T.R. NO. 4

Technical Report

9c

EFFICIENCY OF SEVERAL GENERALISED
LEAST SQUARES ESTIMATORS

by

P.N.Misra

विद्याविनियोगाद्विकासः
IIM
AHMEDABAD

INDIAN INSTITUTE OF MANAGEMENT
AHMEDABAD

EFFICIENCY OF SEVERAL GENERALISED LEAST SQUARES ESTIMATORS

by

P.N.Misra

July 1973

Indian Institute of Management
Ahmedabad

No. 4

To

Chairman (Research)
IIMA

Technical Report

Title of the report .. EFFICIENCY OF SEVERAL GENERALISED LEAST SQUARES ESTIMATORS

Name of the Author ... Prof. P. N. Misra

Under which area do you like to be classified? Econometrics

ABSTRACT (within 250 words)

Econometric theory, as developed till today includes a number of estimation procedures whose small and large sample properties have been analysed at different places by different authors. Looking at these works one can find a few contradictions regarding efficiency of some powerful methods of estimation. Present paper highlights these contradictions and then provides a systematic analysis of xxxxxx asymptotic efficiency of a number of estimation procedures. Effect of orthogonality of explanatory variables on the efficiency of various generalised least squares estimators is also examined.

Please indicate restrictions if any that the author wishes to place upon this note ....

Date 3/8/73

P. N. Misra
Signature of the Author

EFFICIENCY OF SEVERAL GENERALISED LEAST SQUARES ESTIMATORS*

R.N. Misra

1. Introduction

Most of popular methods of estimating economic relationships are based upon Aitken's[1] generalised least squares (GLS) procedure. For example, seemingly unrelated regression equations (SURE) estimator proposed by Zellner[12] is different from its ordinary least squares (OLS) counterpart because it utilizes the knowledge of statistical dependence of disturbances in different regressions. Similarly, two stage least squares (2SLS) method of estimation proposed by Theil[11] or generalized classical linear (GCL) estimator proposed by Basmann[2] (both being equivalent) and three stage least squares (3SLS) estimator proposed jointly by Zellner and Theil[15] may be interpreted as generalised least squares estimators in Aitken's sense. Logically, then, the statistical properties of these estimators should have similar characteristics. For the sake of convenience, we shall henceforth, call the class of estimators namely, GLS, SURE, 2SLS, and 3SLS as generalised estimators.

A survey of econometric literature would reveal that OLS, GLS and SURE estimators are unbiased and 2SLS or GCL and 3SLS estimators are consistent. It may be pointed out that earlier investigation by Zellner[13] showed that, for two equations model, SURE estimator is unbiased provided independent variables in these regressions were orthogonal but later Kakwani[6] found that the estimator is unbiased for M equations model, even without orthogonality assumption, provided the disturbances follow continuous symmetric probability laws. Such uniformity is, however, lacking

---

* An earlier version of this paper was written while the author was at the University of Manchester, U.K.

if we survey the available analyses in respect of efficiency of generalised estimators as compared to their respective OLS counterparts. Cochrane and Orcutt[3] analysed the efficiency of GLS estimator numerically by considering two variable regression model while some other particular cases of temporal dependence have been discussed by Johnston[5]. These analyses indicate that efficiency of GLS estimator depends upon the magnitude of first order autocorrelation. While analysing asymptotic efficiency of SURE estimator Zellner[12] and Zellner and Huang[14] found that gain in its efficiency is substantial provided the set of independent variables in different regressions were orthogonal. Thus, if the latter condition were genuine, then, one may not see any advantage in computing SURE estimator unless one's observations on independent variables in different regressions were orthogonal. However, in actual practice the chance of observing orthogonality is very small. The asymptotic efficiency of 3SLS estimator has been analysed by Zellner and Theil[15] by considering only two equations of the complete system. They assumed one of the equations to be over-identified and the other to be just identified and concluded that 3SLS estimator of the coefficients of over-identified equation does not gain in its efficiency over 2SLS estimator. They arrived at the same conclusion by considering the complete system where one bloc of equations are just-identified and the rest over-identified. In the former case when both the equations are over identified, the authors deduced that 3SLS estimator gains in efficiency over 2SLS estimator provided the correlation coefficient between disturbances in the two equations is not zero. Three points may be noted in this connection. Firstly, the results of asymptotic moment matrices are not properly reported and hence the comparison of appropriate matrices has not been done by the authors Secondly, as we shall see later, the 2SLS and 3SLS estimators themselves are identical in case all the equations are just-identified and therefore, equivalence of their asymptotic moment matrices is obvious in this particular case. Thirdly, in case the complete system is a mixture of just and over-identified equations, then, in fact, the 3SLS and 2SLS estimators do not have identical asymptotic covariance matrix when we consider the over-identified equations only and ignore the rest as concluded by Zellner and Theil [15, pp.63-69]. Consequently, Narayanan's[7] computational procedure of 3SLS estimator too becomes doubtful because the results borrowed from [15] are erroneous. As regards the first point in case of SURE estimator too, the moment matrix of OLS estimator as used for analysing efficiency needs some correction. Further, the effect of the behaviour of independent or predetermined variables (as the case may be) on the efficiency of generalised estimators needs appropriate examination so that the virtues of these estimators are not under-valued due to mistaken apprehensions.

The purpose of this paper is mainly to analyse these issues further and to straighten up a few minor points at appropriate places. In fact, we show that the generalised estimators behave uniformly in respect of their efficiency also as they do in case of their central tendencies. First of all, in Section 2 we provide a brief survey of various generalised estimators in case of single, multiple regressions and simultaneous equations models and try to discover a common feature throughout. Also, in case of simultaneous equations, we analyse 2SLS as well 3SLS estimators under alternative assumptions regarding the identifiability of all or only a part of equations of the complete system. Then, in Section 3, we discuss briefly the concepts of asymptotic covariance matrix and relative efficiency. Section 4 contains derivation of relevant asymptotic covariance matrices while their comparisons are made in Section 5 to decide asymptotically more efficient estimators. The last Section contains a discussion on the effect of orthogonality of explanatory variables on the efficiency of various generalised estimators.

2. A Brief Survey of Generalised Estimator Procedures

We propose to restrict our discussion to linear econometric models only. As the underlying assumptions in case of estimators under consideration are too well known, we would not repeat all of them again in this paper. However, we would specify the assumptions regarding statistical distributions of the distrubance terms in respective cases.

2.2 General Linear Regression Model

We may write the general linear regression model as

(2.1) $y = X\beta + u$

where y is the column vector of observations on the left hand variable, X is the $T \times \Lambda$ matrix of observations on Λ explanatory variables, β is the $\Lambda \times 1$ coefficient vector, u is $T \times 1$ vector of disturbance terms and T is the size of sample.

It is well known that if the disturbances are homoscedastic and temporally independent, i.e.,

(2.2) $$Euu' = \sigma^2 \otimes I$$

where operator E denotes mathematical expectation, $\otimes$ means Kronecker product and I is TxT identity matrix, then, OLS estimator

(2.3) $$b_0 = (X'X)^{-1}X'y$$

is best linear unbiased estimator of the coefficient vector β. On the contrary, if

(2.4) $$Euu' = \begin{bmatrix} v_{11} & \cdots & v_{1T} \\ \cdot & & \\ \cdot & & \\ \cdot & & \\ \cdot & & \\ v_{T1} & \cdots & v_{TT} \end{bmatrix} = V$$

then, instead of b_0, the estimator

(2.5) $$\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}y$$

is best linear unbiased estimator of vector β. The estimator $\hat{\beta}$ is known as Aitken's[1] generalised least squares estimator and its computation depends upon the exact knowledge of the covariance matrix V. Rao[10] has shown that if one uses some estimator V of V and obtains the estimator

(2.6) $$b = (X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}y$$

of β then, b is not necessarily best linear unbiased unless $\hat{V} = V$.

2.2. Set of Linear Regressions Model

Several times we have a set of linear regressions so that the disturbances in different equations exhibit considerable degree of statistical dependence. We may write such a system of M equations as

(2.7) $$y_i = \bar{X}_1 \beta_i + u_i, \quad i = 1, \ldots, M,$$

where y_i is $T\times 1$ vector of observations on the dependent variable, X_i is $T\times \hat{}_i$ matrix of observations on the $\hat{}_i$ explanatory variables, β_i is $\hat{}_i \times 1$ vector of unknown coefficients and u_i is $T\times 1$ vector of disturbance terms. We may, alternatively, express (2.7) as

(2.8) $$y^* = X^* \beta^* + u^*$$

where

(2.9) $$Y^* = \begin{bmatrix} y_1 \\ \cdot \\ \cdot \\ \cdot \\ y_M \end{bmatrix}, \quad X^* = \begin{bmatrix} X_1 \ldots 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \ldots X_M \end{bmatrix} \quad \beta^* = \begin{bmatrix} \beta_1 \\ \cdot \\ \cdot \\ \cdot \\ \beta_M \end{bmatrix} \text{ and } u^* = \begin{bmatrix} u_1 \\ \cdot \\ \cdot \\ \cdot \\ u_M \end{bmatrix}$$

If the disturbances in different regressions are homoscedastic and contemporaneously independent, i.e.,

(2.10) $$Eu_i u_j' = \sigma_{ii} x^I \quad \text{for } i = j$$
$$= 0 \quad \text{for } i \neq j,$$

then, the OLS estimator b^*_0 of β^* in (2.8) could be written as

(2.11) $$b^*_0 = (X^{*\prime} X^*)^{-1} X^{*\prime} y^*$$

which can be seen to be identical to OLS estimator obtained from individual equations in (2.7). On the contrary, if u_i and u_j were independent, i.e.,

(2.12) $$Eu_i u_j = \sigma_{ij} \otimes I$$

for all i and j running from 1 to M, then, GLS estimator

(2.13) $$\beta^* = (X^{*\prime} \Sigma_1^{-1} X^*)^{-1} X^{*\prime} \Sigma_1^{-1} y^*$$

is best linear unbiased estimator β^* where

(2.14) $$\Sigma_1 = \Sigma \otimes I$$

and

(2.15) $$\Sigma = (\sigma_{ij})$$

is MxM matrix. The stimator in (2.13) could be computed only if the matrix Σ were known. Zellner [12] proposed a consistent estimator

(2.16) $$s_{ij} = T^{-1} \hat{u}_i' \hat{u}_j$$

of the element σ_{ij} of the matrix where $\hat{u}_i$ and $\hat{u}_j$ are obtained from

(2.17) $$\hat{u}^* = y^* - X^* b_0^* .$$

Thus, replacing by the matrix

(2.18) $$S = (s_{ij})$$

and writing

(2.19) $$S_1 = S \otimes I$$

we obtain SURE estimator b* of β*as

(2.20) $b^* = (X^{*'}S_1^{-1}X^*)^{-1}X^{*'}S_1^{-1}y^*$

2.3 Simultaneous Equations Model

Let us write a complete system of M structural equations in M jointly dependent and Λ predependent variables as

(2.21) $Y\Gamma + XB = U$

where Y and X are TxM and TxΛ matrices of observations on the jointly dependent and predetermined variables, respectively Γ and B are matrices of structural coefficients and U is the TxM matrix of structural disturbances.

The reduced form of this system is given by

(2.22) $Y = X\Pi + \bar{V}$

where

(2.23) $\Pi = -B\Gamma^{-1}$ and $V = U\Gamma^{-1}$

are ΛxM and TxM matrices of the reduced form coefficients and disturbances, respectively.

The disturbance matrix U can be written as

(2.24) $U = [u_1 \ldots u_M]$

where u_i's (i = 1, ..., M) are each Tx1 column vectors which we assume to be homoschedastic and contemporaneously dependent so that

(2.25) $Eu_iu_j' = \sigma_{ij}^* \otimes I$

for all values of i and j running from 1 to M. Thus, the complete covariance matrix of structural disturbances is

(2.26) $T^{-1}E\ U'U = \Sigma^*$

where Σ^* is defined similar to (2.15) and the disturbances are supposed to be temporally independent.

2.3.1 2SLS or GCL Estimator

It is well known that the identifiability conditions require the exclusion of some variables from each equation of the system. Thus, not all jointly dependent and predetermined variables are represented in each equation. Therefore, suppose $m_i + 1 \leq M$ jointly dependent and $1^{\ell}_{-1} \leq \Lambda$ predetermined variables enter the i-th equation. After rearranging terms we can write the i-th equation of the complete system (2.21) as

$$(2.27) \qquad y_i = Y_i \gamma_i + X_i \beta_i + u_i \ , \ i = 1, \ldots, M$$

where y_i is the vector of T observations on the left hand side jointly dependent variable, Y_i and X_i are Txm_i $Tx\ell_i$ matrices of observations on the right hand jointly dependent and predetermined variables, respectively, γ_i and β_i are coefficient vectors and u_i is the vector of disturbance terms.

The reduced form corresponding to Y_i on the right hand side of (2.27) may be written as

$$(2.28) \qquad Y_i = X \ II_i + \bar{V}_i$$

where II_i and $\bar{V}_i$ are obtained by suitable partioning of the matrices, II and V, respectively.

If we define

$$2.29) \qquad Z_i = [Y_i \ X_i] \text{ and } \delta_i = [\gamma_i \ \beta_i]$$

then, the equation (2.27) may be rewritten as

$$(2.30) \qquad y_i = Z_i \delta_i + u_i .$$

The 2SLS estimator

$$(2.31) \qquad d_{0i} = (Z_i' M^* Z_i)^{-1} Z_i' M^* y_i$$

of δ_i in equation

$$(2.32) \qquad X' y_i = X' Z_i \delta_i + X' u_i$$

is best linear unbiased in Aitken's sense because d_{0i} is simply GLS estimator obtained from (2.32). The matrix M^* in (2.31) is an indempotent matrix of the form

$$(2.33) \qquad M^* = X(X'X)^{-1}X'.$$

It may be noted that equation (2.32) implies premultiplication throughout equation (2.27) by matrix X' which is normally a singular matrix unless $T = \Lambda$. One may, therefore, object to such a procedure. However, the same estimator d_{oi} could be obtained by applying OLS procedure to equation

$$(2.34) \qquad M^* y_i = M^* Z_i \delta_i + u_i$$

which implies that not only right hand jointly dependent variable but left hand variable too in equation (2.27) are corrected for their stochastic parts, arising out of the fact that the equation belongs to a complete system.

Further, the matrix Z_i is of order $T x n_i$ ($n_i = m_i + I_1$) and estimator d_{0i} will exist if

$$(2.35) \qquad \text{Rank } (M^* Z_i) = n_i \leq \Lambda \leq T$$

which is one of the well known assumptions underlying OLS procedure applied to equation (2.34). Incidentally, the right hand side of (2.35) is the order condition of identifiability of equation (2.27). In particular, when (2.27) is just-identified, i.e., $n_i = A$, then,

$$(2.36) \qquad d_{0i} = (X'Z_i)^{-1}X'y_i$$

where use has been made of (2.31) and the fact that the matrix $X'Z_i$ is nonsingular for the just-identified case.

2.3.2 3SLS Estimator

The 2SLS estimator as obtained from equation (2.34) completely ignores the dependence of disturbances in different equations. We can use this information by following exactly the procedure of Section 2.2. We rewrite all the equations in (2.34) as

$$(2.37) \qquad Hy = HZ\delta + u$$

where

$$(2.38) \quad H=\begin{bmatrix} M^* & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & M^* \end{bmatrix}, Z=\begin{bmatrix} Z_1 & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & Z_M \end{bmatrix}, y=\begin{bmatrix} y_1 \\ \vdots \\ y_M \end{bmatrix}, \delta=\begin{bmatrix} \delta_1 \\ \vdots \\ \delta_M \end{bmatrix}, u=\begin{bmatrix} u_1 \\ \vdots \\ u_M \end{bmatrix} \quad \theta \ \delta$$

Then, GLS procedure as applied to equation (2.37) yields the estimator

$$(2.39) \qquad \hat{\delta} = (Z'\Sigma_1^{*-1}HZ)^{-1}Z'\Sigma_1^{*-1}Hy$$

where use has been made of (2.25),

$$(2.40) \qquad \Sigma_1^* = \Sigma^* \otimes I$$

and Σ^* is same as defined in (2.36). In particular, when $\Sigma^*=I$, then, the GLS estimator $\hat{\delta}$ in (2.39) reduces to

$$(2.41)\qquad d_0 = (Z'HZ)^{-1}Z'Hy$$

which is in fact 2SLS estimator of δ and could be obtained by applying OLS procedure to equation (2.37). The estimator in (2.39) could be computed only if the matrix Σ^* were known. Zellner and Theil[15] proposed a consistent estimator S^* of Σ^* where the element s^*_{ij} of S^* is defined as

$$(2.42)\qquad s^*_{ij} = T^{-1}\tilde{u}_i'\,\tilde{u}_j$$

where $\tilde{u}_i$ and $\tilde{u}_j$ are obtained from

$$(2.43)\qquad \tilde{u} = Hy - HZd_0$$

d_0 being same as defined in (2.41). Then, 3SLS estimator d of δ can be defined as

$$(2.44)\qquad d = (Z'S_I^{*-1}HZ)^{-1}Z'S_I^{*-1}Hy$$

where matrix Σ_I^* in (2.39) has been replaced by S_I^* and

$$(2.45)\qquad S_I^* = S^* \otimes I.$$

Thus, we observe a complete analogy between OLS and SURE estimators on the one side and 2SLS and 3SLS estimators on the other provided we consider equations (2.34) for obtaining the latter estimators and note that the regressors in (2.34) are stochastic. Further, we note that the matrix Z is of size $MT \times n$ $(n = \sum_{i=1}^{M} n_i)$;

therefore, nonsingularity of the matrix $(Z'S_I^{*-1}HZ)$ implies that

$$(2.46)\qquad \text{Rank } (Z'S_I^{*-1}HZ) = n \le M\ell \le MT$$

and eclusion of all the identities from the system (2.21) because otherwise the zero disturbances in the identities would render the matrix S^* to be singular. The right hand side of relation (2.46) provides order condition of identifiability of the complete system and would hold true provided (2.35) were true for each equation i.e., each equation of the system were identified.

2.3.3. 3SLS Estimators for Some Special Cases of the Complete System

In this section we shall consider the complete system when all or some of the equations are just identified. First of all, we consider the case when all the equations are just-identified, i.e., $n = M_1$. In that case the matrix ZF will be square and nonsingular where

$$(2.47) \quad F = \begin{bmatrix} X \cdots 0 \\ \\ 0 \quad \cdot X \end{bmatrix}$$

Combining (2.47) with (2.38) and (2.33) we observe that

$$(2.48) \quad H = F\,(F'F)^{-1}F' \;;$$

then, using (2.48) and (2.44) and remembering that ZF is nonsingular we derive after some adjustment.

$$(2.49) \quad d = (F'Z)^{-1}F'y$$

which is same as 2 SLS estimators in (2.36)

Next, we consider the complete system when some of its equations are just-identified and the rest are over-identified. Without any loss of generality let us assume that equations in (2.37) are arranged in such a manner that upper block of M_1 equations are just-identified and the lower block of M_2 equations are over-identified. We also suppose that $M_1+M_2 = M$. Further, let us partition the matrices and vectors Z, F, H, d, y such that upper blocks corresponding to just-identified equations are denoted by Z^*_1, F_1, H_1, d_A, y^*_1 and lower blocs corresponding to over-identified equations are denoted by Z^*_2, F_2, H_2, d_B, y^*_2, respectively. Similarly, the matrix S^* is partitioned as

$$(2.50)\qquad S^{*-1} = \begin{bmatrix} S^{-11} & S^{*12} \\ S^{*21} & S^{*22} \end{bmatrix}$$

and the following notations are adopted for convenience in writing. We write

$$(2.51)\qquad S^{*\,ij}_1 = S^{*ij} \theta I,\ i,j = 1,2.$$

The, we can rewrite the normal equations corresponding to estimator d, defined in (2.44), as

$$(2.52)\qquad \begin{aligned} Z^{*\prime}_1 S^{*\,11}_1 H_1 Z^*_1 d_A + Z^{*\prime}_1 S^{*12}_1 H_2 Z^*_2 d_B &= Z^{*\prime}_1 S^{*11}_1 H_1 y^*_1 + Z^{*\prime}_1 S^{*12}_1 H_2 y^*_2 \\ Z^{*\prime}_2 S^{*21}_1 H_1 Z^*_1 d_A + Z^{*\prime}_2 S^{*22}_1 H_2 Z^*_2\, d_B &= Z^{*\prime}_2 S^{*21}_1 H_1 y^*_1 + Z^{*\prime}_2 S^{*22}_1 H_2 y^*_2 \end{aligned}$$

where use has been made of (2.50) and (2.51). Since matrix $Z^{*\prime}_1 S^{*11}_1 H_1 Z^*_1$ corresponds to just-identified equations only and therefore nonsingular. We can multiply by its inverse throughout the upper equation of (2.52) and solve for d_A in terms of d_B. Then, substituting for d_A in the second equation of (2.52), we obtain after some adjustment.

$$(2.53)\qquad d_B = (Z^{*\prime}_2 G_2 H_2 Z^*_2)^{-1} Z^{*\prime}_2 G_2 H_2 y^*_2$$

where

(2.54) $G_2 = S_1^{*22} - S_1^{*21} F_1 (Z_1^{*\prime} S_1^{*11} F_1)^{-1} Z_1^{*\prime} S_1^{*12}$

and use has been made of (2.48), the nonsingularity of the matrices $F_1' Z_1^*$ and $Z_1^{*\prime} S_1^{*11} F_1$ and

(2.55) $Z_2^{*\prime} S_1^{*21} H_1 Z_1^* (Z_1^{*\prime} S_1^{*11} H_1 Z_1^*)^{-1} Z_1^{*\prime} S_1^{*11} H_1 = Z_2^{*\prime} S_1^{*21} H_1 .$

Further, combining (2.53) with either of the equations in (2.52) we may solve for d_A.

It can be easily verified that

(2.56) $S_1^{*11} F_1 = F_1 S_1^{*11}$

and

(2.57) $S_1^{*12} H_2 = F_1 [S_1^{*12} \otimes (X'X)^{-1} X'] .$

Therefore, using (2.56) and (2.57) we have

(2.58) $G_2 H_2 = S_1^{*22} H_2 - S_1^{*21} F_1 (Z_1^{*\prime} S_1^{*11} F_1)^{-1} Z_1^{*\prime} S_1^{*12} H_2$

$= S_1^{*22} H_2 - S_1^{*21} F_1 (S_1^{*11})^{-1} (Z_1^* F)^{-1} Z_1^{*\prime} F_1 [S_1^{*12} \otimes (X'X)^{-1} X']$

$= S_1^{*22} H_2 - S_1^{*21} (S_1^{*11})^{-1} F_1 [S_1^{*12} \otimes (X'X)^{-1} X']$

$= [S_1^{*22} - S_1^{*21} (S_1^{*11})^{-1} S_1^{*12}] H_2$

It is clear from (2.58) and (2.54) and d_B depends upon S_I^{*11} and S_I^{*12} which in turn depend upon the elements of the matrix S* including those ones also which are computed by using 2SLS estimates of the residuals in the just-identified equations. Accordingly, observations on all the variables involved in the just-identified set of equations are necessarily needed to compute d_B and it will be incorrect to compute 3SLS estimators from the set of over-identified equations only. 2/

3. Concepts of Asymptotic Covariance Matrix and Relative Efficiency

There appears to be some confusion regarding asymptotic covariance matrix and relative asymptotic efficiency in econometric literature. Therefore, it may be useful to state these concepts clearly before proceeding to further analysis.

Suppose T observations are available on a variable x, following normal probability law, and a maximum likelihood estimator b_T is obtained to estimate a parameter Θ of the parent population; then, it is well known (9, Section 5f.2(i)) that b_T is consistent estimator of Θ or alternatively,

(3.1) $\text{Plim } b_T = \Theta$

where 'Plim' means probability limit, i.e., the estimator coverages to Θ in probability. This also implies that asymptotic distribution of b_T collapses around Θ or in other words its variance tends to zero. Thus, if there are more than one consistent estimators, say, b_T and b_T^* of the same parameter Θ and based upon the same T observations on x then, the choice of the better one would be difficult in case of large T because in that case not only biases but

2/ See also Zellner and Theil (15, pp.68-69) and Narayanan (7, pp. 298-306) variances too would be equivalent for both the estimators. On the other hand, we have seldom infinitely large samples in actual practice and it is quite possible that even for finite samples the speed of convergence of distribution of, say b_T is more than that of b_T^* as T tends to increase. This speed could be better judged if we had exact variances of the sampling distributions of both b_T and b_T^*. For, in that case, the relative speed of convergence could have been decided by comparing actual variances for variation in T. Unfortunately, in most cases, actual distributions are not available and at best only asymptotic distributions are known. The problem is, then, to devise a method of deciding the relative speed of convergence from the knowledge of asymptotic distributions. This is done by considering the inflated estimator $T^{\frac{1}{2}}\ (b_T - \theta)$ whose asymptotic distribution (9, Section 5f.2(ii)) is normal with mean zero, variance $[i\ (\theta)]^{-1}$ where

$$(3.2)\quad i(\theta) = -\text{Plim}\ \left[T^{-1}\frac{d^2 l}{d\theta^2}\right],$$

and l is the logarithm of likelihood function based upon the T observations on the variable x. Clearly, i (θ) is a nonstochastic positive quantity. Therefore, even for large samples, the asymptotic variances of the inflated estimators $T^{\frac{1}{2}}\ b_T$ and $T^{\frac{1}{2}} b_T^*$ could be computed with the help of relation (3.2) and then, could be used exactly the same way as in case of exact samples.

It may be observed that, for a given T, variance of B_T is T^{-1} times the variance of inflated estimator $T^{\frac{1}{2}}\ b_T$ and therefore, we can find a number v_a of order of smallness (T^{-1}) so that

$$(3.3)\quad [i(\theta)]^{-1} = \underset{T\to\infty}{\text{Lt}}\ [T\ v_a].$$

The number v_a may be used in the same sense as $[i\ (\theta)]^{-1}$ because the factor T remains same for competing estimators obtained from a given sample. In strict sense, however, v_a should not be named as asymptotic variance of b_T, though, for comparing the speeds of convergence of alternative estimators it serves our purpose. Therefore, to avoid confusion, it would be better if we call v_a as <u>operational asymptotic variance</u> of estimator b_T and reckon that its asymptotic variance is always zero provided it is consistent estimator of O .

These concepts could be extended in a straight forward manner to multivariate situations. Thus, if b denotes a consistent estimator of a Λx1 vector β of unknown coefficients and if we can write

$$(3.4)\quad \left(-\text{Plim}\ T^{-1}\frac{d^2 L}{d_\beta d_\beta{}'}\right)^{-1} = \underset{T\to\infty}{\text{Lt}} \left[T\ V_a\right]$$

then, the $\Lambda \times \Lambda$ matrix V_a is defined as <u>operational asymptotic covariance matrix</u> of estimator b and its elements are each of order of smallness (T^{-1}).

It may be observed that the matrix in (3.4) is, in fact, the limiting value of the sequence $\left\{ET^{\frac{1}{2}}(b-\beta)\ T^{-\frac{1}{2}}(b-\beta)'\right\}$, or, in other words of asymptotic expectation (4, p.116) provided $E(b-\beta)(b-\beta)'$ exists. Therefore, we could define (4,p.118) V_a, alternatively, as

$$(3.5)\quad \text{Plim} \left[T\ (b-\beta)\ (b-\beta)'\right] = \underset{T\to\infty}{\text{Lt}} \left[T\ V_a\right]$$

subject to the restriction that the second moment of exact probability distribution of b is finite.

Further, if estimators b and b* are both consistent and derived from a given sample of observations, then, b is said to be asympto- <u>tically more efficient</u> than b* if

(3.6) $V_a(b^*) \geq V_a(b)$

where $V_a(b)$ and $V_a(b^*)$ are operational asymptotic covariance matrices of b and b*, respectively. Also, by definition

(3.7) $V_a(b^*) \geq V_a(b) \iff V_a(b^*) - V_a(b)$ positive semi-definite.

4. Asymptotic Covariance Matrices

The estimators described in Section 2 could also be interpretted as maximum likelihood estimators provided we assume, additionally, that the disturbance terms in models (2.1), (2.7) and (2.21) follow suitable normal probability laws and consider appropriate version of equations for constructing the likelihood functions. In that case, as is quite well known, the estimators defined in (2.3) and (2.5) may be obtained by determining the values of vector β for which the functions

(4.1) $l_1 = k_1 - \frac{1}{2}(y - X\beta)'(y - X\beta)$

and

(4.2) $l_2 = k_2 - \frac{1}{2}(y - X\beta)'V^{-1}(y - X\beta)$

are maximum where k's are independent of β. If an estimator V of V is known, apriori, or estimated from the sample observations y and X without making any assumption on β then, it can be easily seen that the estimator b, defined in (2.6), is maximum likelihood estimator corresponding to function.

(4.3) $l_3 = k_3 - \frac{1}{2}(y - X\beta)'\hat{V}^{-1}(y - X\beta)$

which can be obtained from(4.2)by substituting $\hat{V}$ - the available dispersion matrix, in place of V- the true dispersion matrix. Alternatively, the estimator b may be called as maximum likelihood estimator using an estimated dispersion matrix.3/ Similarly, SURE,

3/ See also Rao[10] for similar treatment of the theory of least squares using an estimated dispersion matrix.

2SLS and 3SLS estimators, defined in (2.20), (2.31) and (2.44) may be interpretted as maximum likelihood estimators corresponding to likelihood functions.

$$(4.4)\quad l_4 = k_4 - \tfrac{1}{2}(y^* - X^*\beta^*)' S_1^{-1} (y^* - X^*\beta^*)$$

$$(4.5)\quad l_5 = k_5 - \tfrac{1}{2}(X'y_i - X'Z_i\delta_i)'(\sigma_{ii}X'X)^{-1}(X'y_i - X'Z_i\delta_i),$$

and

$$(4.6)\quad l_6 = k_6 - \tfrac{1}{2}(Hy - HZ\delta)' S_1^{*-1} (Hy - HZ\delta)$$

respectively, Thus, it can be easily verified that estimators given in Section 2 are consistent (9, Section 5F.2(i)) provided only consistent estimators of dispersion matrices are used in each case. Accordingly, the asymptotic covariances of these estimators could be worked out as discussed in Section 3.

Throughout the remaining part of this section we shall be concerned with either exact or asymptotic covariance matrices. We denote the former by Q's and the latter by Q*'s. The covariance matrices of b_o and $\hat{\beta}$ are well known and may be written as

$$(4.7)\quad Q_1 = \sigma^2 (X'X)^{-1}$$

and

$$(4.8)\quad Q_2 = (X'V^{-1}X)^{-1},$$

respectively. In case of b, we shall derive operational asymptotic covariance matrix (OACM) because exact distribution of b, in general, is known.

Using (3.4) and (4.3) we can write

$$(4.9)\quad Lt\,(TQ_3^*) = (\text{Plim}\; T^{-1}X'\hat{V}^{-1}X)^{-1} = Lt\,(T^{-1}X'V^{-1}X)^{-1}$$

where use has been of Slutsky's theorem (4, p.118) and

(4.10) $\text{Plim}\ \hat{V} = V.$

The notation Lt in (4.9) and also in later discussion means limit when T approaches to infinity. Thus, the OACM of estimator b_y may be written as

(4.11) $Q_3^* = (X'V^{-1}X)^{-1}.$

In actual practice, one might compute OLS estimator even though the assumption (2.4) were true. In that case the moment matrix of b_o will be different from Q_1, defined in (4.7). Combining (2.3) with (2.1) we observe that

(4.12) $b_o - \beta = (X'X)^{-1} X'u$

and then using (2.4) we obtain 4/

(4.13) $E(b_o - \beta)(b_o - \beta)' = (X'X)^{-1}X'VX(X'X)^{-1} = Q_4.$

The derivation of moment matrices in case of estimators given in Section 2.2 follows exactly in the same way as discussed just above. Thus, if Q_5, Q_6^*, and Q_7 denote covariance matrix of b_o^*, OACM of b^* and covariance matrix of b_o^* when assumption (2.12) is true, then, we have

(4.14) $Q_5 = (X^{*\prime}D^{-1}X^*)^{-1},$

(4.15) $Q_6^* = (X^{*\prime}\ \Sigma_1^{-1}\ X^*)^{-1},$

and

(4.16) $Q_7 = (X^{*\prime}X^*)^{-1}\ X^{*\prime}\ \Sigma_1\ X^*(X^{*\prime}X^*)^{-1},$

4/ Cf (5) p.191. Dhrymes (4a) too has used similar approach while comparing the efficiency of certain structural estimators which appeared after this article was complete.

where

$$(4.17)\quad D = \begin{bmatrix} \sigma_{11} \times I & \dots & 0 \\ \cdot & & \\ \cdot & & \\ \cdot & & \\ 0 & \dots & \sigma_{MM} \otimes I \end{bmatrix}$$

Next, combining (4.5) with (3.4) and (2.33) we get

$$(4.18)\quad Lt(TQ_8^*) = \sigma_{ii}(\text{Plim}T^{-1} Z_i' M^* Z_i)^{-1}$$

$$= \sigma_{ii}\, \text{Plim}(T^{-1}Z_i'X)\text{Plim}(T^{-1}X'X)\text{Plim}(T^{-1}X'Z_i)^{-1}$$

$$= \sigma_{ii}\, T\, [T(Z_i' M^* \bar{Z}_i)^{-1}]$$

where use has been made of Slutsky's theorem (4, p. 118) and

$$(4.19)\quad \bar{Z}_i = EZ_i = E(Y_i \quad X_i) = (X\Pi_i \quad X_i).$$

If we define a matrix A_i whose elements are either unity or zero, so that

$$(4.20)\quad X_i = XA_i\ ,$$

then, we can rewrite (4.19) as

$$(4.21)\quad \bar{Z}_i = XC_i$$

where

$$(4.22)\quad C_i = (\Pi_i \quad A_i).$$

Finally, combining (4.21) with (4.10) we obtain

$$(4.23) \quad Q_8^* = \sigma_{ii} (\bar{Z}_i'\bar{Z})^{-1}$$

It is interesting to note that Q_8^* remains nonstochastic even if the predetermined and jointly dependent variables on the right hand side of model (2.27) are stochastic.[5] At the same time it may be remembered that Q_8^* is the proper OACM of 2SLS estimator d_{oi} provided disturbances in different equations in (2.34) are really independent, as we observed in equation (2.41) while deriving 2 SLS estimators from 3 SLS estimators. This is, however, against the basis of simultaneous equations model in view of the existence of jointly dependent variables. But one can obtain 2 SLS estimators even though assumption (2.25) were true. In that case using (2.41) and (2.37) we obtain

$$(4.24) \quad d_o - \delta = (Z'HZ)^{-1} Z'Hu$$

and the, combining (4.24) with (3.5)[6] and proceeding for further simplification exactly in the same way as we did in (4.18) we get

$$(4.25) \quad Lt(TQ_9^*) = Lt\, [T(\bar{Z}'\bar{Z})^{-1}\bar{Z}' \quad \Sigma_I^*(\bar{Z}'\bar{Z})^{-1}]$$

where

$$(4.26) \quad \bar{Z} = EZ$$

5/ The result reported in (15, p.56) is stochastic. But the same resi;t as om (4.23) above could be derived from asymptotic covariance matrix of 3 SLS estimator provided by the same authors (15,p.60) as a particular case when we replace Ω by an identity matrix, in which case the 3 SLS estimator itself is identical to 2 SLS estimator.

6/ Regarding the validity of (3.5), we may refer to Richardson[6] who has proved, for models containing two endogenous and any number of exogenous variables that $(E(d_o - \delta)\ (d_o - \delta)'$ is finite for adequately identified equations.

From (4.25) we obtain Q^*_9, the OACM of $d_.$ when (2.25) is the true assumption as

$$(4.27) \quad Q^*_9 = (\bar{Z}'\bar{Z})^{-1}\bar{Z}' \Sigma^*_I \bar{Z}(\bar{Z}'\bar{Z})^{-1}$$

Finally, we derive OACM of 3SLS estimator, defined in (2.44). Combining (4.6) with (3.4) we derive

$$(4.28) \quad Lt\ (TQ^*_{10}) = (Plim\ T^{-1}Z'H\Sigma^{*-1}_I Z)^{-1}$$
$$= (LtT^{-1}\bar{Z}' \Sigma^{*-1}_I \bar{Z})^{-1}$$

where simplification has been done similar to that in (4.18) and use has been made of the fact that H is idempotent symmetric matrix and

$$(4.29) \quad S^{*-1}_I H = H S^{*-1}_I$$

besides the fact that S^* is consistent estimator of Σ^*_1 Thus, OACM of estimator d may be written as

$$(4.30) \quad Q^*_{10} = (\bar{Z}' \Sigma^{*-1}_I \bar{Z})^{-1}$$

where $\bar{Z}$ is same as defined in (4.26).

5. Asymptotic Efficiency of Generalised Estimators

In our discussion so far, we considered two alternative estimators corresponding to each one of the three types of models analysed in this paper. In each case, both these estimators are equally acceptable for large samples in view of the fact that their asymptotic distributions collapse around the corresponding true parameters. In such cases we use the criterion of asymptotic efficiency for making choice between the two competing estimators which has been already discussed in Section 3. A careful look at the results in the preceding section would reveal that if we consider the appropriate.[7/] OAC M's for comparing asymptotic efficiency in each case, then, the algebraic forms of expressions are similar in case of all the three models. Therefore, using results, given in Section 4 and Rao [10, Lemma 2c] we may state the theorem

THEOREM: Under the same assumptions as required for derivation of different generalised least squares estimators and the notations developed in the preceding sections we have

(5.1) $Q_4 \geq Q_3^*$

(5.2) $Q_7 \geq Q_6^*$

and

(5.3) $Q_9^* \geq Q_{10}^*$

The equality signs hold only if V Σ, Σ^* are identity matrices.

The results in (5.1) to (5.3) indicate that the generalised estimators are asymptotically more efficient than their OLS counterparts in each case and we do not need any orthogonality assumption regarding the explanatory variables, as Zellner and Huang [14] did for proving the efficiency of SURE estimator. Further, in case of simultaneous equations model, the result (5.3) is true only if all the equations of the system are over identified so that atleast second moments of 2SLS estimators exist in case of each equation. This is because, otherwise, use of (3.5) for deriving Q_9^* would not be justified. Accordingly Zellner and Theil's [15 Section] conclusions, regarding the efficiency of 2SLS and 3SLS estimators.

7/ Note that Zellner and Huang[14, p.306] compared Q_5 with Q_6^*. In fact Q_7 should be compared with Q_6^*.

are misleading; because OACM of coefficient estimates in case of just-identified equations cannot be computed.

6. Effect of Orthogonality on Efficiency

In this section we analyse the effect of orthogonality of explanatory variables in different equations upon the efficiency of various estimators. We consider set of regression equations and simultaneous equations models. In the first instance we assume

(6.1) $X_i'X_j = 0, \; i \neq j \quad i, \quad M,$

which implies that independent variables in different regressions in (2.7) are orthogonal. Combining (6.1) with (4.15) and (4.16) we obtain

$$(6.2)\quad Q_7 - Q_6^* = \begin{pmatrix} (\sigma_{11}^{-1} X_1' X_1)^{-1} - (\sigma^{11} X_1' X_1)^{-} & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & (\sigma_{MM}^{-1} X_M' X_M)^{-1} - (\sigma^{MM} X_M' X_M)^{-1} \end{pmatrix}$$

where σ^{ii} $(i = 1, .., M)$ are elements along the diagonal of the matrix Σ^{-1}. *Thus, gain in efficienty* g_{ii} *of SURE estimator of coefficients* in the i-th equation over corresponding OLS estimator may be written as

(6.3) $g_{ii} = R_i^2 \sigma_{ii} \; (X_i'X_i)^{-1}$

where (9, p. 226)

(6.4) $R_i^2 = 1 - (\sigma_{ii}\sigma^{ii})^{-1}$

which implies that predetermined variables in the whole system are perfectly uncorrelated with each other.

If we denote by Q^*_{11} the O.CM of estimator $\tilde{\delta}_i$ when (6.8) is really false, then, following the procedure of obtaining Q^*_δ we can write

(6.9) $Q^*_{11} = \sigma_{ii}(\bar{Z}_i'XX'\bar{Z}_i)^{-1}\bar{Z}_i' \ XX'XX'\bar{Z}_i \ (\bar{Z}_i' \ XX'\bar{Z}_i)^{-1}$.

Now, using result (5.1) of the theorem we have

(6.10) $Q^*_{11} \geq Q^*_8$,

where use has been of the fact that we can write

(6.11) $Q^*_8 = \sigma_{ii} \left[\bar{Z}_i'X(X'X)^{-1}X'\bar{Z}_i \right]^{-1}$

Furth r, using (4.21) was observed that when (6.8) holds true then,

(6.12) $\bar{Z}_i'XX'\bar{Z}_i = \bar{Z}_i'X(X'X)^{-1}X'\bar{Z}_i = \bar{Z}_i'\bar{Z}_i$.

Therefore, the gain in asymptotic efficiency of d_{oi} over $\tilde{\delta}_i$ is zero when (6.8) is true, because in that case $Q^*_{11} = Q^*_8$ in view of (6.12). But if $X'X \neq I$, that is, the predetermined variables in the complete system are correlated, then, estimator d_{oi} is definitely asymptotically more efficient than $\tilde{\delta}_i$ in view of (6.10). At the same time 3SLS estimators in (2.44) are asymptotically more efficient than 2SLS estimators d_{oi} according to relation (5.3). Thus, we find that inclusion of lagged variables or, in other words, the existence of correlation between different predetermined variables has healthier effect on 2SLS and 3SLS estimators in respect of their asymptotic efficiency.

REFERENCES

(1) Aitken, A.C., "On Least-Squares and Linear Combination of Observations," Proceedings of the Royal Society of Edinburgh, Volume 55, 1934-35, pp.42-48.

(2) Basmann, R.L., "A Generalised Classical Method of Linear Estimation of Coefficients in a Structural Equation," Econometrica, Volume 25, 1957.

(3) Cochrane, D. and G.H. Orcutt, "Application of Least Squares Regressions to Relationships Containing Auto-Correlated Error Terms," Journal of the American Statistical Association, Volume 44, 1949.

(4) Goldberger, A.S., Econometric Theory, John Wiley and Sons, New York, 1964.

(4a) Dhrymes, P.J., Econometrics - Statistical Foundations and Applications, Harper and Row, New York, 1970.

(5) Johnston, J., Econometric Methods, Mcgraw Hill Book Company, New York, 1960.

(6) Kakwani, N.C., "The Unbiasedness of Zellner's Seemingly Unrelated Regression Equations Estimators," Journal of the American Statistical Association, Volume 62, March 1967.

(7) Narayanan, R., "Computation of Zellner-Theil's Three Stage Least-Squares Estimates," Econometrica, Volume 37, April, 1969.

(8) Richardson, D.H., "The Exact Distribution of a Structural Coefficient Estimator," Journal of the American Statistical Association, Volume 63, 1968, pp. 1215-26.

(9) Rao, C.R., Linear Statistical Inference and Its Applications, John Wiley and Sons, New York, 1965.

(10) Rao, C.R., "Least Squares Theory Using an Estimated Disperson Matrix And Its Applications to Measurement of Signals," Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1, 1967, pp. 355-72.

(11) Theil, H., Economic Forecast and Policy, North-Holland Publishing Company, Amsterdam, I

(12) Zellner, A., "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias," Journal of the American Statistical Association, Volume 15, 1962, pp.348-68.

(13) Zellner, A., "Estimators for Seemingly Unrelated Regression Equations: Some Exact Finite Sample Results," Journal of the American Statistical Association, Volume 58, 1963, pp.977-92.

(14) Zellner, A. and D.S. Huang, "Further Properties of Efficient Estimators for Seemingly Unrelated Regression Equations," International Economic Review, Volume 3, 1962, pp.300-313.

(15) Zellner, A. and H. Theil, "Three-Stage Least Squares Simultaneous Estimation of Simultaneous Equations," Econometrica, Volume 30, 1962, pp.54-78.