E\,\sqrt{T}(\hat{\lambda} - 1) = \frac{K^{**} - M^\Delta + 1}{\sqrt{T}} \;,$$ and (A.E. denotes asymptotic expectation)

(2.5.23) asymptotic variance of $\sqrt{T}(\hat{\lambda}-1)$ is

$$\frac{2(K^{**}-M^{\Delta}+1)}{T}.$$

Hence, $\operatorname{plim}\sqrt{T}(\hat{\lambda}-1)=0$ and a fortori, $\operatorname{plim}(\hat{\lambda}-1)=0$.

We have shown that the LIML estimator is a k-class estimator with $k=\hat{\lambda}$. Hence by the condition derived earlier for the consistency of the k-class estimator, i.e., $\operatorname{plim}(k-1)=0$, it follows that the (LIML) estimator is consistent and it has the same asymptotic covariance matrix as the (2SLS) estimator.

In a subsequent chapter based on exact finite sample properties, we shall show that the (LIML) estimators do not possess finite moments. Fuller (1977) presents a modified version of the (LIML) estimator and demonstrates that this modified estimator possesses finite moments and is a member of the k-class family.

Fuller's modified (LIML) estimator is given by:

$$(2.5.24) \qquad \hat{\delta}_1 = \begin{bmatrix}\hat{\beta}_1\\ \hat{\gamma}_1\end{bmatrix} = \begin{bmatrix} Y_1'Y_1-\left(\hat{\lambda}-\frac{\alpha}{T-K}\right)Y_1'MY_1 & Y_1'X_1\\ X_1'Y_1 & X_1'X_1\end{bmatrix}\begin{bmatrix} Y_1y_1-\left(\hat{\lambda}-\frac{\alpha}{T-K}\right)Y_1My_1\\ X_1'y_1\end{bmatrix}$$

where $\hat{\lambda}$ is the smallest root of

$$\left|W_1-\lambda W\right|=0 \qquad \text{and} \quad \alpha>0 \text{ is a fixed real}$$

number.

RELATIONSHIP AMONG K-CLASS ESTIMATORS

Maeshiro (1966) working with a structural equation of the form,

(2.5.25) $$y_{1t} = \beta_{12}\, y_{2t} + \gamma_{11} x_{1t} + \ldots \gamma_{1K} x_{Kt} + u_{1t} \;,$$

deduced firstly that the coefficient estimates of the (2SLS) lie between those of the (OLS) and (LIML) estimates and, secondly that small changes in the sample data are likely to produce widely varying coefficient estimates by the (LIML) method than by the (OLS) and the (2SLS) methods. It should be noted that Maeshiro (1966) worked with the special case where only one endogenous variable appear on the (R.H.S.) of equation (2.5.25).

Oi (1969) working with the more general case where two or more jointly dependent variables appear on the (R.H.S.) of the structural equation, confirmed Maeshiro's results for the estimates of the coefficients of the jointly dependent variables but claimed that no simple inference is possible, except in the special case chosen by Maeshiro, for the estimates of the coefficients of the predetermined variables. Fisher (1966) looked at the sensitivity of the different k-class estimators to specification errors and found that none is uniformly superior to the others.

Kadiyala (1970) proved that the residual sum of squares, S, defined by

(2.5.26) $$S = [y_1 - Z_1\hat{\delta}(k)]^{1}\,[y_1 - Z_1\hat{\delta}(k)]$$

is a monotone increasing function of k for $0 \leq k < k^*$ where k^* is the smallest root of

(2.5.27) $\left| Y_1' M_1 Y_1 - k \, Y_1' M \, Y_1 \right| = 0$. It should be recalled that $\hat{\delta}_1(k)$ is the k-class estimator of $\begin{bmatrix} \beta_1 \\ \gamma_1 \end{bmatrix}$ of the equation

$$y_1 = Y_1 \beta_1 + X_1 \gamma_1 + U_1$$

$$= Z_1 \delta_1 + U_1 \qquad \text{where } Z_1 = [Y_1 \, , X_1]$$

and $\delta_1 = \begin{bmatrix} \beta_1 \\ \gamma_1 \end{bmatrix}$.

In (2.5.27) $\quad M = I - X(X'X)^{-1}X'$

$$M_1 = I - X_1(X_1'X)^{-1}X_1'$$

Kadiyala (1970) used the monotonicity of the residual sum of squares as a criterion for ordering any two consistent k-class estimators. His results indicated that the (2SLS) estimator is preferable over the (LIML) estimator and that if consistency is our criterion for rejecting the (OLS) estimator, then it could be achieved by choosing k arbitrarily close to zero, say $k = 1 - \varepsilon(n)$ where $0 < \varepsilon(n) < 1$ and $\varepsilon(n) \to 0$ as $n \to \infty$. He suggested that an optimal choice of k could be achieved by using an average of the three estimates rather than any one of them.

Farebrother's (1972) work is mainly a refinement of that of Maeshiro and Kadiyala. He showed that there is no noticeable gain by adhering to Kadiyala's suggestion of using the average of the three estimates over the use of any one of them. Farebrother (1972) examined the way in which the k-class estimator varies with k and derived an expression from which the main features of the graph of the k-class estimator can be constructed. Farebrother and Savin (1974) presented an algebraic analysis of the graphs of the k-class estimator and discussed the implications of the algebraic and statistical analysis for the selection of a k-class estimator.

Dhrymes (1969) obtained an identity between Double k-Class (DkC) and (2SLS) estimators which can be put in the following form:

$$\hat{\delta}_1(\mathrm{DkC}) = \delta_1(\mathrm{2SLS}) + \left[Z_1'(I-k_1M)Z_1\right]^{-1} \left[(1-k_2)Z_1'P_x y_1 - (1-k_1)Z_1'MZ\hat{\delta}_1(\mathrm{2SLS})\right] \tag{2.5.28}$$

where

$$P_x = X(X'X)^{-1}X'$$

$$M = I - X(X'X)^{-1}X'$$

V.K. Srivastava and R. Tiwari (1977) presented a simple proof of Dhrymes' identity, (2.5.28).

6. INSTRUMENTAL VARIABLES (IV) ESTIMATORS

It is well-known in the literature that the (2SLS) estimator is an (IV) estimator where $\hat{Y}_1 = Y_1 - \hat{V}$ is used as a set of instrumental variables. As usual, the equation to be estimated is given by

$$(2.6.1)\qquad y_1 = Y_1\beta_1 + X_1\gamma_1 + U_1 \quad ; \quad Z_1 = [Y_1, X_1]$$

$$= Z_1\delta_1 + U_1 \quad ; \quad \delta_1 = \begin{bmatrix} \beta_1 \\ \gamma_1 \end{bmatrix}$$

and

$$\hat{Y}_1 = X(X'X)^{-1} X'Y_1$$

$$Y_1 = \hat{Y}_1 + \hat{V}_1$$

Goldberger (1965) has shown that the k-class estimator $\hat{\delta}_1(k)$ can be interpreted as an (IV) estimator with $W_* = [Y_1 - k\hat{V}_1 , X_1]$ used as instruments. The (IV) estimator of δ_1 of (2.6.1) is given by

$$(2.6.2)\qquad \bar{\delta}_1 = \begin{bmatrix} (Y_1 - k\hat{V}_1)' Y_1 & (Y_1 - k\hat{V}_1)'X_1 \\ X_1'Y_1 & X_1'X_1 \end{bmatrix}^{-1} \begin{bmatrix} (Y_1 - k\hat{V}_1)' y_1 \\ X_1'y_1 \end{bmatrix}$$

and, as before,

$$(2.6.3)\qquad \hat{\delta}_1(k) = \begin{bmatrix} Y_1'Y_1 - k\hat{V}_1'\hat{V}_1 & Y_1'X_1 \\ X_1'Y_1 & X_1'X_1 \end{bmatrix}^{-1} \begin{bmatrix} (Y_1 - kV_1'V_1)' y_1 \\ X_1'y_1 \end{bmatrix}$$

Equations (2.6.2) and (2.6.3) are equivalent:

$$(Y_1 - k\hat{V}_1)'Y_1 = Y_1'Y_1 - k\hat{V}_1'(\hat{Y}_1 + \hat{V}_1) = Y_1'Y_1 - k\hat{V}_1'\hat{V}_1$$

since $\hat{V}_1' \hat{Y}_1 = 0$.

Also, $(Y_1 - k\hat{V}_1)' X_1 = Y_1' X_k$ since $\hat{V}_1' X_1 = 0$.

For the (IV) estimators to be consistent we require

$\text{plim}\ \frac{1}{T} W_*' U_1 = 0$, i.e. we need to show

(2.6.4) $$\text{plim}\ \frac{1}{T}\,(Y_1 - k\hat{V}_1)'\,U_1 = 0 \quad ; \text{ and}$$

(2.6.5) $$\text{plim}\ \frac{1}{T}(X_1'\,U_1) = 0 \ .$$

Now (2.6.5) is always satisfied by the conventional assumptions on the structural equations.

Since $Y_1 - k\hat{V}_1 = Y_1 - k(Y_1 - \hat{Y}_1) = (1-k)Y_1 + k\hat{Y}_1$, we have

$\text{plim}\ \frac{1}{T}(Y_1 - k\hat{V}_1) = \text{plim}(1-k)\ .\ \text{plim}\ \frac{1}{T} Y_1' U_1 + \text{plim}\, k\ .\ \text{plim}\ (\frac{1}{T}\hat{Y}_1' U_1)$.

Now $\text{plim}\ \frac{1}{T}\hat{Y}_1' U_1 = \text{plim}\ \frac{1}{T} Y_1' X(X'X)^{-1}\ .\ \text{plim}\ X' U_1 = 0$, and

$\text{plim}\ \frac{1}{T} Y' U_1 \neq 0$ since disturbances and endogenous variables are in general correlated in simultaneous equation models. Then the k-class estimators are consistent if and only if $\text{plim}(1-k) = 0$.

The asymptotic covariance matrix of the k-class estimator is

$\sigma^2 T\ \text{plim}\ \left[W_*' Z_1\right]^{-1} \left[W_*' W_*\right] \left[Z_1' W_*\right]^{-1}$. If $\text{plim}\ \sqrt{T}\,(k-1) = 0$. this covariance matrix reduces to that of the (2SLS) estimator.

7. INSTRUMENTAL VARIABLES ESTIMATION OF A SINGLE EQUATION

As before, our equation of interest is given by

$$y_1 = Y_1 \beta_1 + X_1 \gamma_1 + U_1$$

(2.7.1) $$= Z_1 \delta_1 + U_1$$ with

reduced form of the (R.H.S.) jointly dependent explanatory variables given by

(2.7.3) $$Y_1 = X\Pi_1 + V_1$$

where $X = [X_1, X_2]$ is a $T \times K$ matrix of observations on all the predetermined variables of the system; Π_1 is a $K \times M^\Delta - 1$ matrix of reduced form disturbances. We also make the usual assumption that

(a) the rows of V_1 are independently and normally distributed with mean zero and non-singular covariance matrix Ω

(b) the matrix $\frac{1}{T} X'X$ converges to a finite positive definite matrix as $T \to \infty$, i.e.

(2.7.3) $$\lim_{T\to\infty} \frac{1}{T} X'X = \Sigma_{X'X}$$

Suppose that we premultiply (2.7.1) by a $T \times (M^\Delta - 1 + K^*)$ matrix of instruments R of rank$(M^\Delta - 1 + K^*)$ to obtain

(2.7.4) $$R'y_1 = R'Z_1\delta_1 + R'U_1$$

Further, suppose that following conditions hold:

(2.7.5) $$\text{plim} \frac{1}{T} R'U_1 = 0$$

$$(2.7.6) \qquad \text{plim}\ \frac{1}{T}\, R'Z_1 = \Sigma_{R'Z_1} \;\; ; \quad \Sigma_{R'Z_1} \quad \text{(nonsingular)}\ .$$

The (I.V.) estimator $\bar{\delta}_1$ of δ_1 is obtained by solving

$$(2.7.7) \qquad R'y_1 = R'Z_1\bar{\delta}_1 \quad \text{to give}$$

$$(2.7.8) \qquad \bar{\delta}_1 = (R'Z_1)^{-1}\, R'y_1 \qquad \text{provided } R'Z_1 \text{ is nonsingular.}$$

The (I.V.) estimator $\bar{\delta}_1$ of δ_1 is consistent since

$$(2.7.9) \qquad \begin{aligned} \text{plim}\ \bar{\delta}_1 &= \delta_1 + \text{plim}\left(\frac{1}{T}\, R'Z_1\right)^{-1} \cdot\ \text{plim}\left(\frac{1}{T}\, R'U_1\right) \\ &= \delta_1 + \Sigma^{-1}_{R'Z_1} \cdot 0 \\ &= \delta_1 \end{aligned}$$

Sargan (1958) showed that

$$(2.7.10) \qquad \sqrt{T}\,(\bar{\delta}_1 - \delta_1) \approx N\left[0,\ \sigma^2\ \text{plim}\left(\frac{R'Z_1}{T}\right)^{-1}\left(\frac{R'R}{T}\right)\left(\frac{Z_1'R}{T}\right)^{-1}\right]$$

Brundy and Jorgenson (1971) proved that a necessary and sufficient condition for the (I.V) estimator given in (2.7.8) to be asymptotically efficient in the sense of attaining the Cramer-Rao bound is that the matrix of instrumental variables R include two subsets, i.e.

$$(2.7.11) \qquad R = [R_1\ , R_2] \qquad \text{such that}$$

$$(2.7.12) \qquad \text{plim}\ \frac{1}{T}\, R_1'\ X = \Pi_1'\ \Sigma_{X'X}$$

and

$$(2.7.13) \qquad \text{plim}\ \frac{1}{T}\, R_2'\ X_1 = \Sigma_{X_1X_1} \qquad \text{where}$$

$$\text{(2.7.14)} \qquad \lim \frac{1}{T} X'X = \Sigma_{X'X} = \begin{bmatrix} \Sigma_{X_1X_1} & \Sigma_{X_1X_2} \\ \Sigma_{X_2X_1} & \Sigma_{X_2X_2} \end{bmatrix}$$

Brundy and Jorgenson (1971) showed that the (2SLS) instruments given by

$$\text{(2.7.15)} \qquad \begin{aligned} &[Y_1 - \hat{V}_1 \; ; \; X_1] \\ &= [X\hat{\Pi}_1 \; ; \; X_1] \\ &= [\hat{Y}_1 \; ; \; X_1] \end{aligned}$$

satisfy the conditions (2.7.5), (2.7.6), (2.6.12) and (2.7.13), where, as pointed out earlier,

$$\text{(2.7.16)} \qquad \hat{V}_1 = [I - X(X'X)^{-1}X']Y_1$$

and

$$\text{(2.7.17)} \qquad \hat{\Pi}_1 = (X'X)^{-1}X'Y_1 = \Pi_1 + (X'X)^{-1}X'V_1$$

are respectively the (OLS) estimates of the reduced-form disturbances V_1 and the reduced-form coefficients Π_1 of (2.7.2). Therefore the (2SLS) is consistent and asymptotically efficient. However,

(2.7.18) $\qquad E(\hat{Y}_1' U_1) \neq 0 \qquad$ since from (2.7.15) and (2.7.17) it can be seen that each row in $\hat{Y}_1$ depends on V_1 and hence through the relationship between the structural and reduced-form disturbances, upon all the components in U_1.

Phillips and Hale (1977) constructed instruments aimed at obtaining an (I.V.) estimator with smaller bias than (2SLS). The (2SLS) instruments in (2.7.15) may be expressed as

(2.7.19)

$$\hat{Y}_1 = \begin{bmatrix} x_1' & 0 & \cdots & 0 \\ 0 & x_2' & \cdots & 0 \\ 0 & . & & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & & x_T' \end{bmatrix} \begin{bmatrix} \hat{\Pi}_1 \\ \hat{\Pi}_2 \\ \vdots \\ \hat{\Pi} \end{bmatrix}$$

$T \times (T \times K)$ $(T \times K) \quad M^{\Delta} - 1$

from $X = \{x_t'\}$ where x_t' is a K element row vector of the t-th observation on the predetermined variables.

Now define $\hat{\Pi}_t$ to be the (OLS) estimate of Π_1 based on $(T-1)$ observations on the predetermined variables when the t-th observation vector is excluded $(t=1,\ldots,T)$. Hence, we could now define a new set of instruments for the jointly dependent variables denoted by

(2.7.20)

$$Y_1^* = \begin{bmatrix} x_1' & 0 & \cdots & 0 \\ 0 & x_2' & \cdots & 0 \\ \vdots & 0 & \cdots & \vdots \\ 0 & \vdots & & x_T' \end{bmatrix} \begin{bmatrix} \hat{\Pi}_1 \\ \hat{\Pi}_2 \\ \vdots \\ \hat{\Pi}_T \end{bmatrix}$$

$T \times (T \times K)$ $(T \times K) \times M^{\Delta} - 1$

The t-th observation of the instruments depends upon all the rows of V_1 except the t-th and hence does not depend upon the t-th component of U_1.

Now since

(2.7.21) $E(\hat{\Pi}_t' x_t U_t) = 0, \quad t=1,\ldots,T$

it follows that

$$(2.7.22)\qquad E(Y_1^*U_1) = E(\hat{\Pi}_1'x_1\ ,\ \hat{\Pi}_2'x_2\ ,\ldots,\ \hat{\Pi}_T'x_T)\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_T \end{bmatrix}$$

$$= E(\hat{\Pi}_1'x_1u_1 + \hat{\Pi}_2'x_2u_2 + \ldots + \hat{\Pi}_Tx_Tu_T)$$

$$= 0 \quad .$$

Thus, by the successive re-estimation of the reduced form parameter, it is possible to construct a set of instruments for the jointly dependent regressors which are contemporaneously independent of the disturbance. Phillips and Hale (1977) have shown that

$$(2.7.23)\qquad Y_1^* = Y_1 - [I - D]^{-1}\hat{V}_1$$

where

$$(2.7.24)\qquad D = \text{diag.}(X(X'X)^{-1}X') \ .$$

Using the fact that

$$(2.7.25)\qquad \lim_{T\to\infty} D = 0 \qquad \text{[see Miller 1974]},$$

it follows that as the sample size gets large, Y_1^* converges to $\hat{Y}_1$, since $\hat{Y}_1 = Y_1 - \hat{V}_1$. Hence the new (IV) estimator has instruments which satisfy the conditions (2.7.5), (2.7.6), (2.7.12) and (2.7.13) and so is consistent and efficient. It was found that the bias of the new (IV) estimator was smaller than that of the (2SLS) in the special case where the total number of predetermined variables in the system exceeded twice the number of variables included in the equation being estimated. This finding seems to indicate that the property of contemparaneous independence between instrumental variables and the disturbance does not play a major role in obtaining (IV) estimators with small bias.

8. THE FIX-POINT OR FIXED-POINT (FP) METHOD

In the (2SLS) approach we observed that if $\hat{Y}_1$, the estimates of the jointly dependent variables, are obtained from the unrestricted reduced-form then $\hat{Y}_1$ can be used as either regressors or as instrumental variables and the resultant estimators of the parameters are the same.

Maddala (1971) pointed out that the two different estimators are obtained when $\hat{Y}_1$, the estimates of the jointly dependent variables, are derived from the restricted reduced form. He referred to these two new estimators as (1) the restricted reduced-form two-stage least-squares (RRF2SLS) and (2) the restricted reduced-form instrumental variables (RRFIV). Both of these estimates have to be computed by an iterative procedure. Maddala (1971) used two such procedures: the solved reduced-form method and the Wold's method or the method of successive iteration. Of these two iterative procedures, Wold's methods is computationally simpler because the structural system need not be solved at each stage of the iteration.

Maddala (1971) examined the convergence properties of the (RRF2SLS) and the (RRFIV) through the use of the two above named iterative procedures. He found that the (RRF2SLS) has several limitations. In some cases it did not converge and in others it gave oscillatory solutions. In some examples the iterations converged to different solutions depending on the starting point. In general, however, he found that the (RRF2SLS) has better convergence properties with the Wold's iteration method than with the solved reduced-form method.

On the other hand, the (RRFIV) estimator often failed to converge when the Wold's iterative procedure was used. But, in all the cases considered, the (RRFIV) by the solved reduced form method converged to one point no matter what variables were chosen as the initial stage instruments.

It should be noted that the (RRF2SLS) method computed by the Wold's iteration procedure is often referred to as the fix-point FP) method. Wold (1965) developed the (FP) method in the context of the generalized independent system (GEID) and was viewed as an alternative to the other simultaneous equation methods.

The Fix-Point (FP) method proposed by Wold (1965) uses a Reformulated Interdependent (REID) system which can be treated as a special case of the Generalized Interdependent (GEID) system. As before, we denote the system of structural equations by

$$(2.8.1) \qquad By_t + \Gamma x_t = u_t \qquad (t=1,2,\ldots,T) \ .$$

Imposing the standard normalization on the diagonal elements of B, the structural equations of (2.8.1) can be written as

$$(2.8.2) \qquad y_t = B^0 y_t + \Gamma x_t + u_t$$ where the diagonal elements of

B^0 are zeros and the matrix $[I - B^0]$ is non-singular.

Solving the structural form (2.8.2) for y_t we obtain the reduced-form

$$(2.8.3) \qquad y_t = [I - B^0]^{-1} \Gamma x_t + [I - B^0]^{-1} u_t \ .$$

Let

$$(2.8.4) \qquad [I - B^0] \Gamma x_t = \bar{y}_t$$

$$(2.8.5) \qquad [I - B^o]^{-1} u_t = \eta_t \ .$$

By introducing $y_t = \bar{y}_t + \eta_t$ on the (R.H.S) of (2.8.2). Wold (1965) reformulates the structural form in the following manner:

$$(2.8.6) \qquad \begin{aligned} y_t &= B^o[\bar{y}_t + \eta_t] + \Gamma x_t + u_t \\ &= B^o\bar{y}_t + \Gamma x_t + \eta_t \qquad \text{(using 2.8.5)} \ . \end{aligned}$$

The form (2.8.6), with the conventional assumption the structural equations, is called the (REID) system specification.

Wold (1965) in his development of the (FP) method introduced the (GEID) system which asserts that η_{it} is uncorrelated with those components of $\bar{y}_t$ and x_t which occur in the i-th reformulated structural equation [see (2.8.6)], but otherwise correlations can occur between η_{it} and the components of $\bar{y}_t$ and x_t which do not occur in the i-th equation. Wold (1966) imposed an even stronger assumption for the (GEID) specification in that $\bar{y}_{it}$ is the conditional expectation of y_{it} for given components of $\bar{y}_t$ and x_t, which occur in the (R.H.S) of the i-th equation.

It should be noted that in the "Classical" specification of the (SEM) Simultaneous Equations Model (2.8.1) the disturbance term in any equation is uncorrelated with all the predetermined variables of the model. Let Π be the matrix of regression coefficients of y_t on the components of x_t. When η_t is independent of x_t we have $\Pi = [I - B^o]^{-1}\Gamma$. For the (GEID) specification, where correlation occur between some elements of η_t and some elements of x_t, the matrices Π and $[I - B^o]^{-1}\Gamma$ are in general different.

Mitchell (1974) in his criticism of the (GEID) specification remarked that there do not seem to be grounds for saying in advance that the disturbance in a given equation is uncorrelated with particular linear combinations of the predertimed variables, yet correlated with other linear combinations of the same variables. [See Watts (1964) for a critical appraisal of the (GEID) Systems].

As pointed out by Mitchell (1974), the (FP) Fix-Point estimation problem can be stated as follows: find the coefficients $\overline{B}^o$ and $\overline{\Gamma}$ and variables $\overline{y}_t$ which simultaneously satisfy

(2.8.6) $$y_t = B^o\overline{y}_t + \Gamma x_t + \eta_t$$

(2.8.7) $$\overline{y}_t = [I - B^o]^{-1}\Gamma x_t$$

(2.8.8) Zero restrictions on B^o and Γ

(2.8.9) Least-squares regression restrictions on (2.8.1),

i.e., in each equation the disturbance η_t must be orthogonal to all non-normalized $\overline{y}_t$ and x_t appearing in that equation with non-zero coefficients.

Unlike the (2SLS) estimator, the (FP) estimator is not defined in terms of unrestricted reduced-form coefficients, since the $\overline{y}_t$'s are determined by the non-zero elements of B and Γ. Wold (1966, 1969) and Mosback and Wold (1970) proposed a non-linear iterative partial least-squares method where the regressors in one approximation are used when the coefficients in the next approximation are obtained, and then a new approximation of the regressors is calculated.

Starting with some initial value of the right hand jointly dependent variables, $y(o)_t$ and the known values of y_t and x_t, we regress y_t on $y(o)_t$ and x_t in each equation, excluding the variables specified to have zero coefficients a priori, to obtain

(2.8.10) $$y_t = B^o(1)y(o)_t + \Gamma(1)x_t + \eta(1)_t \, .$$

Now in each structural equation calculate the predicted values of y_t from these first round estimates applied to x_t and the initial $y(o)_t$ and call them $y(1)_t$.

Therefore,

(2.8.11) $$y(1)_t = B^o(1)\, y(o)_t + \Gamma(1)x_t \, .$$

We return to (2.8.10) and for this round we regress y_t on $y(1)_t$ and x_t and this obtain the second round estimates $B^o(2)$, $\Gamma(2)$. This iterative process is continued until successive values of all the parameters are practically unchanged i.e., if $y_t(j-1)$ be the j-th approximation of $\bar{y}_t$, then the (REID) specification in the j-th approximation is written as

(2.8.12) $$y_t = B^o(j)y(j-1) + \Gamma(j)_t' + \eta(j)_t \, .$$

Again with $y(j-1)_t$ considered as known we apply (OLS) to each structural equation, until we obtain

(2.8.13) $$y(j)_t = B^o y(j-1)_t + \Gamma(j)x_t \, .$$

The fixed point is defined by

(2.8.14) $$y(j-1)_t = y(j)_t = \bar{y}_t$$

$$B^o(j-1) = B^o(j) = \bar{B}^o$$

$$\Gamma(j-1) = \Gamma(j) = \bar{\Gamma}$$

and hence if convergence occurs at the J-th iteration we have $\hat{j} = J$ and

$$(2.8.15) \qquad \hat{y}(j) = [I - B^{o}(j)]^{-1}\,\Gamma(j)x_t \quad \text{as required from (2.8.7).}$$

Wold (1965, 1966) suggested that the initial approximation $y(o)_t$ can be obtained as linear combinations of the vector x_t whereas Mosbaek and Wold (1970) determined the initial approximation by making use of the (OLS) estimates of the matrix Π of the regression coefficients of the components of y_t on the components of x_t. This kind of initial approximation is often referred to as the (2SLS) start.

Questions on the existence and uniqueness of fix-point estimates and on the convergence of suggested iterations have given rise to a number of computing techniques for the (FP) estimator. Mitchell (1974) and Lyttkens (1973) dealt with the following alternative iterative techniques (1) the Recursive Fix-Point Method which has been extensively investigated by Bodin (1969, 1970, 1973); (2) The Fractional Fix-Point Method which has been employed by Agren (1969, 1970, 1972); (3) The Basmann and Bakoney's Iterative Method and (4) The Reduced Fix-Point Method of Wold (1966) and Maddala (1971).

Lyttkens (1973) compared the fix-point estimator with the two-stage least-squares (2SLS), the three-stage least-squares (3SLS) and the Full Information Maximum Likelihood (FIML) estimators and reviewed most of the literature related to the search for asymptotic properties of the fix-point estimator. Mitchell (1974) and Lyttkens (1973) outlined methods for computing the fix-point (FP) estimation of models containing both linear and non-linear identities. Dhrymes and Pandit (1972) proved the consistency of the first iterate of the (RRF2SLS), i.e., the second approximation of the structural parameters obtained by the reduced fix-point method with (2SLS) start. They also considered the asymptotic distribution of the first iterate of the (RRF2SLS) and found that the estimators of some parameters can be more efficient than those of (2SLS) and the estimators of some other parameters can be less efficient than those of (2SLS). Finally, it should be noted that in the (GEID) specification the matrices Π and $[I - B^{0}]^{-1}\Gamma$ are in general different. In this case no known method of obtaining a consistent start of the fix-point iteration has been found.

9. MINIMUM DISTANCE ESTIMATORS (MDE)

From the M structural equations given by $By_t + \Gamma x_t = u_t$ the reduced-form is given by

(2.9.1) $$y_t = \Pi x_t + V_t \qquad (t=1,2,\ldots,T)$$

wherein the overidentified case Π is subject to restrictions and

(2.9.2) $$\Pi = -B^{-1}\Gamma$$

(2.9.3) $$V_t = B^{-1}U_t$$

denote the true coefficients structures as B_o, Γ_o and Π_o respectively or $\Pi_o = -B_o^{-1}\Gamma_o$.

Malinvaud (1966) proposed the Minimum Distance (MD) estimators of $\Pi_T(S)$ of the matrix Π satisfying the restrictions and minimizing

$$(2.9.4) \qquad L_T(S,\Pi) = \frac{1}{T}\sum_{t=1}^{T}(y_t - \Pi x_t)'\,S(y_t - \Pi x_t)$$

where S is a positive definite matrix of order M. The estimators $B_T(S)$ and $\Gamma_1(S)$ corresponding to $\Pi_T(S)$ can be obtained by imposing standard normalization and simultaneously solving the a priori restrictions subject to the equality

$$(2.9.5) \qquad B\Pi_T(S) + \Gamma = 0 \ .$$

Malinvaud (1966) suggested that under fairly general conditions

$$B_T(S) \to B^o \quad \text{and} \quad \Gamma_T(S) \to \Gamma_o$$

and showed that $B_T(S)$ and $\Gamma_T(S)$ attain maximum asymptotic efficiency when S_T tends to $\Omega = (B_o)^{-1}\,\Sigma(B_o')^{-1}$. He also proved that if the matrix

$$S_T \xrightarrow{p} S \qquad \text{then}$$

$$(2.9.6) \qquad \sqrt{T}\left[B_T(S_T) - B_o\right] \quad \text{and} \quad \sqrt{T}\left[\Gamma_T(S_T) - \Gamma_o\right]$$

have a limiting normal distribution and correspondingly that

$$(2.9.7) \qquad \sqrt{T}\left[\Pi_T(S_T) - \Pi_o\right] \quad \text{has a limiting normal distribution.}$$

SIMULTANEOUS LEAST SQUARES (SLS)

Brown (1960) proposed the Simultaneous Least Squares (SLS) method of estimation which is a special case of the Minimum Distance (MD) estimators outlined in (2.9.4) with $S = I$ and the proposed estimators become $B_T(I)$ and $\Gamma_T(I)$ correspondingly.

Nakamura (1960) established the consistency of the (SLS) estimators. Dhrymes (1972) compared the minimands of the (SLS) and Full Information Maximum Likelihood (FIML) and showed that contrary to Browns's assertion, (SLS) is not a Full Information method. In his iterated instrumental variables estimation of (SLS), Dhrymes (1972) pointed out that if the initial estimate of the iteration procedure is consistent then all subsequent iterates are consistent and that (SLS) is asymptotically dominated by (FIML) estimators. (See Dhrymes (1972) for details of proof.)

THE Ω-CLASS ESTIMATORS

Keller (1975) proposed the so-called Ω-class estimator which is a limited information estimator. The Ω-class estimator can be interpreted geometrically as a Minimum Distance estimator (MDE) defined on a symmetric, positive semidefinite index matrix $\bar{\Omega}$. Keller (1975) showed that in the case where $\bar{\Omega}$ corresponds to the covariance matrix Ω of the reduced-form distrubances, and if these disturbances are assumed to be normally distributed, the Ω-class estimator can be interpreted as a Maximum-Likelihood (ML) estimator.

Keller (1975) firstly developed the Ω-class estimator by developing the (ML) maximum likelihood estimators of the structural coefficients as functions of the reduced-form covariance matrix Ω.

From our model of M structural equations

$$(2.9.8) \qquad By_t + \Gamma x_t = U_t \qquad (t=1,2,\ldots,T)$$

we can write (2.9.8) as

$$(2.9.9) \qquad y_t'B' + x_t'\Gamma' = u_t'$$

Now define

$$Y = \begin{bmatrix} y_1' \\ y_2' \\ \vdots \\ y_T' \end{bmatrix}, \quad X = \begin{bmatrix} x_1' \\ x_2' \\ \vdots \\ x_T' \end{bmatrix}, \quad U = \begin{bmatrix} u_1' \\ u_2' \\ \vdots \\ u_T' \end{bmatrix}.$$

Then all M equations at all T sample periods can be written as

$$(2.9.10) \qquad YB' + X\Gamma' = U$$

where Y is $T\times M$, B' is $M\times M$, $X = [X_1, X_2]$ is $T\times K$, Γ' is $K\times M$ and U is $T\times M$. The reduced-form corresponding to (2.9.10) is given by

$$(2.9.11) \qquad Y = X\Pi' + V$$

where $\Pi' = -\Gamma'(B^{-1})'$

$V' = U(B^{-1})'$

As before, the first structural equation is given by

$$(2.9.12) \qquad y_1 = Y_1\beta_1 + X_1\gamma_1 + u_1$$

From Keller (1975), the maximum-likelihood estimators of the structural coefficients as a function of Ω is given by

$$(2.9.13)\qquad \begin{bmatrix} \hat{\hat{\beta}}_1 \\ \hat{\hat{\gamma}}_1 \end{bmatrix}_\Omega = \begin{bmatrix} Y_1'PY_1 - \hat{\hat{\lambda}}\Omega_{11} & Y_1X_1 \\ X_1'Y_1 & X_1'X_1 \end{bmatrix}^{-1} \begin{bmatrix} Y_1'Py_1 - \hat{\hat{\lambda}}\omega, \\ X'_1y_1 \end{bmatrix}$$

where

$$(2.9.14)\qquad P = X(X'X)^{-1}X'$$

$$(2.9.15)\qquad \Omega = \begin{bmatrix} \omega'' & \omega_1' \\ \omega_1 & \Omega_{11} \end{bmatrix}$$

where Ω_{11} is or order $M-1$ and

(2.9.16) $\hat{\hat{\lambda}}$ is the smallest root of the determinantal equation

$$\left| Y'(M_1 - M)Y - \lambda\Omega \right| = 0$$

and

$$M_1 = I - X_1(X_1'X_1)^{-1}X_1'$$

$$M = I - X(X'X)^{-1}X$$

$$Y = [y_1 \quad Y_1].$$

In the (ML) approach Keller pointed out that the normality assumption on the disturbances could be too restrictive and that in practice Ω is rarely known. Therefore, he dropped the likelihood approach and considered (2.9.13) as a class of (MD) Minimum Distance estimators. These (MD) estimators were referred to by Keller Ω-class estimators.

In the context of (MD) estimators, (2.9.13) is considered the estimator corresponding to that of Π which minimizes

$$L_T(\bar{\Omega}^p, \Pi) = \sum_{t=1}^{T} (y_t - \Pi x_t)' \, \bar{\Omega}^p \, (y_t - \Pi x_t) \tag{2.9.17}$$

where $\bar{\Omega}^p$ is the pseudo-inverse (generalized inverse) of the $M \times M$ matrix $\bar{\Omega}$.

The estimators in (2.9.15) is now redefined as a function of $\bar{\Omega}$, which is not necessarily equal to Ω. The resulting estimator for the structural coefficients, i.e., the Ω-class estimator, is given by,

$$\begin{bmatrix} \hat{\hat{\beta}}_1 \\ \hat{\hat{\gamma}}_1 \end{bmatrix}_{\bar{\Omega}} = \begin{bmatrix} y_1'PY_1 - \hat{\hat{\lambda}}\,\bar{\Omega}_{11} & Y_1'X_1 \\ X_1'Y_1 & X_1'X_1 \end{bmatrix}^{-1} \begin{bmatrix} Y_1'Py_1 - \hat{\hat{\lambda}}\,\bar{\omega} \\ X_1'y_1 \end{bmatrix} \tag{2.9.18}$$

and $\hat{\hat{\lambda}}$ is the smallest root of

$$\left| Y(M_1 - M)Y - \lambda\bar{\Omega} \right| = 0 \tag{2.9.19}$$

Keller established the consistency of the Ω-class estimators by showing that under the usual assumptions the Ω-class estimator defined in (2.9.18), (2.9.19) is a consistent estimator of the structural coefficients,

$$\operatorname*{plim}_{T\to\infty} \begin{pmatrix} \hat{\hat{\beta}}_1 \\ \hat{\hat{\gamma}}_1 \end{pmatrix}_{\bar{\Omega}} = \begin{pmatrix} \beta_1 \\ \gamma_1 \end{pmatrix} \tag{2.9.20}$$

and

$$\operatorname*{plim}_{T\to\infty} \frac{\hat{\hat{\phi}}}{T} = 0 \tag{2.9.21}$$

It is interesting to note that in view of (2.9.21) we could demonstrate the asymptotic equivalence of the Ω-class estimator and the k-class estimator which could be defined as

$$(2.9.22)\qquad \begin{bmatrix} \hat{\beta}_1(k) \\ \hat{\gamma}_1(k) \end{bmatrix} = \begin{bmatrix} Y_1'Y_1 - kY_1'MY_1 & Y_1'X_1 \\ X_1'Y_1 & X_1'X_1 \end{bmatrix}^{-1} \begin{bmatrix} Y_1'y_1 - kY_1'My_1 \\ X_1'y_1 \end{bmatrix}$$

The k-class estimator is defined on a scalar k which must converge in probability to one inorder to secure the consistency of the estimators. In the Ω-class, however, the estimators are defined on a symmetric positive semidefinite matrix which can be chosen arbitrarily, under very general conditions, all members of the Ω-class are consistent estimators.

10. THE THREE-STAGE LEAST SQUARES (3SLS) METHOD

Zellner and Theil (1962) developed the (3SLS) as a consistent estimation procedure which takes into account all the equations of the simultaneous equation model under study.

We consider a complete system of M linear stochastic structural equations in M jointly dependent variables and K predetermined variables. The m^{th} structural equation can be written as

$$(2.10.1)\qquad y_m = Y_m\beta_m + X_m\gamma_m + U_m\ , \qquad (m = 1,2,\ldots,M)$$

where y_m is a $T\times 1$ column vector of observations on one of the jointly dependent variables in the m^{th} equation, Y_m is the $T\times(M_m - 1)$ matrix of observations on the explanatory jointly dependent variables, X_m is the $T\times K_m$ matrix of observations on the explanatory predetermined variables, β_m and γ_m are vectors of parameters and U_m is the column vector of T structural disturbances.

Let

$$(2.10.2) \qquad Z_m = [Y_m \, , X_m] \quad \text{and} \quad \delta_m = \begin{bmatrix} \beta_m \\ \gamma_m \end{bmatrix} .$$

From (2.10.2) we can rewrite (2.10.1) as

$$(2.10.3) \qquad y_m = Z_m \delta_m + U_m .$$

In the (3SLS) approach, we write all the equations in a "stacked" form and apply generalized-least-squares method to the system as a whole after making a transformation (as in Zellner's "seemingly unrelated regression").

Premultiply (2.10.3) by X' where X is a $T \times K$ matrix of rank K, then

$$(2.10.4) \qquad X'y_m = X'Z_m\delta_m + X'u_m \qquad (m = 1,2,\ldots,M) .$$

Let

(2.10.5) $\quad E(U_m U_m') = \sigma_{mn} I_T$, i.e., the disturbances in the different equations are contemporaneously correlated but are independent over time. Therefore the variance-covariance matrix of the disturbance in (2.10.4) is given by

$$(2.10.6) \qquad E(X'u_m \; u_m' X) = \sigma_{mm}(X'X)$$

where σ_{mn} is the constant variance of each of the T disturbances in the m-th equation. Then we can apply Generalized-Least-Squares to (2.10.4) to obtain

$$Z_m'X(\sigma_{mm}X'X)^{-1}X'y_m = Z_m'X(\sigma_{mm}X'X)^{-1}X'Z_md_m \tag{2.10.7}$$

from which we derive the (2SLS) estimator

$$d_m = [Z_m'X(X'X)^{-1}X'Z_m]^{-1}\ Z_mX(X'X)^{-1}X'y_m\ . \tag{2.10.8}$$

It should be noted that if the m-th equation is exactly identified, the matrix Z'_mX is square and non-singular. In this case (2.10.7) simplifies to

$$(X'Z_m)d_m = X'y_m\ . \tag{2.10.9}$$

The solution of (2.10.9) written out and rearranged will be seen to be just the indirect least-squares estimates.

(2.10.10)

$$\begin{bmatrix} X'y_1 \\ X'y_2 \\ \vdots \\ X'y_m \end{bmatrix} = \begin{bmatrix} X'Z_1 & 0 \ldots\ldots\ldots & 0 \\ 0 & X'Z_2 \ldots\ldots & 0 \\ \vdots & 0 & \vdots \\ 0 & \ldots\ldots & X'Z_M \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta \end{bmatrix} + \begin{bmatrix} X'U_1 \\ X'U_2 \\ \vdots \\ X'U_M \end{bmatrix}$$

To apply (GLS) to (2.10.10) we need the covariance matrix of the disturbance vector of (2.10.10):

$$V\begin{bmatrix} X'U_1 \\ X'U_2 \\ \vdots \\ X'U_M \end{bmatrix} = \begin{bmatrix} \sigma_{11}X'X & \sigma_{12}X'X \ldots\ldots & \sigma_{1M}X'X \\ \sigma_{21}X'X & \sigma_{22}X'X \ldots\ldots & \sigma_{2M}X'X \\ \vdots & & \\ \sigma_{M1}X'X & \ldots\ldots & \sigma_{MM}X'X \end{bmatrix} \tag{2.10.11}$$

Grouping the σ_{mn} in a matrix Σ, $(m, n=1,2,\ldots,M)$ we have

$$V = \Sigma \otimes (X'X) \tag{2.10.12}$$ and

(2.10.13) $$V^{-1} = \Sigma^{-1} \otimes (X'X)^{-1}$$

where $[\sigma^{mn}] = [\sigma_{mn}]^{-1}$.

In practice, σ_{mn} is not known. Following Zellner and Theil's suggestion, we estimate each structural equation by the (2SLS) method, get the estimated residuals $\hat{U}_m$, estimate σ_{mn} by $S_{mn} = \frac{1}{T}(\hat{U}_m' \hat{U}_n)$, invert this matrix and denote it by S^{mn} and use S^{mn} for σ^{mn} .

Thus, the (3SLS) estimator is given by

(2.10.14)
$$\hat{\delta}_{3SLS} = \begin{bmatrix} \hat{\delta}_1 \\ \hat{\delta}_2 \\ \vdots \\ \hat{\delta}_M \end{bmatrix} = \begin{bmatrix} S^{11}Z_1'PZ & S^{12}Z_1'PZ_2 & \cdots & S^{1M}Z_1'PZ_M \\ S^{21}Z_2'PZ_1 & S^{22}Z_2'PZ_2 & \cdots & S^{2M}Z_2'PZ_M \\ \vdots & \vdots & & \vdots \\ S^{M1}Z_M'PZ_1 & S^{M2}Z_M'PZ_2 & \cdots & S^{MM}Z_M'PZ_M \end{bmatrix}^{-1} \begin{bmatrix} \sum_{m=1}^{M} S^{1m}Z_1'Py_m \\ \sum_{m=1}^{M} S^{2m}Z_2'Py_m \\ \vdots \\ \sum_{m=1}^{M} S^{Mm}Z_M'Py_m \end{bmatrix}$$

where

(2.10.15) $$P = X(X'X)^{-1}X' .$$

The covariance matrix of the (3SLS) estimator is given by the inverse of the (R.H.S) of (2.10.14), i.e.,

(2.10.16) $$[S^{mn}Z_m'PZ_n]^{-1} .$$

It should be noted that the (3SLS) estimators are more efficient than the (2SLS) estimators only if $[\sigma_{mn}]$ is not diagonal. If $\sigma_{mn} = 0$ $(m \neq n)$, then the (3SLS) reduces to the (2SLS).

Writing all the equations compactly as

$$(2.10.17) \qquad y = Z\delta + u \qquad \text{i.e.}$$

$$(2.10.18) \qquad \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix} = \begin{bmatrix} Z_1 & 0 & \dots & 0 \\ 0 & Z_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & & Z_M \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta M \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_M \end{bmatrix}$$

Then

$$(2.10.19) \qquad \hat{\delta}_{2SLS} = \left[Z'(I_M \otimes P)Z\right]^{-1} Z'(I_M \otimes P)y$$

$$(2.10.20) \qquad \hat{\delta}_{GLS} = \left[Z'(\Sigma^{-1} \otimes P)Z\right]^{-1} Z'(\Sigma^{-1} \otimes P)y$$

Therefore

$$(2.10.21) \qquad \hat{\delta}_{3SLS} = \left[Z_1'(\hat{\Sigma} \otimes P)Z\right]^{-1} Z'(\hat{\Sigma} \otimes P)y$$

where $\hat{\Sigma}$ is a consistent estimator of Σ .

It can be shown that $\hat{\delta}_{3SLS}$ is a consistent estimator, i.e.

$$(2.10.22) \qquad \operatorname*{plim}_{T\to\infty} \hat{\delta}_{3SLS} = \delta \ .$$

The consistency of the (3SLS) estimator can be shown as follows: Consider the following set of transformed structural equations

$$(2.10.23) \qquad a = A\delta + q \ , \quad \text{i.e.}$$

$$\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_M \end{bmatrix} = \begin{bmatrix} A_1 & 0 & \dots & 0 \\ 0 & A_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & \dots & & A_M \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_M \end{bmatrix} = \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_M \end{bmatrix}$$

where

$$a_m' = R^{-1}X'y_m$$

(2.10.24) $$A_m = R^{-1}X'Z_m$$

$$q_m = R^{-1}X'u_m$$

where R is a non-singular matrix such that $X'X = RR'$ [rank $X = K$ by assumption and $X'X$ is positive definite, hence there exists a non-singular matrix R such that $X'X = RR'$].

The generalized-least-squares (GLS) estimator of δ in (2.10.23) is given by

(2.10.25) $$\delta_{GLS} = (A' \quad \Theta^{-1} A)^{-1} A \quad \Theta^{-1} a$$

where $\text{cov}(q) = \Theta = \Sigma \otimes I_K$

Therefore,

(2.10.26) $$\hat{\delta}_{3SLS} = (A' \quad \hat{\Theta}^{-1} A)^{-1} A \quad \hat{\Theta}^{-1} a$$ where

$$\hat{\Theta} = \hat{\Sigma} \otimes I_K .$$

Now from (2.10.23) and (2.10.26)

(2.10.27) $$\hat{\delta}_{3SLS} = \delta + (A' \cdot \hat{\Theta}^{-1} A)^{-1} A' \hat{\Theta}^{-1} q .$$

For consistency of δ_{3SLS}, we need to show

(2.10.28) $$\text{plim} \left(\frac{A \quad \hat{\Theta}^{-1} A}{T}\right)^{-1} \frac{A'}{\sqrt{T}} \hat{\Theta}^{-1} \frac{q}{\sqrt{T}}$$ variables.

It should be noted that the probability limit of

$\left(\frac{A\ \hat{\Theta}^{-1}\ A}{T}\right)$ entails a nonstochastic probability limit of the quantities

$$(2.10.2) \qquad \frac{S^{mn}A_m'A_n}{T} = \frac{S^{mn}Z_m'X}{T}\left(\frac{X'X}{T}\right)^{-1}\frac{X'Z_n}{T}$$

and since S^{mn} is a consistent estimator of σ^{mn}, the probability limit of (2.10.29) is a nonstochastic matrix with finite elements (in view of our usual assumptions) i.e., the model contains no lagged endogenous variables and the error terms are mutually independent.

Similarly the probability limit of (2.10.30) $\frac{A'}{\sqrt{T}}\hat{\Theta}^{-1}$ is also a nonstochastic matrix with finite elements. Also,

$$(2.10.31) \qquad \frac{q_m}{\sqrt{T}} = \frac{R^{-1}X'U_m}{\sqrt{T}} = \sqrt{T}\,\frac{R^{-1}X'U_m}{T} \quad .$$

Now $\underset{T\to\infty}{\text{plim}} \frac{X'X}{T}$ exists as a nonsingular matrix (by assumption).

Since $\left(\frac{X'X}{T}\right)^{-1} = T(R^{-1})'\ R^{-1}$, then $\underset{T\to\infty}{\text{plim}} \sqrt{T}\ R^{-1} = R^*$ exists as a matrix with finite elements.

Therefore, (2.10.31) becomes

$$(2.10.32) \qquad R^* \underset{T\to\infty}{\text{plim}} \frac{X'u_m}{T} \qquad \text{where}$$

$$\underset{T\to\infty}{\text{plim}} \frac{X'u_m}{T} = 0 \qquad \text{(by assumption).}$$

Thus, we conclude

$$(2.10.33) \qquad \underset{T\to\infty}{\text{plim}}\ \hat{\delta}_{3SLS} = \delta \ .$$

Zellner and Theil (1962) demonstrated that the asymptotic distribution of $\hat{\delta}_{3SLS}$ is

$$(2.10.34) \qquad \sqrt{T}\left(\hat{\delta}_{3SLS} - \delta\right) \approx N\left[0,\ \underset{T\to\infty}{\text{plim}} \left(\frac{A'\hat{\Theta}^{-1}A}{T}\right)^{-1}\right]$$

Following Dhrymes (1970), we establish the result of (2.10.34) as follows:

Now

$$(2.10.35) \qquad \sqrt{T}\left(\hat{\delta}_{3SLS} - \delta\right) = \left(\frac{A\ \hat{\Theta}^{-1}A}{T}\right)^{-1} \frac{A\ \hat{\Theta}^{-1}q}{\sqrt{T}}$$

From (2.10.35), it is clear that finding the asymptotic distribution of $\sqrt{T}\left(\hat{\delta}_{3SLS} - \delta\right)$ reduces to the question of finding the asymptotic distribution of (2.10.36) $\frac{A\ \hat{\Theta}^{-1}q}{\sqrt{T}}$ which is exactly the asymptotic distribution of

$$(2.10.37) \qquad \frac{A\ \Theta^{-1}q}{\sqrt{T}} \quad \text{where} \quad \Theta = \Sigma \otimes I_K .$$

Now

$$(2.10.38) \qquad A'\Theta^{-1}q = A'(\Sigma \otimes I_K)(I_M \otimes R^{-1}X')U$$

where

$$(2.10.39) \qquad U = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_M \end{bmatrix} .$$

Define

$$(2.10.40) \qquad \bar{A}_m' = A_m R^{-1} X' \quad , \quad \bar{A}' = \text{diag}(\bar{A}_1' , \bar{A}_2' \ldots \bar{A}_M') .$$

Then we see that

(2.10.41) $A'\Theta^{-1}q = \bar{A}'(\Sigma^{-1} \otimes I_T)U$ and further observe that

(2.10.42) $\bar{A}' = A(I_M \otimes R^{-1})(I_M \otimes X')$.

Now define

(2.10.43) $B = \text{diag}[B_1, B_2 \ldots B_m]$ where

$$B_M = A_m' R^{-1} \qquad (m=1, \ldots M)$$

and observe that the quantities

(2.10.44) $$\operatorname*{plim}_{T\to\infty} B = \operatorname*{plim}_{T\to\infty} \frac{A_m'}{\sqrt{T}} \sqrt{T}\ R^{-1} .$$

exist as nonstochastic matrices with finite elements.

Choosing σ^m to be the m-th row of Σ^{-1}, u_T to be the row vector containing the error terms of the M equations of the system at "time" r and x_r' the r-th column of X'. Then we can verify that

(2.10.45) $$\frac{A'\Theta^{-1}q}{\sqrt{T}} = B \sum_{r=1}^{T} \frac{V_r U_r'}{\sqrt{T}} \quad \text{where}$$

(2.10.46) $$V_r = \begin{bmatrix} v_r^{(1)} \\ v_r^{(2)} \\ \vdots \\ v_r^{(M)} \end{bmatrix}, \qquad v_r^{(m)} = x_r' \sigma^m$$

Since B has a well-defined nonstochastic probability limit, then the asymptotic distribution of $\frac{A\,\Theta^{-1}q}{\sqrt{T}}$ is determined by the asymptotic distribution of

(2.10.47) $$\frac{1}{\sqrt{T}} \sum_{r=1}^{T} \frac{V_r u_r'}{\sqrt{T}}$$ where

$V_r'u_r'$, $r = 1, 2, \ldots$ constitute a set of mutually random vectors. The asymptotic distribution of (2.10.47) can be established by an appropriate (C.L.T.) Central Limit Theorem which establishes the asymptotic distribution of $\frac{A\,\theta^{-1}q}{\sqrt{T}}$ and in turn that of $\sqrt{T}(\hat{\delta}_{3SLS} - \delta)$.

11. EFFICIENCY OF THE (2SLS) AND (3SLS) ESTIMATORS

From (2.10.23), we can write a single structural equation as

(2.11.1) $a_1 = A_1\,\delta_1 + q_1$. Then the (2SLS) estimator of δ_1 is simply the (OLS) applied to (2.11.1).

Therefore,

(2.11.2) $$\hat{\delta}_{1(2SLS)} = (A_1'A_1)^{-1}A_1'a_1 .$$

Hence the (2SLS) estimator for all the parameters of the entire system $a = A\delta + q$ is

(2.11.3) $$\hat{\delta}_{(2SLS)} = (A'A)^{-1}A'a = \delta + (A'A)^{-1}A'q$$

and since $\underset{T\to\infty}{\text{plim}} \left(\frac{A_1'A_1}{T}\right)^{-1}$, exists as a matrix with finite nonstochastic elements and

$$\underset{T\to\infty}{\text{plim}} \frac{A'q}{T} = 0 \quad , \qquad (A_m = R^{-1}X'Z_m \;;\; q_m = R^{-1}X'U_m) .$$

Then

(2.11.4) $$\operatorname*{plim}_{T\to\infty} \hat{\delta}_{2SLS} = \hat{\delta} .$$

From Dhrymes (1968)

(2.11.5) $$\sqrt{T}\,(\hat{\delta}_{2SLS} - \delta) \approx N\left[0,\ \operatorname*{plim}_{T\to\infty} \left(\frac{A_1'A}{T}\right)^{-1} \frac{A'\Theta A}{T} \left(\frac{A'A}{T}\right)^{-1}\right]$$

Let C_{2SLS} and C_{3SLS} be the covariance matrices of the asymptotic distribution of the (2SLS) and (3SLS) estimators respectively, where

(2.11.6) $$C_{2SLS} = \operatorname*{plim}_{T\to\infty} \left[\left(\frac{A'A}{T}\right)^{-1} \frac{A'\hat{\Theta}A}{T} \left(\frac{A'A}{T}\right)^{-1}\right]$$

(2.11.7) $$C_{3SLS} = \operatorname*{plim}_{T\to\infty} \left(\frac{A'\hat{\Theta}^{-1}A}{T}\right)^{-1}$$

Define a matrix Q by

(2.11.8) $$Q = (A'A)^{-1}A' - (A'\hat{\Theta}^{-1}A)^{-1}A'\Theta^{-1}$$

and note that $QA = 0$.

Also

(2.11.9) $$\sqrt{T}\,Q = \left[\frac{A'A}{T}\right]^{-1} \frac{A'}{\sqrt{T}} - \left[\frac{A\,\hat{\Theta}^{-1}A}{T}\right]^{-1} \frac{A'}{\sqrt{T}}\,\hat{\Theta}^{-1} .$$

We have established that the matrices on the (RHS) of (2.11.9) have well defined probability limits, therefore, it follows that

(2.11.10) $\operatorname*{plim}_{T\to\infty} \sqrt{T}\,Q = \bar{Q}$ exists as a well-defined matrix of (finite) constants.

From (2.11.8), we have

$$(2.11.11) \qquad \left(\frac{A'A}{T}\right)^{-1} \frac{A'\hat{\Theta}A}{T} \left(\frac{A'A}{T}\right)^{-1} = \left[\left(\frac{A'\hat{\Theta}^{-1}A}{T}\right)^{-1} \frac{A'}{\sqrt{T}} \hat{\Theta}^{-1} + \sqrt{T}\, Q \right] \hat{\Theta} \times \left[\hat{\Theta}^{-1} \frac{A}{\sqrt{T}} \left(\frac{A'\hat{\Theta}A}{T}\right)^{-1} + \sqrt{T}\, Q' \right]$$

Taking the probability limits and noting that $QA = 0$

$$(2.11.12) \qquad \operatorname*{plim}_{T\to\infty} \left[\left(\frac{A'A}{T}\right)^{-1} \frac{A'\hat{\Theta}A}{T} \left(\frac{A'A}{T}\right)^{-1}\right] = \operatorname*{plim}_{T\to\infty} \left(\frac{A\,\hat{\Theta}^{-1}A}{T}\right)^{-1} + \bar{Q}\,\hat{\Theta}\,\bar{Q}'$$

Since $\hat{\Theta}$ is by assumption is positive definite, then

$$(2.11.13) \qquad C = \bar{Q}'\,\hat{\Theta}\,\bar{Q}' \quad \text{is positive semidefinite.}$$

Then from (2.11.12)

$$(2.11.14) \qquad C_{2SLS} - C_{3SLS} = C.$$

Dhrymes (1969) has shown that the (3SLS) is "strictly efficient" relative to the (2SLS) estimator in the sense that the generalized variance of the (3SLS) is strictly less than that of the (2SLS) if and only if

$$(2.11.15) \qquad \operatorname{rank}(\bar{Q}) > 0 \;.$$

It should be noted $C = 0$ when $\sigma_{mn} = 0$ $m \neq n$ i.e., when every equation in the system is exactly identified. In this case the (2SLS) and (3SLS) estimators are equivalent.

We have observed that there is no gain in efficiency of the (3SLS) estimator over the (2SLS) estimator when the variance-covariance matrix is diagonal and/or all the equations of the system are exactly identified.

Srivastava and Tiwari (1978) have obtained somewhat more general conditions under which the (2SLS) and (3SLS) estimators will be identical. They employed a necessary and sufficient condition which C.R. Rao (1968) obtained for $\hat{\delta}_{2SLS}$ and $\hat{\delta}_{GLS}$ to be identical.

Consider

(2.11.16) $$A'\Theta D = 0$$ where

(2.11.17)
$$D = \begin{bmatrix} D_1 & 0 & \cdots & 0 \\ 0 & D_2 & \cdots & \vdots \\ \vdots & \vdots & & \vdots \\ 0 & 0 & & D_M \end{bmatrix}$$

is a $MK \times (MK - \sum_{M=1}^{M} (M_m - 1 + K_m))$ matrix, with full column rank, orthogonal to A and $D_m (m=1,\ldots,M)$ is a $KX(M_{m-1} + K_m)$.

Now

(2.11.18) $A'_m D_m = 0$ by the orthogonality of A and D.

In view of (2.11.17) and (2.11.18), the condition (2.11.16) reduces to the following conditions:

(2.11.19) $\sigma_{mn} A'_n D_n = 0$ for every $m \neq n$ $(m,n=1,2,\ldots M)$.

As pointed out by Zellner and Theil (1962) $\hat{\delta}_{2SLS}$ and $\hat{\delta}_{GLS}$ will be identical if disturbances are uncorrelated. The same result holds true if

(2.11.20) $A'_m D_n = 0$

for every $m \neq n$ which are free from the σ_{mn}'s and hence characterize the situations under which $\hat{\delta}_{2SLS}$ and $\hat{\delta}_{3SLS}$ are identical, since

$\hat{\delta}_{3SLS}$ is a feasible version of $\hat{\delta}_{GLS}$.

From (2.10.25) $A_m = R^{-1}X'Z_m$, the relations (2.11.20) will hold if

(2.11.21) $\quad R^{-1}D_n = 0$, for every n

and since $\quad (X'X) = RR'$,

(2.11.22) $\quad (X'X)^{-1}D_n = 0$, for every n,

which also characterize the equivalence of $\hat{\delta}_{2SLS}$ and $\hat{\delta}_{GLS}$. Now let A^* and D^* be matrices, with nonstochastic elements and full column rank, of order $K \times N$ and $K \times (K-N)$ respectively, such that

(2.11.23) $\quad A^* D^* = 0.$

If V_s denote the vector-space of all K-tuple vectors with real coordinates, then

(2.11.24) $\quad V_s = M(A^*) \oplus M(D^*)$

where $M(A^*)$ and $M(D^*)$ are column spaces of A^* and D^* , respectively and $\oplus$ indicates the direct sum of subspaces. Since the entire space V_s is generated by the columns of A^* and D^* , we can find matrices G_{m1}, G_{m2}, G_{m3} and G_{m4} such that

(2.11.25) $\quad A_m = A^*G_{m1} + D^*G_{m2}$

$\quad D_m = D^*G_{m3} + D^*G_{m4}$

when all A_m's lie in the column space of A^* and all D_m's lie in the column space of D^*. Then G_{m2} and G_{m4} will be null matrices.

Therefore, we have

$$(2.11.26) \qquad A_m = A^*G_{m1}$$

$$D_m = D^*G_{m3}$$

satisfying (2.11.20), i.e. $A_m'D_n = 0$ for all m and n $(m,n=1,2...M)$. It should also be observed that if all the A_m's contain linear combinations of the same variables, $\hat{\delta}_{2SLS}$ and $\hat{\delta}_{3SLS}$ will be idnetical.

Zellner and Theil (1962) proved that if the system of M structural equations consists of a subsystem of M^* equations that are exactly identified and $M-M^*$ equations that are overidentified, the large sample efficiency of the estimates of the parameters in the class of overidentified equations is unaffected if the (3SLS) method is applied to that class alone, ignoring the M^* exactly identified equations. Narayanan (1969) demonstrated that the (3SLS) estimates themselves - not just their sample efficiency - are unaffected if the M^* exactly identified equations are ignored and the (3SLS) method of estimation is applied only to the $M-M^*$ overidentified equations.

12. THE FULL INFORMATION MAXIMUM LIKELIHOOD (FIML) METHOD

The (FIML) method like the (3SLS) is a "system method" in which we estimate the parameters of all the equations simultaneously using all the information on the model.

As was done in (2.9.10), we use the model $By_t + \Gamma x_t = u_t$ $(t = 1,2,\ldots,T)$ to write all M equations at all T sample periods as

(2.12.1) $$YB' + X\Gamma' = U \quad .$$

We assume that

(2.12.2) $u_t \sim N(0\,,\Sigma)$ where Σ is a positive definite matrix.

where Σ is a positive definite matrix. With no loss of generality, we assume exclusion restrictions of B and Γ. There are no restrictions on Σ .

The reduced-form corresponding to (2.12.1) is given by

(2.12.3) $$Y = X\Pi' + V \qquad \text{where}$$

(2.12.4) $$\Pi' = -\Gamma'(B^{-1})' \quad , \quad V = U(B^{-1})'$$

Also

(2.12.5) $$f(u_t) = (2\pi)^{-M/2} \left|\Sigma\right|^{-1/2} \exp(-\tfrac{1}{2}\, u_t' \Sigma^{-1}\, u_t) \ .$$

The joint density of the u's. (for all T observations) is given by

(2.12.6) $$(2\pi)^{-MT/2} \left|\Sigma\right|^{-T/2} \exp\left[-\frac{1}{2} \sum_{t=1}^{T} u_t' \Sigma^{-1} u_t\right]$$

Transforming this density from the unobservable u's to observable y's, we have the joint density of the y's (the endogenous variables) given the x's (the exogenous variables) as:

$$(2.12.7) \qquad \frac{|B|^T}{(2\pi)^{MT/2}|\Sigma|^{T/2}} \exp\left[-\frac{1}{2} \operatorname{tr} \Sigma^{-1}(YB' + X\Gamma')'(YB' + X\Gamma')\right]$$

where from $f(y_t|x_t) = f(u_t)\left|\frac{\partial u_t}{\partial y_t}\right|$

$$= f(u_t)\left|\frac{\partial(By_t + \Gamma x_t)}{\partial y_t}\right|$$

$$= f(u_t)\,|B|$$

where $|B|$ IS THE Jacobian of the transformation from u_t to y_t. Denoting the log-likelihood by L, we have

$$(2.12.8) \qquad L(B';\Gamma';\Sigma) = \text{constant} + \frac{T}{2}\log|\Sigma| + T\log|B'|$$
$$- \frac{1}{2}\left[\operatorname{tr} \Sigma^{-1}(YB' + X\Gamma')'(YB' + X\Gamma')\right].$$

Maximization of $L(B',\Gamma',\Sigma)$ with respect to the a priori restriction on B' and Γ yields the (FIML) estimators.

Since there are no restrictions on Σ, we can replace the estimate of Σ, i.e.,

$$(2.12.9) \qquad \hat{\Sigma} = \frac{1}{T}(YB' + X\Gamma')'(YB' + X\Gamma')$$

in (2.12.8) to get the concentrated likelihood function.

$$(2.12.10) \qquad L^*(B',\Gamma') = \text{const.} - \frac{T}{2}\log|(YB' + X\Gamma')'(YB' + X\Gamma')|$$
$$+ T\log|B'|$$

Since $\log|B'| = \frac{1}{2}\log|BY'YB'| - \frac{1}{2}\log|Y'Y|$ and $|Y'Y|$ is a constant (being a function of observations only) then

$$(2.12.11) \qquad L^*(B',\Gamma') = \text{const} + \frac{T}{2}\log\left|BY'YB'\right| - \frac{T}{2}\log\left|(YB' + X\Gamma')'(YB' + X\Gamma')\right| .$$

It has been established in the literature that we can obtain the same normal equations of the (FIML) estimator as those obtained by maximizing the likelihood function (2.12.8) if we minimize the generalized residual variance $|T'V'V|$ of the reduced form $Y = X\Pi' + V$ subject to the a priori restrictions.

Chow (1968) interpreted the (FIML) estimates as generalized least-squares estimates, in the sense that the generalized variance of $U = YB' + X\Gamma'$ is minimized relative to the generalized variance of the linear combinations YB' of the dependent variables, i.e., the ratio

$$(2.12.12) \qquad \frac{|T'U'U|}{|T^{-1}BY'YB'|} \quad \text{is minimized.}$$

Scharf (1976) used the same arguments as Chow (1968) to derive the normal equations of the (FIML) estimates. He, however, minimized instead

$$(2.12.13) \qquad \frac{|T^{-1}U'U|}{|T^{-1}BY'MYB'|}$$

where $M = I - X(X'X)^{-1}X'$ and (2.12.13) is equivalent to

$$(2.12.14) \qquad \frac{\left|T^{-1} V'V\right|}{\left|T^{-1} \hat{V}' \hat{V}\right|}, \qquad \text{where}$$

$$V = U(B^{-1})' \qquad \text{and}$$

$$\hat{V} = Y - \hat{Y} = Y - X(X'X)^{-1}X'Y$$

the (OLS) estimated-error term of the reduced form. In other words, Scharf (1976) obtained the (FIML) estimates by minimizing the generalized residual variance of the reduced form relative to the estimated residual variance of the reduced form as a function of the structural parameters subject to a priori restrictions.

Multiplying (2.12.13) by $\frac{T}{2}$ and taking logarithms we have

$$(2.12.15) \qquad L = (\tfrac{T}{2})\log\left|T^{-1}U'U\right| - \frac{T}{2}\log\left|T^{-1}BY'MYB'\right|$$

Let

$$(2.12.16) \qquad W = T^{-1}BY'MYB' ,$$

then the (m,n)th element of W is given by

$$(2.12.17) \qquad T^{-1}[y_m' - \beta_m Y_m'] M[y_n' - Y_n\beta_n'] = w_{mn}$$

Let

$$(2.12.18) \qquad S = T^{-1}U'U$$

and the (m,n)th element of S is given by

$$(2.12.19) \qquad T^{-1}[y_m' - \beta_m Y_m' - \gamma_m X_m'] [y_n - Y_n\beta_n' - X_n\gamma_n'] = s_{mn}$$

Taking the partial derivations of L w.r.t. β_i' and γ_i' $(i=1,2,\dots,M)$ we have

$$(2.12.20) \qquad \frac{\partial L}{\partial \beta_i'} = \frac{T}{2} \sum_{m,n=1}^{T} s^{mn} \frac{\partial s_{mn}}{\partial \beta_i'} - \frac{T}{2} \sum_{m,n=1}^{M} w^{mn} \frac{\partial w_{mn}}{\partial \beta_i'}$$

where s^{mn} and w^{mn} are the (m,n)th element of S^{-1} and W^{-1}, respectively.

Following Chow (1968) and Scharf (1976), we have

$$(2.12.21) \qquad \frac{\partial s_{mn}}{\partial \beta_i'} = \begin{cases} 0 & i \neq m,\ i \neq n \\ -T^{-1} Y_i'[y_n - Y_n\beta_n' - X_n\beta_n] & \text{for } i = m \end{cases}$$

$$(2.12.22) \qquad \frac{\partial w_{mn}}{\partial \beta_i'} = \begin{cases} 0, & i \neq m,\ i \neq n \\ -T^{-1} Y_i'M(y_n - Y_n\beta_n') & \text{for } i = m \end{cases}$$

Then,

$$(2.12.23) \qquad \frac{\partial L}{\partial \beta_i'} = -\sum_n s^{in} Y_i'(y_n - Y_n\beta_n' - X_n\gamma_n') + \sum_n w^{in} Y_i'M(y_n - Y_n\beta_n')$$

Similarly,

$$(2.12.24) \qquad \frac{\partial L}{\partial \gamma_i'} = -\sum_n s^{in} X_i'(y_n - Y_n\beta_n' - X_n\gamma_i') .$$

Then

$$(2.12.25) \qquad \frac{\partial L}{\partial \delta_i} = -\sum_n s^{in} Z_i(y_n - Z_n\delta_n) + \sum w^{in} Z_i' M(y_n - Z_n\delta_n)$$

where

$$\delta_i = \begin{bmatrix} \beta_i \\ \gamma_i \end{bmatrix} \qquad Z = [y\ ,\ X] .$$

Then if we put $\frac{\partial L}{\partial \delta_i} = 0 \qquad (i = 1, \ldots, M)$

and

$$\delta = \begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_M \end{bmatrix}$$

we have

$$(2.12.26) \qquad \left[Z_i'(s^{in}I - w^{in}M)Z_n\right]\hat{\delta} = \left[\sum_{n=1}^{M} Z_i'(s^{in}I - w^{in}M)y_n\right]$$

$$i = 1, \ldots, M$$

or

$$(2.12.27) \qquad \left[s^{in}Z_i'(I - \frac{w^{in}}{s^{in}}M)Z_n\right]\hat{\hat{\delta}} = \left[\sum_{n=1}^{M} s^{in}Z_i'(I - \frac{w^{in}}{s^{in}}M)y_n\right]$$

$$i = 1, \ldots, M$$

K-MATRIX-CLASS (KMC) ESTIMATORS

Now Scharf (1976) defined the (KMC) estimators by the normal equations:

$$(2.12.28) \qquad \left(s^{in}Z_i'[I - k_{in}M]Z_n\right)\hat{\hat{\delta}} = \left[\sum_{n=1}^{M} s^{in}Z_i'[I - k_{in}M)y_n\right],$$

$$i = 1, \ldots, M$$

where $k_{in}(i, n = 1, \ldots, M)$ are stochastic or non-stochastic scalars. Comparing the normal equations in (2.12.27) with those of (2.12.28) we observe that $\hat{\hat{\delta}}_{FIML}$ belongs to the (KMC) family with

$$K = (k_{in}) = \frac{w^{in}}{s^{in}} \tag{2.12.29}$$

LINEARIZED MAXIMUM LIKELIHOOD (LML) ESTIMATOR

The normal equations for the (FIML) estimates are clearly non-linear. These equations can be linearized if s^{in} and w^{in} are computed from any consistent estimator of the parameters B and Γ. Rothenberg and Leenders (1964) outlined an interesting procedure for linearizing the FIML normal equations. They differentiated the likelihood function (2.12.8) with respect to the unknown elements of Σ. These partial derivatives are set equal to zero and the resulting equations are used to eliminate the unknown Σ in (2.12.8). This new function thus obtained is called the concentrated likelihood function, (2.12.10). The concentrated likelihood function is differentiated with respect to

$$\delta = \begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_M \end{bmatrix}$$

Let $\ell(\delta)$ denote the resulting gradient of the concentrated likelihood function. Then the (FIML) estimator is obtained from the (nonlinear) equation $\ell(\delta) = 0$.

To obtain the (LML) estimator, choose δ^* as some consistent estimate of δ such that $\delta^* - \delta$ is of order $T^{-1/2}$ in probability i.e.,

$$\sqrt{T}(\delta^* - \delta) \tag{2.12.30}$$

has a limiting distribution with zero mean vector and finite variance covariance matrix. One possible choice of δ^* is the (2SLS) vector $\hat{\delta}_{(2SLS)}$. New equate to zero the Taylor-series linear approximation of $\ell(\delta)$, i.e.,

$$\ell(\delta^*) + H^*(\delta = \delta^*) = 0 \tag{2.12.31}$$

where H^* is the Hessian matrix of the concentrated likelihood function evaluated at δ^*.

From (2.12.32), we obtain

$$\bar{\delta} = \delta^* - (H^*)^{-1}\ell(\delta^*) \tag{2.12.33}$$

which is known as the linearized maximum likelihood (LML) estimator.

LINEAR (KMC) ESTIMATORS

From (2.12.28) it is clear that the normal equations of the (KMC) estimators are non-linear. We can linearize the normal equations of the KMC estimators by computing s^{in} and (if necessary) k_{in} from any consistent estimator the true parameters B and Γ. The solution thus obtained is called the Linear (KMC) estimator, e.g., if we take $\hat{\delta}_{(2SLS)}$, the (2SLS) estimator for s^{in} and set $k_{in} = 1$, $(i,n = 1,2,\ldots,M)$ we obtain the (3SLS) estimator as a special case of the Linear (KMC) estimator.

Then, the i-th reduced-form equation is given by

$$(3.1.7) \qquad y_i = X\pi_i + v_i$$

where y_i is the $T \times 1$ vector of observations on the i-th dependent variable, X is the $T \times K$ matrix of observations on all the predetermined variables, π_i is the $K \times 1$ vector of reduced-form coefficients of the i-th equation, and v_i is the $T \times 1$ vector of distrubances in the i-th reduced-form equation.

The (OLS) estimator of π_i is given by

$$(3.1.8) \qquad \hat{\pi}_i = (X'X)^{-1}X'y_i$$

and its sampling error is given by

$$(3.1.9) \qquad \begin{aligned} \hat{\pi}_i - \pi_i &= (X'X)^{-1}X'y_i - \pi_i \\ &= (X'X)^{-1}X'(X\pi_i + v_i) - \pi_i \\ &= (X'X)^{-1}X'v_i \end{aligned}$$

Also,

$$(3.1.10) \qquad \begin{aligned} \text{plim.}(\hat{\pi}_i - \pi_i) &= \text{plim.}(T^{-1}X'X)^{-1}\ \text{plim.}(T^{-1}X'v_i) \\ &= 0 \end{aligned}$$

since

$$\text{plim}(T^{-1}X'v_i) = 0$$

[the disturbances are uncorrelated with the predetermined variables] and by assumption

$$\text{plim}(Y^{-1}X'X)^{-1} = \Sigma_{XX}^{-1}$$

exists as a nonsingular matrix with finite elements. Therefore,

$$\text{(3.1.11)} \qquad \operatorname{plim} \hat{\pi}_i = \pi_i \quad .$$

Hence $\hat{\pi}_i$ is a consistent estimator of π_i and its asymptotic covariance matrix is given by

$$\text{(3.1.12)} \qquad T^{-1} \operatorname{plim}[\sqrt{T}(\hat{\pi}_i - \pi_i)(\hat{\pi}_i - \pi_i)']$$

$$= T^{-1} \operatorname{plim} T(X'X)^{-1}X'v_i v_i' X(X'X)^{-1} \quad .$$

Now if we consider all M reduced-form equations, i.e. (3.1.4)

$$Y = X\Pi' + V \quad ,$$

the coefficient matrix Π' is consistently estimated by

$$\text{(3.1.13)} \qquad \hat{\Pi}'_{(URF)} = (X'X)^{-1}X'Y$$

or by

$$\text{(3.1.14)} \qquad \hat{\Pi}_{(URF)} = Y'X(X'X)^{-1} \quad .$$

This method of estimating the reduced-form coefficients Π is called the Unrestricted Reduced Form (URF) method. In this method, consistent estimates of the reduced-form coefficients are obtained by applying (OLS) to each reduced-form equation. This unrestricted (OLS) estimator of Π does not, in general, incorporate the a priori information.

It should be noted that in the estimation of reduced or structural form of an econometric model we make use of two types of information: sample information (embodied in the predetermined variables) and identifying restrictions (a priori information). The (URF) estimators use all the sample information but none of the a priori information, i.e.,

$\hat{\Pi}$ does not incorporate the a priori restrictions imposed by the relation $\Pi = -B^{-1}\Gamma$.

2. We now look at a more efficient method of reduced-form estimation called the Restricted (or Derived) Reduced Form (RRF) method which was used by Goldberger, Nagar and Odeh (1961) to obtain reduced-form estimates and forecasts from consistently estimated structural equations.

If $\hat{B}$ and $\hat{\Gamma}$ are estimates of the structural coefficients B and Γ, we may estimate the reduced-form coefficient matrix Π by the Restricted Reduced Form (RRF) coefficient matrix.

$$\hat{\Pi}_{(RRF)} = -\hat{B}^{-1}\hat{\Gamma} \tag{3.2.1}$$

and $\hat{\Pi}_{(RRF)}$ will be consistent since $\hat{B}$ and $\hat{\Gamma}$ are consistent estimates of B and Γ respectively.

If $\hat{\Sigma}$ is a consistent estimate of the structural form disturbance covariance matrix Σ, we may estimate the reduced-form disturbance covariance matrix Ω by the restricted reduced form residual covariance matrix:

$$\hat{\Omega} = \hat{B}^{-1}\hat{\Sigma}(\hat{B}^{-1})' \tag{3.2.2}$$

since

$$\Omega = E(V'V) = E[(B^{-1})U'U(B^{-1})'] = B^{-1}\Sigma(B^{-1})'$$

and $\hat{\Omega}$ will be consistent since $\hat{B}$ and $\hat{\Sigma}$ are consistent estimates of B and Σ respectively.

It should be noted that the Restricted Reduced Form (RRF), $\hat{\Pi}_{RRF}$ will coincide with the Unrestricted Reduced Form, $\hat{\Pi}_{URF}$, only if all the structural equations of the model are exactly identified. Although $\hat{\Pi}_{(URF)}$ and $\hat{\Pi}_{(RRF)}$ are consistent estimators of Π, the latter incorporates more a priori information and will therefore be more efficient, at least asymptotically.

If we obtain consistent estimators $\hat{B}$ and $\hat{\Gamma}$ of B and Γ respectively (say, by Maximum Likelihood or 2SLS) methods) with known asymptotic variance-covariance matrices, we may need to find out the asymptotic variance-covariance matrix of the elements of $\hat{\Pi}_{(RRF)}$. To answer the preceding question, we follow closely the work of Goldberger, Nagar and Odeh (1961).

It should be recalled that in the reduced-form each current endogenous variable is expressed in terms of only predetermined variables and a disturbance. The disturbance is uncorrelated with the predetermined variables so that the conditional expectation of the endogenous variable is given by a linear function of the predetermined variables, i.e.,

$$(3.2.3) \qquad E(y_t|x_t) = E[(\Pi x_t + v_t|x_t)]$$

$$= \Pi x_t$$

since $E(V_t|x_t) = E(v_t) = 0$. A particular reduced form coefficient π_{mk} (say) may be interpreted as the partial derivative (of the conditional expectation) of the current endogenous variables y_{mt} with respect to the predetermined x_{kt}, with all the other x_t's held constant.

The derivation of the asymptotic variances and covariances of all (RRF) coefficients is a problem which calls for going from the variances and covariances of one set of random variables to the variances and covariances of another, functionally related, set of random variables. The following Lemma aids in the solution of the preceding problem:

LEMMA (3.2.A)

Let $\hat{x}$ be the typical item of a sequence of random vectors and let $\hat{y} = f(\hat{x})$ be a vector whose elements are differentiable functions of x. If $\underset{T\to\infty}{\text{plim}}\, E(\hat{x}) = \underset{T\to\infty}{\text{plim}}\, \hat{x} = x$ and

$$T^{-1} \underset{T\to\infty}{\text{plim}} [\sqrt{T}(\hat{x} - x)][\sqrt{T}(\hat{x} - x)]' = T^{-1}\Delta$$

(where Δ is a matrix of finite constants). Then

$$\underset{T\to\infty}{\lim} E(\hat{y}) = \underset{T\to\infty}{\text{plim}}\ \hat{y} = \underset{T\to\infty}{\text{plim}}\ f(\hat{x}) = f(x) = y$$

and

$$T^{-1} \underset{T\to\infty}{\text{plim}} [\sqrt{T}(\hat{y} - y)][\sqrt{T}(\hat{y} - y)]' = T^{-1} L' \Delta L$$

where

$$L = \frac{\partial \hat{y}}{\partial \hat{x}} \quad \text{is evaluated at} \quad \hat{x} = x .$$

(see Goldberger (1964), page 125).

In our case, let α be a column vector of all structural coefficients such that

$$(3.2.4) \qquad \alpha' = [\beta_{11} \cdots \beta_{1M} \gamma_{11} \cdots \gamma_{1K} \beta_{21} \cdots \gamma_{2K} \cdots \beta_{M1} \cdots \beta_{MM} \gamma_{M1} \cdots \gamma_{MK}]$$

with the coefficients of the first structural equation followed by the coefficients of the second, and so on, for the M structural equations.

Let $\hat{\alpha}$ denote the consistent estimates of α obtained by (2SLS), say. (It should be noted that α has $M(M+K)$ components.) That is,

$$\lim_{T\to\infty} E(\hat{\alpha}) = \operatorname*{plim}_{T\to\infty} \hat{\alpha} = \alpha \quad (3.2.5)$$

Let π denote the column vector of all the reduced-form coefficients, i.e., π is a vector of MK components and

$$\pi' = [\pi_{11}\pi_{12}\cdots\pi_{2K}\cdots\pi_{M1}\cdots\pi_{MK}] \quad (3.2.6)$$

with $\hat{\pi}$ being the MK vector of consistent (RRF) coefficients where

$$\hat{\pi} = f(\hat{\alpha}) \qquad \text{i.e.} \qquad \hat{\pi}_i = f_i(\hat{\alpha}) \quad i = 1,2,\ldots. \quad (3.2.7)$$

Let ϕ be the $M(M+K)$-th order square matrix of asymptotic variances and covariances of the elements of $\sqrt{T}(\hat{\alpha}-\alpha)$, i.e.,

$$T^{-1} \operatorname{plim} [\sqrt{T}(\hat{\alpha}-\alpha)][\sqrt{T}(\hat{\alpha}-\alpha)]' = T^{-1}\phi \quad (3.2.8)$$

Then by Lemma (3.2.A):

$$\lim_{T\to\infty} E(\hat{\pi}) = \operatorname{plim} \hat{\pi} = \operatorname{plim} f(\hat{\alpha}) = f(\alpha) = \pi \quad (3.2.9)$$

and

$$\begin{aligned} \psi &= T^{-1} \operatorname{plim} [\sqrt{T}(\hat{\pi}-\pi)][\sqrt{T}(\hat{\pi}-\pi)]' \\ &= T^{-1} D' \phi D \end{aligned} \quad (3.2.10)$$

where

$$D = \frac{\partial\hat{\pi}}{\partial\hat{\alpha}} = \left(\frac{\partial\hat{\pi}_1}{\partial\hat{\alpha}}, \frac{\partial\hat{\pi}_2}{\partial\hat{\alpha}} \cdots\right)_{\hat{\alpha}=\alpha} \quad (3.2.11)$$

Since the matrix D is evaluated at $\hat{\alpha} = \alpha$, then

(3.2.12) $$D = \frac{\partial \pi}{\partial \alpha}$$

Goldberger, Nagar and Odeh (1961) have shown that

(3.2.13) $$\frac{\partial \pi}{\partial \alpha} = (B^{-1})' \otimes \begin{bmatrix} \Pi \\ I_K \end{bmatrix}$$

where B is $M \times M$ and $\begin{bmatrix} \Pi \\ I_K \end{bmatrix}$ is $(M + M) \times K$

so that

(3.2.14) $$\frac{\partial \pi}{\partial \alpha} = \begin{bmatrix} \beta^{11} \begin{bmatrix} \Pi \\ I \end{bmatrix} & \cdots\cdots\cdots & \beta^{1M} \begin{bmatrix} \Pi \\ I \end{bmatrix} \\ \vdots & & \vdots \\ \beta^{1M} \begin{bmatrix} \Pi \\ I \end{bmatrix} & \cdots\cdots\cdots & \beta^{MM} \begin{bmatrix} \Pi \\ \end{bmatrix} \end{bmatrix}$$

In practice D will be unknown, i.e., the true coefficients in (3.2.13) and (3.2.14), are unknown but consistent estimators are provided by $\hat{B}$ and $\hat{\Gamma}$. Also ϕ will be unknown but a consistent estimate $\hat{\Phi}$ of ϕ may be computed. Then

(3.2.15) $$\hat{\psi} = \hat{D}' \hat{\phi} \hat{D}$$

is clearly a consistent estimate of ψ. ($\hat{D}$, the consistent estimate of D will be obtained from the consistent estimates $\hat{B}^{-1}$ and $\hat{\Pi}$.

3. We now look at the problem of predicting or forecasting endogenous variables for given values of the exogenous (predetermined) variables with reference to the reduced form equations.

Suppose the $(K \times 1)$ vector x_* represent the values of the forecast period, where the asterisk denotes the period for which the forecasts are made. Then the forecast values of the M endogenous values are given by the vector

$$\hat{y}_* = \hat{\Pi} x_* \tag{3.3.1}$$

where Π is the $(M \times K)$ matrix of reduced-form coefficients. Previously we obtain consistent estimates of Π by $\hat{\Pi}_{(URF)} = Y'X(X'X)^{-1}$ and by using consistent estimates of the structural coefficients, say $\hat{B}$ and $\hat{\Gamma}$ and setting $\hat{\Pi} = -\hat{B}^{-1}\hat{\Gamma}$.

In this section we derive the variance-covariance matrix when $\hat{\Pi}$ has been obtained by a direct application of (OLS) i.e. $\hat{\Pi}_{(URF)}$. From (3.1.2), the true values of the endogenous variables in the forecast period are

$$y_* = \Pi x_* + v_* \tag{3.3.2}$$

where y_* is the $M \times 1$ vector of endogenous variables [which will be determined by (3.3.2)], x_* is the $K \times 1$ vector of values of the predetermined values for the forecast period and v_* is the $(M \times 1)$ vector of reduced-form disturbances for the forecast period, Π is the $M \times K$ matrix of reduced-form coefficients.

Let the vector of forecast errors be

$$(3.3.3) \qquad \hat{y}_* - y_* = \hat{\Pi}x_* - \Pi x_* - v_*$$

$$= (\hat{\Pi} - \Pi)x_* - v_* \quad .$$

The forecasts are unbiased since $E(\hat{\Pi}) = \Pi$ and $E(v_*) = 0$. The variance-covariance matrix of the forecast error is

$$(3.3.4) \qquad E[\hat{y}_* - y_*)(\hat{y}_* - y_*)'] = \Sigma_{**}$$

Substituting (3.3.3) into (3.3.4) leads to

$$(3.3.5) \qquad \Sigma_{**} = E[(\hat{\Pi} - \Pi)x_*x_*'(\hat{\Pi} \quad \Pi)'] + E(v_*v_*') \ .$$

[The cross-product terms vanish since $\hat{\Pi}$ and v_* are independent on the usual assumptions of serial independence in the structural (and hence reduced-form) disturbances.

If we assume constant variances and non-zero contemporaneous covariances for the reduced-form disturbances, we can set

$$(3.3.6) \qquad E(v_*v_*') = \Sigma_{**} \qquad \text{(say)}.$$

Hence we need only evaluate

$$E[(\hat{\Pi} - \Pi)\ x_*x_*'\ (\hat{\Pi} - \Pi)'] \quad .$$

Now we can write all M reduced form equations as

$$Y = X\Pi' + V$$

where

$$(3.3.7) \qquad V = [v_1, v_2, \ldots, v_M] \quad ; \quad \text{i.e.,} \quad V = U(B^{-1})'$$

is the $(T \times M)$ matrix of reduced form disturbances and v_i $(i=1,2,\ldots M)$

is the column vector for the reduced-form disturbances for Y_i.

From $$Y = X\Pi' + V$$

$$Y' = \Pi X' + V'$$

Therefore,

$$Y'X(X'X)^{-1} = \Pi X'X(X'X)^{-1} + V'X(X'X)^{-1}$$

using $\hat{\Pi} = Y'X(X'X)^{-1}$, we have

(3.3.8) $$\hat{\Pi} = \Pi + V'X(X'X)^{-1}$$

or $$(\hat{\Pi} - \Pi) = V'X(X'X)^{-1} .$$

Therefore

(3.3.9) $$E[\hat{\Pi} - \Pi)\, x_* x_*\, (\hat{\Pi} - \Pi) \quad \text{gives}$$

$$E\begin{bmatrix} v_1'\Lambda'\Lambda v_1 & v_1'\Lambda'\Lambda v_2 & \cdots\cdots & v_1'\Lambda'\Lambda v_M \\ \vdots & & & \\ v_M'\Lambda'\Lambda v_1 & \cdots\cdots & \cdots\cdots & v_M'\Lambda'\Lambda v_M \end{bmatrix}$$

where

(3.3.10) $$\Lambda = x_*'(X'X)^{-1}X' .$$

Consider the terms on the principal diagonal of (3.3.9):

(3.3.11) $$\begin{aligned} E(v_i\Delta'\Lambda v_i) &= E[\mathrm{tr}(v_i\Lambda'\Lambda v_i)] \\ &= E[\mathrm{tr}(\Lambda v_i v_i'\Lambda')] \\ &= \omega_{ii}\ \mathrm{tr}\ (\Lambda\Lambda') \end{aligned}$$

where ω_{ii} is an element of Ω. (Since $E(v_i v_i') = \omega_{ii} I_T$ on the usual assumptions on the reduced form disturbances.) From (3.3.10) we have have

(3.3.12) $$\Lambda\Lambda' = x_*(X'X)^{-1}X'X(X'X)^{-1}x_*$$

$$= x_* (X'X)^{-1} x_*$$

Therefore,

(3.3.13) $$E(v_i\Lambda'\Lambda v_i) = \omega_{ii} x_*(X'X)^{-1}x_*$$

(3.3.14) $$E(v_i\Lambda'\Lambda v_j) = \omega_{ij} x_*(X'X)^{-1}x_*$$

where ω_{ii} and ω_{ij} are elements of

$$E(v_* v_*') = \Sigma_{vv}$$

Thus,

(3.3.15) $$E[(\hat{\Pi} - \Pi)x_* x_*' (\hat{\Pi} - \Pi)'] = x_*'(X'X)^{-1} x_* \Sigma_{vv}$$

Σ_{vv} is unknown but an unbiased estimator for Σ_{vv} is provided by

(3.3.16) $$S_{vv} = \frac{1}{T-K}(Y - X\hat{\Pi}')'(Y - X\hat{\Pi}')$$

$$= \frac{1}{T-K}(Y'Y - \hat{\Pi}X'Y - Y'X\hat{\Pi}' + \hat{\Pi}X'X\hat{\Pi}')$$

$$= \frac{1}{T-K}(Y'Y - \hat{\Pi}X'Y)$$

Since

$$Y'X\hat{\Pi} = Y'X(X'X)X'Y$$

and

$$\hat{\Pi}X'X\hat{\Pi}' = Y'X(X'X)^{-1}X'X(X'X)^{-1}X'Y$$

$$= Y'X(X'X)^{-1}X'Y$$

Now substituting (3.3.15) into (3.3.5) gives

$$\Sigma_{**} = x_*'(X'X)^{-1}x_*\Sigma_{vv} + \Sigma_{vv} \tag{3.3.17}$$

$$= (1 + x_*(X'X)^{-1}x_*)\Sigma_{vv}$$

and the estimated variance-covariance matrix of the error of forecast is given by

$$S_{**} = (1 + x_*(X'X)^{-1}x_*)S_{vv} \quad . \tag{3.3.18}$$

Now we derive the asymptotic-covariance matrix of forecasts when $\hat{\Pi} = -\hat{B}^{-1}\hat{\Gamma}$ i.e., $\hat{\Pi}_{RRF}$. Once again we follow the approach taken by Goldberger, Nagar and Odeh (1961).

For the $(K \times 1)$ vector x_* of values of the predetermined variables for a forecast period, the $(M \times 1)$ vector y_* of the endogenous variables will be determined by

$$y_* = \Pi x_* + v_* \quad . \tag{3.3.19}$$

Let F be the $(M \times MK)$ matrix displaying the values of the predetermined variables for the forecast period.

(3.3.20)

$$F_* = \begin{bmatrix} x_{1*} \cdots x_{K*} & 0 \cdots 0 & \cdots 0 \cdots 0 \\ 0 \cdots 0 & x_{1*} \cdots x_{K*} & \\ \vdots & & \\ 0 \cdots 0 & 0 \cdots 0 & \cdots x_{1*} \cdots x_{K*} \end{bmatrix}$$

so that $F_*\pi = \Pi x_*$ where π is the vector of MK elements defined in (3.2.6).

Therefore

(3.3.21) $$y_* = F_*\pi + v_* \quad ,$$

our forecast will be determined by

(3.3.22) $$\hat{y}_* = \hat{\Pi} x_* \quad ,$$ and therefore

(3.3.23) $$\hat{y}_* = F_*\hat{\pi}$$

The vector of forecast errors is given by

(3.3.24)
$$\begin{aligned} \hat{y}_* - y_* &= (\hat{\Pi} - \Pi)\, x_* - v_* \\ &= F_*\hat{\pi} - F_*\pi - v_* \\ &= F_*(\hat{\pi} - \pi) - v_* \quad . \end{aligned}$$

If we treat x_* as fixed, we find that

$$\operatorname*{plim}_{T\to\infty} E(\hat{\pi} - \pi) = 0 \quad \text{and} \quad E(v_*) = 0.$$

The asymptotic forecast covariance matrix is given by

$$(3.3.25) \qquad \Theta = F_* T^{-1} \operatorname{plim}[\sqrt{T}(\hat{\pi} - \pi)][\sqrt{T}(\hat{\pi} - \pi)]' F_*' + E(v_* v_*')$$

$$= T^{-1} F_* \psi E' + \Omega$$

where ψ was defined in (3.2.10) and if we assume $\hat{\pi}$ is uncorrelated with v_* and v_* is distributed as v_t $(t = 1,2,\ldots,T)$ [Recall $E(v_t v_t' = \Omega]$.

In practice Θ will be unknown since ψ and Ω are unknown. However we may obtain

$$\hat{\Theta} = F_* \hat{\psi} F_*' + \hat{\Omega}$$

as a consistent estimate of Θ

4. By this time it should be clear that there is a statistical difference between a reduced-form system and a structural system. In the reduced form system only predetermined variables appear as explanatory variables, whereas, in the structural system current endogenous variables may (and usually do) appear as explanatory variables. The structural system describes accurately the exact manner in which all current endogenous variables and predetermined variables mutually interact with the specified economic system.

A reduced system, on the other hand, gives only a partial view of this interaction, for it merely describes the way in which the predetermined variables serve to influence the behaviour of the current endogenous variables, after allowance has been made for all interactions among jointly dependent variables.

In the preceding sections of Chapter 3, we discussed the derivation of (asymptotic) variance-covariance matrix of reduced form coefficients and forecasts by both the (2SLS) induced (RRF) estimates and the (URF) estimates. The obvious question should now be whether there is a gain in efficiency of the (2SLS) induced (RRF) estimates over the (URF) estimates.

Dhrymes (1973) provided an answer to this question by proving the following Lemma: (see Lemma 5, Dhrymes, 1973) "Unless (a) the covariance matrix of the structural errors is diagonal, or (b) all equations of the system are exactly identified, (2SLS) induced (RRF) are, asymptotically, neither efficient nor inefficient relative to (URF) estimators." McCarthy (1973) showed that for small samples (in his model has worked with two endogenous variables) moments of the (2SLS) induced (RRF) estimates are in doubt. For his model, he found that forecasts from the (RRF) possess no moments in the case of the over-identified systems.

Since (FIML) and (3SLS) methods lead to asymptotically efficient estimates of B and Γ and since this property carries over to any single-valued functions of B and Γ, it follows that the (3SLS) or (FIML) induced (RRF) estimator leads to a smaller asymptotic variance of reduced form coefficients and forecast error than a (2SLS) or (LIML) induced (RRF) estimator. Dhrymes (1973) (see Lemma 2, Dhrymes, 1973) proved that the (3SLS) induced (RRF) estimator is asymptotically efficient relative to the (2SLS) induced (RRF) estimator.

We consider the derivations by Court (1973) of the asymptotic covariance matrix of the (3SLS) induced (RRF)/ Let us consider the i-th structural equation:

$$y_i = Y_i\beta_i + X_i\gamma_i + u_i \qquad (i = 1,2,\ldots,M)$$

$$= Z_i\delta_i + u_i$$

with reduced form

$$y_i = X\pi_i + v_i$$

We can write a submodel consisting of several structural equations of the form $y_i = Z_i\delta_i + u_i$

(3.4.1)
$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \end{bmatrix} = \begin{bmatrix} Z_1 & 0 & \cdots \\ 0 & Z_2 & \cdots \\ \vdots & & \vdots \\ \cdots & \cdots & \cdots \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \end{bmatrix}$$

Since not all the structural equations have been included in (3.4.1), this submodel can be written concisely as

(3.4.2)
$$y_s = Z\delta + u$$

where Z is a block diagonal matrix, δ is a vector of structural coefficients and u is a vector of the several structural disturbances in (3.4.1). We assume $E(u) = 0$, $E(UU') = \Sigma \otimes I$. In a similar manner we can write the corresponding reduced form equations jointly as

(3.4.3)
$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \end{bmatrix} = \begin{bmatrix} X & 0 & \cdots \\ 0 & X & \cdots \\ \vdots & & \\ \cdots & \cdots & \cdots \end{bmatrix} \begin{bmatrix} \pi_1 \\ \pi_2 \\ \vdots \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ \vdots \end{bmatrix}$$

and concisely as

$$(3.4.4) \qquad y_R = (I \otimes X)\,\pi + v$$

$$(3.4.5) \qquad E(v) = 0 \quad , \quad E(vv') = \Omega \otimes I \quad .$$

Following Court (1973) we estimate the equations (3.4.2) and (3.4.4) jointly by (3SLS). We can write these equations together as

$$(3.4.6) \qquad \begin{bmatrix} y_S \\ y_R \end{bmatrix} = \begin{bmatrix} Z & 0 \\ 0 & I\otimes X \end{bmatrix} \begin{bmatrix} \delta \\ \pi \end{bmatrix} + \begin{bmatrix} u \\ v \end{bmatrix}$$

Let the covariance between u and v in (3.4.2) and (3.4.4) be

$$(3.4.7) \qquad E(vu') = G \otimes I \qquad \text{(say)}$$

then the covariance matrix of disturbances in (3.4.6) is

$$(3.4.8) \qquad E\left\{ \begin{bmatrix} u \\ v \end{bmatrix} [u' \quad v'] \right\} = \begin{bmatrix} \Sigma & G' \\ G & \Omega \end{bmatrix} \otimes I$$

We assume that this covariance matrix is non-singular with inverse given by

$$(3.4.9) \qquad \begin{bmatrix} \Sigma & G' \\ G & \Omega \end{bmatrix}^{-1} \otimes I = \begin{bmatrix} P & Q' \\ Q & R \end{bmatrix} \otimes I$$

To estimate (3.4.6) by the (3SLS) method, we premultiply (3.4.6) by $(I \otimes X')$ where I is an identity matrix, and then apply Generalized Least Squares (GLS) to the transformed system:

$$(3.4.10)\qquad \begin{bmatrix} (I\otimes X')y_S \\ (I\otimes X')y_R \end{bmatrix} = \begin{bmatrix} (I\otimes X')Z & 0 \\ 0 & (I\otimes X')X \end{bmatrix} \begin{bmatrix} \delta \\ \Pi \end{bmatrix} + \begin{bmatrix} (I\otimes X')u \\ (I\otimes X')v \end{bmatrix}$$

The covariance matrix of disturbances in (3.4.10) is given by

$$(3.4.11)\qquad \begin{bmatrix} \Sigma & G' \\ G & \Omega \end{bmatrix} \otimes X'X$$

and its inverse is given by

$$(3.4.12)\qquad \begin{bmatrix} P & Q' \\ Q & R \end{bmatrix} \otimes (X'X)^{-1}$$

Application of (GLS) to (3.4.10) leads to estimators of δ and Π defined by the equations:

$$(3.4.13)\qquad \hat{\delta}_{3SLS} = (Z'[\Sigma^{-1} \otimes X(X'X)^{-1}X']Z)^{-1}Z'[\Sigma^{-1} \otimes X(X'X)^{-1}X']y_s$$

and

$$(3.4.14)\qquad \hat{\Pi}_{3SLS} = [I \otimes (X'X)^{-1}X']y_R - [G\Sigma^{-1} \otimes (X'X)^{-1}X'](y_s - Z\hat{\delta}_{3SLS})$$

Consistent estimators of the covariance matrices of $\hat{\delta}_{3SLS}$ and $\hat{\Pi}_{3SLS}$ are given by

$$(3.4.15)\qquad \text{Est. Cov}(\hat{\delta}) = (Z'[\Sigma^{-1} \otimes X(X'X)^{-1}X']Z)^{-1} = F$$

$$(3.4.16)\qquad \text{Est Cov}(\hat{\Pi}) = (\Omega - G\Sigma^{-1}G') \otimes (X'X)^{-1} + JFJ'$$

where

$$(3.4.17) \qquad J = [G\Sigma^{-1} \otimes (X'X)^{-1}X']$$

(See Court (1973) for more details.)

It should be observed that $\hat{\delta}_{3SLS}$ given in (3.4.13) is the same as that given in Chapter 2 when we applied the (3SLS) technique to the structural equations (3.4.2) while ignoring the reduced-form equations. This result supports the argument of Narayanan (1969), mentioned in Chapter 2, that the exclusion of exactly identified equations from the (3SLS) procedure makes no difference to the estimates of the remaining equations. Here the reduced form equations can be treated as a special case of the exactly identified equations.

Court (1973) demonstrated the efficiency of the (3SLS) induced (RRF) estimates over the (URF) estimates. Let Π^* be the estimator of Π when (OLS) is applied directly to (3.4.4), i.e.,

$$(3.4.18) \qquad \Pi^* = [I \otimes (X'X)^{-1}X']y_R$$

while the covariance matrix of Π^* is consistently estimated by

$$(3.4.19) \qquad \text{Est Cov}(\Pi^*) = \Omega \otimes (X'X)^{-1}$$

Comparing (3.4.16) and (3.4.19) it will be seen that the estimated covariance matrix Π^*, the (URF) estimator, exceeds that of the (3SLS) induced (RRF) estimator Π^*, by a positive semidefinite matrix. (See Court for algebraic details.)

It should also be noted that strict adherence to the (3SLS) procedure would demand that Σ, G and Ω be replaced by consistent estimators. Court (1973) suggested the following estimators of Σ, G and Ω

$$(3.4.20) \qquad \hat{\sigma}_{ij} = \frac{1}{T}(y_i - Z_i\hat{\delta}_{2SLS})'(y_j - Z_j\hat{\delta}_{2SLS})$$

$$(3.4.21) \qquad \hat{g}_{ij} = \frac{1}{T}(y_i - Z_i\hat{\delta}_{2SLS})'(y_j - X\pi_j^*)$$

$$(3.4.22) \qquad \hat{\omega}_{ij} = \frac{1}{T}(y_i - X\pi_i^*)'(y_i - X\pi_j^*)$$

where π_i^* and π_j^* are (OLS) estimators.

5. We now obtain the (3SLS) induced (RRF) estimates for the complete model. Once more the complete model is given by

$$YB' + X\Gamma' = U$$

with reduced-form

$$Y = X\Pi' + V$$

where

$$\Pi' = -\Gamma'(B^{-1})'$$

$$V = U(B^{-1})'$$

We use the same notation for the covariance matrix of the complete model as that used for the submodel in the preceding section. Therefore

$$(3.5.1) \qquad \frac{1}{T} E\left\{\begin{bmatrix} U' \\ V' \end{bmatrix}[U \quad V]\right\} = \begin{bmatrix} \Sigma & G' \\ G & \Omega \end{bmatrix}$$

which has rank at most equal to M where

$$(3.5.2) \qquad G' = \frac{1}{T}E(U'V) = \frac{1}{T}E[U'U(B^{-1})'] = \Sigma(B^{-1})'$$

and

$$(3.5.3) \qquad G = \frac{1}{T}E(V'U) = \frac{1}{T}E[V'VB'] = \Omega B'$$

[since $V = U(B^{-1})'$, then $VB' = U$] .

Therefore,

$$(3.5.4) \qquad \begin{bmatrix} \Sigma & G' \\ G & \Omega \end{bmatrix} = \begin{bmatrix} \Sigma & \Sigma(B^{-1})' \\ \Omega & \Omega \end{bmatrix}$$

The covariance matrix described in (3.5.4) is singular since there exist a nontrivial matrix $\begin{bmatrix} I \\ -B' \end{bmatrix}$ of rank M such that

$$(3.5.5) \qquad \begin{bmatrix} \Sigma & G' \\ G & \Omega \end{bmatrix} \begin{bmatrix} I \\ -B' \end{bmatrix} = \begin{bmatrix} \Sigma - G'B' \\ G - \Omega B' \end{bmatrix}$$

$$= \begin{bmatrix} \Sigma & - \Sigma(B^{-1})' \\ \Omega B' & - \Omega B' \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} .$$

It should be recalled that in the previous section, we assumed that $\begin{bmatrix} \Sigma & G' \\ G & \Omega \end{bmatrix}$ is nonsingular . But since (3.5.1) is singular we cannot carry out the derivation outlined in the previous section.

However, the reduced form estimator developed in the preceding section is clearly defined for the complete model with $y_R = y_s = y$. From (3.4.15) we have .

$$(3.5.6) \qquad \hat{\Pi}_{3SLS} = [I \otimes (X'X)^{-1}X']y - [G\Sigma^{-1} \otimes (X'X)^{-1}X'](y-Z\hat{\delta}_{3SLS})$$

$$= [I \otimes (X'X)^{-1}X']y - [B^{-1} \otimes (X'X)^{-1}X'](y-Z\hat{\delta}_{3SLS})$$

since from (3.5.2) $G\Sigma^{-1} = B^{-1}$ and from (3.5.3)

$$\Omega - G\Sigma^{-1}G' = \Omega - \Omega B'\Sigma^{-1}\Sigma(B^{-1})' = 0 \quad .$$

Hence from (3.4.15) and (3.4.16) we have

$$(3.5.7) \qquad \text{Est Cov}(\hat{\Pi}) = JFJ'$$

We have observed previously that the reduced form coefficients can be estimated by

$$(3.5.8) \qquad \hat{\Pi} = -\hat{B}^{-1}\hat{\Gamma}$$

where $\hat{B}$ and $\hat{\Gamma}$. It should be noted that the Π estimated in (3.5.6) and Π estimated in (3.5.8) are not quite in the same form. Let us denote the Π appearing in (3.5.6) by Π_0. Now Π_0 estimated in (3.5.6) is that which was defined earlier on in Chapter 3, i.e., Π_0 is an MK column vector whereas Π estimated in (3.5.8) is an $M \times K$ matrix. In other words Π_0 is a vector consisting of the columns of Π stacked on top of each other.

Court (1973) used the "Stacking Operator" to demonstrate the equivalence of these two estimators.

Definition: Let $S(D) = d$ where d consists of the columns of D "stacked" on top of each other. Then if M and N are arbitrary matrices conformable for the indicated multiplication,

$$S(MDN) = (N' \otimes M)S(D) = (N' \otimes M)d$$

Hence

$$(3.5.9) \qquad \hat{\Pi}_0 = S(\hat{\Pi}) = S[-\hat{B}^{-1}\hat{\Gamma}]$$

$$= S[(X'X)^{-1}X'X(-\hat{\Gamma}'(\hat{B}^{-1})')]$$

$$= S[(X'X)^{-1}X'(Y\hat{B}' - \hat{U})(\hat{B}^{-1})']$$

$$= S[X'X)^{-1}X'y] - S[(X'X)^{-1}X'\hat{U}(\hat{B}^{-1})']$$

since from

$$YB' + X\Gamma' = U \qquad \text{we have}$$

$$YB + X\hat{\Gamma} = \hat{U}$$

$$-X\Gamma = Y\hat{B}' - \hat{U}$$

Now from the definition of the stacking operator

$$(3.5.10) \qquad \hat{\Pi}_0 = [I\otimes(X'X)^{-1}X']S(Y) - [B^{-1}\otimes(X'X)^{-1}X']S(\hat{U})$$

Now if we write the model as

$$y = Z\hat{\delta} + \hat{u}$$

in the manner of the (3SLS) procedure outlined in Chapter 2, i.e. $S(Y) = y$ and $S(\hat{U}) = \hat{u}$ then (3.5.10) becomes

$$\Pi_0 = [I \otimes (X'X)^{-1}X']y - [B^{-1} \otimes (X'X)^{-1}X']\,\hat{u}$$

$$= [I \otimes (X'X)^{-1}X']y - [B^{-1} \otimes (X'X)^{-1}X'](y - Z\hat{\delta})$$

which demonstrates the equivalence of the two estimators. Finally, it should be noted that the (3SLS) induced (RRF) provides a covariance matrix which is much easier to compute than that of the (2SLS) (RRF) estimates computed by Goldberg, Nagar and Odeh (1961).

6. We now consider an alternative estimator of the reduced-form, namely, the Partially Restricted Reduced Form (PRRF), unlike the (RRF) which incorporates all overidentifying restrictions on all structural equations, uses the overidentifying restrictions on the coefficients of one structural equation at a time.

Let us write one of the equations from the complete model of M structural equations as

$$y = Y\beta + X_1\gamma + u = Z\delta + u \tag{3.6.1}$$

where

$$Z = [Y, X_1]\,, \quad \delta = \begin{bmatrix} \beta \\ \gamma \end{bmatrix} .$$

Also y is a $T \times 1$ vector of observations on the left hand jointly dependent variable, Y and X_1 are matrices of T observations on $m(\leq M-1)$ right-hand jointly dependent variables and $K^*(\leq K)$ predetermined variables respectively, u is a T component vector of structural disturbances and $\delta = \begin{bmatrix} \beta \\ \gamma \end{bmatrix}$ is an $(m + K^*)$ vector of coefficients.

The reduced-form corresponding to y and Y in (3.6.1) is given by

$$(3.6.2) \qquad y = X\pi^* + v = X_1\pi_1^* + X_2\pi_2^* + v = [X_1 \quad X_2]\begin{bmatrix}\pi_1^* \\ \pi_2^*\end{bmatrix} + v$$

$$= X\begin{bmatrix}\pi_1^* \\ \pi_2^*\end{bmatrix} + v$$

and

$$(3.6.3) \qquad Y = X\Pi^* + \overline{V} = [X_1 \ , \ X_2]\begin{bmatrix}\Pi_1^* \\ \Pi_2^*\end{bmatrix} + \overline{V}$$

where $X = [X_1 \ , \ X_2]$ is the $T \times K$ matrix of all the predetermined variables, X_1 is a $T \times K^*$ matrix of predetermined variables included in (3.6.1) and X_2 is a $T \times K^{**}$ matrix of predetermined variables excluded from (3.6.1) and $K = K^* + K^{**}$. In (3.6.2) π^* is a K-component vector of reduced-form coefficients and in (3.6.3) Π^* is a $K \times m$ matrix of reduced-form coefficients.

Since X_1 is a submatrix of X, we can write

$$(3.6.4) \qquad X_1 = X\begin{bmatrix}I \\ 0\end{bmatrix}$$

where I is a $K^* \times K^*$ identity matrix and the null matrix, 0 is $K^{**} \times K^*$. If we substitute (3.6.3) into (3.6.1) we obtain

$$(3.6.5)\qquad y = [X\Pi^* + \bar{V}]\beta + X\begin{bmatrix} I \\ 0 \end{bmatrix}\gamma + u$$

$$= X\begin{bmatrix} \Pi^* & I \\ & 0 \end{bmatrix}\begin{bmatrix} \beta \\ \gamma \end{bmatrix} + [\ \bar{V}\ ,\ 0\ ]\begin{bmatrix} \beta \\ \gamma \end{bmatrix} + u$$

$$= X\,\varphi\,\delta + V_z\delta + u$$

where

$$(3.6.6)\qquad V_z = [\ \bar{V}\ ,\ 0\ ]\ ,$$

0 being a null matrix of order $T \times K^*$ and

$$(3.6.7)\qquad \varphi = \begin{bmatrix} \Pi^* & I \\ & 0 \end{bmatrix}$$

Therefore from (3.6.2) it follows that

$$(3.6.8)\qquad \pi^* = \varphi\,\delta \qquad \text{and} \qquad v = V_z\delta + u\ .$$

The (PRRF) of $\pi^* = \begin{bmatrix} \pi_1^* \\ T_2^* \end{bmatrix}$ is given by

$$(3.6.9)\qquad \hat{\pi}^* = \hat{\varphi}\,\hat{\delta} \qquad \text{where}$$

$$(3.6.10)\qquad \varphi = (X'X)^{-1}X'Z$$

$$= (X'X)^{-1}X'\,[Y\ ,\ X_1]$$

$$= \left[(X'X)^{-1}X'Y\ ,\ (X'X)^{-1}X'X\begin{bmatrix} I \\ 0 \end{bmatrix}\right]$$

$$= \left[(X'X)^{-1}X'Y\ ,\ \begin{matrix} I \\ \cdots \\ 0 \end{matrix}\right]$$

It should be noted that $(X'X)^{-1}X'Y$ is the (OLS) estimatore of

$\Pi^* = \begin{bmatrix} \Pi_1^* \\ \Pi_2^* \end{bmatrix}$ in (3.6.3) and $\hat{\delta} = \begin{bmatrix} \hat{\beta} \\ \hat{\gamma} \end{bmatrix}$ is the (2SLS) estimator of

$\delta = \begin{bmatrix} \beta \\ \gamma \end{bmatrix}$.

From Chapter 2, we can express the (2SLS) estimator of δ in

$y = Y\beta + X_1\gamma + u = Z\delta + u$ as

(3.6.11)
$$\hat{\delta} = \begin{bmatrix} \hat{\beta} \\ \hat{\gamma} \end{bmatrix} = \begin{bmatrix} Y'X(X'X)^{-1}X'Y & Y'X_1 \\ X_1'Y & X_1'X_1 \end{bmatrix}^{-1} \begin{bmatrix} Y'X(X'X)^{-1}X'y \\ X_1'y \end{bmatrix}$$

where

(3.6.12) $$\hat{\beta} = (Y'NY)^{-1}Y'Ny$$

(3.6.13) $$\hat{\gamma} = (X_1'X_1)^{-1}X_1'(y - Y\hat{\beta})$$, where

(3.6.14) $$N = X(X'X)^{-1}X' - X_1(X_1'X_1)^{-1}X_1'$$

is a $T \times T$ independent symmetric matrix such that rank $N = tr.N = K - K^* = K^{**}$. Then the (PRRF) estimator of π^* is denoted by

(3.6.15) $$\hat{\pi}^* = \begin{bmatrix} \pi_1^* \\ \pi_2^* \end{bmatrix} = \begin{bmatrix} (X'X)^{-1}X'Y & \begin{matrix} I \\ \cdots\cdots \\ 0 \end{matrix} \end{bmatrix}$$

and the)PRRF) estimator in (3.6.15) is the same as the (URF) in the case where the structural equation under estimation is exactly identified, i.e., $K = m + K^*$.

$$(3.6.16) \qquad \begin{bmatrix} \pi^* \\ \cdot \\ \pi^* \end{bmatrix} = (X'X)^{-1}X'y \quad .$$

Following Nagar (1959) we can write

$$(3.6.17) \qquad \bar{V} = u\, r' + W$$

which describes the (normally distributed) reduced-form disturbances as consisting of a part which is proportional to the corresponding disturbance of the equation, $y = Y\beta + X_1\gamma + u$; . (viz., ur'; r being a column vector of m components) and a part (viz., W) which has dimensions $T \times m$ and is also normally distributed but independently of the u vector. [See Appendix B of Court and Kakwani (1972) for a proof of Nagar's decomposition (3.6.17)].

It should be recalled that X is assumed to be of rank $K \leq T$ and all its elements are nonstochastic and also that $E(u) = 0$, $E(uu') = \sigma^2 I$, σ^2 being the residual variance.

Then the vector of covariances of the (RHS) variables of (3.6.1) and the disturbances can be expressed as

$$(3.6.18) \qquad q = \frac{1}{T} E \begin{bmatrix} Y'u \\ X'u \end{bmatrix} = \frac{1}{T} \begin{bmatrix} E(\bar{V}'u \\ 0 \end{bmatrix} = \sigma^2 \begin{bmatrix} r \\ 0 \end{bmatrix} .$$

Using (3.6.17), the moment matrix of $\bar{V}$ is given by

$$(3.6.19) \qquad \frac{1}{T} E\,[\bar{V}'V] = \sigma\, r\, r' + \frac{1}{T} E(W'W)$$

and bordering these matrices with K^* rows and K^* columns of zeros, we obtain three square matrices of order $m + K^*$.

$$(3.6.20) \qquad C_1 = \begin{bmatrix} \sigma^2 r\, r' & 0 \\ 0 & 0 \end{bmatrix} ; \quad C_2 = \begin{bmatrix} \frac{1}{T} E(W'W) & 0 \\ 0 & 0 \end{bmatrix}$$

$$(3.6.21) \qquad C = C_1 + C_2$$

[see Nagar (1959) for more details].

Kakwani and Court proved the following theorems:

THEOREM (3.6.A) : The bias to the order T^{-1} of the (PRRF) estimator $\hat{\pi}^*$ of π^* (T being the number of observations) is

$$(3.6.22) \qquad E(\hat{\pi}^* - \hat{\pi}) = L\, \varphi\, Rq \qquad \text{where}$$

$$L = K^{**} - m \;, \quad R = (\varphi' X' X \varphi)^{-1}$$

[φ is defined in (3.6.7) and q is as defined in (3.6.18)].

THEOREM (3.6.B): The moment to the order T^{-2} of the estimator $\hat{\pi}^*$ is given by

$$(3.6.23) \qquad E(\hat{\pi}^* - \pi^*)(\hat{\pi}^* \;\; \pi^*)' = \sigma_v^2\, (X'X)^{-1} - \lambda H$$

$$+ [\lambda \,\mathrm{tr}\, RC + 2L\delta' q\, \mathrm{tr}\, RC_1 + 2L\delta' C_2 Rq]H + \varphi R\, \Theta\, R\varphi$$

where $\qquad \lambda = \sigma^2 + 2\delta' q,$

(3.6.24) $$H = (X'X)^{-1} - \varphi R \varphi$$

$$\Theta = (\sigma^2 L^2 - 2L\lambda)C_1 - L\lambda C_2 - 2LC_2\delta' q$$

and $$\sigma_v^2 = \frac{1}{T} E\,(v'v) = \frac{1}{T} E(V_z\delta + u)'(V_z\delta + u)$$

$$= \delta' C \delta + \sigma^2 + 2\delta' q$$

i.e. σ_v^2, is the reduced form disturbance variance. [Recall that from (3.6.8) $v = V_z\delta + u$].

The first two terms on the (R.H.S.) of (6.3.23) correspond to the asymptotic covariance matrix of $\hat{\pi}^*$ and the remaining terms together make up the correction to the order T^{-2} for small samples. If (6.3.1) is exactly identified then $L = 0$ and φ is a square nonsingular matrix (thus H vanishes), so the moment matrix in (3.6.23) becomes

(3.6.25) $$E(\hat{\pi}^* - \pi^*)(\hat{\pi}^* - \pi^*)' = \sigma_v^2\,(X'X)^{-1}$$

which is the covariance matrix of the (URF) estimator of π^*.

7. If we use π^*, then the (PRRF) forecast value of y_* is given by

$$(3.7.1) \qquad \hat{y}_* = x_*' \hat{\pi}^* \qquad \text{where}$$

$$(3.7.2) \qquad y_* = x_*'\pi^* + V_{z*}\delta + u_*$$

is the true value of the left hand jointly dependent variable in the period of prediction (the asterisk denotes the time period for which the forecasts are made).

In (3.7.1) and (3.7.2) we have

$$(3.7.3) \qquad x_*' = (x_{1*} \ldots\ldots x_{K*})$$

is a vector of observations in the period of prediction $t = *$ on the K predetermined variables and

$$(3.7.4) \qquad V_{z*} = [\ \overline{V}_* \ , 0\ ] \ .$$

As before, we assume that the distribution of the disturbance term for the forecast period is the same as that for the period of observation.

The following results have been derived by Kakwani and Court (1972). The mean forecast error to order T^{-1} is given

$$(3.7.5) \qquad E(\hat{y}_* - y_*) = Lx_*' \varphi Rq$$

and the mean square forecast error to order T^{-2} is given.

$$(3.7.6) \qquad E(\hat{y}_* - y_*)^2 = \sigma_v^2[1 + x_*' (X'X)^{-1}x_*] - \lambda[1 + x_*' Hx_*]$$
$$+ [\lambda \text{ tr } RC + 2L\delta'q \text{ tr } RC_1 + 2L\delta'C_2Rq]x_*' Hx_* + x_*' \varphi R \Theta R \varphi' x_* \ .$$

When equation (3.6.1) is exactly identified, the mean forecast error vanishes to the order T^{-1} and the mean square forecast error to the order T^{-2} becomes

$$E(\hat{y}_* - y_*) = \sigma_v^2 \,(1 + x_*' (X'X)^{-1} x_*)$$

which is the variance of the (URF) error of forecast.

8. Nagar and Sahay (1978) obtained the bias and mean squared error of forecasts from the Partially Restricted Reduced Form (PRRF) in the special case when there are only two endogenous variables present in (3.6.1), i.e. for the case $m = 1$, Knight (1977) proved the existence of finite moments of the (PRRF) estimators. He formulated his existence proof firstly for the special case when two included endogenous variables are present in (3.6.1) and then extended the result for any number of included endogenous variables.

For the case $m = 1$, we rewrite (3.6.1) as

(3.8.1) $$y = y_1\beta + X_1\gamma + u \qquad \text{(we replace } Y \text{ by } y_1\text{)}$$

where y_1 is a column vector $(T \times 1)$ and β is a scalar coefficient. The corresponding reduced-form of y and y_1 may be written as

(3.8.2) $$y = X\pi^* + v = X_1\pi_1^* + X_2\pi_2^* + v$$

$$= X \begin{bmatrix} \pi_1^* \\ \pi_2^* \end{bmatrix} + v$$

$$(3.8.3) \qquad y_1 = X\pi + v_1 = X\pi_1 + X_2\pi_2 + v_1$$

$$= X \begin{bmatrix} \pi_1 \\ \pi_2 \end{bmatrix} + v_1$$

Following Nagar and Sahay (1978) and Knight (1977), we add to our list of assumptions the following:

(3.8.4) The elements of $X = [X_1, X_2]$ are non-stochastic and the columns are orthonormal, i.e.

$$X'X = I_{K\times K}$$

$$X_1'X_1 = I_{K^*\times K^*}$$

$$X_1'X_2 = 0$$

and $\quad N = X(X'X)^{-1}X' - X_1(X_1'X_1)^{-1}X_1'$

$$= X_2X_2'$$

[Recall X is $T\times K$, X_1 is $T\times K^*$, X_2 is $T\times K^{**}$, i.e., $K^{**} = K - K^*$

(3.8.5) The reduced-form dusturbances are assumed to be normally distributed with zero mean and covariance matrix equal to unity.

In view of assumption (3.8.4) we have the (PRRF) estimator $\hat{\pi}^*$ written as

(3.8.6)

$$\hat{\pi}^* = \begin{bmatrix} \hat{\pi}_1^* \\ \\ \pi_2^* \end{bmatrix} = \begin{bmatrix} (X'X)^{-1}X'Y & \begin{matrix} I \\ \cdots\cdots \\ 0 \end{matrix} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \\ \hat{\gamma} \end{bmatrix}$$

$$= \begin{bmatrix} X'Y & \begin{matrix} I \\ \cdots\cdots \\ 0 \end{matrix} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \\ \hat{\gamma} \end{bmatrix}$$

$$= \begin{bmatrix} X_1'Y & I \\ X_2'Y & 0 \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{\gamma} \end{bmatrix}$$

$$\begin{bmatrix} X'Y\,\hat{\beta} + \hat{\gamma} \\ \\ X_2'Y\,\hat{\beta} \end{bmatrix}$$

It should be recalled that in (3.6.12) and (3.6.13)

$$\hat{\beta} = [Y'N\,Y]^{-1}Y'Ny \qquad \text{and}$$

$$\hat{\gamma} = [X_1'X_1]^{-1}\ X_1'(y - Y\hat{\beta})$$

Now in the case of equation (3.8.1), where $m = 1$ and Y is replaced by y_1, we have

(3.8.7) $$\hat{\beta} = \frac{y_1'Ny}{y_1'Ny_1} \qquad \text{and}$$

(3.8.8) $$\hat{\gamma} = X_1'(Y - y_1\hat{\beta}) \qquad \text{since} \quad X_1'X_1 = I\ .$$

Therefore from (3.8.8) we have

(3.8.9) $$\hat{\pi}_1^* = X_1'y_1\hat{\beta} + \hat{\gamma}$$

and

$$\hat{\pi}_2^* = X_2'y_1\hat{\beta}$$

i.e., $$\hat{\pi}_1^* = X_1'y_1 \frac{y_1'Ny}{y_1'Ny_1} + X_1'\left(y - y_1 \frac{y_1'Ny}{y_1'Ny_1}\right)$$

$$= X_1'y .$$

Also

$$\hat{\pi}_2^* = \hat{\beta}\, X_2'y_1$$

$$= \frac{y_1'Ny}{y_1'Ny_1} X_2'y_1 = \frac{y_1'X_2X_2'y}{y_1'X_2X_2'y_1} X_2'y_1$$

Using $N = X_2X_2'$.

Define

(3.8.10) $$Z = X_2'y \;,\quad Z_1 = X_2'y_1 \;,$$ then

(3.8.11) $$\hat{\pi}^* = \left[\frac{Z_1'\,Z}{Z_1'\,Z_1}\right] Z_1 .$$

where z and z_1 are $K^{**}\times 1$ vectors. Therefore

(3.8.12) $$\hat{\pi} = \begin{bmatrix} \hat{\pi}^* \\ \hat{\pi}^* \end{bmatrix} = \begin{bmatrix} X_1'y \\ \left(\frac{Z_1'\,Z}{Z_1'\,Z_1}\right) Z_1 \end{bmatrix}$$

From assumptions (3.8.5) we have

(3.8.13) $$\bar{y} = E(y) = X_1\pi_1^* + X_2\pi_2^*$$

(3.8.14) $$\bar{y}_1 = E(y_1) = X_1\pi_1 + X_2\pi_2 \quad .$$

Therefore, we have

(3.8.15) $$y \sim N(\bar{y}\ ,\ I_T) \ ;\ y_1 \sim N(\bar{y}_1\ ,\ I_T)$$

and y and y_1 are mutually independent since

(3.8.16) $$E[(y - \bar{y})\ (y_1 - \bar{y}_1)'] = 0 \quad .$$

It should be noted that the elements of z and z_1 are independently normally distributed with

(3.8.17) $$\bar{z} = E(z) = X_2' E(y) = X_2' \bar{y} = X_2'(X_1\pi_1^* + X_2\pi_2^*) = \pi_2^* \qquad \text{and}$$

(3.8.18) $$\bar{z}_1 = E(z_1) = X_2' E(y_1) = X_2' \bar{y}_1 = X_2'(X_1\pi_1 + X_2\pi_2) = \pi_2$$

(Note that $X_2'X_1 = 0$).

Also

(3:8.19) $$E(z - \bar{z})(z - \bar{z})' = E(z_1 - \bar{z}_1)(z_1 - \bar{z}_1)' = I$$

and

(3.8.20) $$E(z - \bar{z})(z_1 - \bar{z}_1)' = 0$$

where I is the $K^{**} \times K^{**}$ unit matrix and 0 is the $K^{**} \times K^{**}$ null matrix.

From (3.8.9)

$$(3.8.21) \qquad E(\hat{\pi}_1^*) = E(X_1'y) = X_1' E(y) = X_1'(X_1\pi_1^* + X_2\pi_2^*) = \pi_1^*$$

since $X_1'X_1 = I$ and $X_1'X_2 = 0$.

Let z_{1j} and x_i denote the i-th element of z_1 and z respectively, then the k-th element of the vector

$$\hat{\pi}_2^* = \left(\frac{Z_1' \ Z_1}{Z_1' \ Z_1}\right) Z_1 \qquad \text{is given by}$$

$$(3.8.22) \qquad \hat{\pi}^*_{2(k)} = \frac{\sum_{i=1}^{K_2} (Z_{1i} Z_i)}{\sum_{i=1}^{K_2} Z^2_{1i}} \; Z_{1(k)}$$

Now the r-th moment of $\hat{\pi}^*_{2(k)}$ is given by

$$(3.8.23) \qquad E\left[\hat{\pi}^*_{2(k)}\right] = E\,\frac{\left(\sum_{i=1}^{K_2} Z_{1i}Z_i\right)^r}{\left(\sum_{i=1}^{K_2} Z^2_{1i}\right)^r}\, Z^r_{1(k)}$$

(Note that we have simplified the notation by writing K_2 for K^{**})
[See Knight (1977) for an existence proof of (3.8.23)].

Following Nagar and Sahay (1978) we have from (3.8.22)

$$(3.8.24) \qquad E\left(\hat{\pi}^*_{2(k)}\right) = E\left(Z_{1(k)}\middle|\sum_{i=1}^{K_2} Z_{1i}^2\right) E\, Z_{(k)}$$

$$+ \sum_{i \neq k=1}^{K_2} E\left(Z_{1i}Z_{1(k)}\middle|\sum_{i=1}^{K_2} Z_{1i}^2\right) EZ_i$$

Now since the elements of Z_1 are independently normal i.e. $Z_{1i}, \ldots Z_{1K_2}$ are independently normal and $E(z_{1i}) = \bar{z}_{1i} = \bar{z}_{1i}$ and $Vz_{1i} = 1$, $i = 1, \ldots K_2$ then

$$(3.8.25) \qquad W = \sum_{i=1}^{K_2} Z_{1i}^2$$

is distributed according to non-central chi-square distribution with K_2 degrees of freedom and non-centrality parameter,

$$(3.8.26) \qquad \Theta = \frac{1}{2} \sum_{i=1}^{K_2} \bar{Z}_{1i}^2 .$$

Nagar and Ullah (1973) worked out the following expectations:

$$(3.8.27) \qquad E(Z_{1i}^2\, W^{-1}) = \frac{1}{2}\, e^{-\Theta}[z_{1i}f_{12} + f_{01}$$

$$(3.8.28) \qquad E(Z_{1i}Z_{1k}W^{-1}) = \frac{1}{2}\, \bar{z}_{1i}\bar{z}_{ik}\, e^{-\Theta}\, f_{12},\ i \neq k$$

where W and Θ are given in (3.8.25) and (3.8.26) respectively. Nagar and Sahay (1978) used the results outlined in (3.8.27) and (3.8.28). [See Appendix of Nagar and Sahay (1978) for details] to show that

$$(3.8.29) \qquad E(\hat{\pi}^*_2) = e^{-\Theta}(\Theta f_{12} + \frac{1}{2}\, f_{01})\pi^*_2$$

where

(3.8.30) $$\Theta = \frac{1}{2}\,\overline{y}_1 N\,\overline{y}_1 = \frac{1}{2}\,\overline{y}_1\; X_2X_2'\,\overline{y}$$

[since $N = X_2X_2'$ and $\overline{y}_1 = E(y_1)$] is a certain non-centrality parameter.

From (3.8.18) $E(Z_1) = \overline{Z}_1 = X_2'\,\overline{y}_1$

where $\overline{Z}_1$ is a $(K^{**} \times 1)$ or $(K_2 \times 1)$ (we use K_2 for K^{**}) vector.

Therefore
$$\overline{Z}_1'\,\overline{Z}_1 = (X_2'\,\overline{y}_1)^1\,(X_2'\,\overline{y}_1)$$
$$= \overline{y}_1'\,X_2X_2'\,\overline{y}_1$$
$$= \overline{y}_1'\,N\,\overline{y}_1$$

Also since $\overline{Z}_1$ has elements $\overline{Z}_{11}\,\overline{Z}_{12}\ldots\overline{Z}_{1K_2}$, then

$$\overline{Z}_1'\,\overline{Z}_1 = \begin{bmatrix}\overline{Z}_{11} & \overline{Z}_{12}\ldots Z_{1K_2}\end{bmatrix}\begin{bmatrix}\overline{Z}_{11}\\ \overline{Z}_{12}\\ \vdots\\ \overline{Z}_{1K_2}\end{bmatrix}$$

$$= \sum_{i=1}^{K_2} \overline{Z}_{1i}^2 .$$

Therefore, from (3.8.30) $\Theta = \frac{1}{2}\,\overline{y}_1'\,N\,\overline{y}_1$

$$= \frac{1}{2}\sum_{i=1}^{K_2} \overline{Z}_{1i}^2 .$$

In (3.8.27) (3.8.28) and (3.8.29), f_{12} and f_{01} have been obtained from

$$(3.8.31) \qquad f_{ac} = \frac{\Gamma\left(\frac{K_2}{2} + a\right)}{\Gamma\left(\frac{K_2}{2} + c\right)} \, {}_1F_1\left(\frac{K_2}{2} + a \, , \frac{K_2}{2} + c \, ; \theta\right)$$

by setting $a = 1$, $c = 2$, and $a = 0$, $c = 1$ respectively and ${}_1F_1(\cdot)$ is a confluent hypergeometric function.

DEFINITION:*

$$(3.8.32) \qquad {}_1F_1(p, q; x) = \sum_{j=0}^{\infty} \frac{\frac{\Gamma(p+j)}{\Gamma(p)}}{\frac{\Gamma(q+j)}{\Gamma(q)}} \frac{x^j}{j!}$$

$$= \frac{\Gamma(q)}{\Gamma(p)} \sum_{j=0}^{\infty} \frac{\Gamma(p+j)}{\Gamma(q+j)} \frac{x^j}{j!}$$

*[See Chapter 4 of Y.L. Luke, Vol. I. (1969) or L.J. Slater (1960) for details on confluent hypergeometric functions].

9. In this section we look at the (PRRF) forecast errors for the special case $m = 1$, i.e., for (3.8.1);

$$y = y_1\beta + X_1\gamma + u$$

where β is a scalar and the reduced form corresponding to y and y_1 given by (3.8.2) and (3.8.3) respectively. As in (3.7.3) let

$$x_*' = [x_{1*}, x_{2*} \ldots x_{K*}]$$

denote a vector of observations for the forecast period $t = *$ on the K predetermined variables. Then the forecast value of the left-hand endogenous variable in (3.8.1) is denoted by

$$\hat{y}_* = x_*' \begin{bmatrix} \hat{\pi}_1^* \\ \hat{\pi}_2^* \end{bmatrix} \qquad \text{where} \tag{3.9.1}$$

$$y_* = x_*' \begin{bmatrix} \hat{\pi}_1^* \\ \hat{\pi}_2^* \end{bmatrix} + v_* \tag{3.9.2}$$

is the true value of the lefthand endogenous (jointly dependent) variable in the period of forecast and its estimated counterpart is given by

$$y_* = x_*' \begin{bmatrix} \hat{\pi}_1^* \\ \hat{\pi}_2^* \end{bmatrix} + \hat{v}_* \tag{3.9.3}$$

From (3.9.1) and (3.9.3) we have

$$(3.9.4) \qquad \hat{y}_* - y_* = - \hat{v}_*$$

$$= x_*' \begin{bmatrix} \hat{\pi}_1^* - \pi_1^* \\ \hat{\pi}_2^* - \pi_2^* \end{bmatrix} - v_*$$

Therefore

$$(3.9.5) \qquad \hat{v}_* = v_* - x_*' \begin{bmatrix} \hat{\pi}_1^* & -\pi_1^* \\ \hat{\pi}_2^* & \pi_2^* \end{bmatrix}$$

Partition

$$x_* = \begin{bmatrix} x_{*(1)} \\ x_{*(2)} \end{bmatrix}$$

so that $x_{*(1)}$ is the column vector of observations on variables in X_1 in the forecast period specified and $x_{*(2)}$ is the column vector of observations on variables in X_2 for the specified forecast period.

Then the following results have been obtained by Nagar and Sahay (1978):

$$(3.9.6) \qquad E(\hat{y}_* - y_*) = - E(\hat{v}_*) \qquad \text{from (3.9.4)}$$

$$= E \left\{ x_*' \begin{bmatrix} \hat{\pi}_1^* & \pi_1^* \\ \hat{\pi}_2^* & \pi_2^* \end{bmatrix} \right\} \quad \text{since} \quad E(v_*) = 0$$

[i.e. we assume that the distribution of the disturbance term for the forecast period is the same as that for the period of observation].

Therefore

$$E(\hat{y}_* - y_*) = E\left\{\begin{bmatrix} x'_{*(1)} & x'_{*(2)} \end{bmatrix} \begin{bmatrix} \hat{\pi}_1^* - \pi_1^* \\ \hat{\pi}_2^* - \pi_2^* \end{bmatrix}\right.$$

$$= E\left\{x'_{*(1)} \hat{\pi}_1^* - x_{*(1)}\pi_1^* + x_{*(2)}\hat{\pi}_2^* - x_{*(2)}\pi_2^*\right]$$

$$= x_{*(2)} E(\hat{\pi}_2^* - \pi_2^*)$$

since $E(\hat{\pi}_1^*) = \pi_1^*$ from (3.8.21)

or

$$E(\hat{y}_* - y_*) = e^{-\Theta} (\Theta f_{12} + \frac{1}{2} f_{01} - 1) x'_{*(2)} \pi_2^*$$

where Θ and f_{ac} have been defined previously (3.9.7) for θ large and positive an asymptotic approximation to the bias of the (PRRF) forecast is given by

$$E(\hat{y}_* - y_*) \approx \frac{1-K_2}{2} \frac{1}{\Theta} x'_{*(2)} \pi_2^* , K_2 > 1$$

where terms of lower order than $\frac{1}{\Theta}$ have been omitted.

Also the exact mean squared error of the (PRRF) forecast of the lefthand jointly dependent variable in (3.8.1) is given by

(3.9.8) $$E(\hat{y}_* - y_*)^2 = 1 + x'_{*(1)}x_{*(1)} + x'_{*(2)}Dx_{*(2)}$$

where $x_{*(1)}$ and $x_{*(2)}$ are as defined in (3.9.5) and

$$D = E(\hat{\pi}_2^* - \pi_2^*)(\hat{\pi}_2^* - \pi_2^*)'$$

$$= \Big[(\tfrac{1}{2}\Theta + \beta^2 \Theta^2)\, I + (1 + \tfrac{K_2}{4} + \tfrac{1}{2}\Theta\beta^2)\pi_2 \pi_2'$$

$$+ \beta\,\Theta\,\pi_2\,\pi_2^{*} + \beta\,\Theta\,\pi_2^*\,\pi_2' \Big]$$

$$+ (\tfrac{1}{2}\Theta + \beta^2\Theta^2)\,\pi_2\,\pi_2'\,e^{-\Theta} f_{24}$$

$$+ \Big[(\tfrac{1}{2} + \tfrac{K_2}{4} + \tfrac{1}{2}\Theta\beta^2)\, I + \tfrac{1}{2}\pi_2^*\pi_2' \Big] e^{-\Theta} f_{02}$$

$$- 2e^{-\Theta}\,\Theta\, f_{12}\,\pi_2^*\,\pi_2'^* - e^{-\Theta} f_{01}\,\pi_2^*\,\pi_2'^* + \pi_2^*\,\pi_2'^*$$

[From (3.8.17), $E(z) = \bar{z} = \pi_2^*$ and
from (3.8.18 $E(z_1) = \bar{z}_1 = \pi_2$]

and Θ and f_{ac} are as defined previously.

For large values of Θ and $K_2 > 2$, the asymptotic approximation to the (M.S.E.) of forecast is given by

$$(3.9.9) \qquad E(\hat{y}_* - y_*)^2 \approx 1 + x_{*(1)}'\,x_{*(1)} + \beta^2\,x_{*(2)}'\,x_{*(2)}$$

$$+ \Big[\Big(\frac{1+\beta^2}{2} - K_2\,\beta^2\Big)\, x_{*(2)}'\,x_{*(2)}$$

$$+ \frac{1-\beta^2}{2}\, x_{*(2)}'\,\pi_2'\,\pi_{*(2)} \Big] \frac{1}{\Theta}$$

where terms of lower order thatn $\frac{1}{\Theta}$ have been neglected.

CHAPTER 4

DISTURBANCE-VARIANCE ESTIMATION

1. Once more we consider the system of M structural equations in M jointly dependent variables and K predetermined variables. Suppose we denote the equation (3.6.1) as

$$y = Y\beta + X_1\gamma + u \tag{4.1.1}$$

and write the reduced-form corresponding to the explanatory jointly dependent variable as

$$Y = X\Pi + V \quad . \tag{4.1.2}$$

The (OLS) estimate of V in (4.1.2) is given by

$$\hat{V} = Y - X(X'X)^{-1}X'Y \tag{4.1.3}$$

and using this, the k-class estimator proposed by Theil is given by

$$\begin{bmatrix} Y'Y - k\hat{V}'\hat{V} & Y'X_1 \\ X_1'Y & X_1'X_1 \end{bmatrix} \begin{bmatrix} \hat{\beta}(k) \\ \hat{\gamma}(k) \end{bmatrix} = \begin{bmatrix} Y' - k\hat{V}' \\ X_1' \end{bmatrix} y \tag{4.1.4}$$

where k is an arbitrary scalar.

The structural disturbances of the entire system are assumed to be normally distributed with mean zero and independent over time. Thus, we write

$$V = ur' + W. \tag{4.1.5}$$

as in (3.6.17). We estimate the disturbance vector u in (4.1.1) by

$$u_e = y - [Y \; , \; X_1] \begin{bmatrix} \hat{\beta}(k) \\ \hat{\gamma}(k) \end{bmatrix} \quad . \tag{4.1.6}$$

In addition we assume that k is nonstochastic and differ from 1 to the order of $O(T^{-1})$. This assumption implies

$$k = 1 + \frac{h}{T} \quad , \tag{4.1.7}$$

h being nonstochastic and independent of T [see Nagar (1959)].

Nagar (1961) obtained the bias to the order $O(T^{-1})$ in probability of the disturbance-variance estimator,

$$\hat{\sigma}_k^2 = \frac{1}{T} u_e' u_e \quad . \tag{4.1.8}$$

The bias of the estimator (4.1.8) to the order $O(T^{-1})$ in probability is given by

$$B(\hat{\sigma}_k^2) = \sigma^2 \left[(2h - 2L + 3)\mathrm{tr}QC_1 + \mathrm{tr}QC_2 - \frac{1}{T}(m + K_2) \right] \tag{4.1.9}$$

where $K_2 = K^{**}$, $L = K_2 - m$, C_1 and C_2 are as defined in (3.6.20) and (3.6.21) respectively and

$$Q = \begin{bmatrix} \Pi' X' X \Pi & \Pi' \bar{X}' X_1 \\ X_1' X \Pi & X_1' X_1 \end{bmatrix}^{-1} \tag{4.1.10}$$

Srivastava (1971) claimed that bias alone is not enough to judge the appropriateness of the estimator (4.1.8), unless it is analysed in conjunction with the mean squared error (MSE). Thus, he proceeded to compute the (MSE) of the estimator (4.1.8) to the same order of approximation, i.e., $0(T^{-1})$. He denoted the (MSE) by

$$M(\hat{\sigma}_k^2) = 2\sigma^4 \left[2 \text{ tr } QC_1 + \frac{1}{T}\right] \tag{4.1.11}$$

Dhrymes (1969) proposed a modification of the estimator (4.1.8) and denoted the modified estimator by

$$\bar{\sigma}_k^2 = \frac{1}{L} U_e' [X(X'X)^{-1}X'] U_e \quad . \tag{4.1.12}$$

Srivastava (1971) computed the bias and (MSE) of the estimator $\bar{\sigma}_k^2$ in (4.1.12) to the order $0(T^{-1})$ in probability and denoted them by

$$B(\bar{\sigma}_k^2) = \frac{\sigma^2}{L} [h^2 - L^2 + L] \text{ tr } QC_1 \tag{4.1.13}$$

and

$$M(\bar{\sigma}_k^2) = \frac{2\sigma^4}{L} [1 - 2(L-1)\text{tr } QC_1] \quad . \tag{4.1.14}$$

In his comparison of the two estimators $\hat{\sigma}_k^2$ and $\bar{\sigma}_k^2$, Srivastava (1971) observed that the estimator $\bar{\sigma}_k^2$ yields a smaller (MSE) than the estimator $\hat{\sigma}_k^2$ if

$$\text{tr } QC_1 > \frac{1}{2[2L-1]} (1 - \frac{1}{T}) \tag{4.1.15}$$

and that the (MSE) of $\bar{\sigma}_k^2$ decreases as L increases. This implies that the larger the number of predetermined (exogenous variables excluded (from the equation under estimation) over the number of explanatory jointly dependent variables, the smaller is the (MSE) of the disturbance-variance estimator $\bar{\sigma}_k^2$.

2. Later, Srivastava and Tiwari (1976) developed a similar estimator for the covariance of the disturbances and computed the bias and Mean Squared Error (MSE) with reference to (2SLS) and (3SLS) methods of estimation.

We denote the complete system of M structural equations by

$$YB' + X\Gamma' = U \tag{4.2.1}$$

with reduced-form given

$$Y = X\Pi' + V \tag{4.2.2}$$

U is a $(T \times M)$ matrix of structural disturbances assumed to be independent over time and normally distributed with zero and dispersion matrix.

$$\frac{1}{T} E(U'U) = \Sigma = [\sigma_1, \sigma_2, \ldots, \sigma_M] = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1M} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2M} \\ \vdots & \vdots & & \vdots \\ \sigma_{M1} & & & \sigma_{MM} \end{bmatrix} \tag{4.2.3}$$

As was done previously, we denote the i-th structural equation by

$$(4.2.4)\qquad y_i = Y_i\beta_i + X_i\gamma_i + u_i$$

$$= Z_i\delta_i + u_i \qquad (i = 1,2,\ldots,M)\ .$$

Then the (2SLS) estimator of δ_i can be expressed as

$$(4.2.5)\qquad \hat{\delta}_{i(2SLS)} = (Z_i'\, P\, Z_i)^{-1} Z_i P y_i$$

where $\quad P = X(X'X)^{-1}X'\ .$

The disturbance vector u_i is then estimated by

$$(4.2.6)\qquad \hat{u}_{i(2SLS)} = y_i - Z_i\hat{\delta}_{i(2SLS)}\ .$$

A consistent estimator of σ_{ij} based on the (2SLS) method is given by s_{ij} where

$$(4.2.7)\qquad s_{ij} = \frac{(y_i - Z_i\hat{\delta}_i)'\,(y_j - Z_j\hat{\delta}_j)}{T}$$

$$= \frac{\hat{U}_{i(2SLS)}'\ U_{j(2SLS)}}{T} \qquad (i,j = 1,2,\ldots,M)$$

Let $\bar{\rho}_{ij}$ be the estimator of ρ_{ij} the coefficient of correlation between the disturbances of the i-th and j-th equation where

$$(4.2.8)\qquad \bar{\rho}_{ij} = \frac{s_{ij}}{(s_{ii}\,s_{jj})^{1/2}}$$

In a similar manner let

$$(4.2.9)\qquad \hat{U}_{3SLS} = y - Z\,\hat{\delta}_{(3SLS)}$$

where

(4.2.10) $$y = Z\delta + u$$

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix}; \quad Z = \begin{bmatrix} Z_1 & 0 & \cdots & 0 \\ 0 & Z_2 & \cdots & \\ \vdots & \vdots & & \\ 0 & 0 & & Z_M \end{bmatrix}, \quad \delta = \begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_M \end{bmatrix}$$

$$U = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_M \end{bmatrix}$$

and the (3SLS) estimator of δ is given by

(4.2.11) $$\hat{\delta}_{3SLS} = \left[Z'(S^{-1} \otimes I)(I \otimes P)Z \right]^{-1} Z'(S^{-1} \otimes I)(I \otimes P)y$$

where

(4.2.12) $$S = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1M} \\ s_{21} & s_{22} & \cdots & s_{2M} \\ \vdots & \vdots & & \vdots \\ s_{M1} & s_{M2} & \cdots & s_{MM} \end{bmatrix}.$$

and $\otimes$ denotes the Kronecker Product.

A consistent estimator of σ_{ij} based on the 3SLS method is given by

(4.2.13) $$\hat{\sigma}_{ij} = \frac{\hat{u}_{i(3SLS)}' \hat{u}_{j(3SLS)}}{T}$$

where $\hat{U}_{i(3SLS)}$ and $\hat{U}_{j(3SLS)}$ are the i-th and j-th subvectors of $\hat{U}_{3SLS}$ respectively. Denote the (3SLS) estimator of ρ_{ij} by

$$(4.2.14) \qquad \hat{\rho}_{ij} = \frac{\hat{\sigma}_{ij}}{(\hat{\sigma}_{ii}\,\hat{\sigma}_{jj})^{1/2}} .$$

We can write the reduced-form corresponding to the (RHS) (explanatory) jointly dependent variables in equation (4.2.4) by

$$(4.2.15) \qquad Y_i = X\Pi_i + V_i$$

where Π_i and V_i are $K \times m_i$ and $T \times m_i$ respectively; Π_i and V_i are submatrices of Π and V respectively.

Using the relation $V = U(B^{-1})'$ we can write

$$(4.2.16) \qquad [V_i \quad 0] = UG_i$$

where V_i and 0 are $T \times m_i$ and $T \times K_i^*$ respectively and G_i is $M \times n_i$, $n_i = (m_i + K_i^*)$, a matrix of constants. Then we can write

$$(4.2.17) \qquad Z_i = [Y_i \,, X_i] = [X\Pi_i \quad X_i] + UG_i = A_i + UG_i,$$

where

$$(4.2.18) \qquad A_i = [X\Pi_i \,, X_i]$$

Now Let

(4.2.19)

$$A = \begin{bmatrix} A_1 & 0 & \dots & 0 \\ 0 & A_2 & \dots & 0 \\ \vdots & & & \vdots \\ 0 & & & A_M \end{bmatrix}$$

and

(4.2.20)

$$G = \begin{bmatrix} G_1 & 0 & \dots & 0 \\ 0 & G_2 & & 0 \\ \vdots & & & \vdots \\ 0 & & & G_M \end{bmatrix}$$

then

(4.2.21) $$Z = A + (I \otimes U)G$$

If we define

(4.2.22)

$$\phi = [A'(\Sigma^{-1} \otimes IA]^{-1}$$

$$\Theta = \phi A'(\Sigma^{-1} \otimes I)$$

$$\Delta = (\Sigma^{-1} \otimes I) - (\Sigma^{-1} \otimes I)A\phi A'(\Sigma^{-1} \otimes I)$$

and introduce

(4.2.23) $$Q_i = (A_i' A_i)^{-1}$$

$$\sigma_i = \begin{bmatrix} \sigma_{1i} \\ \sigma_{2i} \\ \vdots \\ \sigma_{Mi} \end{bmatrix}$$ the i-th column of Σ -

and

$$d_{ij} = G_j' \sigma_{ij} \qquad (i,j = 1,2,\ldots,M)$$

Then the following results have been derived by Srivastava and Tiwari (1976).

THEOREM 4.2.A. The bias and mean squared error, to the order $O(T^{-1})$ in probability of the (2SLS) estimator s_{ij} defined in (4.2.7) are given by

$$(4.2.24) \qquad E(s_{ij} - \sigma_{ij}) = \frac{1}{T}\Big[\mathrm{tr}\ Q_i A_i' A_j Q_j A_j' A_i) - (n_i + n_j)\Big]\sigma_{ij}$$
$$- (K - n_i - 1)d_{ii}' Q_i d_{ij} - (K-n_j-1)d_{jj}' Q_j d_{ji}$$
$$+ (\mathrm{tr}\ G_j \Sigma G_j Q_j A_j' A_i Q_i)\sigma_{ij}$$

$$(4.2.25) \qquad E(s_{ij} - \sigma_{ij})^2 = \frac{1}{T}(\sigma_{ii}\sigma_{jj} + \sigma_{ij}^2) + \sigma_{ii} d_{ij}' Q_i d_{ij}$$
$$+ \sigma_{jj} d_{ji}' Q_j d_{ji}$$
$$+ 2\sigma_{ij} d_{ji}' Q_j A_j' A_i Q_i d_{ij}$$

THEOREM 4.2.B. The bias and mean squared error, to the order $O(T^{-1})$ in probability of the (3SLS) estimator $\hat{\sigma}_{ij}$ defined by (4.2.13) are given by

$$(4.2.26) \qquad E(\hat{\sigma}_{ij}-\sigma_{ij}) = \frac{1}{T}(\mathrm{tr}\ A_j' A_i \phi_{ij}) - d_{ji}' Q_j d_{jj} - d_{ij} Q_i d_{ii}$$
$$- \sum_{m,n}^{M} (d_{ji}' H_{jmn} + \tilde{d}_{ij}' H_{imn}) d_{mn} + (\mathrm{tr}\ G_i' \Sigma_i \phi_{ji})$$

$$(4.2.27) \qquad E(\hat{\sigma}_{ij}-\sigma_{ij})^2 = \frac{1}{T}(\sigma_{ii}\sigma_{jj} + \sigma_{ij}^2) + d_{ij}'\phi_{ii}d_{ij} + d_{ji}\phi_{jj}d_{ji} + 2d_{ji}'\phi_{ji}d_{ij}$$

where

$$(4.2.28) \qquad H_{\ell mn} = (\mathrm{tr}\,\Delta_{mn}P)\phi_{\ell m} - 2\Theta_{\ell m}\Theta_{mn}' + \Theta_{\ell n}(A_m Q_m - \Theta_{mm}') \qquad (\ell = i,j)$$

and $\phi_{\ell m}$ is the (ℓ m)-th submatrix, of order $n_\ell \times n_m$ of ϕ being partitioned as

$$(4.2.29) \qquad \phi = \begin{bmatrix} \phi_{11} \cdots\cdots \phi_{1M} \\ \phi_{M1} \cdots\cdots \phi_{MM} \end{bmatrix}$$

The matrices Δ_{mn} and $\Theta_{\ell m}$ are similarly defined with references to Δ and ϕ respectively.

The change in the mean squared error is given by

$$(4.2.30) \qquad E(s_{ij}-\sigma_{ij})^2 - E(\hat{\sigma}_{ij}-\sigma_{ij})^2 = d_{ij}'(\sigma_{ii}Q_i - \phi_{ii})d_{ij} + d_{ji}'(\sigma_{jj}Q_j - \phi_{jj})d_{ji} + 2d_{ji}'(\sigma_{ij}Q_jA_j'A_iQ_i - \phi_{ji})d_{ij}$$

[using (4.2.25) and (4.2.27)].

Now we consider the estimation of variance σ_{ii}. From Theorem 4.2.A and Theorem 4.2.B we have to order $O(T^{-1})$ in probability.

$$(4.2.31) \qquad E(s_{ii} - \sigma_{ii}) = -\frac{n_i}{T}\sigma_{ii} - 2(K - n_i - 1)d_{ii}'Q_i d_{ii} + (\text{tr } G_i'\Sigma G_i Q_i)\sigma_{ii}$$

$$(4.2.32) \qquad E(s_{ij} - \sigma_{ii})^2 = \frac{2}{T}\sigma_{ii}^2 + 4\sigma_{ii}d_{ii}'Q_i d_{ii}$$

$$(4.2.33) \qquad E(\hat{\sigma}_{ii} - \sigma_{ii}) = -\frac{1}{T}(\text{tr } Q_i^{-1}\phi_{ii}) - 2d_{ii}'Q_i d_{ii} - 2\sum_{m,n}^{M} d_{ii}'H_{imn}d_{mn} + (\text{tr } G_i'\Sigma G_i\phi_{ii})$$

$$(4.2.34) \qquad E(\hat{\sigma}_{ii} - \sigma_{ii})^2 = \frac{2}{T}\sigma_{ii}^2 + 4d_{ii}'\phi_{ii}d_{ii} \ .$$

The gain in efficiency is given by

$$(4.2.35) \qquad E(s_{ii} - \sigma_{ii})^2 - E(\hat{\sigma}_{ii} - \sigma_{ii})^2 = 4d_{ii}'(\sigma_{ii}Q_i - \phi_{ii})d_{ii}$$

which is non-negative since the matrix $\sigma_{ii}Q_i - \phi_{ii}$ is non-negative definite. [See Appendix of Srivastava and Tiwari for proof.]

Srivastava and Tiwari (1976) also computed the bias and mean squared error, both to order $O(T^{-1})$ in probability of the (2SLS) estimator $\bar{\rho}_{ij}$ and the (3SLS) estimator $\hat{\rho}_{ij}$ defined in (4.2.8) and (4.2.14) respectively.

3. In this section we utilize the small-disturbance asymptotic approach introduced by Kadane (1970, 1971) to investigate the disturbance-variance estimators. Brown, Ramage and Srivastava (1972) and later Kadane and Ramage (1974) summarized Kadane's approach as follows:

"The random variables corresponding to a particular statistical procedure (here, disturbance-variance estimating technique) are approximated by other random variables close to the original variables in probability. The approximating random variables are derived from a Taylor-series expansion in powers of a (small) scalar multiple σ of the system of structural disturbances. The properties of the approximating random variables, such as their distribution, mean, and second-moment matrix, are examined in order to investigate properties of the original procedure. No claim is made that the moments of the approximating random variables are close to the moments of the approximated random variables. The claim is rather that when σ (where σ is a scalar multiple of the system of structural disturbances) is small, the first two asymptotic moments will give useful measures of location and dispersion of the approximated distribution."

We can write the complete structural form of the model as

$$YB' + X\Gamma' = \sigma U \tag{4.3.1}$$

where σ is a scalar assumed to approach zero; the rows of U are assumed to be mutually independent with zero means and covariance matrix Σ. The reduced-form of the models is

$$(4.3.2) \qquad Y = -X\Gamma'(B^{-1})' + \sigma U(B^{-1})'$$
$$= X\Pi' + \sigma V$$

The covariance matrix Ω of the reduced-form disturbances is given by

$$(4.3.3) \qquad \Omega = (B^{-1}) \Sigma (B^{-1})' \quad .$$

We consider the first structural equation of the system (4.3.1) given by

$$(4.3.4) \qquad y = Y_1\beta + X_1\gamma + \sigma u$$
$$= [Y_1 \ , X_1]\delta + \sigma u$$

where $\delta = \begin{bmatrix} \beta \\ \gamma \end{bmatrix}$ is $(m_1 \times K^*) \times 1$;

$[y \ , Y_1]$ is $T \times (1 \times m_1)$ are the included endogenous variables;and X_1 is the $(T \times K_1)$ are the included exogenous variables; u is a $T \times 1$ vector of disturbances with finite variance ψ . With no loss of generality, the equations (4.3.1) are taken to be so arranged that Y and X can be partitioned according to the variables included and excluded from (4.3.4) i.e.

$$(4.3.5) \qquad Y = [y, y_1, Y_2] \quad \text{is} \quad T \times (1 + m_1 + m_2)$$
$$X = [X_1, X_2] \quad \text{is} \quad T \times (K^* + K^{**}) \quad .$$

The variance parameter to be estimated in equation (4.3.4) is $\sigma^2 \psi$, where ψ is the leading diagonal element of Σ . The general

This thesis contains no page 153

double-k-class estimators of the parameters δ in (4.3.4) is given by

(4.3.6)

$$\delta_{k_1k_2} = \left\{\left[Y_1 , X_1\right]^1 (I - k_1\bar{P}_x)\left[Y_1 , X_1\right]\right\}^{-1} \left[Y_1 , X_1\right]^1 (I - k_2\bar{P}_x)y$$

where $\bar{P}_x = [I - P_x] = I - X(X'X)^{-1}X^1$ is the matrix of the projection orthogonal to the subspace $M(X)$ spanned by the columns of X. If in (4.3.6) $k_1 = k_2 = k$ we have the k-class estimator $\hat{\delta}_k$ of δ and similarly setting $k_1 = 1 - h^2$ and $k_2 = 1 - h$ leads to the h-class estimator of δ. The most frequently encountered estimators are members of the k-class: $k = 0$ gives the (OLS) estimator, $k = 1$ gives the (2SLS) and a random value of k given by

(4.3.7)
$$\ell_1 = \min_{\beta_*}\left\{\frac{\| \bar{P}_x[y,Y_1]\beta_* \|^2}{\| \bar{P}_x[y,Y_1]\beta_* \|^2}\right\}$$

$$= \frac{\| \bar{P}_x[y,Y_1]\hat{\beta}_* \|^2}{\| \bar{P}_x[y,Y_1]\hat{\beta}_* \|^2}$$

leads to the (LIML) estimator of δ (where $\|\cdot\|$ is the Euclidean Norm).

It is well known that all the proposed estimators of $\sigma^2\psi$ are based on the residuals from k-class estimation. Brown, Ramage and Srivastava (1972) suggested that previously proposed estimators $\sigma^2\psi$ can be treated as special cases of the following:

(4.3.8) (i) $\frac{1}{\alpha} \left\| y - [Y_1 , X_1]\hat{\delta}_* \right\|^2$

which is simply the sum of the squared residuals, i.e.

$$u_k = [y - (Y_1 \;, X_1)\hat{\delta}_k] \;,$$

normalized by a non-random scalar α. Srivastava (1972) first studied this estimator in its general form, using small disturbance asymptotic methods.

$$(4.3.8) \quad (ii) \qquad \frac{1}{\alpha} \left\| \overline{P}_X (y - [Y_1 \;, X_1]\hat{\delta}_k \right\|^2$$

which is motivated as follows:

From (4.3.3) $\Sigma = B\,\Omega\,B'$, and from (4.3.4), the first column of β is written as

$$\begin{bmatrix} -1 \\ \beta \\ 0 \end{bmatrix}$$

where β is $m_1 \times 1$; 0 is $m_2 \times 1$. Since ψ is leading diagonal element of Σ,

$$\sigma^2 \psi = \begin{bmatrix} -1 \\ \beta \\ 0 \end{bmatrix}' \sigma^2 \Omega \begin{bmatrix} -1 \\ \beta \\ 0 \end{bmatrix}$$

$$= \sigma^2 \omega_{11} - 2\sigma^2 \omega_1' \beta + \beta' \sigma^2 \Omega_{11} \beta$$

where

$$(4.3.9) \qquad \Omega = \begin{bmatrix} \omega_{11} & \omega_1 & * \\ \omega_1' & \Omega_{11} & * \\ * & * & * \end{bmatrix} \quad \text{is } (1+m_1+m_2) \times (1+m_1+m_2).$$

The estimator (4.3.8) (ii) of the parameter $\sigma^2\psi$ results from estimating β by its k-class estimator $\hat{\beta}_k$ and $\sigma^2\Omega$ by the function of the reduced-form residuals, $\frac{1}{\alpha} Y'\bar{P}_X Y$, i.e.,

$$(4.3.10) \qquad \hat{\sigma}^2\omega = \hat{\sigma}^2\omega_{11} - 2\hat{\sigma}^2\omega_1 + \beta_k' \hat{\sigma}^2 \Omega_{11} \beta_k$$

$$= \frac{1}{\alpha}\left[Y'\bar{P}_X y - 2y'\bar{P}_X Y_1 \beta_k + \beta_k' Y_1' \bar{P}_X Y_1 \beta_k\right]$$

$$= \frac{1}{\alpha}\left\| \bar{P}_X(y - [Y_1 , X_1]\hat{\delta}_k)\right\|^2$$

Special cases of the estimator (4.3.8) (ii) were dealt with by Basmann (1959) for $k = 1$ (2SLS) and $\alpha = T - K$ (T being the number of observations) and by Basmann and Richardson (1969) for $k = 1$, $\alpha = T$.

$$(4.3.8) \text{ (iii)} \qquad \frac{1}{\alpha}\left\| P_X(y - [Y_1 , X_1]\hat{\delta}_k)\right\|^2$$

which is a special case of the estimator proposed by Dhrymes (1969), i.e. for $k = 1$ (2SLS) and $\alpha = L = (K^{**} - m_1) = (K_2 - m_1)$, the degree of overidentification of (4.3.4). (As before we use K_2 for K^{**}.)

4. In order to treat the problem of disturbances-variance estimation in a systematic manner and to make comparisons among competing estimators, Brown, Ramage and Srivastava (1972) constructed the (S_1, S_2)- and R-class of disturbance-variance estimators.

The (S_1, S_2)-class estimators of $\sigma^2 \psi$ is formed as follows:

(4.4.1) $$\psi_k = \frac{1}{\alpha} \hat{u}'\hat{u}$$

where α is a scalar (non-random) and

(4.4.2) $$\hat{u} = S_1 y - S_2[Y_1, X_1]\hat{\delta}$$

and S_1 and S_2 are specified $(T \times T)$ symmetric matrices. In addition, S_1 and S_2 should be so chosen that

(4.4.3) $$(S_1 - S_2)A_1 = 0$$

where

(4.4.4) $$A_1 = [X\Pi_1 , X_1]$$

is $T \times (m_1 + K^*)$ and Π_1 is defined implicitly by

(4.4.5) $$\Pi = [\Pi_o , \Pi_1 , \Pi_2] \quad \text{is} \quad K \times (1 + m_1 + m_2).$$

Two special cases of the (S_1, S_2)-class which satisfy the condition (4.4.3) have been examined by Brown, Srivastava and Ramage:

(4.4.6) (a) $$S_1 - S_2 = \overline{P}_x$$

(4.4.6) (b) $$S_1 = S_2$$

The motivation for (4.4.6) (a) is that the columns of $\overline{P}_X$ span the subspace orthogonal to $M(X)$, the column space of X.

A particular convenient specilization of (4.4.) (a) has been specified as:

$$S_1 - S_2 = \overline{P}_X \tag{4.4.7}$$

$$S_2 = (cP_X + \overline{c}\,\overline{P}_X)$$

where c is a scalar and $\overline{c} = (1 - c)$. The second subclass (4.4.6) (b) leads to the R-class estimator: Let

$$S_1 = S_2 \tag{4.4.8}$$

$$R = \frac{1}{\alpha} S_1^2$$

The general R-class estimator is given by

$$\psi_k = u_k' R\, u_k \tag{4.4.9}$$

where u_k is the k-class residuals, i.e.

$$u_k = y - [Y_1, X_1]\hat{\delta}_k \tag{4.4.10}$$

In order to facilitate comparisons among the estimators (4.3.8) (i)-(iii) a further specialization of the R-class estimator has been developed:

$$R = \frac{1}{\alpha}(c\, P_X - \overline{c}\,\overline{P}_X) \tag{4.4.11}$$

where c is a scalar and $\overline{c} = (1 - c)$. This two parameter case includes the estimators (4.3.8)(i) - (iii):

For $c = \bar{c}$, we have (4.3.8) (i); for $c = 0$, we have (4.3.8) (ii) and for $c = 1$ we have (4.3.8) (iii).

5. Before stating the principal result, we introduce the following definitions:

(4.5.1)

$$Y_1 = X\Pi_1 + \sigma V_1$$

$$A_1 = [X\Pi_1, X_1] \quad \text{is } T \times (m_1 + K^*)$$

$$\bar{V} = [V_1, 0] \quad \text{is } T \times (m_1 + K^*)$$

$$Q = (A_1'A_1) \quad \text{is } (m_1 + K^*) \times (m_1 + K^*)$$

$$M_k = \bar{P}_{A_1} - k\bar{P}_X$$

$$a = \operatorname{tr} \bar{P}_X = T - K$$

$$L = (K^{**} - m_1) \text{ or } (K_2 - m_1)$$

$$q = \psi^{-1}\left[\frac{1}{T} E(\bar{V}'u)\right] \quad \text{is } (m_1 + K^*) \times 1$$

$$G = qq'$$

$$\Omega^* = \psi^{-1}\left[\frac{1}{T} E(\bar{V}'\bar{V})\right] \quad \text{is } (m_1 + K^*) \times (m_1 + K^*)$$

$$D = \Omega^* - G$$

Brown, Ramage and Srivastava (1972) also used the following conventions:

(4.5.2) For any square matrix Δ ,

$$< \Delta > = \Delta + \Delta'$$

For any matrix Z of full column rank,

$$P_Z = Z(Z'Z)^{-1}Z'$$

$$\overline{P}_Z = (I - P_Z)$$

The following relations have been used extensively:

$$\overline{P}_X A_1 = 0 \tag{4.5.3}$$

$$\overline{P}_X \overline{P}_{A_1} = \overline{P}_X = \overline{P}_{A_1} \overline{P}_X$$

<u>THEOREM 4.5.A</u>. When σ is small, the general (S_1, S_2)-class disturbance-variance estimator ψ_k defined in (4.4.1) is approximated closely by $\psi_k(.)$ i.e. $\psi_k - \psi_k(.) = O_p(\sigma^5)$.

In order to approximate ψ_k defined in (4.4.1), a series expansion in powers of the scalar multiple σ of the system of disturbances is first developed for $\hat{U}$ defined in (4.4.2). Brown, Ramage and Srivastava (1972) utilized the small-disturbance approximation to the k-class estimation errors, $e_k = (\hat{\delta}_k - \delta)$, developed by Kadane (1971).

Substituting for y from (4.3.4) and for $[Y_1 \; X_1]$ leads to

$$e_k = (\hat{\delta}_k - \delta) \tag{4.5.4}$$

i.e.

$$e_k = \left[[Y_1, X_1]'(I - k\bar{P}_X)[Y_1, X_1]\right]^{-1}[Y_1, X_1]'(I - k\bar{P}_X)y - \delta$$

$$= \left[(A_1 + \sigma\bar{V})'(I - k\bar{P}_X)(X + \sigma\bar{V})\right]^{-1}(A_1 + \sigma\bar{V})'(I - k\bar{P}_X)\sigma u$$

$$= \sigma e_k(1) + \sigma^2 e_k(2) + \sigma^3 e_k(3) + 0_p(\sigma^4)$$

$$= e_k(1) + 0_p(\sigma^4)$$

where

$$e_k(1) = Q A_1'u$$

$$e_k(2) = Q\bar{V}M_k u - QA_1'\bar{V}QA_1'u$$

$$e_k(3) = - Q < A_1'\bar{V} > Q\bar{V}'M_k u$$

$$- Q\bar{V}'M_k\bar{V}QA_1'u + QA_1'(\bar{V}QA_1')^2 u$$

Using this approximation for e_k in (4.4.2) leads to

$$(4.5.5) \qquad \hat{U} = (S_1 - S_2)y + S_2\left\{y - [Y_1, X_1]\hat{\delta}_k\right\}$$

$$= (S_1 - S_2)y - S_2\left\{[Y_1, X_1]e_k - \sigma u\right\}$$

$$= (S_1 - S_2)\left\{(A_1 + \sigma\bar{V})\delta + \sigma u\right\}$$

$$- S_2\left\{(A_1 + \sigma\bar{V})e_k - \sigma u\right\}$$

$$= \hat{U}(0) + \sigma\hat{u}(1) + \sigma^2\hat{u}(2) + \sigma^3\hat{u}(3)$$

$$+ \sigma^2\,\hat{u}(4) + 0_p(\sigma^6)$$

where

$$\hat{U}(o) = (S_1 - S_2)A_1'\delta$$

$$\hat{U}(1) = (S_1 - S_2)(u + \overline{V}\delta) + S_2 u - S_2 A_1 Q A_1' u$$

$$= (S_1 - S_2)(u + \overline{V}\delta) + S_2 \overline{P}_{A_1} u$$

$$\hat{U}(2) = - S_2 \overline{V} Q A_1' u - S_2 A_1 Q \overline{V}' M_k u$$

$$+ S_2 A_1 Q A_1' \overline{V} Q A_1' u$$

$$= - S_2 [A_1 Q \overline{V}' M_k u + \overline{P}_{A_1} \overline{V} Q A_1' u]$$

$$\hat{U}(3) = - S_2 \Big[\overline{V} Q \overline{V}' M_k u - (\overline{V} Q A_1')^2 u$$

$$- A_1 Q < A' \overline{V} > Q \overline{V}' M_k u$$

$$- A_1 Q \overline{V}' M_k \overline{V} Q A' u$$

$$+ A_1' Q A_1' (\overline{V} Q A_1')^2 u \Big] .$$

Then

(4.5.6)
$$\psi_k = \frac{1}{\alpha} \hat{U}' \hat{U}$$

$$= \psi_k(o) + \sigma \psi_k(1) + \sigma^2 \psi_k(2) + \sigma^3 \psi_k(3)$$

$$+ \sigma^4 \psi_k(4) + 0_p(\psi^6)$$

where

$$\psi_k(0) = \frac{1}{\alpha}\,\delta' A_1'(S_1 - S_2)^2 A_1\delta$$

$$\psi_k(1) = \frac{1}{\alpha}\Big[2\delta' A_1'(S_1 - S_2)S_2\bar{P}_{A_1} u$$

$$+ 2\delta' A_1'(S_1 - S_2)^2(u + \bar{V}\delta)\Big]$$

$$\psi_k(2) = \frac{1}{\alpha}\Big\{- 2\delta' A_1'(S_1 - S_2)S_2 A_1 Q\bar{V}' M_k u$$

$$- 2\delta' A_1'(S_1 - S_2)S_2\bar{P}_{A_1}\bar{V}QA_1' u$$

$$+ 2(u + \bar{V}\delta)'(S_1 - S_2)S_2\bar{P}_x u$$

$$+ (u + \bar{V}G)^1 (S_1 - S_2)^2 (u + \bar{V}\delta) + u'\bar{P}_x S_2^2 \bar{P}_x u\Big\}$$

See Brown, Ramage and Srivastava (1972) for further details.

COROLLARY 4.5.A.

The bias and mean-squared error of the approximation $\psi_k(\cdot)$ are given by

$$(4.5.7) \qquad E[\psi_k(\cdot) - \sigma^2 \psi] = d_o + \sigma^2 d_2 + \sigma^2 d_4 + O(\sigma^6)$$

and

$$(4.5.8) \qquad E[\psi_k(\cdot) - \sigma^2 \psi]^2 = D_o + \sigma^2 D_2 + \sigma^4 D_4 + O(\sigma^6)$$

where

$$d_o = E\psi_k(o) = \frac{1}{\alpha}\,\delta' A_1'(S_1 - S_2)^2 A_1 \delta$$

$$d_2 = E[\psi_k(2) - \psi]$$

$$d_4 = E\,\psi_k(4)$$

$$D_o = d_o^2$$

$$D_2 = 2d_o d_2 + E[\psi_k(1)]^2$$

$$D_4 = 2d_o d_4 + 2E[\psi_k(1)\,\psi_k(3)] + E[\psi_k(2) - \psi]^2$$

[See Appendix 2, page 56 of Brown, Ramage and Srivastava (1972) for details].

It follows from Theorem (4.5.A.) and Corollary (4.5.A) that any estimator for which $\hat{u}(0) = (S_1 - S_2)A_1'\delta$ vanishes has a smaller asymptotic mean-squared error than any estimator for which $\hat{u}(o) \neq 0$.

THEOREM (4.5.B)

For the special case of Theorem (4.5.A) and Corollary (4.5.A) where $(S_1 + S_2)A_1 = 0$, the bias and mean-squared error of the approximation $\psi_k(\cdot)$ to the (S_1, S_2)-class estimator ψ_k defined in (4.4.1) are as follows:

(4.5.9) $$E[\psi_k(\cdot) - \sigma^2\psi] = \sigma^2 d_2 + O(\sigma^4)$$

(4.5.10) $$E[\psi_k(\cdot) - \sigma^2\psi]^2 = \sigma^2 D_4 + O(\sigma^6)$$

(See page 56 of Appendix 2 of Brown, Ramage and Srivastava (1972) for proof).

6. THEOREM 4.6.A (The General R-class)

The the special case of Theorem 4.5.A and Corollary 4.5.A where

$$S_1 = S_2$$

$$R = \frac{1}{\alpha} S_1^2$$

the bias and mean-squared error approximation $\psi_k(\cdot)$ to the R-class estimator ψ_k defined in (4.4.9), i.e., $\psi_k = u_k' R u_k$, are as follows:

$$E[\psi_k(\cdot) - \sigma^2 \psi] = \sigma^2 h_2 + \sigma^4 h_4(k) + 0(\sigma^6) \tag{4.6.1}$$

and

$$E[\psi_k(\cdot) - \sigma^2 \psi]^2 = \sigma^4 H_4 + \sigma^6 H_6(k) + 0(\sigma^8) \tag{4.6.2}$$

where

$$h_2 = \psi \cdot a_o$$

$$\begin{aligned} h_4 = \psi^2 \Big\{ & a_1(k)(\mathrm{tr}\ Q\ G) + a_2(k)(\mathrm{tr}\ Q\ D) \\ & + a_3(k)(\mathrm{tr}\ Q\ A_1'\ R\ A_1 Q_1 G) \\ & + a_4(k)(\mathrm{tr}\ Q\ A_1'\ R\ A_1\ Q\ D)\Big\} \end{aligned}$$

and (separating non-stochastic k and (LIML) cases) .

$$a_o = [\mathrm{tr}\, \overline{P}_{A_1} R - 1]$$

$$a_1(k) = [-(2\, \mathrm{tr}\, M_k + 1)(\mathrm{tr}\, \overline{P}_{A_1} R) + 4k(\mathrm{tr}\, \overline{P}_X R)]$$

$$a_1(\ell_1) = [3(\mathrm{tr}\, \overline{P}_{A_1} R)]$$

$$a_2(k) = [-(\mathrm{tr}\, \overline{P}_{A_1} R) + 2k(\mathrm{tr}\, \overline{P}_X R)]$$

$$a_2(\ell_1) = [-(\mathrm{tr}\, \overline{P}_{A_1} R) + 2(1 + L/a)(\mathrm{tr}\, \overline{P}_X R)]$$

$$a_3(k) = [(\mathrm{tr}\, M_k)^2 + 2\, \mathrm{tr}\, M_k^2]$$

$$a_3(\ell_1) = 0$$

$$a_4(k) = [\mathrm{tr}\, M_k^2]$$

$$a_4(\ell_1) = [L(a + L)/(a - 2)]$$

[see (4.5.1) and (4.3.7)].

Also

$$H_4 = \psi^2 \left\{ a_o^2 + 2\, \mathrm{tr}\, (\overline{P}_{A_1} R)^2 \right\}$$

$$\begin{aligned} H_6(k) = \psi^3 \Big\{ & B_1(k)\, \mathrm{tr}(QG) + A_2(k)(\mathrm{tr}\, QD) \\ & + B_3(k)(\mathrm{tr}\, QA_1'RA_1QG) + B_4(k)\,(\mathrm{tr}\, QA_1'RA_1QD) \\ & + B_5(k)(\mathrm{tr}\, \overline{P}_{A_1} R\, A_1 Q\, G\, Q\, A_1'R) \\ & + B_6(k)\ (\mathrm{tr}\, \overline{P}_{A_1} R\, A_1 Q\, D\, Q\, A_1'R) \\ & + B_7(k)\ (\mathrm{tr}\, \overline{P}_X R\, A_1 Q\, G\, Q\, A_1'R) \\ & + B_8(k)\ (\mathrm{tr}\, \overline{P}_X R\, A_1 Q\, D\, Q\, A_1'R) \Big\} \end{aligned}$$

where (separating the non-stochastic k and LIML cases).

$$B_1(k) = 2\Big[a_0 a_1(k) - 2(\mathrm{tr}\,\overline{P}_{A_1} R - 2k\ \mathrm{tr}\,\overline{P}_x R)(\mathrm{tr}\,\overline{P}_{A_1} R)$$

$$- 2(2\ \mathrm{tr}\,M_k + 3)\mathrm{tr}(\overline{P}_{A_1} R)^2 + 16k\ \mathrm{tr}(\overline{P}_{A_1} R\ \overline{P}_x R)\Big]$$

$$B_1(\ell_1) = 2\left[a_0 a_1(\ell_1) + 2(\mathrm{tr}\,\overline{P}_{A_1} R)^2 + 10\ \mathrm{tr}(\overline{P}_{A_1} R)^2\right]$$

$$B_2(k) = 2\left[a_0 a_2(k) - 2\ \mathrm{tr}(\overline{P}_{A_1} R)^2 + 4k\ \mathrm{tr}(\overline{P}_{A_1} R\ \overline{P}_x R)\right]$$

$$B_2(\ell_1) = 2\Big[a_0 a_2(\ell_1) + \frac{4}{a}\,[\mathrm{tr}\,\overline{P}_{A_1} R - (1 + L/(a+2))\ \mathrm{tr}\,\overline{P}_x R]\ \mathrm{tr}\,\overline{P}_x R$$

$$- 2\mathrm{tr}(\overline{P}_{A_1} R)^2 + 4(1 + (L + 2/a)\mathrm{tr}\ (\overline{P}_{A_1} R\ \overline{P}_x R)$$

$$- (8/a)(1 + L/(a+2))\mathrm{tr}(\overline{P}_x R)^2\Big]$$

$$B_3(k) = 2\Big[a_0 a_3\cancel{(k) + 4(\mathrm{tr}\,M_k + 2)(\mathrm{tr}\,\overline{P}_{A_1} R)}$$

$$- 4k(\mathrm{tr}\,M_k + 2(2 - k)(\mathrm{tr}\,\overline{P}_x R)\Big]$$

$$B_3(\ell_1) = 0$$

$$B_4(k) = 2\left[a_0 a_4(k) + 2(\mathrm{tr}\,\overline{P}_{A_1} R) - 2k(2 - k)(\mathrm{tr}\,\overline{P}_x R)\right]$$

$$B_4(\ell_1) = 2\Big[a_0 a_4(\ell_1) + 2(a + 2L + 2)/(a - 2)(\mathrm{tr}\,\overline{P}_{A_1} R)$$

$$- 2(1 + (L+2)/(2 + L/a)(a - 2)(\mathrm{tr}\,\overline{P}_x R)\Big]$$

$$B_5(k) = 4\left[(\mathrm{tr}\,M_k)(\mathrm{tr}\,M_k + 4) + 2\mathrm{tr}\,M_k^2 + 8\right]$$

$$B_5(\ell_1) = 0$$

$$B_6(k) = 4[\mathrm{tr}\,M_k^2 + 2]$$

$$B_6(\ell_1) = 4[(L+2)\cdot(a + L + 2)/a - 2)]$$

$$B_7(k) = -16k[\mathrm{tr}\, M_k + 2(2 - k)$$

$$B_7(\ell_1) = 0$$

$$B_8(k) = -8k(2 - k)$$

$$B_8(\ell_1) = -8[1 + (L + 2)(2 + L/a)(a - 2)]$$

[See (4.5.1) for definition of a, L, G, M_k, A_1, and D].

<u>COROLLARY 4.6.A.</u> (A parameter special case)

When $R = \frac{1}{\alpha}(c\, P_x + \bar{c}\, \bar{P}_x)$ [see (4.4.11)]

$$E[\psi_k(.) - \sigma^2 \psi] = \sigma^2 d_2 + \sigma^4 d_4(k) + 0(\sigma^6) \tag{4.6.3}$$

$$E[\psi_k(.) - \sigma^2 \psi]^2 = \sigma^4 D_4 + \sigma^6 D_6(k) + 0(\sigma^8) \tag{4.6.4}$$

where

$$d_2 = \frac{\psi}{\alpha}\left\{a_0(c) - \alpha\right\}$$

$$d_4 = \frac{\psi^2}{\alpha}\left\{a_1(c,k)(\mathrm{tr}\, Q\, G) + a_2(c,k)(\mathrm{tr}\, Q\, D)\right\}$$

and (separating the non-stochastic k and LIML cases)

$$a_0(c) = (\bar{c}\, a + c\, L)$$

$$a_1(c,k) = [-(2\mathrm{tr}\, M_k + 1)(\bar{c}\, a + c\, L) + 4k\,\bar{c}\, a + c((\mathrm{tr}\, M_k)^2 + 2(\mathrm{tr}\, M_k^2))]$$

$$a_1(c,\ell_1) = 3(\bar{c}\, a + c\, L)$$

$$a_2(c,k) = [-(\bar{c}\ a + c\ L) + 2k\ \bar{c}\ a + c(\text{tr}\ M_k^2)]$$

$$a_2(c,\ell_1) = [-(\bar{c}\ a + c\ L) + (a+L)(2\ \bar{c} + cL/(a-2))]$$

Also

$$D_4 = \frac{\psi^2}{\alpha^2}\left\{(a_o(c)-\alpha)^2 + 2(\bar{c}^2 a + c^2 L)\right\}$$

$$D_6(k) = \frac{\psi^2}{\alpha^2}\left\{B_1(c,k)(\text{tr}\ Q\ G) + B_2(c,k)(\text{tr}\ Q\ D)\right\}$$

where (separating non-stochastic k from LIML)

$$\begin{aligned} B_1(c,k) = 2\Big[a_o(c)-\alpha)a_1(c,k) &- 2(\bar{c}\ a + c\ L)^2 \\ &+ 4(c(\text{tr}\ M_k + 2) + \bar{c}\ k\ a))(\bar{c}\ a + c\ L) \\ &- 2(2\text{tr}\ M_k + 3)(\bar{c}^2 a + c^2 L) \\ &+ 4k\ \bar{c}\ a\left\{(4\ \bar{c} - c(\text{tr}\ M_k + 2(2-k))\right\}\Big] \end{aligned}$$

$$\begin{aligned} B_1(c,k) = 2\Big[a_o(c)-\alpha)a_1(c,\ell_1) &+ 2(\bar{c}\ a + cL)^2 \\ &+ 10(\bar{c}^2 a + c^2 L)\Big] \end{aligned}$$

$$B_2(c,k) = 2\Big[(a_o(c)-\alpha)a_2(c,k) + 2\bar{c}\ a(\bar{c}(k-1) + c(1-k)^2\Big]$$

$$\begin{aligned} B_2(c,\ell_1) = 2\Big[(a_o(c) &- \alpha_2(c,\ell_1) \\ &+ 2(2\bar{c} + c(a+2L+2)/(a-2))\ (\bar{c}\ a + cL) \\ &- 2(\bar{c}^2 + c^2 L) - 2c\ \bar{c}(a + (L+2)(2a+L)/(a-2)\Big]. \end{aligned}$$

The asymptotic moment results for the estimators (4.3.8)(i) - (iii) are obtained by choosing $c = \bar{c}$; $c = 0$, and $c = 1$, respectively, in Corollary (4.6.A). Based on the leading term D_4 of the mean-squared error result in Corollary (4.6.A), the best α for each case and the corresponding mean-squared error are as follows:

TABLE (4.6.A)

c	ESTIMATOR	BEST α	ASYMPTOTIC MSE
$c = \bar{c}$	$\frac{1}{\alpha} \left\| y - [Y_1, X_2]\hat{\delta}_k \right\|^2$	$(T - K + L + 2)$	$2\sigma^4\psi^2/T-K+L+2) + O(\sigma^6)$
$c = 0$	$\frac{1}{\alpha} \left\| \bar{P}_X(y - [Y_1, X_2]\hat{\delta}_k) \right\|^2$	$(T - K + 2)$	$2\sigma^4\psi^2/(T-K+2) + 0(\sigma^6)$
$c = 1$	$\frac{1}{\alpha} \left\| P_X(y - [Y_1, X_1]\hat{\delta}_k) \right\|^2$	$(L + 2)$	$2\sigma^4\psi^2/(L + 2) + 0(\sigma^6)$

CHAPTER 5

FINITE SAMPLE PROPERTIES

The properties of consistency and asymptotic normality are commonly shared by all the limited information as well as full information estimators. Although the asymptotic properties of various estimators have been studied extensively, they do not provide a basis for choosing the estimation method when dealing with finite samples. Due to mathematical intractability there are as yet few exact results on the small sample properties of the various estimators.

Basmann [1961, 1963] and Kabe [1963, 1964] were among the pioneers in the theoretical investigation of the exact finite - sample properties of various estimation methods in a system of simultaneous equations. Specifically, they derived the exact finite-sample probability density function of the (2SLS) and (OLS) estimators for certain special systems composed of at most three equations. Bergstrom [1962] derived the exact density function of the (OLS) and (2SLS) estimators for the simple Keynesian model.

In particular, when the predetermined variables are exogenous, two endogenous variables occur in the relevant equation, and the coefficient of one endogenous variable is specified to be one, the exact distribution of the coefficient of one endogenous variable has been obtained by Richardson [1968] and Sawa [1969] in the case of (2SLS) and by Mariano and Sawa (1972) in the case of limited information maximum likelihood (LIML). Takeuchi [1970] and Sawa (1971), using the same situation, derived the exact sampling moments (first and second) of the (OLS) and (2SLS) estimator, and the exact finite-sample moments of Theil's k-class estimators for $o \le k \le 1$, respectively. Nagar and Ullah [1974] computed the mean of the (2SLS) estimator in the case where there are three endogenous variables in the equation under estimation.

Following the approach of Kabe [1963], Richardson [1968] and Sawa [1969] derived the exact distribution of the (OLS) and (2SLS) estimator by making a change of variable in the non-central Wishart distribution. This method of derivation involves the cumbersome task of integrating a non-cnetral Wishart distribution, and the results obtained are usually too complicated for making any meaningful comparison between the properties of various estimators.

Under the assumption that all the predetermined variables are exogenous and the equation to be estimated admits an arbitrary number of endogenous variables, Mariano [1972] proved that even moments of the (2SLS) estimator estimator exist if and only if the order is less than $K_2 - M_1 + 1$, and that even moments of the (OLS) estimator exist if and only if the order is less than $T - K_1 - M_1 + 1$. ($M_1 + 1$ is the number of included endogenous variables, K_1 and K_2 indicate the number of included and excluded exogenous variables respectively. T denotes the sample size.)

1. THE CASE OF ~~TWO~~ INCLUDED ENDOGENOUS VARIABLES.

The scalar paramter β in the structural equation

(5.T.1) $\quad y_1 = \beta y_2 + X_1\gamma_1 + U \quad$ is to be estimated.

Here y_1 and y_2 are (column) vectors of T observations on two endogenous variables, X_1 is a T x K_1 matrix of observations on K_1

included endogenous variables, U is a $T \times 1$ vector of disturbances, and γ_1 is a column vector of K_1 parameters.

The two reduced - form equations are:

$$(5.1.2) \qquad y_1 = X\pi_1 + v_1 = X_1\pi_{11} + X_2\pi_{21} + v_1$$

$$(5.1.3) \qquad y_2 = X\pi_2 + v_2 = X_1\pi_{12} + X_2\pi_{22} + v_2$$

Equations (5.1.2) and (5.1.3) can be expressed compactly as

$$(5.1.4) \qquad Y = X\Pi + v \quad \text{where}$$

$$Y = [y_1, y_2], \qquad V = [v_1, v_2].$$

$$X = [X_1, X_2]$$

is a $T \times K$ matrix of observations on all the exogenous variables in the system. X_2 is a $T \times K_2$ matrix of observations on K_2 exogenous variables which have been excluded, a priori, from equation (5.1.1) π_1 and π_2 are $K \times 1$ vectors of reduced - form coefficients, v_1 and v_2 are $T \times 1$ vectors of disturbances.

In addition to the conventional assumptions, we add the following:

Assumption 1: Each row of $[v_1, v_2]$ is independently and identically distributed two-dimensional normal variate with mean vector 0 and positive definite variance - covariance matrix

$$(5.1.5) \qquad \Omega = \begin{bmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{bmatrix}$$

Assumption 2: The $T \times K$ matrix X consists of known numbers, is of rank K, and $T>K$. To relate (5.1.1) to (5.1.4), we partition Π into K_1 and K_2 $(K_2 = K - K_1)$ rows, respectively and into two columns:

(5.1.6)
$$\Pi = \begin{bmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{bmatrix}$$

If we multiply (5.1.4) on the right by $[1, -\beta]'$, we obtain (5.1.1) where $\gamma_1 = \pi_{11} - \beta\pi_{12}$ and $U = v_1 - \beta v_2$. In order that (5.1.1) be properly written with X_2 omitted,

(5.1.7) $\pi_{21} = \beta\pi_{22}$

Assumption 3: The matrix $[\pi_{21} \;\; \pi_{22}]$ is of rank one and π_{22} has at least one non-zero component.

The components of U in (5.1.1) are independently normally distributed with mean 0 and variances

(5.1.8) $\sigma^2 = \omega_{11} - 2\beta\omega_{12} + \beta^2\omega_{22}$

2.

In this section we state some exact finite sample properties of the k-class estimator for $0\le k\le 1$, [See Sawa (1971) for details of proofs], and for double k-class estimators when k_1 lies between -1 and 1 but k_2 is unrestricted, [See Dwivedi and Srivastava (1980)].

In addition to our assumptions we adopt the following convention:

(5.2.1) $X_1'X_2 = 0$

Under the assumptions and the convention stated in (5.2.1), the k-class estimator $\hat{\beta}_k$ of β in (5.1.1) is given by the first component of

(5.2.2) $$\begin{bmatrix} y_2'y_2 - k\hat{V}_2'\hat{V}_2 & y_2'X_1 \\ X_1y_2 & X_1'X_2 \end{bmatrix}^{-1} \begin{bmatrix} y_2'y_1 - kV_2'y_1 \\ X_1'y_1 \end{bmatrix}$$

where $\hat{V}_2$ is the projection of y_2 into the space spanned by X, namely

(5.2.3) $\hat{V}_2 = My_2$ where

(5.2.4) $$M = I_T - X(X'X)^{-1}X'$$
$$= I_T - X_1(X_1'X_1)^{-1}X_1' - X_2(X_2'X_2)^{-1}X_2'$$

After a little manipulation, we can write

(5.2.5) $$\hat{\beta}_k = \frac{y_2' P_k y_1}{y_2' P_k y_2}$$ where

(5.2.6) $$P_k = (1 - k)M + X_2(X_2'X_2)^{-1}X_2$$

for k=0, k=1, the (OLS) and (2SLS) estimates of β are obtained.

To simplify proofs of theorems, we perform the following linear transformation on the two endogenous variables $[y_1, y_2]$:

(5.2.7) $$[y_1^\circ, y_2^\circ]f = [y_1, y_2]$$

to reduce the variance - covariance matrix, Ω, to an identity matrix, i.e., $\Omega = I$, is added to our assumptions.

The 2 x 2 matrix f is defined as:

(5.2.8) $$f = \begin{bmatrix} \alpha & o \\ bd & b \end{bmatrix}$$ where f is

a non-singular lower triangular matrix such that

(5.2.9) $\Omega = f'f$

(5.2.10) $\quad b = \sqrt{\omega_{22}}\ ; \quad d = \frac{\omega_{12}}{\omega_{22}}\ ; \quad a = \sqrt{\omega_{11} - \frac{\omega_{12}^2}{\omega_{22}}}$.

As pointed out by Sawa (1971), d is a regression coefficient of V_{1t} on V_{2t} and a^2 is the conditional variance of V_{1t} given V_{2t}, t=1.......T.

In terms of the transformed variables, we have an equivalent expression for the structural equation (5.1.1):

(5.2.11) $\quad y_1^\circ = \beta y_2^\circ + X_1\gamma_1^\circ + u^\circ,$ where

(5.2.12) $\quad \beta^\circ = \frac{b(\beta - d)}{a},$

$\gamma_1^\circ = a^{-1}\gamma_1$ and $u^\circ = a^{-1}u$

Also

(5.2.13) $\quad \hat{\beta}_k = d + \frac{a}{b}\hat{\beta}_k^\circ$ where $\hat{\beta}_k$ and $\hat{\beta}_k^\circ$

are the k-class estimators of β in (5.1.1) and β° in (5.2.11) respectively.

Sawa (1971) used the following LEMMA as a basis for deriving exact moments of the k-class extimator, $0 \leq k \leq 1$:

LEMMA (5.2.A):

Let Z_1 be an almost everywhere positive random variable and Z_2 be an arbitrary random variable. Suppose there exists a joint moment generating function of Z_1 and Z_2:

(5.2.14) $\quad \phi(\Theta_1, \Theta_2) = E[\exp(\Theta_1 Z_1 + \Theta_2 Z_2)]$

for $\Theta_1 < \varepsilon$ and $|\Theta_2| < \varepsilon$ where $\varepsilon > 0$ (ε is a constant). Then the m^{th} order moment of Z_2/Z_1 is given by

(5.2.15) $\quad \frac{1}{\Gamma(m)} \int_{-\infty}^{0} (-\Theta_1)^{m-1} \left[\frac{\partial^m \phi(\Theta_1, \Theta_2)}{\partial \Theta_2^m}\right]_{\Theta_2 = 0} d\Theta_1$

provided it either exists or is infinite, where m is a positive integer.

PROOF: By assumption, it is permissible to reverse the order of integration with respect to Θ_1 and Θ_2. Thus,

(5.2.15) is written as

$$(5.2.16) \quad \frac{1}{\Gamma(m)} \int_{-\infty}^{0} (-\Theta_1)^{m-1} \frac{\partial^m}{\partial \Theta_2^m} \Big[E(\exp.(\Theta_1 Z_1 + \Theta_2 Z_2) \Big]_{\Theta_2 = 0} d\Theta_1$$

$$= \frac{1}{\Gamma(m)} E\left[\left(\frac{\partial^m}{\partial \Theta_2^m} \exp.\Theta_2 Z_2\right)_{\Theta_2 = 0} \int_{-\infty}^{0} (-\Theta_1)^{m-1} \exp.\, \Theta_1 Z_1 \, d\Theta_1 \right]$$

$$= \frac{1}{\Gamma(m)} E\left[Z_2^m \int_0^\infty y^{m-1} e^{-Z_1 y} dy \right]$$

where $\quad y = -\Theta_1, \qquad dy = -\, d\Theta_1$

$$= \frac{1}{\Gamma(m)} E\left[Z_2^m \frac{\Gamma(m)}{Z_1^m} \right] = E\left[\frac{Z_2^m}{Z_1^m} \right]$$

$$= E\left[\left(\frac{Z_2}{Z_1}\right)^m \right],$$

since in general, for $m > 0$, and $p > 0$

$$\int_0^\infty t^{p-1} e^{-at} dt = \frac{\Gamma(p)}{a^p}$$

To apply the preceding procedure to evaluate the exact moments of the k-class estimator $\hat{\beta}_k$ in (5.2.5), the quadratic form $y_2' P_k y_2$ in (5.2.5) is required to be non-negative definite.

LEMMA (5.2.B): The $T \times T$ matrix P_k is non-negative definite if and only if $k \leq 1$ [See Sawa (1971) for proof].

Before stating the results, we introduce the following notations and functions:

(5.2.17) $g_1 = \frac{T-K_1}{2}$

(5.2.18) $g_2 = \frac{T-K}{2}$

(5.2.19) $c = \frac{\pi_{22}'(X_2'X_2)\pi_{22}}{2\omega_{22}}$

(if $X_1'X_2 \neq 0$, the definition of c should be changed as

$$\pi_{22}'X_2'[I-X_1(X_1'X_1)^{-1}X_1']X_2\pi_{22},$$

the noncentrality or concentration parameter).

(5.2.20) $g(x;\ k,\ p,\ q) = \frac{2}{(1-2x)^{p-q}[1-2(1-k)x]^q} \exp.\left[-c + \frac{c}{1-2x}\right]$

where $c > 0,\ p \geq q \geq 1;\ 0 \leq k \leq 1.$

(5.2.21) $G(k,\ c;\ p,\ q) = \int_{-\infty}^{0} g(x;\ k,\ p,\ q)dx$

The function G has the following powere series representation:

Lemma (5.2.C) for $0 \leq k < 1$,

(5.2.22) $G = \begin{cases} e^{-c} \sum_{i=0}^{\infty} (q)_i\ k^i \frac{\Gamma(p-1)}{\Gamma(p+i)}\ {}_1F_1(p-1;\ p+i;\ c) & \text{if } p > 1 \\ \infty, \text{ otherwise} \end{cases}$

or equivalently

(5.2.23) $G = \begin{cases} e^{-c} \sum_{i=0}^{\infty} \frac{c^i\Gamma(p-1+i)}{i!\,\Gamma(p+i)}\ {}_2F_1(1,\ q;\ p+i,\ k) & \text{if } p > 1 \\ \infty, \text{ otherwise.} \end{cases}$

and for k=1

(5.2.24) $G = \begin{cases} e^{-c} \frac{\Gamma(p-q-1)}{\Gamma(p-q)}\ {}_1F_1(p-q-1;\ p-q;\ c) & \text{if } p-q > 1 \\ \infty, \text{ otherwise} \end{cases}$

where the symbol $(\alpha)_i$ denotes the quantity

$$(\alpha)_0 = 1,\ (\alpha)_i = \frac{\Gamma(\alpha+i)}{\Gamma(\alpha)} = \alpha(\alpha+1)\ldots.(\alpha+i-1) \qquad i=1,2\ldots$$

${}_1F_1$ and ${}_2F_1$ denote the hypergeometric functions

$$(5.2.25) \quad {}_1F_1(r;s;x) = \sum_{h=0}^{\infty} \frac{(r)_h}{h!(s)_h} x^h,\ s > 0,\ |x| < \infty.$$

$$(5.2.26) \quad {}_2F_1(r,t;s;x) = \sum_{h=0}^{\infty} \frac{(r)_h(t)_h}{h!(s)_h} x^h,\ x > 0,\ |x| < 1$$

(See Sawa [1971], Appendix B, for a proof of Lemma 2C.).

THEOREM 5.2.(A):

The first order moment of the k-class estimator of β° in equation (5.2.11) for $0 \leq k \leq 1$ exists; it is given by

$$(5.2.27) \quad E(\hat{\beta}_k^\circ) = \beta^\circ\ c \cdot G\ (k,c;g_1+1,g_2)$$

THEOREM 5.2.(B):

The second-order moment of k-class estimator of β° in equation (5.2.11) for $0 \leq k \leq 1$ exists, provided when $0 \leq k \leq 1$ we have $T-K_1 \geq 3$ and when k=1 we have $K_2 \geq 3$. The second-order moment of $\hat{\beta}_k^\circ$ is given by:

$$(5.2.28) \quad E(\hat{\beta}_k^{\circ 2}) = [c+2\beta^{\circ 2}c^2]\ H\ (k,c;g_1+1,g_2)$$
$$+ (g_1-g_2+\beta^{\circ 2}c)\ H\ (k,c;g_1,g_2)$$
$$+ (1-k)^2 g_2\ H\ (k,c;g_1,g_2+1)$$

where

$$(5.2.29) \quad H(k,c;p,q) = -\frac{1}{2}\frac{\partial}{\partial c} G(k,c;p,q)$$
$$= \frac{1}{2}[G(k,c;p,q) - G(k,c;p+1,q)]$$

It should be noted that for k=0 or k=1, the function G can be simplified to the confluent hypergeometric function (5.2.25). Then moments are represented in terms of confluent hypergeometric functions.

From Theorems 5.2.(A) and 5.2.(B) and the relation

$$\hat{\beta}_k = d + \frac{a}{b}\hat{\beta}_k^\circ \qquad \text{and}$$

$$\hat{\beta}_k^2 = d^2 + 2\,d\,\frac{a}{b}\hat{\beta}_k^\circ + \frac{a^2}{b^2}\hat{\beta}_k^{\circ 2}$$

we define the following results:

COROLLARY 5.2.(A):

The first-order moment of the k-class estimator of β in (5.1.1) for $0 \le k \le 1$ exists; it is given by

(5.2.30) $d + (\beta-d)c\,G\,(k,c;g_1+1,g_2)$

which is obtained from

(5.2.31) $E(\hat{\beta}_k) = d + \frac{a}{b}E(\hat{\beta}_k^\circ)$

$$= d + \frac{a}{b}\left[\frac{b(\beta-d)}{a}\right]c\,G\,(k,c;g_1+1,g_2)$$

using (5.2.12).

COROLLARY 5.2.(B):

The second-order moment of the k-class estimator of β in (5.1.1) exists, provided when $0 \le k \le 1$ we have $T - K_1 \ge 3$ and for $k=1$ we have $K_2 \ge 3$;

(5.2.32) $E(\hat{\beta}_k^2) = d^2 + 2d\,\frac{a}{b}\,E\,(\hat{\beta}_k^\circ) + \frac{a^2}{b^2}\,E\,(\hat{\beta}_k^{\circ 2})$

(5.2.33) $= d^2 + 2d\,\frac{a}{b}\,h_1 + \frac{a^2}{b^2}\,h_2$

where h_1 and h_2 are respectively the right-hand side of (5.2.27) and (5.2.28) with $\frac{b}{a}(\beta-d)$ replacing β°.

COROLLARY 5.2.(C):

The following statements hold concerning the bias of the k-class estimator β_k of β in (5.1.1) for $0 \leq k \leq 1$.

(1) $\hat{\beta}_k$ is unbiased if and only if β=d.

(2) If $\beta \neq d$; $\hat{\beta}_k$ is biased in the same direction for all $0 \leq k \leq 1$ which is opposite to the sign of $(\beta - d)$.

Sawa (1971) also established the result that for k>1, the k-class estimator of β in (5.1.1) does not possess finite moments of any order. i.e.,

$$E\,[|\beta_k|] = \infty \tag{5.2.34}$$

The double k-class estimator, proposed by Nagar (1962) of β and γ_1 in (5.1.1) is given by

(5.2.35)
$$\begin{bmatrix} \hat{\beta}_1(\text{Dkc}) \\ \\ \gamma_1(\text{DkC}) \end{bmatrix} = \begin{bmatrix} y_2^1(I_T - k_1 M)y_2 & y_2^1 X_1 \\ & \\ X_1^1 y_2 & X_1^1 X_1 \end{bmatrix}^{-1} \begin{bmatrix} y_2^1(I_T - k_2 M)y_1 \\ \\ X_1^1 y_1 \end{bmatrix}$$

where k_1 and k_2 are scalars characterizing the estimator and M is as defined in (5.2.4).

Following Dwivedi and Srivastava (1980) the double k-class estimator of β in (5.1.1) can be written as

$$\hat{\beta}(\text{Dkc}) = \frac{y_2^1\, A\, y_1}{y_2^1\, A\, y_2} + (k_1 - k_2)\,\frac{y_2^1\, M\, y_1}{y_2^1\, A\, y_2} \tag{5.2.36}$$

$$= \hat{\beta}(k) + (k_1 - k_2)\,\frac{y_2^1\, M\, y_1}{y_2^1\, A\, y_2}$$

where $\hat{\beta}(k)$ denotes the k-class estimator of β with k_1 as the characterizing scalar and

$$A = [1-k_1][I_T - X_1(X_1'X_1)^{-1}X_1'] + k_1X_2(X_2'X_2)^{-1}X_2'$$

Dwivedi and Srivastava (1980) analysed the properties to the double k-class estimators when k_1 lies between -1 and 1 but k_2 is unrestricted. Utilizing the fact that the first two moments of double-k-class estimators are continuous functions of both the characterizing scalars k_1 and k_2, they found that it is possible to choose k_2 given k_1 such that the estimator specified by them is unbiased and that there always exists a value of k_2 given k_1 which will provide a double k-class estimator with $k_1 \neq k_2$ having smaller mean squared error than that of k-class estimator with k_1 as the characterizing scalar.

Before presenting their main results, we introduce the following function:

$$\psi_x(p;q;r) = e^{-c} \sum_{\alpha=0}^{\infty} \sum_{j=0}^{\infty} (x\alpha+1)k_1^{\alpha} \frac{\Gamma(g_1+j+p-1)\ \Gamma(g_2+\alpha+q)}{\Gamma(g_1+j+\alpha+r)\ \Gamma(g_2)} \frac{c^j}{j!} \tag{5.2.37}$$

where p, q, r and x are non-negative integers and $-1 \leq k < 1$, and g_1, g_2 and c are as defined in (5.2.17), (5.2.18) and (5.2.19).

THEOREM (5.2.C):

The mean of the double k-class estimator of β in (5.1.1) with $-1 \leq k_1 < 1$, is given by

$$E[\hat{\beta}(Dkc)] = \frac{\omega_{12}}{\omega_{22}} + (\beta - \frac{\omega_{12}}{\omega_{22}})c\ \psi_0\ (1;0;1) + (k_1-k_2)\ \frac{\omega_{12}}{\omega_{22}}\ \psi_0\ (1;1;1) \tag{5.2.38}$$

provided $T - K_1 \geq 1$.

Theorem (5.2.D):

Assuming $T - K_1 \geq 3$ the second moment of the double-k-class estimator of β in (5.1.1) for $-1 \leq k < 1$ is

$$(5.2.39)\quad E[\hat{\beta}^2(Dkc)] = (\frac{\omega_{12}}{\omega_{22}})^2 [1+2(k_1-k_2)\psi_0(1;1;1) + (k_1-k_2)^2\psi_1(1;2;2)]$$

$$+ \frac{\omega_{11.2}}{2\omega_{22}} [(1-k_2)^2\psi_1(0;1;1) + (g_1-g_2)\psi_1(0;0;1) + c\psi_1(1;0;2)]$$

$$+ \frac{c}{2} (\beta - \frac{\omega_{12}}{\omega_{22}})^2 [\psi_1(0;0;1) + 2c\ \psi_1(1;0;2)]$$

$$+ 2c \frac{\omega_{12}}{\omega_{22}} (\beta - \frac{\omega_{12}}{\omega_{22}}) [\psi_0(1;0;1) + (k_1-k_2)\psi_1(1;1;2)]$$

where from Anderson and Sawa (1973)

$\frac{\omega_{12}}{\omega_{22}}$ is the regression of a component of v_1 on the corresponding component of v_2 and $\omega_{11.2} = \omega_{11} - \frac{\omega_{12}^2}{\omega_{22}}$ is the variance of the residual of the component of v_1 from the regression of the corresponding component of v_2.

THEOREM (5.2.E):

The expressions for first and second moment of the double-k-class estimator of β in (5.1.1) with $k_1=1$ are:

$$(5.2.40)\quad E[\hat{\beta}(Dkc)] = \frac{\omega_{12}}{\omega_{22}} + (\beta - \frac{\omega_{12}}{\omega_{22}})c\ \phi(0;1) + g_2 (1-k_2) \frac{\omega_{12}}{\omega_{22}}\ \phi(1;0)$$

and

$$\text{(5.2.41)} \quad E[\hat{\beta}^2(Dkc)] = \left(\frac{\omega_{12}}{\omega_{22}}\right)^2 [1+2g_2(1-k_2)\phi(1;0) + g_2(g_2+1)(1-k_2)^2\phi(2;0)]$$
$$+ \frac{\omega_{11.2}}{\omega_{22}} [g_2(1-k_2)^2\phi(2;0) + \phi(1;0)]$$
$$+ \frac{c}{2} \left(\beta - \frac{\omega_{12}}{\omega_{22}}\right)^2 [\phi(1;1) + 2c\phi(0;2)]$$
$$+ 2c \frac{\omega_{12}}{\omega_{22}} \left(\beta - \frac{\omega_{12}}{\omega_{22}}\right) [\phi(0;1) + g_2(1-k_2)\phi(0;1)]$$

where $(g_1-g_2) \geq \frac{3}{2}$ and for non-negative integers p and q

$$\text{(5.2.42)} \quad \phi(p,q) = e^{-c} \sum_{j=0}^{\infty} \frac{\Gamma(g_1-g_2+j-p)}{\Gamma(g_1-g_2+j+q)} \cdot \frac{c^j}{j!}$$

Setting $k_1=k_2$ in (5.2.38), (5.2.39), (5.2.40) and (5.2.41), we get the corresponding results for the k-class estimator with k_1 as the characterizing scalar. It is interesting to note that the corresponding expressions thus obtained, i.e., $E[\hat{\beta}(k)]$ and $E[\beta^2(k)]$ agree with those given in (5.2.30) and (5.2.32) for $0 \leq k \leq 1$ if the following equivalence for $G(\cdot)$ and $H(\cdot)$ is noted:

Sawa's Notation	Dwivedi and Srivastava's Notation	
	$0 < k_1 < 1$	$k=1$
$G(k_1;c;g_1+1,g_2)$	$\psi_0(1;0;1)$	$\phi(0;1)$
$H(k_1;c;g_1,g_2)$	$\frac{1}{2}\psi_1(1;0;2)$	$\frac{1}{2}\phi(0;2)$
$H(k_1;c;g_1,g_2)$	$\frac{1}{2}\psi_1(0;0;1)$	$\frac{1}{2}\phi(1;1)$
$H(k_1,c;g_1,g_2+1)$	$\frac{1}{2}\psi_1(0;1;1)$	$\frac{1}{2}\phi(2;0)$

along with the result

$$(5.2.43)\qquad c\phi(0;2) + (g_1-g_2)\phi(1;1) = \phi(1;0)$$

Utilizing the evaluation of $\lim_{z\to 1-} {}_1F_1(\alpha,\beta;\gamma;z)$ for real $(\alpha-\beta-\gamma)>0$ as developed in LEBEDEV (1965, pp. 243-246), they found that $E[\hat{\beta}(Dkc)]$ is a continuous function of k_1 and k_2 similarly $E[\hat{\beta}^2(Dkc)]$ is also a continuous function of k_1 and k_2 provided $(g_1-g_2)\geq\frac{3}{2}$.

From

$$(5.2.44)\qquad \lim_{k_1\to 1-} \psi_0(1;0;1) = \phi(0;1)$$

$$\lim_{k_1\to 1-} \psi_0(1;1;1) = g_2\phi(1;0)$$

they found the bias of the double-k-class estimator of β is given by

$$(5.2.45)\qquad E[\hat{\beta}(Dkc) - \beta] = (\beta - \frac{\omega_{12}}{\omega_{22}})\ [c\psi_0(1;0;1)-1]$$

$$+ (k_1-k_2)\frac{\omega_{12}}{\omega_{22}}\psi_0(1;1;1)$$

$$\text{if } -1 \leq k < 1$$

and

$$E[\hat{\beta}(Dkc) - \beta] = (\beta - \frac{\omega_{12}}{\omega_{22}})\ [c\psi_0(0;1)-1]$$

$$+ g_2\ (1-k_2)\ \frac{\omega_{12}}{\omega_{22}}\phi(1;0) \quad \text{if } k_1=1$$

Now (5.2.45) vanishes when for $-1 \leq k_1 < 1$

$$(5.2.46)\qquad k_1 = k_2 \quad\text{and}\quad \beta = \frac{\omega_{12}}{\omega_{22}}$$

or

$$(5.2.47)\qquad k_2 = k_1 + \frac{\omega_{22}}{\omega_{12}}(\beta - \frac{\omega_{12}}{\omega_{22}})\left[\frac{c\psi_0(1;0;1)-1}{\psi_0(1;1;1)}\right];$$

and $T \times K_2$ matrices with columns as the orthorgonal characteristic vectors corresponding to the root 1 of idempotent matrices $X_1(X_1'X_1)^{-1}X_1'$ and $X_2(X_2'X_2)^{-1}X_2'$ respectively, and Q_3 is a $T \times (T-K)$ matrix orthogonal to Q_1 and Q_2 with $Q_3'Q_3 = I_{T-K}$.

Further, we define

(5.3.20) $$Q_2'Y = A\ , \quad q_2'y = a$$
$$Q_3'Y = B\ , \quad q_3'y = b$$

By construction of Q and the definition of P, it follows that

(5.3.21) $$Q'PQ = \begin{bmatrix} 0 & 0 & 0 \\ 0 & I_{K_2} & 0 \\ 0 & 0 & (I-k)I_{T-K} \end{bmatrix}$$

Thus we have

(5.3.22) $$\begin{aligned} Y'PY &= Y'Q \cdot Q'PQ \cdot Q'Y \\ &= A'A + (1-k)B'B \end{aligned}$$

(5.3.23) $$\begin{aligned} Y'Py &= Y'Q \cdot Q'PQ \cdot Q'y \\ &= A'a + (1-k)B'b \end{aligned}$$

so that from (5.3.18) the k-class estimator of β can now be written as

(5.3.24) $$\hat{\beta}(k) = [A'A+(1-k)B'b]^{-1}[A'a+(1-k)B'b]\ .$$

Utilizing the normality of the disturbances, it can be shown that a, b, A and B are independently distributed with a and b following multivariate normal distributions:

(5.3.25) $$a \sim MND\ (\alpha, I_{K_2})$$

(5.3.26) $\quad b \sim MND(0, I_{T-K})$

Also A'A follows a non-central Wishart distribution while B'B follows a central Wishart distribution.

If p is any arbitrary m-dimensional vector with positive elements and k is nonstochastic, consider the first absolute moment of $p'\hat{\beta}(k)$. For this purpose, let z denote the vector obtained from $[A'a+(1-k)B'b]$ by replacing each element in it by its absolute element.

Thus, we have

(5.3.27) $\quad E(|p'\hat{\beta}(k)|) = E(p^{*\prime}z)$

where

(5.3.28) $\quad p^{*\prime} = p'[A'A+(1-k)B'B]^{-1}$

Since the conditional distribution of $p'\hat{\beta}(k) = p^{*\prime}[A'a+(1-k)B'b]$ given A and B is univariate normal with mean $p^{*\prime}A'\alpha(=\mu$, say) and variance $p^{*\prime}[A'A+(1-k)^2B'B]p^*(=\sigma^2$, say), we observe that [see Sawa (1971), equation 6.11]

(5.3.29) $\quad E_{a,b}(|p'\hat{\beta}(k)|\,|A,B) \geq (\frac{2}{\pi})^{\frac{1}{2}}\,\sigma\, e^{-\mu^2/2\sigma^2}$

From [Rao (1965), (If. 1.1), p. 48], we have

(5.3.30) $\quad \underset{A^*p}{\text{Sup}} \dfrac{(p^{*\prime}A'\alpha)^2}{p^{*\prime}A'Ap^*} = \alpha'\alpha$

so that

(5.3.31) $\quad \dfrac{\mu^2}{\sigma^2} \leq \dfrac{(p^{*\prime}A'\alpha)^2}{p^{*\prime}A'Ap^*} \leq \alpha'\alpha$

using (5.3.31) along with the result

(5.3.32) $\quad \sigma^2 \geq (1-k)\sqrt{p^{*\prime}B'Bp^*}$

we find from (5.3.29)

(5.3.33) $$E_{a,b}(|p'\hat{\beta}(k)|\,|A,B) \geq (\tfrac{2}{\pi})^{\frac{1}{2}}(1-k)\sqrt{p^{*\prime}B'Bp^{*}}\cdot e^{-\frac{1}{2}\alpha'\alpha}$$

and thus

(5.3.34) $$E(|p'\hat{\beta}(k)|\,|B) = E_A\{E_{a,b}(|p'\hat{\beta}(k)|A,B\}|B\}$$

$$\geq (\tfrac{2}{\pi})^{\frac{1}{2}}(1-k)e^{-\frac{1}{2}\alpha'\alpha}\, E_A(\sqrt{p^{*\prime}B'Bp^{*}}\,|B)$$

Now recalling that A'A follows a noncentral Wishart distribution and is independent of B'B, we have

(5.3.35) $$E_A(\,p^{*\prime}B'Bp^{*}|B)$$

$$= E_A(\{p'[A'A+(1-k)B'B]^{-1}\;B'B[A'A+(1-k)B'B]^{-1}p\}^{\frac{1}{2}}|B)$$

$$= \int\{p'[A'A-(k-1)B'B]^{-1}\;B'B[A'A-(k-1)B'B]^{-1}p\}^{\frac{1}{2}}f(A'A)dA'A$$

which is infinite for $k>1$; see also [Sawa (1971) Footnote 12, p. 666]. Here $f(A'A)$ denotes the noncentral Wishart density function.

Combining (5.3.34) and (5.3.35), it follows that the conditional expectation of $|p'\hat{\beta}(k)|$ given B does not exist for $k>1$ and hence the expectation of $|p'\hat{\beta}(k)|$ for $k>1$ does not exist. This implies the expectation of $p'\hat{\beta}(k)$ for $k>1$ is not finite, but p is any arbitrary vector so that expectation of $\hat{\beta}(k)$ and similarly expectation of $\hat{\gamma}(k)$ are infinite for $k>1$. This establishes the nonexistence of moments of k-class estimator when k is nonstochastic and $k>1$.

The characterizing scalar k in the case of the (LIML) is stochastic and is given by [see Kadane equation 4, p. 726]

(5.3.36) $$k = k_0$$

where

$$k_0 = \min_{\beta} \frac{(y-Y\beta)'\overline{P}_{X_1}(y-Y\beta)}{(y-Y\beta)'\overline{P}_{X}(y-Y\beta)}$$

where

$$\bar{P}_{X_1} = I_T - X_1(X_1'X_1)^{-1}X_T'$$

(5.3.37) $$\bar{P}_X = I_T - X(X'X)^{-1}X'$$

From the relations $\bar{P}_{X_1}$, $\bar{P}_X$,

(5.3.38) $$k_0 = 1 + \min_{\beta} \frac{(y-Y\beta)'X_2(X_2'X_2)^{-1}X_2'(y-Y\beta)}{(y-Y\beta)'\bar{P}_X(y-Y\beta)}$$

$$= 1 + \varepsilon \quad \text{(say)}$$

where ε is a non-negative scalar.

Thus we observe that the characterizing scalar in the case of (LIML) assumes a value larger than 1 and therefore the associated estimator will not possess finite moments.

4.

Carter (1976) derived the exact distribution of the Instrumental Variables (I.V.) estimator when the instruments are non-stochastic. Once again we consider equation (5.1.1) i.e.,

$$y_1 = \beta y_2 + X_1\gamma_1 + u.$$

Let $[W,X_1]$ be the set of non-stochastic instruments where W is a T x 1 vector which is a non-stochastic linear combination of the columns of X; $W = X\eta$, where η is a K x 1 (non-stochastic) vector chosen so that the matrix

$$[W,X_1]'[y_2,X_1] \text{ is non-singular.}$$

The Instrumental Variables (I.V.) estimator of the structural coefficients in (5.1.1) is given by

(5.4.1) $\{[W,X_1]' [y_2,X_1]\}^{-1} [W,X_1]' y_1$

i.e.,

(5.4.2) $$\begin{bmatrix} \hat{\beta} \\ \gamma_1 \end{bmatrix} = \begin{bmatrix} W'y_2 & W'X_1 \\ X_1'y_2 & X_1'X_1 \end{bmatrix}^{-1} \begin{bmatrix} W'y_1 \\ X_1'y_1 \end{bmatrix}$$

The instrumental Variables (I.V.) normal equations are

(5.4.3) $W'y_2 \hat{\beta} + W'X_1\hat{\gamma}_1 = W'y_1$

(5.3.4) $X'y_2 \hat{\beta} + X_1'X_1\hat{\gamma}_1 = X_1'y_1$

Subsituting,

(5.4.5) $\hat{\gamma}_1 = (X_1'X_1)^{-1}X_1y_1 - (X_1'X_1)^{-1}X_1'y_2\hat{\beta}$

[using (5.4.4)]

into (5.4.3) yields

(5.4.6) $W'y_1 = W'y_2\hat{\beta} + W'X_1(X_1'X_1)^{-1}X_1'y_1 - W'X_1(X_1'X_1)^{-1}X_1'y_2\hat{\beta}$

i.e.,

$$W'y_1 - W'X_1(X_1'X)^{-1}X_1'y_1 = \hat{\beta}[W'y_2-W'X_1(X_1'X_1)^{-1}X_1'y_2]$$

Therefore,

(5.4.7) $$\hat{\beta} = \frac{W'[I_T-X_1(X_1'X_1)^{-1}X_1']y_1}{W'[I_T-X_1(X_1'X_1)^{-1}X_1]y_2}$$

$$= \frac{W'M_1y_1}{W'M_1y_2} = \frac{z_1}{z_2} \quad \text{(say)}$$

where

$$M_1 = [I_T - X_1(X_1'X_1)^{-1}X_1']$$

The elements of W and M are by assumption non-stochastic. Thus, we have

(5.4.8) $\quad E(Z_1) = WM_1X\pi_1 = \mu_1 \quad$ (say)

$E(Z_2) = WM_1X\Pi_2 = \mu_2 \quad$ (say)

(Using the reduced forms of y_1 and y_2.)

and

(5.4.9) $\quad E(Z_1-\mu_1)^2 = \omega_{11}\, W'M_1W = \sigma_1{}^2 \quad$ (say

$E(Z_2-\mu_2)^2 = \omega_{22}\, W'M_1W = \sigma_2{}^2 \quad$ (say)

$E(Z_1-\mu_1)(Z_2-\mu_2) = \omega_{22}\, W'M_1W = \sigma_{12} \quad$ (say)

From the normality assumption (Assumption 1, Section I, Chapter 5) we find that Z_1 and Z_2 are from the bivariate normal population

(5.4.10)
$$f(z_1z_2) = \frac{1}{2\Pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \text{ exp. } \{\frac{-1}{2(1-\rho^2)} \left[\left(\frac{Z_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{Z_1-\mu_1}{\sigma_1}\right)\left(\frac{Z_2-\mu_2}{\sigma_2}\right) + \left(\frac{Z_2-\mu_2}{\sigma_2}\right)^2 \right]\}$$

with $\rho = \dfrac{\sigma_{12}}{\sigma_1\sigma_2}$ and $\Pi = 3.1415.....$ (It has no relation to the reduced form coefficients.)

From Feller (1932) the density function for the ratios of normal variables like (5.4.7) is given by:

(5.4.11)
$$f(\tilde{\beta}) = \frac{1}{\Pi\alpha} \{\sigma_1\sigma_2\sqrt{1-\rho^2} \text{ exp. } \left[\frac{-1}{2(1-\rho^2)} \left(\frac{\mu_2^2}{\sigma_2^2} - \frac{2\rho\mu_1\mu_2}{\sigma_1\sigma_2} + \frac{\mu_1{}^2}{\sigma_1{}^2}\right)\right]$$
$$+ \frac{b}{\sqrt{a}} \text{ exp. } \left[\frac{-m^2}{2}\right] \int_0^{\frac{b}{\sigma_1\sigma_2\sqrt{a(1-\rho^2)}}} \text{exp. } [-\tfrac{1}{2}t^2]dt\}$$

where

$$a = \sigma_1{}^2 - 2\hat{\beta}\sigma_{12} - \hat{\beta}^2\sigma_2{}^2$$

$$b = (\mu_1\sigma_{12}-\mu_2\sigma_1^{\,2}) + (\mu_2\sigma_{12}-\mu_1\sigma_2^{\,2})\hat{\beta}$$

$$m = \frac{\mu_1-\mu_2\hat{\beta}}{\sqrt{a}}$$

and

$$t = \frac{aZ_2 - \rho\sigma_1\mu_1 + (\rho\sigma_1\mu_2-\sigma_2\mu_1)\hat{\beta} - \mu_2\sigma_2\hat{\beta}^2}{\sigma_1\sqrt{a(1-\rho^2)}}$$

It should be noted that for $\mu_1 = \mu_2 = 0$, $\sigma_1^{\,2} = \sigma_2^{\,2} = 1$ and $\rho = 0$, (5.4.11) reduces to the Cauchy distribution.

Feller (1932) pointed out that the distribution given by (5.4.11) has no finite moments of any order. Thus, the exact distribution of the Instrumental Variables (I.V.) estimator using non-stochastic instruments to estimate an equation of the form (5.1.1) has no finite moments for any degree of overidentification. Sawa (1969) showed that the (OLS) and (2SLS) estimator of β in (5.1.1) posses finite moments up to order $T - K_1 - 1$ and $K_2 - 1$ respectively. Hence the (2SLS) estimator is perferable over the I.V. estimator in the case of (5.1.1).

Geary (1930) approximated the distribution given in (5.4.11) by one which does have moments. Carter (1976), using Geary's approximation, pointed out that in the special case where a multiple of y_2 has a standard deviation less than one-third as large as its mean (for any sample size or degree of over-identification) the distribution of the I.V. estimator is closely approximated by a distribution whose moments exist and whose mean is nearly equal to β.

5. THE EXACT MEAN OF (2SLS) ESTIMATOR.

The Case of three Endogenous Variables:

Let the equation of interest be denoted by

(5.5.1) $$y = \beta_1 y_1 + \beta_2 y_2 + X_1 \gamma_1 + U,$$

where y, y_1 and y_2 are $T \times 1$ vectors of observations on the jointly dependent (endogenous) variables, X_1 is the $T \times K_1$ matrix of observations on the predetermined (exogenous) variables, β_1 and β_2 are scalar parameters, γ_1 is a vector parameter and U is a $T \times 1$ vector of disturbances.

The reduced form corresponding to y, y_1 and y_2 can be written as:

(5.5.2) $$y = X\pi + v$$

(5.5.3) $$y_1 = X\pi_1 + v_1$$

(5.5.4) $$y_2 = X\pi_2 + v_2$$

where $X = [X_1, X_2]$ π, π_1 and π_2 are $K \times 1$ vectors of reduced from parameters; v, v_1 and v_2 are $T \times 1$ vectors of reduced form disturbances.

Assumption:

The random vector $[v', v_1', v_2']$ is distributed as multivariate normal with zero mean and positive definite covariance matrix

(5.5.5) $$\Omega \otimes I_T \quad \text{where} \quad \Omega = \begin{bmatrix} \omega_{11} & \omega_{12} & \omega_{13} \\ \omega_{21} & \omega_{22} & \omega_{23} \\ \omega_{31} & \omega_{32} & \omega_{32} \end{bmatrix}$$

and $\otimes$ denotes the Kronecker product.

From the above assumption, it can be deduced that the random vector $[y', y_1', y_2']$ is also distributed as multivariate normal with mean $[\bar{y}', \bar{y}_1', \bar{y}_2']$ and covariance matrix $\Omega \otimes I$, where

$$(5.5.6) \qquad \bar{y} = E(y) = X\pi;$$

$$(5.5.7) \qquad \bar{y}_1 = E(y_1) = X\pi_1$$

$$(5.5.8) \qquad \bar{y}_2 = E(y_2) = X\pi_2 \, .$$

The two-stage least squares estimators of β_1 and β_2 in (5.5.1), denoted by $\hat{\beta}_1$ and $\hat{\beta}_2$ respectively, are given by

$$(5.5.7) \qquad \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} y_1'Ny_1 & y_1'Ny_2 \\ y_2'Ny_1 & y_2'Ny_2 \end{bmatrix}^{-1} \begin{bmatrix} y_1'Ny \\ y_2'Ny \end{bmatrix}$$

where

$$(5.5.8) \qquad N = X(X'X)^{-1}X' - X_1(X_1'X_1)^{-1}X_1'$$

is a T x T symmetric idempotent matrix, and

$$(5.5.9) \qquad \text{rank } N = \text{trace } N = K_2.$$

Nagar and Ullah (1974) showed that the mean value of the (2SLS) estimator depends on the variance-covariance structure of the endogenous variables in the case where the equation to be estimated has three endogenous variables. Sawa (1969) pointed out that a similar situation holds in the case of the equation having two endogenous variables.

Following Nagar and Ullah (1974), we make the following linear transformation:

$$(5.5.10) \qquad [y^*, y_1^*, y_2^*]T' = [y, y_1, y_2]$$

where T is given by the nonsingular upper triangular matrix such that $\Omega = TT'$,

$$(5.5.11) \qquad T = \begin{bmatrix} t_{11} & t_{12} & t_{13} \\ 0 & t_{21} & t_{23} \\ 0 & 0 & t_{33} \end{bmatrix}$$

where

$$(5.5.12) \qquad t_{33} = \sqrt{\omega_{33}} \quad ; \quad t_{23} = \frac{\omega_{23}}{\sqrt{\omega_{33}}}$$

$$t_{22} = \sqrt{\omega_{22} - t_{23}^2} \quad ; \quad t_{12} = \frac{12 - t_{13}\, t_{23}}{\sqrt{\omega_{22} - t_{23}^2}}$$

$$t_{11} = \sqrt{\omega_{11} - t_{12}^2 - t_{13}^2}$$

Under the transformation (5.5.8) identifiability restrictions are preserved and the elements of y^*, y_1^* and y_2^* are independently normally distributed with means

$$(5.5.13) \qquad [\bar{y}^*, \bar{y}_1^*, \bar{y}_2^*] = [\bar{y}, \bar{y}_1, \bar{y}_2](T^{-1})'$$

and covariance matrices

$$(5.5.14) \qquad \begin{aligned} E(y^*-\bar{y}^*)(y^*-\bar{y}^*)' &= E(y_1^*-\bar{y}_1^*)(y_1^*-\bar{y}_1^*)' \\ &= E(y_2^*-\bar{y}_2^*)(y_2^*-\bar{y}_2^*)' \\ &= I \\ E(y^*-\bar{y}^*)(y_1^*-\bar{y}_1^*)' &= E(y_1^*-\bar{y}_1^*)(y_2^*-\bar{y}_2^*)' \\ &= E(y^*-\bar{y}^*)(y_2^*-\bar{y}_2^*)' \\ &= 0 \end{aligned}$$

Let the transformed structural equation be denoted by

(5.5.15) $$y^* = \beta_1^* y_1^* + \beta_2^* y^*_2 + X_1\gamma_1^* + U^*,$$

where

(5.5.16) $$\beta_1^* = (\beta_1 t_{22} - t_{12})t_{11}^{-1}$$

$$\beta_2^* = (\beta_1 t_{23} + \beta_2 t_{23} - t_{13})t_{11}^{-1}$$

and the (2SLS) estimators of β_1^* and β_2^*, denoted by $\hat{\beta}_1^*$ and $\hat{\beta}_2^*$ respectively, are given by

(5.5.17) $$\begin{bmatrix} \hat{\beta}_1^* \\ \hat{\beta}_2^* \end{bmatrix} = \begin{bmatrix} y_1^{*\prime}Ny_1^* & y_1^{*\prime}Ny_2^* \\ y_2^{*\prime}Ny_1^* & y_2^{*\prime}Ny_2^* \end{bmatrix}^{-1} \begin{bmatrix} y_1^{*\prime}Ny^* \\ y^{*\prime}Ny^* \end{bmatrix}$$

where N is the same as in (5.5.8).

The (2SLS) estimators $\hat{\beta}_1^*$ and $\hat{\beta}_2^*$ are linearly related to $\hat{\beta}_1$ and $\hat{\beta}_2$ as follows:

(5.5.18) $$\hat{\beta}_1^* = (\hat{\beta}_1 t_{22} - t_{12})t_{11}^{-1} \quad \text{or}$$

$$\hat{\beta}_1^* = [\hat{\beta}_1^* t_{11} + t_{12}]t_{22}^{-1}$$

(5.5.19) $$\hat{\beta}_2^* = [\hat{\beta}_1 t_{23} + \hat{\beta}_2 t_{23} - t_{13}]t_{11}^{-1} \quad \text{or}$$

$$\hat{\beta}_2 = [(\hat{\beta}_2^* t_{11} + t_{13}) - t_{23}\,(\hat{\beta}_1^* t_{11} + t_{12})_{22}^{-1}]t_{33}^{-1}.$$

Since rank N = trace N = K_2, there exists a T x K_2 matrix R of K_2 orthonormal vectors such that*

(5.5.20) $$N = RR' \text{ and } R'R = I.$$

Then we can rewrite the (2SLS) estimators $\hat{\beta}_1^*$ and $\hat{\beta}_2^*$ in (5.5.17) as:

---

* See page 13 of "An Introduction to Linear Statistical Models" by Graybill, N.Y., McGraw-Hill, 1961.

$$(5.5.21)\qquad \begin{bmatrix} \hat{\beta}_1^* \\ \\ \hat{\beta}_2^* \end{bmatrix} = \frac{1}{D}\begin{bmatrix} (Z_2'Z_2)(Z_1'Z) - (Z_1'Z_2)(Z_2'Z) \\ \\ (Z_1'Z_1)(Z_2'Z) - (Z_1'Z)(Z_1'Z) \end{bmatrix}$$

where

$$(5.5.22)\qquad D = (Z_1'Z_1)(Z_2'Z_2) - (Z_1'Z_2)^2$$

$$(5.5.23)\qquad Z = R'y \ ; \quad Z_1 = R'y_1 \ , \quad Z_2 = R'y_2$$

It should be noted that the elements of Z, Z_1 and Z_2 being linear functions of y, y_1 and y_2 respectively, are independently normally distributed with means given by $\bar{Z}$, $\bar{Z}_1$ and $\bar{Z}_2$, and covariance matrices I, respectively. (See Nagar and Ullah, 1974).

Let Z_i, Z_{1i} and Z_{2i} be the i^{th} element of Z, Z_1 and Z_2 respectively. Similarly, let $\bar{Z}_i$, $\bar{Z}_{1i}$ and $\bar{Z}_{2i}$ be the i^{th} element of $\bar{Z}$, $\bar{Z}_1$, and $\bar{Z}_2$ respectively. $(i = 1,2,\ldots, K_2)$ Then

$$(5.5.24)\qquad E(Z_i) = \bar{Z}_i \ ; \quad E(Z_i^2) = 1 + \bar{Z}_i^2$$

$$E(Z_{1i}) = \bar{Z}_{1i} \ ; \quad E(Z_{1i}^2) = 1 + \bar{Z}_{1i}^2$$

$$E(Z_{2i}) = \bar{Z}_{2i} \ ; \quad E(Z_{2i}^2) = 1 + \bar{Z}_{2i}^2$$

The means of $\hat{\beta}_1^*$ and $\hat{\beta}_2^*$ [as obtained by Nagar and Ullah (1974)] using (5.5.21) are given by

$$(5.5.25)\qquad E(\hat{\beta}_1^*) = E\left[\frac{(Z_2'Z_2)Z_1'}{D}\right]E(Z) - E\left[\frac{(Z_1'Z_2)Z_2'}{D}\right]E(Z)$$

$$E(\hat{\beta}_2^*) = E\left[\frac{(Z_1'Z_1)Z_2}{D}\right]E(Z) - E\left[\frac{(Z_1'Z_2)Z_1'}{D}\right]E(Z)$$

Let the elements of Z, Z_1 and Z_2 are stochastically independent, (5.5.25) can be written as:

$$(5.5.26)\qquad E(\hat{\beta}_1^*) = \sum_{i\neq j}^{K_2} E\left(\frac{Z_{2i}^2 Z_{1j}}{D}\right)\bar{Z}_j - \sum_{i\neq j}^{K_2} E\left(\frac{Z_{1i}Z_{2i}Z_{2j}}{D}\right)\bar{Z}_j$$

and

$$(5.5.27)\qquad E(\hat{\beta}_2^*) = \sum_{i\neq j}^{K_2} E\left(\frac{Z_{1i}^2 Z_{2j}}{D}\right)\bar{Z}_j - \sum_{i\neq j}^{K_2} E\left(\frac{Z_{1i}Z_{2i}Z_{1j}}{D}\right)\bar{Z}_j$$

Using the relations (5.5.18) and (5.5.19), we can write the mean value of the (2SLS) estimators $\hat{\beta}_1^*$ and $\hat{\beta}_2^*$ as:

$$(5.5.28)\qquad E(\hat{\beta}_1) = [t_{11}\; E\hat{\beta}_1^* + t_{12}]t_{22}^{-1}$$

and

$$(5.5.29)\qquad E(\hat{\beta}_2) = [t_{11}\; E(\hat{\beta}_1^* + t_{13}) - t_{23}\; E(\hat{\beta}_1)]t_{33}^{-1}$$

The relations (5.5.28) and (5.5.29) demonstrate that the mean value of the (2SLS) estimator depends on the variance - covariance structure of the endogenous variables.

As pointed out by Nagar and Ullah (1974)*, the expression $D = (Z_1'Z_1)(Z_2'Z_2) - (Z_1'Z_2)^2$ in (5.5.22) is the determinant of the matrix (generalized variance) given by

$$(5.5.30)\qquad \begin{bmatrix} Z_1'Z_1 & Z_1'Z_2 \\ Z_2'Z_1 & Z_2'Z_2 \end{bmatrix}$$

where the elements of (5.5.30) are distributed according to the non-central Wishart distribution with the matrix of noncentrality parameters given by

$$(5.5.31) \qquad Q^* = \begin{bmatrix} q_1 & q_2 \\ & \\ q_2 & q_1' \end{bmatrix}$$

where

$$q_1 = \sum_{i=1}^{K_2} \overline{Z}_{1i} \; ; \quad q_1' = \frac{\sum_{i=1}^{K_2} \overline{Z}_{2i}^2}{2} \; , \quad q_2 = \frac{\sum_{i=1}^{K_2} \overline{Z}_{1i}\overline{Z}_{2i}}{2}$$

and means sigma matrix I and K_2 degrees of freedom.

Recall that

$$Z_1 = \begin{bmatrix} Z_{11} \\ \vdots \\ Z_{1K_2} \end{bmatrix} \quad \text{and} \quad Z_2 = \begin{bmatrix} Z_{21} \\ \\ Z_{2K_2} \end{bmatrix}$$

are $K_2 \times 1$ column vectors.

After computing the expectation of the generalized variance (see 5.5.30) and its partials derivatives with respect to the noncentrality parameters, the *following results are obtained:

$$(5.5.32) \qquad E(\hat{\beta}_1^*) = 4\beta_1^* h_2 \Lambda_3 + \overline{Z}'\overline{Z}_1[(K_2-1)\Lambda_1 + 2h_2 \Lambda_2]$$

$$(5.5.33) \qquad E(\hat{\beta}_2^*) = 4\beta_2^* h_2 \Lambda_3 + \overline{Z}'\overline{Z}_2[(K_2-1)\Lambda_1 + 2h_2 \Lambda_2]$$

where

$$\Lambda_1 = p_{0,2}^* - h_1 p_{1,1}^* - \frac{3}{2} p_{1,0}^* - h_2 p_{2,0}^*$$

(5.5.34)

$$\Lambda_2 = h_1 p_{2,1}^* - p_{1,2}^* + \frac{5}{2} f_{2,0}^* - h_2 \hat{p}_{3,0}^*$$

---

* See Appendix A and B of Nagar and Ullah (1974).

$$\Delta_3 = p^*_{0,3} - h_1 p^*_{1,2} + h_2 p^*_{2,1} - \frac{5}{2} p^*_{1,1}$$

(5.5.35) $$p^*_{i,j} = \sum_{r=0}^{\infty} \frac{(-h_2)}{r!} g^*_{r+i,j} \ ;$$

and

$$g^*_{r+i,j} = \frac{(K_{2/2} - 1 + r + i)_j \ (K_{2/2} - 1)_{r+i} (1)_{r+i} \ e^{-h_1}}{(K_2 - 2)(K_2 - 3)(K_{2/2} + 2(r+i))_j [(K_2 - 1)/2]_{r+i}}$$

times $${}_1F_1\left(\frac{K_2}{2} - 1 + r + i + j;\ \frac{K_2}{2} + 2(r+i) + j;\ h_1\right)$$

and

$$h_1 = \frac{(\overline{Z}_1' Z_1 + \overline{Z}_2' \overline{Z}_2)}{2}$$

$$h_2 = \frac{1}{4} [(\overline{Z}_1' \overline{Z}_1)(\overline{Z}_2' \overline{Z}_2) - (\overline{Z}_1' \overline{Z}_2)^2$$

also $(a)_n = a(a+1), \ldots (a+n-1)$.

CHAPTER 6

PROPERTIES OF ESTIMATORS THROUGH ASYMPTOTIC EXPANSIONS

1. LARGE-SAMPLE APPROXIMATIONS

In this section we consider the approach pioneered mainly by Nagar [1959, 1962] in which the sampling error of an estimator is expressed as the sum of an infinite series of random variables, successive terms of which are of decreasing order of sample size in probability. The claim is then made that large sample properties of the estimator under consideration can be approximated (to the desired order of sample size) by the properties of the first few terms of the infinite series.

As in chapter 4, let the structural equation of interest be

$$(6.1.1) \qquad y = Y\beta + X_1\gamma + u$$

where y is a column vector of T observations on the jointly dependent variable, Y is a $T \times m_1$ matrix of observations on the m_1 explanatory jointly dependent variables, X_1 is a $T \times K_1$ matrix of observations on the K_1 explanatory predetermined variables, u is the $T \times 1$ disturbance vector and β and γ are unknown vector parameters.

$$(6.1.2) \qquad X = [X_1, X_2]$$

is a $T \times K$ matrix of all the predetermined variables in the system, while X_2 is a $T \times K_2$ matrix of predetermined varaibles which occur in the system but do not appear in (6.1.1).

From (6.1.1) the reduced form corresponding to the Y's on the right-hand side (R.H.S) is given by

$$(6.1.3) \qquad Y = X\Pi + V = X_1\Pi_1 + X_2\Pi_2 + V$$

$$= \bar{Y} + V$$

and the (OLS) estimate of V is given by

(6.1.4) $\hat{V} = Y - X(X'X)^{-1}X'Y = P^{\circ}Y$ (say)

The general k-class estimator of $\begin{bmatrix}\beta\\ \gamma\end{bmatrix}$ is given by

(6.1.5) $$\begin{bmatrix} Y'Y - k\,\hat{V}'\hat{V} & Y'X_1 \\ X_1'Y & X_1'X_1 \end{bmatrix} \begin{bmatrix} \hat{\beta}(k) \\ \hat{\gamma}(k) \end{bmatrix} = \begin{bmatrix} Y'-k\hat{V}' \\ X_1' \end{bmatrix} y$$

[For k=0, k=1, we obtain the (OLS) and (2SLS) estimators of β and γ respectively.]

We reintroduce some of the notations used in chapter 4 before stating the results

(6.1.6) $V = ur' + W$

where r is a column vector of m_1 constants and W is a matrix with elements distributed normally but indepentently of u and with mean zero.

(6.1.7) $u_e = u - [Y, X_1]\, e_k$

where u_e is the estimated disturbance vector, and e_k is the sampling error,

$$e_k = \begin{bmatrix} \hat{\beta}(k) \\ \gamma(k) \end{bmatrix} - \begin{bmatrix} \beta \\ \gamma \end{bmatrix}$$

i.e.,

(6.1.8) $$e_k = \begin{bmatrix} Y'Y-k\hat{V}'\hat{V} & Y'X_1 \\ X_1'Y & X_1'X_1 \end{bmatrix}^{-1} \begin{bmatrix} (Y-k\hat{V})' \\ X_1' \end{bmatrix} u$$

Similarly, the sampling error of the (2SLS) estimator is given by e_1 where e_1 is given by

$$(6.1.9) \quad \begin{bmatrix} Y'Y - \hat{V}\hat{V} & Y'X_1 \\ X_1'Y & X_1'X_1 \end{bmatrix}^{-1} \begin{bmatrix} (Y - \hat{V})' \\ X_1' \end{bmatrix} u$$

Theorem 6.1.(A)

The bias to the order T^{-1} (T denotes sample size) of the estimator

$$\begin{bmatrix} \hat{\beta}(k) \\ \hat{\gamma}(k) \end{bmatrix} \text{ of } \begin{bmatrix} \beta \\ \gamma \end{bmatrix}$$

of (6.1.1) is given by

$$(6.1.10) \quad E(e_k) = [-h + L - 1]Qq$$

where

$$L = K - [m_1 + K_1]$$

$$(6.1.11) \quad k = 1 + \frac{h}{T}$$

where h is a nonstochastic number independent of T. [See Nagar, (1959)]

$$(6.1.12) \quad q = \frac{1}{T}\begin{bmatrix} E(V'u) \\ 0 \end{bmatrix} = \sigma^2 \begin{bmatrix} r \\ 0 \end{bmatrix}$$

σ^2 being the variance of the disturbance of (6.1.1)

$$(6.1.13) \quad Q = \begin{bmatrix} \bar{Y}'\bar{Y} & \bar{Y}'X_1 \\ X_1'\bar{Y} & X_1'X_1 \end{bmatrix}^{-1} = (A_1'A_1)^{-1}$$

where $\bar{Y}$ is as defined in (6.1.3) and

$$(6.1.14) \quad A_1 = [\bar{Y}, X_1] = [X\Pi, X_1]$$

Corollary 6.1.A

The bias of the (2SLS) estimator of $\begin{bmatrix}\beta\\ \gamma\end{bmatrix}$ to be order T^{-1} is given by

$$E(e_1) = (L-1)Qq \tag{6.1.15}$$

Corollary 6.1.B

The bias vanishes for

$$k = 1 + \frac{L-1}{T} \tag{6.1.16}$$

which provides, to the order T^{-1}, an unbiased estimator of $\begin{bmatrix}\beta\\ \gamma\end{bmatrix}$.

From (6.1.6) we have

$$\frac{1}{T} E(V'V) = \sigma^2 rr^1 + \frac{1}{T} E(W'W) \tag{6.1.17}$$

For convenience, we reintroduce once more the notation uses in chapter 4.

$$M^* = X(X'X)^{-1}X' \tag{6.1.18}$$

$$d = (\bar{Y} M_1 \bar{Y})^{-1} \bar{Y} M_1 \text{ , where} \tag{6.1.19}$$

$$M_1 = I - X_1(X_1'X_1)^{-1}X_1'$$

$$B = A_1 \; Q \; A_1 = A_1(A_1'A_1)A_1 \tag{6.1.20}$$

$$V_* = [V,0] \tag{6.1.21}$$

$$C_1 = \begin{bmatrix} \sigma^2 \; r'r & 0 \\ 0 & 0 \end{bmatrix} \tag{6.1.22}$$

$$C_2 = \begin{bmatrix} \frac{1}{T} E(W'W) & 0 \\ 0 & 0 \end{bmatrix} \tag{6.1.22}$$

$$C = \begin{bmatrix} \frac{1}{T} E(V'V) & 0 \\ 0 & 0 \end{bmatrix} = C_1 + C_2 \tag{6.1.23}$$

Theorem 6.1.B

The moment matrix, to the order T^{-2}, of the estimator $\begin{bmatrix}\hat{\beta}(k)\\ \hat{\gamma}(k)\end{bmatrix}$ around the paramater vector $\begin{bmatrix}\beta\\ \gamma\end{bmatrix}$ is given by:

(6.1.24) $$E(e_k e_k') = \sigma^2 Q[I+R^*]$$

where R^* is a matrix of order T^{-1}:

(6.1.25) $$R^* = [(2h-2L+3)tr(C_1Q) + tr(C_2Q) \cdot I + \{(h-L+2)^2 + 2(h+1)\} C_1Q + (2h-L+2) C_2Q$$

Corollary (6.1.C):

The moment matrix, to the order T^{-2}, of the (2SLS) estimator $\begin{bmatrix}\hat{\beta}\\ \hat{\gamma}\end{bmatrix}$ vector $\begin{bmatrix}\beta\\ \gamma\end{bmatrix}$ is given by (6.1.24) where R^* is now defined as:

(6.1.26) $$R^* = [-(2L-3)tr(C_1Q) + tr(C_2Q)] \cdot I + \{(L-2)^2+2\} C_1Q - (L-2) C_2Q$$

As pointed out by Nagar (1959), for the choice of the "best k" we minimize

(6.1.27) $$|E(e_k e_k')| = \sigma^2 |Q| \cdot |I+R^*| = \sigma^2 |Q| \cdot (1+trR^*)$$

to the order T^{-2}, for variations in k or h. The h-value which minimizes the determinant value of the moment matrix (6.1.24) is given by

(6.1.23) $$\dot{h} = K - 2(M_1+K_1) - 3 - \frac{tr(C_2Q)}{tr(C_1Q)}$$

The proofs of Theorems (6.1:A) and (6.1.B) utilize the following procedure:

Using the relations (6.1.3), (6.1.14), (6.1.18) and (6.1.21) we can write

(6.1.29) $$e_k = [Q^{-1} + A_1 V_* + V_*' A_1 + (1-k)V_*' + kV_*' M^* V_*]^{-1}$$

times $$[A_1' + (1-k)V_*' + kV_*' M^*]U$$

$$= [I + Q\{A_1' V_* + V_*' A_1 + (1-k)V_*' V_* + kV_*' M^* V_*\}]^{-1} Q$$

times $$[A_1' U + (1-k)V_*' u + kV_*' M^* u]$$

$$= Q[A_1' u + (1-k)V_*' u + kV_*' M^* u]$$

$$- Q[A_1' V_* + V_*' A_1 + (1-k)V_*' V_* + kV_*' M^* V_*]Q$$

times $$[A_1' u + (1-k)V_*' u + kV_*' M^* u] + \ldots\ldots$$

Using the relation $k = 1 + \frac{h}{T}$ and neglecting terms of higher order of smallness then T^{-1}, we obtain

(6.1.30) $$e_k = Q[A_1' u - \frac{1}{T} hV_*' u + V_*' M^* U]$$

$$- QA_1' V_* QA_1' U - QV_*' A_1 QA_1' U.$$

The preceding derivation assumes the validity of the expansion

(6.1.31) $$[I + Q\nabla]^{-1}$$

where

(6.1.32) $$\nabla = A_1' V_* + V_*' A_1' + (1-k)V_*' V_* + k(V_*' M^* V_*)$$

(As $Q\nabla$ is of order $T^{-\frac{1}{2}}$ in probability, we find that the successive terms of the expansion are of decreasing order in T.) [For details of proofs of Theorems (6.1.A) and (6.1.B), see Nagar (1959).]

Srinivasan (1970) questioned the validity of Nagar's approach, and through a series of contrived examples demonstrated that this approach can be misleading in that (a) it can yield an estimate for finite sample bias (to the specified order of sample size) that differs from the true finite-sample bias; (b) it may suggest that the bias is infinite while the true value is finite; (c) it may result in finite valued expressions for bias while moments of the exact sampling distribution are infinite. This does not suggest that the results obtained by Nagar are nessarily invalid, only that further investigation is needed to establish the validity of the procedure.

2.

Mikhail (1972) on the findings of the Monte Carlo experiments argued that Nargar's approximation of the bias of the 2SLS estimator would be better approximated to the order T^{-2} instead of to the order T^{-1}.

The 2SLS bias to the order T^{-2}

From (6.1.9) e_1 could be expressed as

(6.2.1) $$e_1 = [Q^{-1} + A_1'V_* + V_*'A_1 + V_*'M^*V_*]^{-1}$$

times $$[A_1'u + V_*'M^*U]$$

$$= [I + Q\{A_1'V_* + V_*'A_1 + V_*'M^*V\}]^{-1}$$

times $$Q[A_1'u + V_*'M^*u]$$

$$= [I - Q\nabla + Q\nabla Q\nabla - Q\nabla Q\nabla Q\nabla + \ldots\ldots]$$

times $$Q[A_1'u + V_*'M^*u],$$

where ∇ is as defined in (6.1.32) with $k = 1$ and ∇ is the symmetric matrix of $O(T^{-\frac{1}{2}})$.

Taking expectations, and leaving out terms which involve odd moments, we get:

(6.2.2) $$\begin{aligned} E(e_1) = E\,[&QA_1'u + QV_*'M^*u - QA_1'V_*QA_1'u - QV_*'A_1QA_1'u] \\ &- QV_*'M^*V_*QV_*'M^*u + QA_1'V_*QV_*'M^*V_*QA_1'u \\ &+ QV_*'A_1QV_*'M^*V_*QA'u + QV_*'M^*V_*QA_1'V_*QA_1'u \\ &+ QV_*'M^*V_*QV_*'A_1QA_1'u + QA_1'V_*QA_1'V_*QV_*'M^*u \\ &+ QA_1'V_*QV_*'A_1QV_*'M^*u + QV_*'AQA_1'V_*QV_*'M^*u \\ &+ QV_*'A_1QV_*'A_1QV_*'M^*u - QA_1'V_*QA_1'V_*QA_1'V_*QA_1'u \\ &- QA_1'V_*QA_1'V_*QV_*'A_1QA_1'u - QA_1'V_*QV_*'A_1QA_1'V_*QA_1'u \\ &- QA_1'V_*QV_*'A_1QV_*'A_1QA_1'u - QV_*'A_1QA_1'V_*QA_1'V_*QA_1'u \\ &- QV_*'A_1QA_1'V_*QV_*'A_1QA_1'u - QV_*'A_1QV_*'A_1QA_1'V_*QA_1'u \\ &- QV_*'A_1QV_*'A_1QV_*'A_1QA_1'u\} + O(T^{-2}) \end{aligned}$$

After taking the expected values of all the terms in (6.22) [see Mikhail, (1972) for details], we obtain

(6.2.3) $$\begin{aligned} E(e_1) = (L-1)Qq &+ \{[L(4-L)-1]\,(\mathrm{tr}\;Q\;C_1)I \\ &+ (L-1)\,(\mathrm{tr}\;Q\;C_2)I - 2QC_1 \end{aligned}$$

$$+ [L(3-L)-2]Q\,C_2\}Qq$$

where the first term on the (R.H.S.) of (6.2.3) is Nagar's expression for the expectation of the first four terms in square brackets of the expression (6.2.2). L is the same as defined in (6.1.10).

From the relations $C = C_1 + C_2$, (6.1.22) and (6.1.23)

$$QCQq = QC_1Qq + QC_2Qq$$

$$= \frac{1}{\sigma^2} Qqq'Qq + QC_2Qq$$

$$= tr(QC_1)Qq + QC_2Qq$$

Thus, the expression (6.2.3) simplifies to

$$(6.2.4) \qquad E(e_1) = (L-1)Qq + (L-1)tr(QC)Qq$$

$$- (L-1)(L-2)QCQq$$

or

$$(6.2.5) \qquad E(e_1) = (L-1)[I + tr(QC)I - (L-2)QC]Qq$$

As stated in chapter 2, the double-k-class estimator $d = \begin{bmatrix}\hat{\beta}\\ \hat{\gamma}\end{bmatrix}$ of $\delta = \begin{bmatrix}\beta\\ \gamma\end{bmatrix}$ of (6.1.1) is given by

$$(6.2.6) \qquad \begin{bmatrix} Y' - k_2\hat{V}' \\ X_1' \end{bmatrix} y = \begin{bmatrix} Y'Y - k_1\hat{V}'\hat{V} & Y'X_1 \\ X_1'Y & X_1'X_1 \end{bmatrix} d$$

where k_1 and k_2 are two arbitary real numbers

Following Nagar (1962), the scalars k_1 and k_2 are non-stochastic and they differ from 1 to the order T^{-1}.

That is, we can write

$$(6.2.7) \qquad k_1 = 1 + \frac{h_1}{T}, \quad k_2 = 1 + \frac{h_2}{T}$$

(neglecting terms of higher order of smallness of T^{-1}.) Both h_1 and h_2 are non-stochastic real numbers, independent of T

If we rewrite equation (6.1.1) as

$$(6.2.8) \qquad y = (Y - \frac{k_1-k_2}{1-k_2}\hat{V})\beta + X_1\gamma + u^*$$

$$= (Y - \frac{k_1-k_2}{1-k_2}\hat{V}X_1)\delta + u^*$$

[In (6.2.8), the case $k_2 = 1$ is treated as the limit for $k_2 \to 1$.].

$$(6.2.9) \qquad \text{where } \delta = \begin{bmatrix}\beta\\ \gamma\end{bmatrix}, \quad u^* = u + \frac{k_1-k_2}{1-k_2}\hat{V}\beta$$

$$= u + \frac{k_1-k_2}{1-k_2}p^0V_*\delta$$

where p^0, $\hat{V}$ are defined in (6.1.4) and V_* is defined in (6.1.21).

Then combining (6.2.8) with (6.2.6), we get the sampling error of the double-k-class estimator denoted by e, where

$$(6.2.10) \qquad e = d - \delta = \begin{bmatrix} Y'Y - k_1\hat{V}'\hat{V} & Y'X_1 \\ X_1'Y & X_1'X_1 \end{bmatrix}^{-1} \begin{bmatrix} Y' - k_2\hat{V}' \\ X_1' \end{bmatrix} u^*$$

As was shown in (6.1.29), we can obtain a similar expansion of e i.e.,

$$(6.2.11) \qquad e = [Q^{-1} + \{A_1'V_* + V_*'A_1 + (1-k_1)V_*'V_* + k_1V_*'M^*V_*\}]^{-1}$$

times $\qquad [A_1' + (1-k_2)V_*' + k_2V_*'M^*]U^*$

$$= [I + Q\ A_1'V_* + V_*'A_1 + (1-k_1)V_*'V_* + k_1V_*'M^*V_*\ ]^{-1}Q$$

times $\qquad [A_1'U^* + (1-k_2)V_*'U^* + k_2V_*'M^*U^*]$

Neglecting terms of higher order of smallness than T^{-1} and using the relation (6.2.7), we got

$$(6.2.12) \qquad e = Q[A_1'U^* - \frac{h_2}{T}V_*'M^*U^* - A_1'V_*QA_1'U^* - V_*'A_1QA_1'U^*]$$

where it should be noted that $A_1'U^* = A'U$, $V_*'M^*U^* = V_*'M^*U$.

To obtain $E(e)$ from (6.2.12), write

$$U^* = U + \frac{h_1-h_2}{h_2}V_*\delta - \frac{h_2-h_1}{h_2}M^*V_*\delta$$

Theorem 6.2.A

The bias to the order T^{-1} of the double-k-class estimator d of the parameter vector δ in (6.1.1) is given by

$$(6.2.13) \qquad E(e) = (-h_2+L-1)Qq - (h_1-h_2)QC\delta$$

where e is the sampling error $e = d - \delta$.

Theorem 6.2.B

The moment matrix, to the order T^{-2}, of the double-k-class estimator d around the parameter vector δ is given by

$$(6.2.14)\qquad E(ee') = \sigma^2[(1+\bar{a}_0)Q + \bar{a}_1QC_1Q + a_2QC_2Q + \bar{a}_3(Qqn' + nq'Q) + \bar{a}_4nn']$$

where

$$(6.2.15)\qquad a_0 = -2(h_1-h_2)\frac{n'q}{\sigma^2} - [-h_2 + 2L-3 + 2(h_1-h_2)\frac{\delta'q}{\sigma^2}] \text{ times trace } C_1Q + \text{trace } C_2Q.$$

$$a_1 = 2(h_1+1) + [(h_1-h_2)\frac{\delta'q}{\sigma^2} - h_2 + L-2]^2$$

$$a_2 = 2h_1 - L + 2$$

$$a_3 = \frac{h_1-h_2}{\sigma^2}[(h_1-h_2)\frac{\delta'q}{\sigma^2} - h_2 + L-2]$$

$$a_4 = (\frac{h_1-h_2}{\sigma^2})^2$$

$$n = QC_2\delta$$

In the proofs of Theorems (6.2.A) and (6.2.B) ∇ defined in (6.1.32) is now defined with $k = k_1$.

For Theorem (6.2.8), the expression ee' is given by

$$(6.2.16)\qquad ee' = [I + Q\nabla]^{-1}\, QSQ\, [I + Q\nabla]^{-1}$$

where

$$(6.2.17)\qquad S = [A_1'U + (1-k_2)V_*'U^* + k_2V_*'M^*U]$$

$$\text{times}\qquad [U'A_1 + (1-k_2)(U^*)'V_* + k_2U'M^*V_*]$$

[See Nagar, (1962) for details of proofs.]

From the double-k-class estimator we can derive the h-class estimators of (6.1.1) by setting

$$(6.2.18) \qquad k_1 = 1-h^2, \quad k_2 = 1-h.$$

It should be noted that the h used in (6.2.18) has no relationship to the h used in (6.1.1) and (6.2.7).

Further, as for k_1 and k_2, we assume that h is non-stochastic and that it differs from 0, to the order T^{-1}, i.e.,

$$(6.2.19) \qquad h = \frac{\alpha^*}{T} \quad , \quad \alpha^*$$

being non-stochastic real number independent of T. Therefore, from

$$(6.2.20) \qquad k_1 = 1 + \frac{h_1}{T} = 1 - h_2 = 1 - \frac{\alpha^{*2}}{T^2} \quad ,$$

we obtain $h_1 = \frac{-\alpha^{*2}}{T}$ which is zero to Nagar's order of approximation, and, from

$$k_2 = 1 + \frac{h_2}{T} = 1 - h = 1 - \frac{\alpha^*}{T} \quad ,$$

we have

$$h_2 = -\alpha^*$$

Using the relations obtained in (6.2.20) and Theorem (6.2.A), we obtain the bias, to the order T^{-1}, of the h-class estimator, i.e.,

$$(6.2.21) \qquad E(e_h) = (\alpha^* + L-1)Qq + \alpha^*QC\delta.$$

where e_h is the sampling error of the h-class estimator.

The moment matrix, to the order T^{-2}, of the h-class estimator is given by

$$(6.2.22)\qquad E[e_h e_h'] = \sigma^2[(1+b_0)Q + b_1 QC_1Q + b_2QC_2Q] + b_3(Qqn'+nq'Q) + b_4nn']$$

where

$$(6.2.23)\qquad b_0 = \frac{-2\alpha^* q'n}{\sigma^2} - (2\alpha^*+2L-3+2\alpha^* \frac{\delta' q}{\sigma^2}) \text{ trace } C_1Q + \text{trace } C_2Q$$

$$b_1 = 2 + (\alpha^* \frac{\delta' q}{\sigma^2} + \alpha^* + L-2)^2$$

$$b_2 = -(L-2)$$

$$b_3 = \frac{\alpha^*}{\sigma^2} (\alpha^* \frac{\delta' q}{\sigma^2} + \alpha^* + L-2)$$

$$b_4 = \frac{\alpha^{*2}}{\sigma^2}$$

$$n = QC_2\delta.$$

3.

In this section we look at the small-disturbance asymptotic moment matrix of k-class estimates of parameters in different equations. Knowledge of the cross-section moment matrices is useful for problems in which the properties of linear combinations of parameters in different equations of the system are of interest, e.g., when studying the significance of a linear combination (sum or difference, say) of coefficients in different equations.

The results given in this section will be mainly those obtained by

G.F. Brown (1974). The results obtained by Theil (1970) who computed the large-sample limiting covariance matrix of the (2SLS) estimates of coefficients in two different equations, and of Nagar and Gupta (1970), who derived the large-sample asymptotic moment matrix to the order T^{-2}, are derived as the limit as $T\to\infty$ of the small-disturbance expressions obtained by G.F. Brown (1974) and Kadane (1971).

The procedure used in deriving the cross-section moment matrix is based upon the small-disturbance methods employed by Kadane (1971). A discussion of Kadane's methods was outlined in chapter 4. The approximating random variables employed here are derived from a Taylor series expansion in the powers of a (small) scalar multiple σ of the structural disturbances of the system.

We reintroduce the structural model of the complete system of T observations on M linear structrual equations in M current endogenous variables and K exogenous (predetermined) variables. Let the model be denoted by

$$(6.3.1) \qquad YB' + X\Gamma' = \sigma U,$$

where σ is a small positive scalar. As was stated previously, the reduced form of (6.3.1) is given by

$$(6.3.2) \qquad Y = X\Pi' + \sigma V.$$

Then let us denote the i^{th} structural equation of the system in (6.3.1) by

$$(6.3.3) \qquad y_i = Y_i\gamma_i + X_i\gamma_i + \sigma u_i$$

where

y_i and u_i are $(T\times 1)$

Y_i is $(T\times M_i)$; X_i is $(T\times K_i)$

β_i is $(M_i\times 1)$ and γ_i is $(K_i\times 1)$

It follows from (6.3.3) that Y_i and X_i are submatrices of Y and X, respectively. Following Nagar and Gupta (1970), we write

(6.3.4) $\quad Y_i = YP_{1i}\ ; \quad X_i = XP_{2i}$

where P_{1i} is $(M\times M_i)$ and P_{2i} is $(K\times K_i)$. The elements of P_{1i} and P_{2i} are either zero or one, and are given by the structure of the model.

It should be noted that

(6.3.5)
$$
\begin{aligned}
[Y_i, X_i] &= [YP_{1i}, XP_{2i}] \\
&= [X\Pi' P_{1i}, XP_{2i}] + \sigma U[(B^{-1})' P_{1i}, 0] \\
&= X[\Pi' P_{1i}, P_{2i}] + \sigma U[(B^{-1})' P_{1i}, 0] \\
&\overset{\text{def}}{=} XC_i + \sigma UD_i \\
&\overset{\text{def}}{=} Z_i + \sigma V_i
\end{aligned}
$$

where

(6.3.6)
$$
\begin{aligned}
C_i &= [\Pi' P_{1i}, P_{2i}], \quad D_i = [(B^{-1})' P_{1i}, 0] \\
&= [-\Gamma'(B^{-1})' P_{1i}, P_{2i}]
\end{aligned}
$$

[from 6.3.1 and 6.3.2 $\Pi' = -\Gamma'(B^{-1})'$, $V = U(B^{-1})'$]

As in Nagar and Gupta (1970), for arbitrary i and j define:

(6.3.7)
$$
\begin{aligned}
\Omega_{ij} &= D_i' \Sigma D_{ji} \\
q_{ij} &= D_i' \sigma_j
\end{aligned}
$$

where

$$(6.3.8) \qquad \Sigma = \frac{1}{T} E(U'U) = [\sigma_1, \sigma_2, \ldots, \sigma_M]$$

$$= \begin{bmatrix} \sigma_{11} & \sigma_{12} & & \sigma_{1M} \\ \sigma_{21} & \sigma_{22} & & \sigma_{2M} \\ \vdots & & & \\ \sigma_{M1} & \sigma_{M2} & & \sigma_{MM} \end{bmatrix}$$

Denote the k-class estimator of the parameters $\begin{bmatrix}\beta_i \\ \gamma_i\end{bmatrix}$ in (6.3.3) by $\begin{bmatrix}\beta_i \\ \gamma_i\end{bmatrix}_{k_i}$

where

$$(6.2.9) \qquad \begin{bmatrix}\beta_i \\ \gamma_i\end{bmatrix}_{k_i} = \{[Y_i, X_i]'(I - k_i \overline{P}_X)[Y_i, X_i]\}^{-1}$$

times $\qquad [Y_i, X_i]'(I - k_i \overline{P}_X) y_i$

where

$$\overline{P}_X = I - X(X'X)^{-1}X'$$

The sampling error of the k-class estimator of (6.3.3) is given by

$$(6.3.10) \qquad e_i = \begin{bmatrix}\beta_i \\ \gamma_i\end{bmatrix}_{k_i} - \begin{bmatrix}\beta_i \\ \gamma_i\end{bmatrix}$$

are closely approximated in probability by the random variables $e_i(\cdot)$.

From Kadane (1971)

$$(6.3.11) \qquad e_i(\cdot) = \sigma Q_i C_i' X' U_i + \sigma^2 \{Q_i D_i' U' (I - k_i \overline{P}_X) u_i$$

$$- Q_i < C_i' X' U D_i > Q_i C_i' X' u_i\}$$

$$+ \sigma^3 \{Q_i < C_i' X' U D_i > Q_i < C_i' X' U D_i > Q_i C_i' X' u_i$$

$$- Q_i\langle C_i'X'UD_i\rangle Q_iD_i'U'(I-k_i\bar{P}_x)u_i$$

$$- Q_iD_i'U'(I-k_i\bar{P}_x)UD_iQ_iC_i'X'U_i\} + O_p(\sigma^4)$$

In (6.3.11), O_p denotes order in probability,

$$(6.3.12) \qquad Q_i = (Z_i'Z_i)^{-1} = (C_iX'XC_i)^{-1}$$

$$(6.3.13) \qquad \langle A\rangle = A + A'$$

for any square matrix A.

Before stating the results we introduce the following notation:

$$(6.3.14) \qquad Q_{ij} = Z_i'Z_j = C_i'X'XC_j$$

$$(6.3.15) \qquad W_{ij} = q_{ij}q_{ji}' + \sigma_{ij}\Omega_{ij} \quad \text{(see 6.3.7)}$$

$$(6.3.16) \qquad L_i = (K-K_i) - M_i = K - [M_i+K_i],$$

(the degree of overidentification in the ith equation)

$$(6.3.17) \qquad S_{ki} = (1-k_i)(T-K) + L_i - 1$$

Theorem 6.3.A

If the ith and jth equations are estimated by k-class estimators with fixed parameters k_i and k_j, respectively, then the small-disturbance asymptotic cross-section matrix is given by

$$(6.3.18) \qquad E(e_i(\cdot)e_j(\cdot)']$$

$$= \sigma^2\{\sigma_{ij}Q_iQ_{ij}\} + \sigma^4\{[\sigma_{ij}tr(Q_iQ_{ij}Q_j\Omega_{ji})$$

$$- S_{ki}\, tr(Q_iq_{ii}q_{jj}')$$

$$- S_{k_j} \operatorname{tr}(Q_j q_{jj} q'_{ji})] \, Q_i Q_{ij} Q_j + [\operatorname{tr}(Q_j Q_{ij} Q_j Q'_{ij})$$

$$+ S_{k_i} + S_{k_j}$$

$$- (T-k_i k_j (T-K))] \, Q_i W_{ij} Q_j + [S_{k_i} S_{k_j}] \, Q_i q_{ii} q'_{ji} Q_j$$

$$+ [-S_{k_i}] \, Q_i W_{ii} Q_i Q_{ij} Q_j + [-S_{k_j}] \, Q_i Q_{ij} Q_j W_{jj} Q_j$$

$$+ Q_i W_{ij} Q_j Q'_{ij} Q_i Q_{ij} Q_i + Q_i Q_{ij} Q_j q_{ji} q'_{ij} Q_i Q_{ij} Q_j$$

$$+ Q_i Q_{ij} Q_j Q'_{ij} Q_i W_{ij} Q_j \} + O_p(\sigma^6)$$

The result given in (6.3.18) can be considered as a general case of other specific cases in literature:

(1) The σ^2 order term $\sigma_{ij} Q_j Q_{ij} Q_j$ is the large sample asymptotic covariance matrix given by Theil (1970).

(2) If i=j in (6.3.18) we obtain the result given in Theorem 2 of Kadane (1971).

(3) Substituting $k_i = k_j = 1$ in (6.3.18) gives the result of Nagar and Gupta (1970) found as $T \to \infty$ for (2SLS) estimators.

(4) Letting $k_i = 1 + \frac{h_i}{T}$, $k_j = 1 + \frac{h_j}{T}$ for fixed h_i and h_j, and retaining terms to order T^{-2} extends their large-sample results to all k-class members with parameters of the form considered in Theorem 6.3.A.

[See Brown, Kadane and Ramage (1974) for the derivation of the small-disturbance asymptotic bias and mean-squared error of all double-k-class estimators with parameters either (arbitrarily) fixed or random.]

Theorem (6.3)(B)

If the ith equation is estimated by the k-class estimator with fixed parameters k_i and the jth equation is estimated by Limited Information Maximum Likelihood (LIML), then the asymptotic cross-equation moment matrix is given by

(6.3.19)
$$
\begin{aligned}
E[e_i(\cdot)\, e_j(\cdot)'] = {} & \sigma^2\{\sigma_{ij} Q_i Q_{ij} Q_j\} \\
& + \sigma^4\{[\sigma_{ij}\{tr(Q_i Q_{ij} Q_j \Omega_{ji}) \\
& - S_{k_i}\, tr(Q_i q_{ii} q_{ij}') \\
& + tr(Q_i q_{jj} q_{ji}')]\, Q_i Q_{ij} Q_j + [tr(Q_i Q_{ij} Q_j Q_{ij}') + k_i L_j \\
& - (M_i + K_i) - 2]\, Q_i W_{ij} Q_j + [-S_{k_i}] Q_i q_{ii} q_{jj}' Q_j \\
& + [-S_{k_i}] Q_i W_{ii} Q_i Q_{ij} Q_j + Q_i Q_{ij} Q_j W_{jj} Q_j \\
& + Q_i W_{ij} Q_j Q_{ij}' Q_{ij} Q_j + Q_i Q_{ij} Q_j q_{ij} q_{ji}' Q_i Q_{ij} Q_j \\
& + Q_i Q_{ij} Q_j Q_{ij}' Q_i W_{ij} Q_j \\
& + 2\, \frac{\sigma_{ij}}{\sigma_{ji}} [1 + (M_i + K_i) - k_i L_j - tr(Q_i Q_{ij} Q_j Q_{ij}')]
\end{aligned}
$$

times
$$
Q_i q_{ij} q_{jj}' Q_j
$$
$$
+ [-2]\, \frac{\sigma_{ij}}{\sigma_{ji}}\, Q_i Q_{ij} Q_j Q_{ij}' Q_i q_{jj} q_{jj}' Q_j\} + O_p(\sigma^5)
$$

Theorem (6.3)(C)

If both the ith and jth equation are estimated by LIML, then the asmptotic cross-equation moment matrix is given by:

$$(6.3.20)\quad E[e_i(\cdot)\, e_j(\cdot)^1] = \sigma^2\{\sigma_{ij} Q_i Q_{ij} Q_j\}$$

$$+ \sigma^4 [\sigma_{ij}\ tr(Q_i Q_{ij} Q_j \Omega_{ji}) + tr(Q_i q_{ii} q'_{ij})$$

$$+ tr(Q_j q_{jj} q'_{ji})] Q_i Q_{ij} Q_j + [tr(Q_i Q_{ij} Q_j Q'_{ij})$$

$$- (K+2) + L_i + L_j] Q_i W_{ij} Q_j + [1 - L_i L_j] Q_i q_{ii} q'_{jj} Q_j$$

$$+ Q_i W_{ii} Q_i Q_{ij} Q_j + Q_i Q_{ij} Q_j W_{jj} Q_j$$

$$+ Q_i W_{ij} Q_j Q'_{ij} Q_i Q_{ij} Q_j + Q_i Q_{ij} Q_j q_{ji} q'_{ij} Q_i Q_{ij} Q_j$$

$$+ Q_i Q_{ij} Q_j Q'_{ij} Q_j W_{ij} Q_j$$

$$+ 2\,\frac{\sigma_{ij}}{\sigma_{jj}} [1 - L_j + (\tilde{M}_i + K_i) - tr(Q_i Q_{ij} Q_j Q'_{ij})] Q_i q_{ij} q'_{jj} Q_j$$

$$+ 2\,\frac{\sigma_{ij}}{\sigma_{ii}} [1 - L_i + (M_j + K_j) - tr(Q_i Q_{ij} Q_j Q'_{ij})$$

times $\quad Q_i q_{ii} q'_{ji} Q_j$

$$- 2\,\frac{\sigma_{ij}}{\sigma_{jj}} Q_i Q_{ij} Q_j Q_{ij} Q_i q_{ij} q'_{jj} Q_j$$

$$- 2\,\frac{\sigma_{ij}}{\sigma_{ii}}\ Q_j q_{ii} q'_{ji} Q_j Q'_{ij} Q_i Q_{ij} Q_j$$

$$+ G(i,j)\} + 0_p(\sigma^5).$$

where

$$(6.3.21)\qquad G(i,i) = \left[\frac{L_i(L_i+2)}{T-K-2}\right] \sigma_{ii} Q_i \Omega_{ii} Q_i$$

$$+ \left[L_i(L_i+2) - \frac{L_i(L_i+2)}{T-K-2}\right] Q_i q_{ii} q_{ii}' Q_i$$

and for $i \neq j$ $\sigma_{ii}\sigma_{ji} - \sigma_{ij}^2 \equiv \Delta \neq 0$,

$$(6.3.22)\qquad G(i\ j) = Q_i q_{ii} q_{jj}' Q_j [L_i L_j \{(-2+\frac{1}{T-K})\ \frac{\sigma_{ii}\sigma_{ij}^2\sigma_{jj}}{\Delta^2}$$

$$+ \left(\frac{\sigma_{ii}\sigma_{jj}}{\Delta}\right)^2 + \frac{\sigma_{ii}\sigma_{ij}\sigma_{jj}}{\Delta} H\}$$

$$+ 2(\mathrm{tr}\ Q_i Q_{ij} Q_j Q_{ij}') - (T-K))$$

times $$\{(-2+\frac{1}{T-K})\ \frac{\sigma_{ij}^4}{\Delta^2} + \frac{\sigma_{ii}\sigma_{ij}^2\sigma_{jj}}{\Delta^2} + \frac{\sigma_{ij}^3}{\Delta} H\}]$$

$$+ Q_i q_{ii} q_{ji}' Q_j [L_i L_j \{\frac{\sigma_{ij}^3\sigma_{jj}}{\Delta^2} - \frac{\sigma_{ii}\sigma_{ij}^2}{\Delta} H$$

$$- \frac{\sigma_{ii}\sigma_{ij}\sigma_{jj}^2}{\Delta^2}\ \frac{1}{T-K}\ \}$$

$$+ 2(\mathrm{tr}\ Q_i Q_{ij} Q_j Q_{ij}') - (T-K))\ \{\frac{\sigma_{ij}^5}{\sigma_{ii}\Delta^2}$$

$$- \frac{\sigma_{ij}^2\sigma_{jj}}{\Delta} H - \frac{\sigma_{ij}^2\sigma_{jj}}{\Delta^2}\ \frac{1}{T-K}\ \}] + Q_i q_{ij} q_{ji}' Q_j$$

times $$-[L_i L_j\ \{- \frac{\sigma_{ii}\sigma_{ij}^2\sigma_{jj}}{\Delta^2} + \frac{\sigma_{ii}\sigma_{ij}\sigma_{jj}}{\Delta} H + \left(\frac{\sigma_{ii}\sigma_{jj}}{\Delta}\right)^2 \frac{1}{T-K}\}$$

$$+ 2(\mathrm{tr}(Q_i Q_{ij} Q_j Q_{ij}') - \{(T-K))\ \frac{\sigma_{ij}^4}{\Delta^2} + \frac{\sigma_{ij}^3}{\Delta} H$$

$$+ \frac{\sigma_{ii}\sigma_{ij}^{2}\sigma_{jj}}{2}\frac{1}{T-K}\}] + Q_i q_{ij} q_{jj} Q_j [L_i L_j \{\frac{\sigma_{ii}\sigma_{ij}^{2}}{\Delta^2}$$

$$- \frac{\sigma_{ij}^{2}\sigma_{jj}}{\Delta} H - \frac{\sigma_{ii}^{2}\sigma_{ij}\sigma_{jj}}{\Delta^2}\frac{1}{T-K}\}$$

$$+ 2(tr(Q_i Q_{ij} Q_j Q_{ij}') - (T-K))\{\frac{\sigma_{ij}^{5}}{\sigma_{ij}\Delta^2} - \frac{\sigma_{ii}\sigma_{ij}^{2}}{\Delta} H$$

$$- \frac{\sigma_{ii}\sigma_{ij}^{2}}{\Delta^2}\frac{1}{T-K}] + Q_i \Omega_{ij} Q_j [H\{\sigma_{ii}\sigma_{jj} L_i L_j$$

$$+ 2\sigma_{ij}^{2}\, tr(Q_i Q_{ij} Q_j Q_{ij}') - 2\sigma_{ij}^{2}(T-K)\}]$$

where H is given by

$$H = \frac{1}{2}\frac{1}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{jj}}}\sum_{q=\frac{T-K}{2}}^{\infty}\frac{1}{q}\frac{\Gamma(q)}{\Gamma(\frac{T-K}{2})\Gamma(q-\frac{T-K}{2}+1)}\left(\frac{1}{2}\right)^{q} \qquad (6.3.23)$$

[See Appendix A, of G.F. Brown (1974) for proofs of Theorems (6.3.A), (6.3.B) and (6.3.C).]

4. ASYMPTOTIC EXPANSIONS OF DISTRIBUTION FUNCTIONS.

In chapter 5 we looked at the exact distributions of estimators in the special case where the structural equation of interest contains two endogenous variables. Unfortunately, the expressions derived for the exact distributions are usually too complicated to interpret meaningfully. Mariano (1973), working with the case of two endogenous variables, derived approximations to the distribution functions of k-class (k non-stochastic) estimators. These approximations have been obtained by expressing the k-class estimators in terms of mutually independent bivariate normal random vectors and then applying Taylor Series expansions to the derived expressions. A similar approach was used by Anderson (1974) and Sargan and Mikhail (1971): Anderson obtained an approximation to the distribution funtion of (LIML) estimators, while Sargan and Mikhail obtained approximations to the distributions of Instrumental Variables (I.V.) estimates.

Anderson and Sawa (1973) used an approach based on the asymptotic expansions of characteristic function to obtain asymptotic expansions for the distributions of k-class estimators. Mariano (1973) holds the sample size fixed as the noncentrality parameter increases, while for Anderson and Sawa (1973), the sample size increases with the noncentrality parameter.

Following Mariano (1973), we characterize the k-class estimators as functions of noncentral Wishart matrices. Let the equation of interest be noted by:

(6.4.1) $$y_1 = Y_1\beta + X_1\gamma + u$$

with $[y, Y_1]$ the $Tx(m_1+1)$ matrix of included endogenous variables, X_1 is the TxK_1 matrix of included predetermined variables, U is the (Tx1) vector of distrubances, and β and γ are vectors of the unknown parameters.

The reduced form of the m_1+1 endogenous variables in (6.4.1) is given by

(6.4.2) $$Y = X\Pi + V = X_1\Pi_1 + X_2\Pi_2 + v$$

where $Y = [y_1, Y_1]$ $X = [X_1, X_2]$ is the TxK matrix of all the predetermined variables, X_2 is the TxK_2 matrix of predetermined variables excluded from (6.4.1), Π_1 is $K_1x(m_1+1)$ and Π_2 is $K_2+(m_1+1)$ derived from Π, the $Kx(m_1+1)$

matrix of reduced form coefficients.

In addition to the usual assumptions we consider the rows of V to be mutually independent and identically distributed as normal random vectors with zero mean and positive definite $(m_1+1) \times (m_1+1)$ covaraiance matrix, Ω.

Under these assumptions, the k-class estimator of β in (6.4.1) simplifies to

$$\hat{\beta}(k) = [Y_1 M_k Y_1]^{-1} (Y_1 M_k y_1) \tag{6.4.3}$$

where

$$M_k = (1-k)\,[I - X_1(X_1'X_1)^{-1}X_1'] + k[X(X'X)^{-1}X' - X_1(X_1'X_1)^{-1}X_1'] \tag{6.4.4}$$

Let

$$A = [Y'M_0Y] = \begin{bmatrix} y_1' \\ Y_1' \end{bmatrix} [I - X_1(X_1'X_1)^{-1}X_1'] \, [y_1, Y_1] \tag{6.4.5}$$

$$= \begin{bmatrix} y_1'M_0y_1 & Y_1'M_0Y_1 \\ Y_1'M_0y_1 & Y_1'M_0Y_1 \end{bmatrix} = \begin{bmatrix} a_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

Let

$$B = [Y'M_1Y] = \begin{bmatrix} y_1'M_1y_1 & y_1'M_1Y_1 \\ Y_1'M_1y_1 & Y_1'M_1Y_1 \end{bmatrix} = \begin{bmatrix} b_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \tag{6.4.6}$$

and let

$$C = kB + (1-k)A = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & C_{22} \end{bmatrix} \tag{6.4.7}$$

From (6.4.5) (6.4.6) and (6.4.7) the (OLS), (2SLS) and k-class estimators of β in (6.4.1) are given by:

$$\tilde{\beta} = (A_{22})^{-1}A_{21} \tag{6.4.8}$$

$$\hat{\beta} = (B_{22})^{-1}B_{21} \tag{6.4.9}$$

and

$$\hat{\beta}(k) = C_{22}^{-1}C_{21}, \tag{6.4.10}$$

respectively.

From our assumptions, it should be noted that the rows of Y are mutually independent random vectors with common covariance matrix Ω. Also, M_0 and M_1 are symmetric indempotent matrices with ranks $T-K_1$ and K_2 respectivley.

Following Mariano (1972, 1973) we have the results:

$$A = [Y'M_0Y] \sim W_{m_1+1}(T-K_1,\Omega;D) \tag{6.4.11}$$

$$B = [Y'M_1Y] \sim W_{m_1+1}(K_2,\Omega;D) \tag{6.4.12}$$

where

$$D = \Pi_2'X_2'[I-X_1(X_1'X_1)^{-1}X_1']X_2\Pi_2, \tag{6.4.13}$$

and for $W = A-B$, W is independent of B,

$$W \sim W_{m_1+1}(T-K,\Omega;0) \tag{6.4.14}$$

and in terms of W and B,

$$C = B + (1-k)W. \tag{6.4.15}$$

To achieve a reduction to canonical form, partition Ω and D as follows:

$$(6.4.16)\qquad \Omega = \begin{bmatrix} \omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix}$$

$$(6.4.17)\qquad D = \begin{bmatrix} d_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix},$$

where Ω_{22} and D_{22} are both $m_1 x m_1$. Let ψ be a $m_1 x m_1$ non-singular matrix such that

$$(6.4.18)\qquad \psi\Omega_{22}\psi' = I,$$

and

$$(6.4.19)\qquad \psi D_{22}\psi' = Q \quad \text{(say)}$$

where Q is a $m_1 x m_1$ diagonal matrix whose main diagonal elements are the characteristic roots of $\Omega_{22}^{-1}D_{22}$ arranged in increasing order. Also let

$$(6.4.20)\qquad \sigma^2 = \omega_{11} - 2\beta'\Omega_{21} + \beta'\Omega_{22}\beta$$

and

$$(6.4.21)\qquad \rho = \frac{\psi}{\sigma}(\Omega_{21}-\Omega_{22}\beta).$$

Then from Mariano (1973)

$$(6.4.22)\qquad \hat{\beta}(k) = \beta + \psi' C_{22}^{*-1} c_{21}$$

where

$$C^* = B^* + (1-k)W^* = \begin{bmatrix} c^*_{11} & c_{12} \\ c_{21} & c^*_{22} \end{bmatrix} \tag{6.4.23}$$

$$B^* \sim W_{m_1+1}\,(K_2,\Omega^*;D^*) \tag{6.4.24}$$

$$W^* \sim W_{m_1+1}\,(T-K,\Omega^*;0) \tag{6.4.25}$$

where

$$\Omega^* = \begin{bmatrix} 1 & \rho' \\ \rho & I \end{bmatrix} \tag{6.4.26}$$

$$D^* = \begin{bmatrix} 0 & 0 \\ 0 & Q \end{bmatrix} \tag{6.4.27}$$

5. For the case of two included endogenous variables β, the coefficients vector, is now a scalar. Hence, the k-class estimator of β is now expressed as

(6.5.1) $$\hat{\beta}(k) \quad \frac{y_2' M_k y_1}{y_2' M_k y_2}$$

Following (6.4.22), the k-class estimator in canonical form is given by

(6.5.2) $$\hat{\beta}(k) = \beta + \frac{\sigma}{\sqrt{\omega_{22}}} \hat{\beta}^*(k) \qquad \text{where}$$

$$\hat{\beta}^*(k) = \frac{c^*_{21}}{c^*_{22}} \quad .$$

Making the corresponding change in notation (for the case of two included endogenous variables) we have

(6.5.3) $$\Omega = \begin{bmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{bmatrix}$$

(6.5.4) $$\rho = \frac{\omega_{21} - \beta\omega_{22}}{\sigma\sqrt{\omega_{22}}}$$

which is the coefficient of correlation between the right hand-side endogenous variable and the disturbance term in the equation being estimated, i.e.

$$y_1 = \beta y_2 + X_1\gamma_1 + u \ .$$

From the relations (6.4.23), (6.4.23) and (6.4.25), we write $\hat{\beta}^*(k)$ defined in (6.5.2) as

$$(6.5.5) \qquad \hat{\beta}^*(k) = \frac{\sum_{i=1}^{r} x_i^* y_i^* + (1-k) \sum_{j=1}^{s} u_j^* v_j^*}{\sum_{i=1}^{r} y_i^{*2} + (1-k) \sum_{j=1}^{r} v_j^{*2}}$$

for $r = K_2$ and $s = T - K$ where the (x_i^*, y_i^*)'s and (u_j^*, v_j^*)'s are mutually independent bivariate normal with common covariance matrix.

$$(6.5.6) \qquad \Omega^* = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

and zero mean vectors except for (x_i^*, y_i^*) where

$$(6.5.7) \qquad E(x_i^*, y_i^*) = \begin{cases} (0,0), & i = 1,2,\ldots,T-1 \\ (0,\mu); & i = r, \end{cases}$$

Let $\phi(x)$ and $\Phi(x)$ denote the standard normal density and distribution functions evaluated at x and, let the variables x_i, y_i, u_j and v_j be such that

$$(5.6.8) \qquad x_i^* = \sqrt{1-\rho^2}\, x_i + \rho y_i \qquad i = 1,2,\ldots r$$

$$(6.5.9) \qquad y_i^* = \begin{cases} y_i & i = 1,2,\ldots.r \\ y_i + \mu & i = r \end{cases}$$

$$(6.5.10) \qquad u_j^* = \sqrt{1-\rho^2}\, u_j + \rho v_j \qquad j = 1,2,\ldots.s$$

$$(6.5.11) \qquad v_j^* = v_j \qquad j = 1,2,\ldots.s \; .$$

The x_i, y_i, u_j and v_j are mutually independent standard normal variables and μ^2 is the concentration or noncentrality parameter given by

$$(6.5.12) \qquad \mu^2 = \frac{\pi_{22}' X_2' [I - X_1(X_1'X)^{-1}X_1']X_2 \pi_{22}}{\omega_{22}}$$

where π_{22} is the vector of coefficients of excluded endogenous variables in the reduced-form equation for the right-hand side endogenous variable. In terms of the variables x_i, y_i, u_j and v_j.

Mariano (1973) defines $\hat{\beta}*(k)$ as

$$(6.5.13) \qquad \hat{\beta}*(k) = \frac{n_1^{*2}}{\mu^2}\Big\{\sum_{i=1}^{r} y_i \sqrt{1-\rho^2}\, x_i + \rho y_i) + \mu(\sqrt{1-\rho^2}\, x_r + \rho y_r + (1-k)\sum_{j=1}^{s} v_j\,(\sqrt{1-\rho^2}\, u_j + \rho v_j)\Big\}$$

$$(6.5.14) \qquad = \rho + \frac{n_1^2}{\mu^2}\Big\{\sqrt{1-\rho^2}\,\Big[\sum_{i=1}^{r} x_i y_i + \mu x_r + (1-k)\sum_{j=1}^{s} u_j v_j\Big] - \rho\mu(y_r + \mu)\Big\}$$

$$(6.5.15) \qquad = \rho + \frac{\sqrt{1-\rho^2}\, z n_1^2}{\mu n_2} - \frac{\rho(y_r + \mu)n_1^2}{\mu}$$

where

$$(6.5.16) \qquad n_1^2 = \frac{\mu^2}{\sum_{i=1}^{r-1} y_i^2 + (y_r + \mu)^2 + (1-k)\sum_{j=1}^{s} v_j^2}$$

(6.5.17) $$n_2^2 = \frac{\mu^2}{\sum_{i=1}^{r-1} y_i^2 + (y_r+\mu)^2 + (1-k)^2 \sum_{j=1}^{s} v_j^2}$$

and Z is a standard normal random variable independent of $y_1, y_2, \ldots, y_r, v_1, v_2, \ldots, v_s$. It should be noted that from our assumptions and the definition of μ^2 in (6.5.12) that $\frac{\mu^2}{T}$ constant as $T \to \infty$. Also from (6.5.13) and (6.5.17) and our assumptions about the x_i's and the y_i's, it follows that as T is kept fixed and $\mu \to \infty$ that

(6.5.18) $$\text{plim}\,(\mu\,\hat{\beta}^* - \sqrt{1-\rho^2}\,x_r - \sigma y_r) = 0 \quad .$$

Since $(\sqrt{1-\rho^2}\,x_r - \rho y_r)$ is a standard normal random variate, this implies that $\mu\,\hat{\beta}^*(k)$ has a limiting standard normal distribution i.e., using (6.5.2).

LEMMA (6.5.4): For fixed T, the limiting distribution of $\mu \frac{\sqrt{\omega_{22}}}{\sigma}(\hat{\beta}(k) - \beta)$ is a standard normal distribution as $\mu \to \infty$.

Before presenting the main results, we introduce the following definitions:

(6.5.19) $$N^* = \frac{n_2}{\sqrt{1-\rho^2}}\left\{\frac{\mu}{n_1^2}\left(\frac{a}{\mu} - \rho\right) + \rho(y_r + \mu)\right\}$$

$$\bar{N}^* = \frac{1}{\sqrt{1-\rho^2}} \left\{ -\frac{\rho}{\mu} \left[\sum_{i=1}^{r-1} y_i^2 + (1-k) \sum_{j=1}^{s} v_j^2 \right] + \left(\frac{a}{\mu} - \rho\right) y_r + a \right\} \tag{6.5.20}$$

where a is any arbitrary real number.

Let $$\varepsilon^{-2} = 1 - \frac{2a\rho}{\mu} + \frac{a}{\mu^2} .$$

Then from (6.5.15), (6.5.17) and (6.5.19), (using pr to denote probability)

$$\mathrm{pr}(\mu\, \hat{\beta}^*(k) \le a) = \mathrm{pr}(z \le N^*) \tag{6.5.21}$$

$$= \mathrm{pr}\left\{ z < \bar{N}^* + 0\left(\frac{1}{\mu^2}\right) \right. \quad \text{as } \mu \to \infty \tag{6.5.22}$$

$$= \mathrm{pr}\left\{ z' < a\varepsilon - \frac{\rho\varepsilon}{\mu} \left[\sum_{i=1}^{r-1} y_i^2 + (1-k) \sum_{j=1}^{s} v_j^2 \right] \right\} + 0\left(\frac{1}{\mu^2}\right) \quad \text{as } \mu \to \infty \tag{6.5.23}$$

$$= E\Phi\left[a\varepsilon - \frac{\rho\varepsilon}{\mu} \left(\sum_{i=1}^{r-1} y_i^2 + (1-k) \sum_{j=1}^{s} v_j^2 \right) \right] + 0\left(\frac{1}{\mu^2}\right) \quad \text{as } \mu \to \infty \tag{6.5.24}$$

$$= \Phi(a) + \frac{\rho}{\mu} \phi(a) [a^2 - r + 1 + (1-k)s] + 0\left(\frac{1}{\mu^2}\right) \quad \text{as } \mu \to \infty . \tag{6.5.25}$$

It should be noted that for the equation (6.5.22), the function $\overline{N}^*$ may be obtained from N^* in succeeding steps. First, delete the term with the factor $\frac{1}{\mu^2}$ from the factor of N^* in the braces in (6.5.19). Then approximate n_1 by using Taylor-Series theorem, by restricting the region $\{y : |y_r| \leq (1-\eta)\mu\}$ for fixed $0 < \eta < 1$.

In (6.5.23) Z' is standard normal variable independent of $y_1 \ldots y_{r-1}, v_1, \ldots, v_s$ and equation (6.5.23) is obtained by a straightforward manipulation of the inequality $Z \leq \overline{N}^*$. Equation (6.5.24) follows immediately from (6.5.23) and (6.5.25) is obtained by using a Taylor series expansion to expand the leading term (6.5.24) about a as well as to expand the expression ε about unity. (See Mariano (1972, 1973) for further details.)

<u>THEOREM 6.5.A:</u> In the case of two included endogenous variables in the equation being estimated, let the sample size T be fixed. Then an approximation to the distribution function of the k-class estimator is given by

$$\text{(6.5.26)} \qquad \mathrm{pr}\left\{(\hat{\beta}(k) - \beta) \leq \frac{a\sigma}{\mu\sqrt{\omega_{22}}}\right\} = \Phi(a) + \frac{\rho}{\mu}\phi(a)$$

$$\text{times } [a^2 - K_2 + 1 + (1-k)\,T - K)] + O\left(\frac{1}{\mu^2}\right) \quad \text{as } \mu \to \infty$$

where a is an arbitrary real number,

$$\rho = \frac{1}{\sigma\sqrt{\omega_{22}}}(\omega_{21} - \beta\,\omega_{22})$$

$$\sigma^2 = \omega_{11} - 2\beta\omega_{21} + \beta^2\omega_{22} \qquad \text{and}$$

$$\mu^2 = \frac{1}{\omega_{22}} \left\{ \pi_{22}' \, X_2'[I - X_1(X_1'X_1)^{-1}X_1'] \, X_2 \, \pi_{22} \right\} .$$

<u>THEOREM 6.2.B</u>: In the case of two included endogenous variables in the equation to be estimated, an approximation to the 2SLS distribution function is given by

$$(6.5.27) \qquad \mathrm{pr}\left\{ (\hat{\beta} - \beta) \le \frac{a\sigma}{\mu\sqrt{\omega_{22}}} \right\} = \Phi(a) + \frac{\rho}{\mu}\phi(a)(a^2 - K_2 + 1) + O(\frac{1}{\mu^2}) \qquad \text{as } \mu \to \infty .$$

<u>THEOREM 6.5.C</u>: In the case of two endogenous variables present in the equation to be estimated, let $\tilde{\beta}$ be the OLS estimator of β and let the sample size T be fixed. Then

$$(6.5.28) \qquad \mathrm{pr}\left\{ (\tilde{\beta} - \beta) \le \frac{a\sigma}{\mu\sqrt{\omega_{22}}} \right\} = \Phi(a) + \frac{\rho}{\mu}\phi(a) \text{ times } [a^2 - T + K_1 + 1] + O(\frac{1}{\mu^2}) \qquad \text{as } \mu \to \infty .$$

with a, ρ, μ^2 and σ defined previously.

It has been shown in (6.5.18) and (6.5.19) that for non-stochastic k, with T fixed, that $\mu\hat{\beta}^*(k)$ converges in probability to a standard unit normal and random variable as $\mu^2 \to \infty$. Mariano (1975) has shown that the preceeding statement implies that $\hat{\beta}^*(k)$ converges in probability to zero and hence, as $\mu^2 \to \infty$ with fixed T, all members of the non-stochastic k-class of estimators converge in probability to the true parameter value by virtue of (6.5.2), i.e.

$$\hat{\beta}(k) = \beta + \frac{\sigma}{\sqrt{\omega_{22}}} \hat{\beta}^*(k) .$$

Mariano (1975) obtained a sufficient condition under which similar conclusions hold for the stochastic k-class estimators. His results on asymptotic behaviour are derived for $\mu^2 \to \infty$ while T (the sample size) either stays fixed or increases indefinitely. More precisely his results are expressed by the following proposition.

PROPOSITION (6.5.A)

Let $\mu^2 \to \infty$ by having either $T \to \infty$ or $\ell^2 \to \infty$ or both. If $\text{plim}\ (\sqrt{T}(1-k)/\ell) = 0$, then

$$\frac{\mu\sqrt{\omega_{22}}}{\sigma}(\hat{\beta}(k) - \beta)$$

converges in distribution to a standard unit normal and $\hat{\beta}$ converges in probability to β.

Here

(6.5.29) $\qquad \mu^2 = T\ell^2$, i.e.,

$$\ell^2 = \frac{\pi_{22}\ X_2'\ [I - X_1(X_1'X_1)^{-1}X_1']\ X_2\pi_{22}}{T\ \omega_{22}}$$

(ℓ^2 is the concentration parameter or noncentrality parameter normalized for sample size).

An outline of the proof of Proposition 6.5.A. as given by Mariano (1975) is as follows:

From (6.5.13) and (6.5.16), let

$$\hat{\beta}^*_{(k)} = \frac{H}{G} \qquad \text{where} \tag{6.5.30}$$

$$H = \sum_{i=1}^{r} y_i(\sqrt{1-\rho^2})x_i + \rho y_i) + \mu(\sqrt{1-\rho^2})\, x_r + \rho y_r) + (1-k)\sum_{j=1}^{s} v_j(\sqrt{1-\rho^2})\, \mu_j + \rho v_j) \tag{6.5.31}$$

$$G = \sum_{i=1}^{r-1} y_i^2 + (y_r + \mu)^2 + (1-k)\sum_{j=1}^{s} v_j^2 \tag{6.5.32}$$

From (6.5.30), we write

$$\mu\,\hat{\beta}^*(k) = \frac{H/\mu}{G/\mu^2} \qquad \text{where} \tag{6.5.33}$$

H can be defined more compactly as

$$H = P^* + \mu\, Q^* + (1-k)R_T \tag{6.5.34}$$

where P^*, μQ^* and $(1-k)R_T$ correspond, respectively, to the first, second and third terms in (6.5.31), so that

$$Q^* \sim N(0,1) \qquad \text{(see 6.5.18)} \tag{6.5.35}$$

$$R_T \sim \sum_{j=1}^{s} r_j, \qquad r_j \text{ iid}(\rho, 1-\rho^2). \tag{6.5.36}$$

and Q^* and R_T are independent.

From (6.5.31) and (6.5.35), we have

(6.5.37) $$\frac{1}{\mu}(P^* + \mu Q^*) \xrightarrow{a.e.} Q^* \quad \text{as} \quad \mu \to \infty .$$

where $\xrightarrow{a.e.}$ indicates converges almost everywhere, and by the Central Limit Theorem (CLT),

(6.5.38) $$\frac{R_T - T\rho}{\sqrt{T}\sqrt{1-\rho^2}} \xrightarrow{d} N(0,1) \quad \text{as} \quad T \to \infty ,$$

where $\xrightarrow{d}$ indicates coverges in distribution.

Now let $\mu^2 \to +\infty$ with T staying fixed, then from (6.5.34) and (6.5.37), H/μ converges in probability to Q^* if

(6.5.39) $$\text{plim}\,[(1-k)/\mu] = 0 \quad \text{as} \quad \mu \to \infty \text{ (with } T \text{ fixed).}$$

In the case where $\mu^2 \to \infty$ while $T \to \infty$ simultaneously, we write

(6.5.40) $$\frac{(1-k)R_T}{\mu} = \sqrt{(1-\rho^2)}\left\{\frac{1-k}{\ell}\left(\frac{R_T - T\rho}{\sqrt{T}\sqrt{1-\rho^2}} + \frac{\sqrt{T}\rho}{\sqrt{1-\rho^2}}\right)\right\}$$

[From (6.5.29) $\ell = \frac{\mu}{\sqrt{T}}$].

Then from (6.5.37) and (6.5.38) it follows that H/μ converges in distribution to Q^* if

(6.5.41) $$\text{plim}\,[\sqrt{T}(1-k)/\ell] = 0 \quad \text{as} \quad \mu \to \infty \quad \text{and} \quad T \to \infty .$$

Using $\ell^2 = \frac{\mu^2}{T}$, the relations (6.5.39) and (6.5.41) can be combined into

(6.5.42) $$\operatorname{plim}\,[\sqrt{T}\,(1 - k)/\ell] = 0 \qquad \text{as } \mu^2 \to \infty$$

i.e., as either $T \to \infty$ or $\ell^2 \to \infty$ or both.

For the denominator in (6.5.33), we have

(6.5.43) $$\operatorname{plim}\,(G/\mu^2) = 1 + (\operatorname{plim}\,[(1 - k)/\ell^2])$$

$$\text{times } (\operatorname{plim}\,[(\sum_{j=1}^{s} v_j^2/T])$$

$$= 1 + \alpha \operatorname{plim}\,[(1 - k)/\ell^2]$$

where $\alpha \to$ unity if $T \to \infty$ and

(6.5.44) $$\alpha = \frac{1}{T}\sum_{j=1}^{s} v_j^2 \qquad \text{if } T \text{ is fixed as } \mu^2 \to \infty\,.$$

It follows from (6.5.38) that if condition (6.5.41) holds, then

(6.5.45) $$\operatorname{plim}\,(G/\mu^2) = 1 \qquad \text{as } \mu^2 \to \infty\,.$$

Further, Mariano (1975) showed that the conclusions of Proposition 6.5.A. hold for the (LIML) estimator of β

6. We consider the equation (5.1.1) and for convenience denote it by

$$(6.6.1) \qquad y_1 = \beta y_2 + X_1\gamma_1 + U \qquad \text{with reduced form}$$

$$(6.6.2) \qquad Y = X\Pi + V , \qquad \text{where} \quad Y = [y_1 , y_2] .$$

Let R denote the nonsingular matrix $(X'X)$; i.e.

$$(6.6.3) \qquad R = (X'X) = \begin{bmatrix} X_1' \\ X_1' \end{bmatrix} [X_1 , X_2]$$

$$= \begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix} = \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix}$$

The usual regression estimate of Π in (6.6.2) is given by

$$(6.6.4) \qquad S = (X'X)^{-1}X'Y = R^{-1}X'Y = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}$$

Let $\hat{\Omega}$ denote an estimate for Ω as defined in (5.1.5) i.e.

$$(6.6.5) \qquad \hat{\Omega} = \frac{1}{T} C \qquad \left(\text{or } \frac{1}{T-K} C\right) , \qquad \text{where}$$

$$(6.6.6) \qquad C = Y'Y - S'RS$$

$$= Y'Y - Y'X(X'X)^{-1}(X'X)(X'X)^{-1}XY$$

$$= Y' [I - X(X'X)^{-1}X']Y .$$

The (2SLS) method of estimation uses (5.1.7), i.e., $\pi_{21} = \beta\pi_{22}$ in terms of the estimates S_{21} and S_{22}. Let

$$R_{22.1} = R_{22} - R_{21} R_{11}^{-1} R_{12} \tag{6.6.7}$$
$$= X_2'X_2 - X_2'X_1(X_1'X_1)^{-1}X_1'X_2$$
$$= X_2' [I - X_1(X_1'X_1)^{-1}X_1']X_2$$

which is the second-order moment matrix of the residuals of the excluded exogenous variables from their regression on the included endogenous variables. The covariance matrix of $(S_{21} - \beta S_{22})$ is

$$\sigma^2(R_{22.1}) \qquad (\sigma^2 \text{ is defined in (5.1.8)}). \tag{6.6.8}$$

As pointed out by Anderson and Rubin (1949), the (LIML) estimate of β, denoted by $\hat{\beta}_{LIML}$, is the value of β that minimizes the quadratic form,

$$(S_{21} - \beta S_{22})' R_{22.1} (S_{21} - \beta S_{22}) \tag{6.6.9}$$

relative to the estimate of σ^2, i.e., the vlaue of of β that minimizes the rates of the quadratic forms,

$$\frac{(S_{21} - \beta S_{22})' R_{22.1} (S_{21} - \beta S_{22})}{\hat{\omega}_{11} - 2\beta\hat{\omega}_{12} + \beta\hat{\omega}_{22}} \tag{6.6.10}$$

where $\hat{\omega}_{11}$, $\hat{\omega}_{12}$ are $\hat{\omega}_{22}$ are corresponding estimates of ω_{11}, ω_{12} and ω_{22} (see 5.1.8).

The (2SLS) estimate of the parameter β, denoted by $\hat{\beta}$, is the value of β that minimizes the numerator of (6.6.10). Thus

(6.6.11) $$\hat{\beta} = \frac{S'_{21}\ R_{22.1}\ S_{22}}{S'_{22}\ R_{22.1}\ S_{22}}$$

A general class of estimates that include both the (OLS) and (2SLS) estimates can be denoted by

(6.6.12) $$\frac{S'_{21}\ R_{22.1}\ S_{22} + \alpha c_{12}}{S'_{22}\ R_{22.1}\ S_{22} + \alpha c_{22}}$$

where $0 \le \alpha \le 1$, c_{ij} is the (i,j)the element of C as defined in (6.6.6). It should be noted that with α replaced by $(1-k)$, the resulting k-class estimates are obtained. (OLS) and (2SLS) estimates correspond to the case where $\alpha = 1$ and $\alpha = 0$, respectively.

We look at the asymptotic expansions of the distributions of the (LIML) and (2SLS) estimates obtained by Anderson and Sawa (1973) and by Anderson (1974). The expansions are given in terms of the noncentrality parameter which increases as the sample size, T, increases.

The noncentrality parameter μ^2 can now be expresses as

(6.6.13) $$\mu^2 = \frac{\pi'_{22}\ R_{22.1}\ \pi_{22}}{\omega_{22}}$$

Following Anderson (1974), the distribution of the matrix $[S_{21}, S_{22}]$ is multivariate normal with expected value $[\pi_{21}, \pi_{22}]$ and covariances

(6.6.14) $$E(S_{2i} - \pi_{2i})(S_{2j} - \pi_{2j})' = \omega_{ij}\ R^{-1}_{22.1}, \quad [i, j=1,2] .$$

The matrix

$$\psi = \begin{bmatrix} S'_{21} \\ S'_{22} \end{bmatrix} R_{22.1} \; [S_{21} \; , \; S_{22}] \tag{6.6.15}$$

is a non-central Wishart distribution with K_2 degrees of freedom, covariance matrix Ω , and means sigma matrix

$$\begin{bmatrix} \pi'_{21} \\ \pi'_{22} \end{bmatrix} R_{22.1} \; [\pi_{21} \, , \pi_{22}] \tag{6.6.16}$$

$$= [\beta\pi_{22} \, , \pi_{22}]' \; R_{22.1} \; [\beta\pi_{22} \, , \pi] \; ,$$

from the relation, $\pi_{21} = \beta\pi_{22}$.

We can further write (6.6.16) as

$$\pi'_{22} R_{22.1} \; \pi_{22} \begin{bmatrix} \beta^2 & \beta \\ \beta & 1 \end{bmatrix} \; .$$

The matrix C defined in (6.6.6) has a central Wishart distribution with $T-K$ degrees of freedom and covariance matrix Ω . The distributions and asymptotic expansions of distributions are obtained from the distributions of Ω and C .

The distribution of $\hat{\beta}_{(2SLS)}$ depends on the sample size, T, only through $\pi'_{22} R_{22.1} \pi_{22}$ since the distribution of Ψ depends on the sample size only through $\pi'_{22} R_{22.1} \pi_{22}$. The asymptotic expansion of $\hat{\beta}_{(2SLS)}$ is, in terms of this noncentrality parameter, increasing. The cumulative distribution of the 2SLS estimate $\hat{\beta}_{(2SLS)}$ is given by

$$\text{(6.6.17)} \qquad \text{pr}\left\{\frac{\sqrt{\pi_{22}' R_{22.1} \pi_{22}}}{\sigma}(\hat{\beta}_{2SLS} - \beta) \le w\right\} = \Phi(w)$$

$$- \frac{c}{q}[w^2 - (K_2 - 1]\phi(w)$$

$$+ \frac{1}{2q^2}\left\{[K_2 - 1 - (K_2 - 1)^2 c^2]w\right.$$

$$\left. + (2K_2 c^2 - 1)w^3 - c^2 w^5\right\}\phi(w) + O(q^{-3}),$$

and approximate density is

$$\text{(6.6.18)} \qquad \left\{1 + \frac{c}{q}[w^3 - (K_2+1)w] + \frac{1}{2q^2}(K_2 - 1) - K_2 - 1)^2 c^2\right.$$

$$+ [(K_2^2 + 4K_2 + L)c^2 - (K_2 + 2)]w^2$$

$$\left. + [1 - (2K_2 + 5)c^2]w^4 + c^2 w^6\right\}\phi(w)$$

(6.6.19) where $\phi(\ )$ and $\Phi(\ )$ are the density function and the cumulative distribution function of the standard normal distribution

$$\text{(6.6.20)} \qquad q^2 = \pi_{22}' R_{22.1} \pi_{22} \frac{\sigma^2}{|\Omega|}$$

and

$$\text{(6.6.21)} \qquad c = \frac{\beta\omega_{22} - \omega_{12}}{\sqrt{|\Omega|}} \quad ,$$

The cumulative distribution of $\hat{\beta}_{(LIML)}$ as obtained by Anderson (1974) is given by

$$\text{(6.6.22)} \qquad \operatorname{pr}\left\{ \frac{\sqrt{\pi_{22}' R_{22.1} \pi_{22}}}{\sigma} (\hat{\beta}_{LIML} - \beta) \le w \right\} = \Phi(w)$$

$$+ \left\{ -\frac{c}{q} w^2 + \frac{1}{2q^2} [-(K_2 - 1)w \right.$$

$$+ (2c^2 - 1)w^3 - c^2 w^5$$

$$+ \frac{c}{6q^3} [3(K_2 - 1)w^2 - (6c^2 + 3K_2 - 12)w^4$$

$$\left. + (7c^3 - 3)w^5 - c^2 w^8] \right\} + \phi(w) + 0(q^{-4}) .$$

The approximate density from (6.6.23) is

$$\text{(6.6.23)} \qquad \left\{ 1 + \frac{c}{q} (w^3 - 2w) + \frac{1}{2q^2} [-(K_2 - 1) + (K_2 - 4 + 6c^2)w^2 \right.$$

$$+ (1 - 7c^2)w^4 + c^2 w^6] + \frac{c}{6q^3} [6(K_2 - 1)w$$

$$- (24c^2 + 15K_2 - 51)w^3 + (48c^2 + 3K_2 - 30)w^5$$

$$\left. - (15c^2 - 3)w^7 + c^2 w^9] \right\} \phi(w) .$$

In the cumulative distribution of $\hat{\beta}_{LIML}$, Anderson (1974) assumes that $\dfrac{\pi_{22}' R_{22.1} \pi_{22}}{T\omega_{22}}$ is bounded. This assumption implies that T increases as fast as the noncentrality parameter $\dfrac{\pi_{22}' R_{22.1} \pi_{22}}{\omega_{22}}$.

In both the (2SLS) and (LIML) estimates, the leading term is the standard normal distribution, and the other terms are products of the standard normal density and polynomials in W. In the case of $K_2 = 1$

(i.e. the case of exact identification) the asymptotic expansions of the two estimates are equivalent.

CHAPTER 7

A REVIEW OF SOME MONTE CARLO STUDIES

1. Since the samples with which we deal in practice are rather small, it would be of great interest to inquire into the small-sample properties of various estimators. We have seen that (2SLS) and (LIML) estimators are asymptotically equivalent - the same is true with respect to (3SLS) and (FIML) estimators. The Theorems on these asymptotic properties do not shed too much light on how these estimators will behave when the sample size is of small or moderate size. Then, there is the question of the efficiency of estimators. From the literature, it is clear that (3SLS) and (FIML) are both asymptotically more efficient relative to the (2SLS) and (LIML) estimates, respectively. Again, we are not so sure that this condition prevails when dealing with small-samples.

In any limited information technique, e.g. (2SLS) or (LIML), we focus attention on the explicit specification of the equation being estimated. On the other hand, in the full information methods, we have to rely on the specification of every equation in the system under study. With this fact in view, we would like to know how the various estimators behave when misspecification errors are present. In addition, we would like to know whether (OLS) is, in small samples, sufficiently inferior to other simultaneous equation methods so as to justify its exclusion as an estimating procedure for structural coefficients.

Apart from the problems mentioned above, typical difficulties are faced in estimation when implications such as autocorrelated disturbances, missing observations and errors in observations are present. They pose a variety of problems that are as yet unsolved. The problem becomes much more complicated when two or more of these complications affect the system simultaneously. Further, we may be interested in similar questions about reduced form coefficients and in the area of forecasting and testing of hypotheses.

These questions are of great practical significance, to which, in general no precise analytical answer can be furnished. Nevertheless, some tentative results can be derived through direct or modified Monte Carlo experiments.

2. In general, a Monte Carlo experiment is essentially an empirical method for gaining insight about the probability distribution of a statistic. A number of researchers have used Monte Carlo experiments to determine the small-sample properties of different estimators. Although, still in its infancy, this branch of econometrics has grown concurrently with the capability of high speed computing machines to permit numerous replications at low cost. The Monte Carlo method allows for a large number of samples to be drawn from a known propulation; sample statistics can then be computed in various ways and compared with the true population parameters.

The direct Monte Carlo experiment is usually conducted along the following lines: An artificial structure is set up, consisting of model $By_t + \Gamma x_t = u_t$ (say) with reduced form $y_t' = -B^{-1}\Gamma x_t' + B^{-1}u_t$ and $(t = 1,\ldots,T)$ with known parameters B and Γ. The probability distribution of the disturbance vector, u_t, is specified. A set of values of exogenous variables are chosen. The known structure and the known values of exogenous variables, together with random drawings from the specified disturbance probability distribution are used to generate a sample of observations. Various estimating techniques are used to obtain sample estimates of the population parameters.

The procedure is repeated to obtain a number of samples of size T (say). For each of these samples the reduced form equations are used to generate a set of values of the jointly dependent variables y_t and the predetermined variables for $t = 1,.....,T$. In this way, we obtain a large number of artificial samples of data for y_t and x_t (for $t = 1,....,T$), that have been generated by the artificial structure set up initially.

The artificial sample data for y_t and x_t are used to estimate the parameters B and Γ by each method in question. When the estimated values of the parameters have been computed by each method for each sample, we can then compare the sampling distributin of each estimation method with the value of the parameters. In this way, we could decide which of the estimation methods will be given preference for that parameter with a sample size T.

3. In this section we look at the results of some direct Monte Carlo studies. Johnston (1972) reviewed, in some detail, the direct Monte Carlo studies conducted by Ladd (1956), Wagner (1958), Basmann (1958), Neiswagner and Yancey (1959), Klien (1960), Nagar (1960), Quandt (1965), Summers (1965), Cragg (1966, 1967) and Simister (1969).

Johnston's review revealed that there is a fairly general agreement on the following points:

(a) The (OLS) estimates generally display the greatest finite sample bias of all the estimators considered; the consistent estimators [e.g., 2SLS, 3SLS, FIML] show finite sample bias and the variation of the bias among consistent estimators is not significant enough to favour one estimator over another. Cragg (1967) considered six degrees of multicollinearity in the exogenous variables. His results showed that even in this case, (OLS) has a larger bias relative to that of the consistent estimators. His results also indicated that multicollinearity produced a substantial increase in the bias of consistent estimators.

(b) Among the estimators, (OLS) frequently has the smallest variance (measured around the mean).

(c) When the (OLS) method is used there is always the likelihood of making incorrect inferences about the true values of structural coefficients. (See Summers (1965) and Cragg (1966) for details.)

4. Here we review some of the results obtained by Summers (1965). The hypothetical model used by Summers (1965) was the same as that used by Neiswagner and Yancey (1959). The model considered is given by

$$(7.4.1)\quad (a)\quad y_{1t} = \beta_{12}\, y_{2t} + \gamma_{11}\, x_{1t} + \gamma_{12}\, x_{2t} + \gamma_{10} + u_{1t}$$

$$(b)\quad y_{2t} = \beta_{21}\, y_{1t} + \gamma_{23}\, x_{3t} + \gamma_{24}\, x_{4t} + \gamma_{20} + u_{2t}$$

The two equation model considered in (7.4.1) consists of two over-identified equations and could readily admit several economic interpretations.

There are, however, a few shortcomings in the model (7.4.1). Firstly, the (2SLS) applied to this model produces zero bias to the order T^{-1} and T^{-2} (T, being the number of observations) secondly, Basmann (1958), using the model (7.4.1), derived the exact sample frequency functions of Generalized Classical Linear (GCL) estimators and found that the means exist while moments of higher order than the first do not.

In the case of Summers (1965) the y's and x's in (7.4.1) represent jointly dependent and predetermined variables respectively. The error vectors $[(u_{1t}, u_{2t}) : t = 1, \ldots, T]$ are specified to be mutually independent and identically distributed as $N(0, \Sigma)$ where

$$(7.4.2)\qquad \Sigma = \begin{bmatrix} 400, & 200 \\ 200, & 400 \end{bmatrix} = 400 \begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{bmatrix}$$

The reduced form model of (7.4.1) is given by

$$(7.4.3)\qquad \begin{aligned} y_{1t} &= \pi_{10} + \pi_{11} x_{1t} + \pi_{12} x_{2t} + \pi_{13} x_{3t} + \pi_{14} x_{4t} + v_{1t} \\ y_{2t} &= \pi_{20} + \pi_{21} x_{1t} + \pi_{22} x_{2t} + \pi_{23} x_{3t} + \pi_{24} x_{4t} + v_{2t} \end{aligned}$$

where the π's are functions of the β's and the γ's, and the v's are linear combinations of the u's.

Summers (1965) conducted 12 experiments using the model or variants of the model in (7.4.1). Fifty samples of size 20 or size 40 were used. Five different sets of parameters were used; four were used with a single sample of size 20, but the fifth was used with two sample sizes, T = 20, T = 40. It is difficult to summarize such a conprehensive study. As such, only some of the features of this study will be outlined.

Summers (1965) compared (OLS) , (LIML), Least Squares, No Restrictions (LSNR), (2SLS) and (FIML) structural estimates. Reduced form estimates were also computed and conditional forecasts of the two endogenous variables based on the five sets of reduced form estimates [i.e. (OLS), (LIML) (LSNR), (2SLS) and (FIML)] were compared.

For some comparisons he used the mean square error (MSE) as a criterion of quality of the estimators, and for some comparisons the Root Mean Square Error (RMSE) was used. These two criteria gave similar results except in a few cases where the (LIML) estimates were involved. The (FIML) estimator usually stood out as the best structural estimator except in the misspecification experiments where (LIML) and (FIML) estimators were worse than the (2SLS) estimators. In fact, the (2SLS) estimators were the least affected by multicollinearity and misspecification. In the case of conditional forecasts of endogenous variables (OLS) was almost always substnatially worse, and the other four methods of reduced form estimation yielded forecasts of very similar quality.

(7.4.4) Least Squares, No Restrictions (LSNR): In this method (OLS) is applied to each reduced-form equation separately. Consistent estimates of the reduced-form coefficients are obtained, but the estimates are not efficient because the method ignores any a priori information specified in the model. The (LSNR) method is computationally simple and is relatively insensitive to structural specification error.

(7.4.5) Mean Square Error (MSE): If $\hat{\beta}$ is the estimate of β, then

$$\begin{aligned} \text{MSE} &= E[(\hat{\beta} - \beta)^2] \\ &= E[(\hat{\beta} - E\beta)^2 + (E\hat{\beta} - \beta)^2] \\ &= \text{variance plus square of bias.} \end{aligned}$$

The square root of the (MSE) is called the root mean square error (RMSE).

In Summers (1965) we consider the experiment (4A). Let

(7.4.6)
$$\delta_1 = \begin{bmatrix} \beta_{12} \\ \gamma_{10} \\ \gamma_{11} \\ \gamma_{12} \end{bmatrix} \quad \text{and} \quad \delta_2 = \begin{bmatrix} \beta_{22} \\ \gamma_{20} \\ \gamma_{23} \\ \gamma_{24} \end{bmatrix}$$

(see 7.4.1).

In experiment (4A), the sample size $T = 20$ and the number of samples $N = 50$. The predetermined variables were almost uncorrelated; their correlation matrix is specified by

(7.4.7)

$$\mathrm{Corr}(x) = \begin{array}{c|cccc} & X_1 & X_2 & X_3 & X_4 \\ \hline X_1 & 1 & .078 & .16 & .38 \\ X_2 & & 1 & .017 & -0.057 \\ X_3 & & & 1 & .31 \\ X_4 & & & & 1 \end{array}$$

In the experiment (4A),

(7.4.8)

$$\delta_1 = \begin{bmatrix} -1.3 \\ -149.5 \\ 0.8 \\ 0.7 \end{bmatrix} \quad \text{and} \quad \delta_2 = \begin{bmatrix} 0.4 \\ -149.6 \\ 0.6 \\ -0.4 \end{bmatrix}$$

A general summary of his results for Experiment (4A) is outlined in the following table:

TABLE 7.4 (A)

The Bias and RMSE of Estimators (for model correctly specified)

	$b(\hat{\delta}_1)$	RMSE $(\hat{\delta}_1)$	$b(\hat{\delta}_2)$	RMSE $(\hat{\delta}_2)$
LIML	.0276	.132	.0118	.182
	.2019	22.3	-4.032	75.000
	-0.0279	.173	.0084	.669
	-0.0022	.105	.0118	.062
2SLS	.0326	.133	-0.0005	.175
	.3508	22.3	-3.190	75.1
	-0.0322	.173	.0078	.668
	-0.0040	.105	.0075	.061
FIML	.0280	.131	.0099	.181
	.3471	22.0	-8.484	71.8
	-0.0279	.164	.0488	.628
	-0.0038	.103	.0045	.062
OLS	.1051	.170	-0.1249	.218
	2.565	22.2	7.876	72.7
	-0.098	.197	-0.0174	.640
	-0.0290	.110	-0.0410	.072

From the preceding results we find that both (2SLS) and (LIML) estimators do not have too much significant difference between them. It would seem that the (FIML) estimators are slightly more efficient in the sense that their (RMSE) are slightly smaller than those of the (2SLS) or (LIML) estimators. The (OLS) estimators seem to be inferior to the others.

We consider Experiment (5A) where Summers (1965) attempted to test the sensitivity of the various estimators to misspecification errors. Here he used $T = 20$ and $N = 50$, and

$$(7.4.9) \qquad \delta_1 = \begin{bmatrix} -.7 \\ -149.5 \\ 0.8 \\ 0.7 \end{bmatrix} \quad \text{and} \quad \delta_2 = \begin{bmatrix} 0.4 \\ -149.6 \\ 0.6 \\ -0.4 \end{bmatrix}$$

In this misspecification experiment the data on y_1 and y_2 in (7.4.1) were generated by the model defined by (7.4.1) (a) and

$$y_{1t} + \beta_{22} y_{2t} + \gamma_{21} x_{1t} + \gamma_{23} x_{3t} + \gamma_{24} x_{4t} + \gamma_{20} = u_{2t} .$$

In this misspecification model, the second equation is just-identifiable instead of being overidentified.

The results of Experiment (5A) are given in the following table:

TABLE 7.4 (B)

Bias and RMSE of Estimators (model misspecified)

	$b(\hat{\delta}_1)$	RMSE $(\hat{\delta}_1)$	$b(\hat{\delta}_2)$	RMSE $(\hat{\delta}_2)$
LIML	.0081	.078	.2871	.364
	.5413	23.2	31.01	77.2
	-0.0164	.176	.1154	.659
	.0116	.082	- 0.0632	.108
2SLS	.0126	.078	.1669	.241
	.7452	23.2	42.22	77.7
	- .0204	.176	.0643	.609
	.0092	.082	- 0.0188	.071
FIML	.0363	.101	.3396	.491
	13.09	31.10	30.11	92.7
	- 0.1443	.284	.1027	.783
	.0186	.600	- 0.0823	.164
OLS	.0704	.099	.0325	.144
	3.680	22.9	54.58	83.5
	- 0.0742	.183	.0099	.582
	- 0.0217	.083	.0307	.664

As expected, Table 7.4 (B) shows that the FIML is considerably more sensitive to misspecification errors. The (2SLS) estimators seemed to be the least affected by misspecification. This is to be expected, since the specification of equations other than the one under study is of little consequence in Limited Information Methods. On the other hand, in the Full Information Methods the structure of every equation is explicitly utilized in the estimation of every other equation. Thus, misspecification in any of the equations will affect the estimators of every parameter of the system.

5. In general, Monte Carlo studies attempting to simulate small-sample properties of simultaneous equation estimators have traditionally assumed that the disturbances are independent drawings from a multivariate normal distribution. Raj (1980), in his Monte Carlo study of small-sample properties of simultaneous equation estimators employed the "direct simulation technique" (meaning that no refinement is exercised in the choice and use of random numbers) and considered four alternative forms of two-parameter error distributions: (a) normal (b) uniform (c) lognormal, and (d) Laplace or double exponential.

Specifically, Raj's two experiments employed two sets of 1000 generated samples of 20 observations each on an overidentified model consisting of two structural equations and an identity. The hypothetical used is given by

$$(7.5.1) \qquad -y_1 + \beta_{21} y_2 + \gamma_{11} x_1 + \gamma_{21} x_2 + u_1 = 0$$

$$(7.5.2) \quad \beta_{12} y_1 - y_2 + \beta_{23} y_3 + \gamma_{22} x_2 + \gamma_{32} x_3 + \gamma_{42} x_4 + u_2 = 0$$

$$(7.5.3) \quad y_1 - y_2 - y_3 + x_3 + x_5 + x_6 = 0$$

where the y's are endogenous variables, the x's are exogenous variables, and the u's are random tisturbances. As pointed out by Raj (1980), this model is basically a modification of the overidentified model used by Summers (1965). [See 7.4.1.]

Raj (1980) studied four estimators: (a) OLS, (b) 2SLS, (c) 3SLS, and (d) FIML. His simulation experiments were conducted with a view to determining (a) whether the small-sample rankings of econometric estimators of both structural parameters and forecasts of endogenous variables, according to the criteria of bias and dispersion, are different for different forms of error distributions and (b) whether the small-sample rankings correspond to the well-known asymptotic properties of structural estimators.

Although the conclusions of the Monte Carlo study will strictly apply to the particular model used, generalizations can be made to other models. Raj's study is of special importance since hardly any analytical results regarding the existence of moments are available when the disturbances are nonnormally distributed. Although some knowledge of the existence of moments of econometric estimators is important for a well-formulated Monte Carlo study, nevertheless, inferences based on some basic measures, such as median and quartile deviation of the sampling distribution, can provide useful information even when moments of the corresponding estimator do not exist.

TABLE 7.5.A

Parameter Combinations Used In The Two Sampling Experiments

Structural Parameters

Experiment	β_{21}	γ_{11}	γ_{21}	β_{12}	β_{32}	γ_{22}	γ_{32}	γ_{42}
I	-.90	-2.00	.50	2.50	1.90	2.80	-2.20	-.60
II	-.90	-2.00	.50	2.50	1.90	2.80	-2.20	-.60

Elements of Moment Matrix

Experiment	σ_{11}	σ_{12}	σ_{22}
I	1.81	-1.79	27.77
II	8.84	-1.79	24.46

Table (7.5.A) shows the two sets of numerical values of structural parameters and the elements in the moment matrix of structural disturbances. The moment matrices of the reduced-form disturbances in the two experiments had the following values.

(7.5.4)

$$\Omega_1 = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \\ 1 & -1 & 2 \end{bmatrix}$$

(7.5.5)

$$\Omega_2 = \begin{bmatrix} 2.20 & 1.97 & .23 \\ 1.97 & 3.82 & -1.85 \\ .23 & -1.85 & 2.08 \end{bmatrix}$$

As pointed out earlier, Raj's Monte Carlo experiments utilized two sets of 1,000 generated samples of 20 observations each. The observations on the six exogenous variables used in the simulations were independent random drawings from the uniform distribution in the range -17.321 to 17.321 . The set of 20 observations on the 6 exogenous variables were kept fixed in repeated samples.

The variance covariance matrix of the exogenous variables was given by

(7.5.6)

$$\text{Var}(X) = \begin{bmatrix} 100.30 & & & & & \\ -.039 & 99.621 & & & & \\ .096 & -2.43 & 99.893 & & & \\ 3.264 & -1.646 & 3.220 & 102.388 & & \\ .327 & -.125 & .123 & .603 & 100.354 & \\ .346 & .006 & .142 & 3.775 & .373 & 100.391 \end{bmatrix}$$

[see Raj (1980) for the method employed for the generation of reduced-form disturbances.]

The main findings of Raj's Monte Carlo experiments can be summarized as follows. The small-sample rankings of (OLS, (2SLS), (3SLS) and (FIML) estimators of structural coefficients according to parametric and non-parametric measures of bias, dispersion, and dispersion including bias are, except in a few instances, invariant to the form of the error distribution. (FIML), except for the mean squared error criterion, is the most efficient, whereas (OLS) is the least efficient of the four estimators of structural coefficients for all four error distributions. Also, (OLS) has the largest bias, while (FIML) has the smallest bias among the four estimators of the structural coefficients for all four error distributions. The most biased (OLS) estimator of the structural coefficients retains the Gauss Markov property of minimum variance. The large bias of (OLS), however, more than offset the small variance, so that (OLS) has the largest mean squared errors of the four structural estimators.

In the case where the four estimators was judged according to their ability to predict the mean values of each endogenous variable conditional on the exogenous variables, it was found that FIML, with a few exceptions, is the most efficient, and (OLS) is the most biased conditional predictors of the mean values of endogenous variables. [See Raj 1980 for more details.]

6. MODIFIED MONTE CARLO METHODS

As pointed out by Mikhail (1972) direct Monte Carlo studies conducted to simulate the small-sample properties of estimators of simultaneous equations often lead to indeterminate and sometimes contradictory results. These discrepancies are perhaps due mainly to sampling errors inherent in the simulation process used.

Makhail (1972) applied Antithetic and Control Variates to obtain a modified Monte Carlo technique for simulating small-sample properties of simultaneous equation estimators.

The direct Monte Carlo method (for a particular case) can be summarized as follows: suppose we wish to determine Θ which represents the bias of an estimator. Then Θ is a population parameter and through our experiments we seek to determine accurate estimates of it (e.g. in a minimum mean square error sense). Each replication of the experiment yields one observation and frequently, the chosen parameter is the mean value of a large number of random replications (N, say). The sample mean is unbiased, with a variance which decreases as a function of $\frac{1}{N}$ and, in the absence of other information, would be the minimum mean square error of Θ. This, however, is not a valid argument for computer simulated experiments since the basic random numbers are known and can be reused as antithetic varieties [see Hammersley and Hadscomb (1964) and Henry and Trevedi (1972) for further details].

For example, in the case where ξ_t (say) is $N(0, \sigma^2)$, one possible antithetic pair is provided by $[\xi_t, -\xi_t]$ since they are perfectly negatively correlated and hence exactly offset each other's variability without affecting the unbiasedness of the outcome. Every trial is thus performed twice, once using $[\xi_1, \xi_2, \ldots, \xi_T]$ and once with $[-\xi_1, -\xi_2, \ldots, -\xi_T]$, and the resulting estimates $\hat{\theta}_1$ and $\hat{\theta}_2$ (say) are averaged to yield

$$(7.6.2) \qquad \bar{\theta} = \frac{\hat{\theta}_1 + \hat{\theta}_2}{2}.$$

Since

$$(7.6.3) \qquad \text{Var}\,(\bar{\theta}) = \frac{1}{4}\,[\text{var}\,\hat{\theta}_1 + \text{var}\,\hat{\theta}_2 + 2\,\text{cov}\,(\hat{\theta}_1, \hat{\theta}_2)].$$

By forcing the covariance to be negative through our choice of error terms, var $(\bar{\theta})$ is reduced below the random sampling outcome.

Recall that in the direct Monte Carlo experiments we usually confine ourselves to random numbers α_i (say) uniformly distributed between zero and one. In the procedure described above, we can now use the set $\alpha_i^* = 1 - \alpha_i$ to get an antithetic estimator θ'' (say) which is negatively correlated with θ' (say), the initial Monte Carlo estimator of θ (say).

Thus, the numbers α_i^* will again be distributed in the range (0, 1) and the random normal deviates corresponding to the number α_i^* will have the same mean and the same variance as the random number deviates corresponding to the number α_i.

The information obtained from the two mutually antithetic estimates is then combined to give a better knowledge of the parameters of the distribution.

Mikhail (1972) altered the equation 7.4.1 (a), (b) by adding two more exogenous variables, one in each equation in order to counteract the problem of zero bias and the nonexistence of finite second moments for the (2SLS) estimator.

The new model is denoted by

$$(7.6.4) \qquad y_{1t} = \beta_{12}y_{2t} + \gamma_{11}x_{1t} + \gamma_{12}x_{2t} + \gamma_{15}x_{5t} + \gamma_{10} + u_{1t}$$

$$y_{2t} = \beta_{21}y_{1t} + \gamma_{23}x_{3t} + \gamma_{24}x_{4t} + \gamma_{26}x_{6t} + \gamma_{20} + u_{2t}$$

The exogenous variables used in the simulation process were all random drawings from the uniform distribution

$$(7.6.5) \qquad dF(x_i) = \frac{1}{b_i - a_i}\, dx_i$$

in the ranges 5 to 10, 5 to 5; 15 to 30, 2 to 8, 8 to 22 and 10 to 14. A sample of 20 observations generated in this fashion was kept fixed in repeated samples. [See Mikhail (1972) for a table of the means and covariance matrix of the exogenous variables.]

In his comparison between direct simulation and the two-antithetic method, where comparisons are made of the bias, variance, mean square error and mean absolute error. [See Tables 1, 2, 3 and 4 of Mikhail, 1972.] The results listed in his tables indicate the two-antithetic method performed better than the direct simulation method in the estimation of the bias. In estimating the variance and the mean absolute error, the two-antithetic method was not significantly better than the direct simulation method. Mikhail (1972) also applied Control Variates (CV) for two- and three-stage least squares and FIML for static models, as opposed to dynamic models, [see Hendry and Harrison (1974) for details] and obtained significant efficiency gains for estimating both the means and variances.

The principle underlying Control Variates is as follows: The basic idea is to find an auxiliary statistic Θ^* (say) such that Θ^* and $\hat{\Theta}$ (the original Monte Carlo estimator of Θ) are positively correlated, but the distribution (at least for the first few moments) of Θ^* should be obtained analytically. Then instead of investigating $\hat{\Theta}$, the direct simulation or Monte Carlo estimator of Θ, we use

$$\Theta^{**} = \hat{\Theta} - \Theta^* + E(\Theta^*)$$

by simulating both $\hat{\Theta}$ and Θ^* and computing Θ^{**} by subtraction. It should be noted that

$$E(\Theta^{**}) \equiv E(\hat{\Theta}) ,$$

which is what we wish to estimate. However,

$$\mathrm{Var}\,(\Theta^{**}) = \mathrm{Var}\,(\hat{\Theta}) + \mathrm{Var}\,(\Theta^*) - 2\,\mathrm{Cov}\,(\hat{\Theta}, \Theta^*)$$

which will be less than $\mathrm{Var}\,(\hat{\Theta})$ if

$$\mathrm{Cov}\,(\hat{\Theta}, \Theta^*) > \frac{1}{2}\,\mathrm{Var}\,(\Theta^*).$$

The major problem here is in locating an appropriate Θ^*. Table 1 of Makhail (1972) showed that the control variate method is better than the antithetic method for estimating the variance, mean square error and absolute mean square error.

In the case where combined control/antithetic variates were used, they were found to be superior in estimating dispersions but give exactly the same results on the bias as that obtained by the two-antithetic method.

REFERENCES

1. Agren, A. and H. Wold, On the structure and estimation of general interdependent systems. In Multivariate Analysis 2 (P.R. Krishnaiah, ed.), pp. 543-565, New York, Academic Press, 1969.

2. Amemiya, T., On the Use of Principle Components of Independent Variables in Two State Least Squares Estimation, International Economic Review, 7, (1966), pp. 283-303.

3. Anderson, T.W., An Introduction to Multivariate Statistical Analysis, New York, John Wiley and Sons, Inc., 1958.

4. Anderson, T.W., An Asymtoptic Expansion of the Distribution of the Limited Information Maximum Likelihood Estimate of a Coefficient in a Simultaneous Equation System, Journal of the American Statistical Association, 69, (1974), pp. 565-573.

5. Anderson, T.W., Asymototic Expansions of the Distribution of Estimates in Simultaneous Equations for Alternative Parameter Sequences, Econometrica, 45, (1977), pp. 509-518.

6. Anderson, T.W. and H. Rubin, Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations, Annals of Mathematical Statistics, 20, (1949), pp. 46-63.

7. Anderson, T.W. and H. Rubin, The Asymptotic Properties of Estimates of the Parameters of a Single Equation in a Complete System of Stochastic Equations, Annals of Mathematical Statistics, 21, (1950), pp. 570-582.

8. Anderson, T.W. and T. Sawa, Distributions of Estimates of Coefficients of a Single Equation in a Simultaneous System and their Asymptotic Expansions, Econometrica, 41, (1973), pp. 683-714.

9. Andreson, T.W. and T.Sawa, Two-Stage Least Squares: In which Direction should the Residuals be Minimized? Journal of the American Statistical Association, 72, (1977), pp. 187-192.

10. Basmann, R.L., The Computation of Generalized Classical Estimates of Coefficients in a Structural Equation, Econometrica, 27, (1959), pp. 72-81; Errata, Econometrica, 28, (1960), p. 170.

11. Basmann, R.L., A Note on the Exact Finite Sample Frequency Functions of Generalized Classical Linear Estimators in a Leading Three-Equation Case, Journal of the American Statistical Association, 58, (1963), pp. 161-171.

12. Basmann, R.L., A Note on the Exact Finite Sample Frequency Functions of Generalized Classical Linear Estimators in Two Leading Overidentified Cases, Journal of the American Statistical Association, 56, (1961), pp. 619-636.

13. Basmann, R.L., An Expository Note on the Estimation of Simultaneous Structural Equations, Biometrics, 16, No. 3, (1960), pp. 464-480.

14. Basmann, R.L., On Finite Sample Distributions of Generalized Classical Linear Identifiability Test Statistics, Journal of the American Statistical Association, 55, (1960), pp. 650-659.

15. Basmann, R.L. and L.T. Bakony, Letter to the Editor, Econometrica, 29, (1961).

16. Bergstrom, A.R., The Covariance Matrix of the Limited Information Maximum Likelihood Estimator, Econometrica, 40, (1972), pp. 899-900.

17. Bronsard, L.S. and A.P. Barten, 2SLS Estimation with Shifts in Structural Form, Econometrica, (1970), pp. 938-941.

18. Brown, T.M., Simultaneous Least Squares: A Distribution Free Method of Equation System Structure Estimation, International Economic Review, 1, (1960), pp. 173-191.

19. Brown, T.M., Simultaneous Least Squares and Invariance under Changes of Units of Measurement, International Economic Review, 8, (1967), pp. 97-102.

20. Brown, G.F., The Small-Disturbance Asymptotic Moment Matrix of k-Class Estimates of Parameters of Different Equations in a Complete System of Simultaneous Linear Equations, Ann. Inst. Statist. Math., 27, (1975), pp. 463-472.

21. Brown, G.F., J.B. Kadane and J.G. Ramage, The Asymptotic Bias and Mean-Squared Error of Double k-Class Estimators when the Disturbances are Small, International Economic Review, 15, (1974), pp. 667-679.

22. Brown, G.F., J.G. Ramage and V.K. Srivastava, Disturbance-Variance Estimation in Simultaneous Equation Systems, Technical Report, No. 68, Dept. of Statistics, Carnegie-Mellon University, (1972).

23. Brundy, J.M. and D.W. Jorgenson, Efficient Estimation of Simultaneous Equations by Instrumental Variables, Review of Economics and Statistics, 53, (1971), pp. 207-224.

24. Byron, R.P., Limited Information Using the Newton Solution, Australian Economic Papers, (1972), pp. 112-115.

25. Carter, R.A.L., The Exact Distribution of an Instrumental Variables Estimator, International Economic Review, 17, (1976), pp. 228-233.

26. Carter, R.A.L. and A.L. Nagar, Coefficients of Correlation for Simultaneous Equation Systems, Journal of Econometrics, 6, (1977), pp. 39-50.

27. Chow, C.G., A Comparison of Alternative Estimators for Simultaneous Equations, Econometrica, 32, (1964), pp. 532-553.

28. Chow, C.G., Two Methods of Computing Full-Information Maximum Likelihood Estimates in Simultaneous Stochastic Equations, International Economic Review, 9, (1969), pp. 100-112.

29. Christ, C.F., Simultaneous Equation Estimation: Any Verdict Yet? Econometrica, 28, (1960), pp. 835-845.

30. Christ, C.F., Econometric Models and Methods, New York, John Wiley and Sons, Inc., (166).

31. Court, R.H., Efficient Estimation of the Reduced-Form from Incomplete Econometric Models, Review of Economic Studies, 40, (1973), pp. 411-418.

32. Court, R.H. and N.C. Kokwani, Reduced-form Coefficient Estimation and Forecasting from a Simultaneous Equation Model, Australian Journal of Statistics, 14, (1972), pp. 143-160.

33. Cragg, J.G., On the Sensitivity of Simultaneous Equation Estimators to the Stochastic Assumptions of the Models, Journal of the American Statistical Association, 61, (1966), pp. 136-151.

34. Cragg, J.G., On the Relative Small Sample Properties of Several Structural Equation Estimators, Econometrica, 35, (1967), pp. 89-110.

35. Cragg, J.G., Some Effects of Incorrect Specification on the Small-Sample Properties of Several Simultaneous-Equation Estimators, International Economic Review, 9, (1968), pp. 63-91.

36. Cragg, J.G., Small-Sample Performances of Various Simultaneous-Equation Estimators in Estimating the Reduced-Form, Metroeconomica, 19, (1967), pp. 76-93.

37. Dent, W., Information and Computation in Simultaneous Equations Estimation, Journal of Econometrics 4, (1976), pp. 89-95.

38. Dhrymes, P.J., Asymptotic Properties of Simultaneous Least Squares Estimators, International Economic Review, 13, (1972), pp. 201-211.

39. Dhrymes, P.J., Small-Sample and Asymptotic Relations between Maximum-likelihood and Three-stage Least Squares, Econometrica, 41, (1973), pp. 357-364.

40. Dhrymes, P.J. and H. Erlat, Asymptotic Properties of Full Information Estimators in Dynamic Autoregressive Simultaneous Equation Models, Journal of Econometrics, 2, (1974), pp. 247-260.

41. Dhrymes, P.J., Econometrics: Statistical Foundations and Applications, New York, Springer-Verlag, Inc., (1974).

42. Dhrymes, P.J., Restricted and Unrestricted Reduced Forms: Asymptotic Distribution and Relative Efficiency, Econometrica, 41, (1973), pp. 119-134.

43. Dhrymes, P.J. and V. Pandit, The Asymptotic Distribution of the First Iterate of the Iterated Two-stage Least Squares Estimates, Journal of the American Statistical Association, (1971).

44. Dhrymes, P.J., An Identity Between Double k-Class and Two-Stage Least Squares Estimators, International Economic Review, 10, (1969), pp. 114-117.

45. Dhrymes, P.J., Alternative Asymptotic Tests of Significance and Related Aspects of 2SLS and 3SLS Estimated Parameters, Review of Economic Studies, 30, No. 106, (1969), pp. 213-226.

46. Drettakis, E.G., Missing Data in Econometric Estimation, Review of Economic Studies, (1973), pp. 537-545.

47. Dowling, J.M. and F.R. Glahe, Readings in Econometric Theory, Colorado: Colorado Associated University Press, 1970.

48. Dwivedi, T.D., Nonexistence of Moments of Limited Information Maximum Likelihood Estimator in Simultaneous Equations, Dept. of Math., Concordia University, Montreal, 1980.

49. Dwivedi, T.D. and V.K. Srivastava, Exact Finite Sample Properties of Double k-Class Estimators in Simultaneous Equations, Dept. of Math., Concordia University, Montreal, 1980.

50. Eisenpress, H., Note on the Computation of Full-Information Maximum Likelihood Estimates of Coefficients of a Simultaneous System, Econometrica, 30, (1962), pp. 343-348.

51. Fait, R.C., The estimation of simultaneous equations models with logged endogenous variables and first order serially correlated errors, Econometrica, 38, (1970), pp. 507-516.

52. Fait, P.C., Efficient estimation of simultaneous equations with autoregressive errors by instrumental variables, Review of Economics and Statistics, 54, (1972), pp. 444-449.

53. Farebrother, R.W., The graph of a k-class estimator, The Review of Economic Studies, 41, (1974), pp. 533-538.

54. Farebrother, R.W. and N.E. Savin, The graph of the k-class estimator (an algebraic and statistical interpretation), Journal of Econometrics, 2, (1974), pp. 373-388.

55. Fisher, F.M., Generalizations of the Rank and Order Conditions for Identifiability, Econometrica, 27, (1959), pp. 431-447.

56. Fisher, F.M., On the Cost of Approximate Specification in Simultaneous Equation Estimation, Econometrica, 29, (1961), pp. 139-170.

57. Fisher, F.M., The Identification Problem in Econometrics, New York, McGraw-Hill Book Company, (1966).

58. Fisher, F.M., The Relative Sensitivity to Specification Error of Different k-Class Estimators, Journal of the American Statistical Association, 61, (1966), pp. 345-356.

59. Fisher, F.M., Restrictions on the Reduced Form and the Rank and Order Conditions, International Economic Review, 7, (1966), pp. 77-82.

60. Fisher, F.M., Approximate Specification and the Choice of a k-Class Estimator, Journal of the American Statistical Association, 62, (1967), pp. 1265-1276.

61. Fisher, F.M. and B.M. Mitchell, The Choice of Instrumental Variables in the Estimation of Economy-wide Econometric Models: Some Further Thoughts, International Economic Review, 11, (1970), pp. 226-234.

62. Fisk, P.R., Stochastically Dependent Equations, Statistical Monograph, 21, New York, Charles Griffin, 1967.

63. Fuller, W.A., Some properties of a modification of the Limited Information Estimator, Econometrica, 45, No. 4, (1977), pp. 939-953.

64. Geary, R.C., The Frequency Distribution of the Quotient of Two Normal Variates, Journal of the Royal Statistical Society, 53, part 3, 1930, pp. 442-446.

65. Goldberger, A.S., Econometric Theory, New York, John Wiley and Sons, Inc., 1964.

66. Goldberger, A.S., An Instrumental Variable Interpretation of k-Class Estimation, Econometric Journal, 13, (1965), pp. 424-431.

67. Goldberger, A.S., A.L. Nagar and H.S. Odeh, The Covariance Matrices of Reduced-Form Coefficients and of Forecasts for a structural Econometric Model, Econometrica, 29, (1961), pp. 556-573.

68. Goldberger, A.S. and I. Olkin, A Minimum-Distance Interpretation of Limited-Information Estimation, Econometrica, 39, (1971), pp. 635-639.

69. Graybill, F.A., An Introduction to Linear Statistical Models, Vol. I, New York, McGraw-Hill Book Company, Inc., 1961.

70. Green, R.J., Alternative significance test for (2SLS) estimated parameters: Some Monte Carlo Evidence, Metroeconomica, 4, (1973), pp. 181-201.

71. Gujarati, D., Basic Econometrics, New York, McGraw-Hill Book Company, 1978.

72. Haavelmo, T., The Statistical Implications of a System of Simultaneous Equations, Econometrica, 11, (1943), pp. 1-12.

73. Haavelmo, T., The Probability Approach to Econometrics, Supplement to Econometrica, 12, (1944).

Hammersley, J.M. and D.C. Handscomb, Monte Carlo Methods, London, Methuen, 1964.

74. Hausman, J.A., An instrumental variable approach to full-information estimators for linear and certain non-linear econometric models, Ecnoometrica, 43, (1975), pp. 325-349.

75. Hausman, J.A., Errors in Variables in Simultaneous Equation Models, Journal of Econometrics, 5, (1977), pp. 389-401.

76. Hendry, D.F., On asymptotic theory and finite sample experiments, Economica, 60, (1973), pp. 210-217.

77. Hendry, D.F., The Structure of Simultaneous Equations Estimators, Journal of Econometrics, 4, (1976), pp. 51-88.

78. Hendry, D.F. and P.K. Trivedi, Maximum likelihood estimation of difference equations with moving average errors: A simultaneous study, The Review of Economic Studies, 39, (1972), pp. 117-145.

79. Hendry, D.F. and R.W. Harrison, Monte Carlo methodology and the small sample behaviour of ordinary and two-stage least squares, Journal of Econometrics, 2, (1974), pp. 151-174.

80. Hildreth, C., Simultaneous Equations: Any Verdict Yet? Econometrica, 28, (1960), pp. 846-854.

81. Hymans, S.A., Simultaneous Confidence Intervals in Econometric Forecasting, Econometrica, 36 (1968), pp. 18-30.

82. Johnston, J., Econometric Methods (second edition), New York, McGraw-Hill Book Company, 1972.

83. Judge, G.G. and Thomas A. Yancey, The Use of Prior Information in Estimating the Parameters of Economic Relationships, Metroeconomica, 21, (1969), pp. 97-139.

84. Kabe, D.G., A note on the exact distributions of the estimators (GCL) estimators in two leading overidentified cases, Journal of the American Statistical Association, 58, (1963), pp. 535-537.

85. Kabe, D.G., On the exact distributions of the (GCL) estimates in a leading three-equation case, Journal of the American Statistical Association, 59, (1964), pp. 881-894.

86. Kadane, J.B., Testing the Overidentifying Restrictions when the Disturbances are Small, Journal of the American Statistical Association, 65, (1970), pp. 182-185.

87. Kadane, J.B., Comparison of k-Class Estimators when the Disturbances are Small, Econometrica, 39, No. 5, (1971), pp. 723-737.

88. Kadiyala, K.R., An exact small-sample property of the k-class estimators, Econometrica, 38, (1970), pp. 930-932.

89. Kakwani, N.C., The Unbiasedness of Zellner's Seemingly Unrelated Regression Equation Estimators, Journal of the American Statistical Association, 62, (1967), pp. 141-142.

90. Kakwani, N.C., A Note on the Unbiasedness of a Mixed Regression Estimator, Econometrica, 36, (1968), pp. 610-611.

91. Kelejian, H.H., Two stage least squares and econometric systems linear in parameters but non-linear in endogenous variables, Journal of the American Statistical Association, 66, (1971), pp. 373-374.

158. Srivastava, V.K., Disturbance Variance Estimation in Simultaneous Equations by k-Class Method, Ann. Inst. Stat. Math. 32, (1971), pp. 437-449.

159. Srivastava, V.K., Maximum-Likelihood and Least-Squares Approaches to Simultaneous Estimation of Simultaneous Equations, Trab. Est. Invest. Oper., 22, (1971), pp. 79-88.

160. Srivastava, V.K., Disturbance Variance Estimation in Simultaneous Equations when Disturbances are Small, Journal of the American Statistical Association, 67, (1972), pp. 164-168.

161. Srivastava, V.K., Three-Stage Least Squares and Generalized Double k-Class Estimators: A Mathematical Relationship, International Economic Review, 12, (1971), pp. 312-316.

162. Srivastava, V.K. and R. Tiwari, Two-Stage and Three-Stage Least Squares Estimation of Dipression Matrix of Disturbances in Simultaneous Equations, Ann. Inst. Stat. Math., 28, (1976), pp. 411-428.

163. Srivastava, V.K. and R. Tiwari, Efficiency of Two-Stage and Three-Stage Least Squares Estimators, Econometrica, 46, (1978), pp. 1495-1498.

164. Srivastava, V.K. and R. Tiwari, A Simple Derivation of the Identity Connecting Double k-Class and Two-Stage Least Squares Estimators, Sankhya: The Indian Journal of Statistics, 39, Series C, Pt. 2, (1977), pp. 1-2.

165. Summers, R., A Capital Intensive Approach to Small Sample Properties of Various Simultaneous Equation Estimators, Econometrica, 33, (1965), pp. 1-41.

166. Takeuchi, K., Exact sampling moments of the ordinary least squares, instrumental variables and two-stage least squares estimators, International Economic Review, 11, (1970), pp. 1-12.

167. Tamura, H., and W.A. Spivey, Generalized Simultaneous Equation Models, International Economic Review, 11, (1970), pp. 216-223.

168. Theil, H., Principles of Econometrics, New York, John Wiley and Sons, Inc., 1971.

169. Ulla, A. and A.L. Nagar, Exact moments of two-stage least squares estimator and their approximations, Working Paper, Delhi School of Economics, (1972).

170. Ullah, A. and A.L. Nagar, Note on the Approximate Skewness and Kurtosis of the (2SLS) estimator, Indian Economic Review, 8, (1973), pp. 69-80.

171. Ullah, A. and A.L. Nagar, The Exact Mean of the (2SLS) Estimator of the Structural Parameters in an Equation having Three Endogenous Variables, Econometrica, 42, (1974), pp. 749-758.

172. Vandaele, W. and A. Morgan, On Testing Hypotheses in Simultaneous Equation Models, Journal of Econometrics, 2, (1974), pp. 55-65.

173. Wagner, H.M., A Monte Carlo Study of Estimates of Simultaneous Linear Structural Equations, Econometrica, 26, (1958), pp. 117-133.

174. Wold, H., A Fix-Point Theorem with Econometric Background, Ark. Mat., 6, (1966), pp. 209-240.

175. Wold, H. and E. Lyttkens, Nonlinear Iterative Partial Least Squares (NIPALS) Estimation Procedures, Bull. Intern. Statist. Inst., 43 (1), (1969), pp. 29-51.