9

ASYMPTOTIC PROPERTIES OF SIMULTANEOUS LEAST SQUARES ESTIMATORS

by

Phoebus J. Dhrymes

Discussion Paper Number 9

ASYMPTOTIC PROPERTIES OF SIMULTANEOUS LEAST SQUARES ESTIMATORS*

Phoebus J. Dhrymes

1. Introduction and Summary.

In an interesting paper [2], over a decade ago, Brown had suggested an estimation procedure for the standard simultaneous equations model which he termed simultaneous least squares (SLS). A paper by Nakamura [6] asserted the consistency of SLS. Since that time, however, no systematic attempt has been made to study the efficiency of this estimator vis-a-vis other commonly employed estimators, except in desultory Monte Carlo fashion. In part this is due to the fact that the asymptotic distribution of the SLS estimator has not been obtained in the literature.

We shall derive, below, this asymptotic distribution and show the relation of SLS to the two stage least squares (2SLS) and three stage least squares (3SLS) estimators; hence, to limited and full information maximum likelihood (LIML, FIML) estimators as well.

Contrary to assertions in [2] SLS is not a full information estimator, if by the latter we mean one that takes into account the stochastic dependence of the system's structural errors. It is a full information procedure only in the sense that it estimates all structural parameters simultaneously. In fact, the best way to understand the essence of SLS is in terms of the general linear model

---

*The research on which this paper is based was in part supported by NSF grant GS 2289 at the University of Pennsylvania and was completed during the author's visit at the University of California, Los Angeles.

(1) $$y = X\beta + u, \quad E(u) = 0, \quad Cov(u) = \Sigma .$$

If in (1) we obtain the estimator

(2) $$\tilde{\beta} = (XRX)^{-1}X'Ry$$

where R is a symmetric matrix which, if nonstochastic it is not necessarily Σ^{-1}, and again if stochastic it is not necessarily true that plim $R = \Sigma^{-1}$, then one has an exact analog of the relation between SLS and 2SLS, 3SLS. Clearly the estimator in (2) may or may not be efficient relative to the ordinary least squares (OLS) estimator of β depending on the proximity of R (or of its probability limit) to Σ^{-1}. This is exactly the relation of SLS to 2SLS. Thus SLS would be (asymptotically) inefficient relative to 3SLS, except in very special circumstances, and cannot be uniformly ranked relative to the 2SLS estimator in terms of (asymptotic) efficiency.

2. Asymptotic Properties of the SLS Estimator.

Consider the standard structural econometric model

(3) $$y_{t\cdot} = y_{t\cdot}B + x_{t\cdot}C + u_{t\cdot} , \qquad t = 1,2,\ldots,T$$

where $y_{t\cdot} = (y_{t1}, y_{t2}, \ldots, y_{tm})$, $x_{t\cdot} = (x_{t1}, x_{t2}, \ldots, x_{tG})$ are, respectively, the vectors of observation on the m current endogenous and G predetermined variables of the system; the matrices B, C contain the structural parameters of the problem. The system is assumed to be identified through a priori restrictions so that it will be known that some elements of B and C are zero.

The sequence $\{u_{t.} : t = 1,2,\ldots\}$ is assumed to be one of independent identically distributed (i.i.d.) random variables obeying

$$(4) \qquad E(u_{t.}') = 0 \qquad Cov(u_{t.}') = \Sigma .$$

It is convenient, but not necessary, to assume that the predetermined variables are nonstochastic and that the second moment matrices of the variables of the problem possess nonsingular limits.

If we impose the a priori restrictions, the observations on the ith structural equation may be written as

$$(5) \qquad y_{\cdot i} = Y_i\beta_{\cdot i} + X_i\gamma_{\cdot i} + u_{\cdot i}$$

where $y_{\cdot i}$ is the ith column of

$$(6) \; Y = (y_{\cdot 1}, y_{\cdot 2}, \ldots, y_{\cdot m}) = (y_{ti}) \quad t = 1,2,\ldots,T, \; i = 1,2,\ldots,m \; .$$

If we put

$$(7) \; X = (x_{\cdot 1}, x_{\cdot 2}, \ldots, x_{\cdot G}) = (x_{tj}) \quad t = 1,2,\ldots,T, \; j = 1,2,\ldots,m$$

then Y_i, X_i in (5) are appropriate submatrices of Y, X respectively; $\beta_{\cdot i}$, $\gamma_{\cdot i}$ are the ith columns of B, C, respectively after suppression of elements known to be zero. Finally $u_{\cdot i}$ is the ith column of

(8) $U = (u_{\cdot 1}, u_{\cdot 2}, \ldots, u_{\cdot m}) = (u_{ti}) \quad t = 1,2,\ldots,T, \; i = 1,2,\ldots,m$.

The model in (3) may be written more compactly as

(9) $$Y = YB + XC + U$$

from which we can obtain the reduced form

(10) $$Y = X\Pi + V, \; \Pi = C(I-B)^{-1}, \; V = U(I-B)^{-1} .$$

The SLS estimator is obtained by minimizing

(11) $$\sum_{t=1}^{T} (y_{t\cdot} - x_{t\cdot}\Pi)(y_{t\cdot} - x_{t\cdot}\Pi)' = \mathrm{tr}(Y - X\Pi)'(Y - X\Pi)$$

with respect to the unknown elements of C and B.

It is easy to show from (11), that SLS is essentially a limited information procedure. To see this recall that 2SLS is obtained by minimizing

$$\mathrm{tr}\; A'\tilde{Z}'\tilde{Z}A$$

while 3SLS is obtained by minimizing

$$\mathrm{tr}\; \Sigma^{-1}A'\tilde{Z}'\tilde{Z}A$$

the essential difference between them being that 3SLS takes into account the correlation structure of the error terms but 2SLS does not. In the above

$$(12)\quad A = \begin{pmatrix} I - B \\ -C \end{pmatrix}, \quad Z = (Y,X), \quad \tilde{Z} = (\tilde{Y},X), \quad \tilde{Y} = NY, \quad N = X(X'X)^{-1}X' .$$

The minimand of SLS may be rewritten more suggestively as

$$(13)\quad \mathrm{tr}(Y - X\Pi)'(Y - X\Pi) = \mathrm{tr}\, K^{-1}A'Z'ZA, \quad K = (I - B')(I - B)$$

A "full information" analog of SLS would have to be obtained from

$$(14)\quad \mathrm{tr}\, \Omega^{-1}(Y - X\Pi)'(Y - X\Pi) = \mathrm{tr}\, \Sigma^{-1}A'Z'ZA$$

where Ω is the covariance matrix of the reduced form errors

$$(15)\quad \Omega = (I - B')^{-1}\, \Sigma(I - B)^{-1}$$

by minimizing with respect to A, given a prior estimate of Σ. If the structural errors were normal then (14) gives the exponential of the likelihood function, and the "full information" analog of SLS is clearly inefficient relative to FIML since it fails to take into account the Jacobian of the transformation from the structural errors to the jointly dependent variables of the system. Worse than that, the estimator thus obtained, is inconsistent, as may be easily verified.

In obtaining the normal equations for the SLS estimator it is more convenient to group the unknown parameters according to

$$(16)\quad \delta = (\gamma', \beta')', \quad \gamma = (\gamma'_{\cdot 1}, \gamma'_{\cdot 2}, \ldots, \gamma'_{\cdot m})', \quad \beta = (\beta'_{\cdot 1}, \beta'_{\cdot 2}, \ldots, \beta'_{\cdot m})'$$

instead of the customary grouping, $\delta = (\delta'_{\cdot 1}, \delta'_{\cdot 2}, \ldots, \delta'_{\cdot m})'$, $\delta_{\cdot i} = (\beta'_{\cdot i}, \gamma'_{\cdot i})'$. Observations on the model in (3) or (5) may then be written in the compact form

$$(17)\qquad y = Z^{*}\delta + u$$

where

$$(18)\quad y = (y'_{\cdot 1}, y'_{\cdot 2}, \ldots, y'_{\cdot m})', \quad X^{*} = \text{diag}(X_1, X_2, \ldots, X_m), \quad Y^{*} = \text{diag}(Y_1, Y_2, \ldots, Y_m)$$

$$Z^{*} = (X^{*}, Y^{*}), \quad u = (u'_{\cdot 1}, u'_{\cdot 2}, \ldots, u'_{\cdot m})' .$$

We shall now write down the normal equations of SLS, relegating the details of their derivation to the appendix. Thus, we have

$$(19)\qquad f(\delta) = Z^{*\prime}(K^{-1} \otimes I_T)Z^{*}\,\delta - Z^{*\prime}(K^{-1} \otimes I_T)y + \begin{pmatrix} 0 \\ V^{*\prime} \end{pmatrix}(K^{-1} \otimes I_T)(y - Z^{*}\delta)$$

where

$$(20)\qquad V^{*} = \text{diag}(V_1, V_2, \ldots, V_m), \quad V_i = Y_i - X\Pi_i, \qquad i = 1, 2, \ldots, m .$$

Remark 1: Equation (19) suggests a possibly efficient algorithm for obtaining the SLS estimator - as a solution of the equation $f(\delta) = 0$ -

by iteration. Thus, suppose an initial consistent estimator is available for δ, say $\tilde{\delta}_{(0)}$. We thus compute the quantities $\tilde{K}^{-1}$, $\tilde{V}_i = Y_i - \tilde{Y}_i$, $\tilde{Y}_i = X\tilde{\Pi}_i$ and obtain the first iterate

$$(21). \qquad \tilde{\delta}_{(1)} = [\tilde{Z}^{*\prime}(\tilde{K}^{-1} \otimes I_T)Z^*]^{-1} [\tilde{Z}^{*\prime}(\tilde{K}^{-1} \otimes I_T)y]$$

where

$$(22) \qquad \tilde{Z}^* = (X^*,\tilde{Y}^*),\ \tilde{Y}^* = \mathrm{diag}(\tilde{Y}_1,\tilde{Y}_2,\ldots,\tilde{Y}_m) .$$

It is then clear that given $\tilde{\delta}_{(1)}$ we can obtain $\tilde{\delta}_{(2)}$ and continue until convergence is attained. Of course, the conditions under which convergence holds will have to be established. <u>It is clear that the converging iterate is the SLS estimator</u>.

Remark 2: It is evident from (21) that all iterates - beyond the initial consistent estimator $\tilde{\delta}_{(0)}$ - have an interpretation as instrumental variables (I.V.) estimators with instrumental matrix $\tilde{Z}^{*\prime}(\tilde{K}^{-1} \otimes I)$.

Remark 3: The expression in (21) elucidates the comments made at the end of section 1. Thus, we see that SLS proceeds analogously with the 3SLS estimator except that it uses the irrelevant matrix $\tilde{K}^{-1}$ instead of $\tilde{\Sigma}^{-1}$ and "purges" the current endogenous variables in each equation by using the <u>restricted reduced form residuals</u>. This is the price paid by not considering the dependence of the structural errors of the system. We remind the reader that 3SLS has an interpretation as an

I.V. estimator obeying a relation like (21) in which $\tilde{K}^{-1}$ is replaced by $\tilde{\Sigma}^{-1}$ and $\tilde{Y}^*$ is obtained from the unrestricted reducted form by ordinary least squares. Thus, although SLS has the appearance (and the computational burden) of a full information estimator, it is essentially a limited information one.

Remark 4: The initial consistent estimator $\tilde{\delta}_{(0)}$ may be easily obtained from equations (5) by using as instruments, in addition to the included predetermined variables, other (excluded) predetermined variables. In view of the identifiability condition it will always be possible to obtain $\tilde{\delta}_{(0)}$. In fact there is a multiplicity of such estimators. For the reasons above, perhaps SLS is more aptly named the restricted reduced form iterated instrumental variables (RRFIIV) estimator.

From (21) we immediately have

LEMMA 1. The iterates of the RRFIIV estimator are consistent provided

$$(23) \qquad \lim_{T \to \infty} \frac{\bar{Z}^{*\prime}(K^{-1} \otimes I_T)\bar{Z}^*}{T} = \Psi$$

exists as a nonsingular matrix, where

$$(24) \quad \bar{Z}^* = (X^*, \bar{Y}^*), \; \bar{Y}^* = \mathrm{diag}(\bar{Y}_1, \bar{Y}_2, \ldots, \bar{Y}_m), \; \bar{Y}_i = X\Pi_i, \quad i = 1, 2, \ldots, m \; .$$

Proof: It will be sufficient to prove the consistency of $\tilde{\delta}_{(1)}$. Thus,

(25) $$\tilde{\delta}_{(1)} - \delta^0 = [\tilde{Z}^{*\prime}(\tilde{K}^{-1} \otimes I_T)Z^*]^{-1}\tilde{Z}^{*\prime}(\tilde{K}^{-1} \otimes I_T)u$$

and the result follows immediately from the consistency of $\tilde{\delta}_{(0)}$ and the nonstochastic character of the predetermined variables[1].

COROLLARY 1. If the RRFIIV (SLS) estimator can be obtained through the iteration of (21), then it is consistent.

Proof. Obvious from Lemma 1.

Remark 5: Although Corollary 1 qualifies the consistency property, in fact Brown [2] assumes iterative convergence beginning from an unspecified initial condition. Nakamura [2] takes convergence of the Brown process as given or, at any rate, assumes that the minimizing solution has been somehow obtained. Thus, the corollary above may be viewed as an alternative proof of consistency.

LEMMA 2. All iterates have the same asymptotic distribution.

Proof. It is clear that for any $i \geq 1$

(26) $$\operatorname*{plim}_{T\to\infty} \sqrt{T}\,(\tilde{\delta}_{(i)} - \delta^0) = \Psi^{-1} \operatorname*{plim}_{T\to\infty} \frac{1}{\sqrt{T}}\, \overline{Z}^{*\prime}(K^{-1} \otimes I_T)u \,.$$

Thus, asymptotically,

---

[1] Actually independence of, or even uncorrelatedness with the elements of u would be sufficient to establish consistency. In (25) δ^0 is the "true" parameter vector.

$$\sqrt{T}\,(\tilde{\delta}_{(i)} - \delta^0) \sim \sqrt{T}\,(\tilde{\delta}_{(1)} - \delta^0) \qquad i = 2,3,\ldots$$

Remark 6: Lemma 2 does not imply that "iteration does not matter." It states that it does not matter for "large samples." Evidently, small sample properties will differ according as the estimator we obtain satisfies $f(\hat{\delta}) = 0$ or not. What we my deduce, however, is that, if we terminate the iteration after k steps, the resulting estimator has the same properties as the converging iterate, if one exists.

The asymptotic distribution is given by

LEMMA 3. The RRFIIV (SLS) estimator has, asymptotically, the distribution

$$\sqrt{T}\,(\hat{\delta} - \delta^0) \sim N(0,\Phi) \tag{27}$$

where

$$\Phi = \Psi^{-1}\left[\lim_{T\to\infty} \frac{\bar{Z}^{*\prime}(K^{-1}\otimes I_T)(\Sigma\otimes I_T)(K^{-1}\otimes I_T)\bar{Z}^{*}}{T}\right]\Psi^{-1} . \tag{28}$$

Proof. Utilizing the result in (26) it will be sufficient to obtain the asymptotic distribution of $1/\sqrt{T}\ \bar{Z}^{*\prime}(K^{-1}\otimes I_T)u$. To this effect see, for example, [3, p. 254] or [4, Ch. 3]. Now define the matrices

$$R_s^{(i)} = {}_i x'_{s\cdot}\, k^*_{i\cdot}\ , \quad Q_s^{(i)} = {}_i\bar{y}'_s\, k^*_{i\cdot}\ , \qquad i = 1,2,\ldots,m,\ s = 1,2,\ldots,T \tag{29}$$

where ${}_i x'_{s\cdot}$ is sth column of X'_i, ${}_i\bar{y}'_{s\cdot}$ is sth column of $\bar{Y}_i$ and each contains, respectively, G_i and m_i elements; evidently, $k^*_{i\cdot}$ is the ith row of K^{-1}. Let

$$(30) \qquad R_s = \begin{bmatrix} R_s^{(1)} \\ \vdots \\ R_s^{(m)} \end{bmatrix}, \quad Q_s = \begin{bmatrix} Q_s^{(1)} \\ \vdots \\ Q_s^{(m)} \end{bmatrix}, \quad P_s = \begin{bmatrix} R_s \\ Q_s \end{bmatrix}.$$

Noting that P_s is a matrix of dimension $\sum_{i=1}^{m} (m_i + G_i) \times m$ we may thus write

$$(31) \qquad \frac{1}{\sqrt{T}} \bar{Z}^{*\prime}(K^{-1} \otimes I_T)u = \frac{1}{\sqrt{T}} \sum_{s=1}^{T} P_s u'_s .$$

But $\{P_s u'_{s\cdot} : s = 1,2,\ldots\}$ is a sequence of mutually independent nonidentically distributed random variables with mean zero and covariance matrix $P_s \Sigma P'_s$. We further note that

$$(32) \qquad \lim_{T\to\infty} \frac{1}{T} \sum_{s=1}^{T} P_s \Sigma P'_s = \lim_{T\to\infty} \frac{1}{T} [\bar{Z}^{*\prime}(K^{-1} \otimes I_T)(\Sigma \otimes I_T)(K^{-1} \otimes I_T)\bar{Z}^*]$$

which we may assert to be a nonsingular matrix with finite elements. Let $F(\cdot)$ be the common distribution of the vectors u_s. Then we have that

$$(33)\quad \lim_{T\to\infty} \frac{1}{T} \sum_{s=1}^{T} \int_{|P_s u_s| > \eta\sqrt{T}} |P_s u_s|^2 dF(u_s)$$

$$\leq \lim_{T\to\infty} \frac{1}{T} \sum_{s=1}^{T} \|P_s\|^2 \int_{\|P_s\||\varphi| > \eta\sqrt{T}} |\varphi|^2 dF(\varphi)$$

where $\|P_s\|^2$ is a norm for the matrix P_s defined by

$$(34)\quad \|P_s\|^2 = \operatorname{tr} P_s' P_s \ .$$

Since the predetrmined variables are nonstochastic and second moments exist we have no difficulty in establishing that

$$\lim_{T\to\infty} \frac{1}{T} \sum_{s=1}^{T} \|P_s\|^2$$

is a finite quantity. Let

$$(35)\quad \alpha = \sup_s \|P_s\|, \ \alpha > 0 \ .$$

The condition $\alpha > 0$ simply requires that for some time s <u>not all</u> predetermined variables are zero. Thus,

$$(36)\quad \lim_{T\to\infty} \frac{1}{T} \sum_{s=1}^{T} \|P_s\|^2 \int_{\|P_s\||\varphi| > \eta\sqrt{T}} |\varphi|^2 dF(\varphi)$$

$$\leq \lim_{T\to\infty} \frac{1}{T} \sum_{s=1}^{T} \|P_s\|^2 \int_{|\varphi| > (\eta/\alpha)\sqrt{T}} |\varphi|^2 dF(\varphi) \ .$$

But the last integral converges to zero because the structural errors have finite variances. Consequently, the Lindeberg-Feller condition, see [4, Ch. 3], is satisfied and thus

$$(37) \qquad \frac{1}{\sqrt{T}} \bar{Z}^{*\prime}(K^{-1} \otimes I_T)u \sim N\left(0, \lim_{T\to\infty} \frac{1}{T} \sum_{s=1}^{T} P_s \Sigma P_s'\right) .$$

From (25) it then follows immediately that

$$(38) \qquad \sqrt{T}\,(\hat{\delta} - \delta^0) \sim N(0,\Phi) \qquad \text{q.e.d.}$$

Remark 7: In the proof above certain conditions were invoked, viz., those in (32) and (35). Instead of asserting (32) one ought to show how it follows from the assumptions made on the predetermined variables. Such an activity, however, has only elegance to recommend it and it lies outside the main objectives of this paper.

The proof above strictly speaking applies to the iterates of equation (21) and would give the distribution of the SLS estimator only when the iteration converges. If the equation $f(\hat{\delta}) = 0$ has a consistent solution - which must be obtained by a procedure other than the one given in (21) - then an alternative argument can establish the asymptotic distribution of the estimator. Thus, by the mean value theorem, write

$$(39) \qquad f(\hat{\delta}) = f(\delta^0) + \frac{d}{d\delta} f(\delta^*)\,(\hat{\delta} - \delta^0)$$

where δ^* lies between δ^0 and $\hat{\delta}$. One can then show that

$$(40) \qquad \operatorname*{plim}_{T \to \infty} \frac{1}{T} \frac{d}{d\delta} f(\delta^*) = \Psi .$$

Moreover, by establishing the asymptotic distribution of $1/\sqrt{T}\ f(\delta^0)$ one obtains the result given in Lemma 3.

Remark 8: If we waive the condition that the predetermined variables are nonstochastic and, in particular, if we admit lagged endogenous variables a certain complication will arise. Thus in (31) we would no longer be dealing with a sequence of independent random vectors. However, if we obtain the final form of the model - see [4, Ch. 12] - we can express the lagged endogenous variables in terms of the exogenous variables and a certain rational lag in the structural errors. Because we shall impose a stability condition on the model it will be possible to deduce <u>mutatis mutandis</u>, exactly the same results as in Lemma 3 by the application of a central limit theorem on m-dependent variables. For extensive applications of this approach, see [5].

We have therefore established

THEOREM 1. Consider the model in (3) subject to the assumptions and conditions given in section 2. Then

i. The RRFIIV (SLS) estimator has an interpretation as an instrumental variables estimator.

ii. It may be obtained by iteration, as given in equation (21), provided the iteration converges.

iii. If we begin the process with an initial consistent estimator then all subsequent iterates are consistent.

iv. All iterates have the same asymptotic distribution.

v. The asymptotic distribution of the converging iterate (if one exists), in fact of every iterate, is given by $\sqrt{T}\,(\hat{\delta} - \delta^0) \sim N(0,\Phi)$ where Φ is as given in (28).

3. Comparison with 2SLS and 3SLS Estimators.

Having obtained the asymptotic distribution of RRFIIV it is now rather simple to appraise its efficiency relative to that of 3SLS or full information maximum likelihood (FIML) estimators. The asymptotic distribution of 3SLS is given by

$$(41) \qquad \sqrt{T}\,(\hat{\delta}_{3SLS} - \delta^0) \sim N(0,C_3)$$

where

$$(42) \qquad C_3 = \lim_{T\to\infty} \left[\frac{\overline{Z}^{*\prime}(\Sigma^{-1} \otimes I_T)\overline{Z}^{*}}{T} \right]^{-1}$$

We have

LEMMA 4. RRFIIV(SLS) estimators are asymptotically inefficient relative to 3SLS.

<u>Proof</u>. We show that the difference of the two covariance matrices is positive semidefinite. Now, define D by

$$(43) \qquad [\bar{Z}^{*\prime}(K^{-1} \otimes I_T)\bar{Z}^*]^{-1}\bar{Z}^{*\prime}(K^{-1} \otimes I_T) = [\bar{Z}^{*\prime}(\Sigma^{-1} \otimes I_T)\bar{Z}^*]^{-1}\bar{Z}^{*\prime}(\Sigma^{-1} \otimes I_T) + D$$

and observe that $D\bar{Z}^* = 0$. Multiply both sides of (43) by $(\Sigma \otimes I_T)$ to obtain

$$(44) \quad [\bar{Z}^{*\prime}(K^{-1} \otimes I_T)\bar{Z}^*]^{-1}\bar{Z}^{*\prime}(K^{-1} \otimes I_T)\ (\Sigma \otimes I_T)$$
$$= [\bar{Z}^{*\prime}(\Sigma^{-1} \otimes I_T)\bar{Z}^*]^{-1}\bar{Z}^{*\prime} + D(\Sigma \otimes I_T) .$$

Post multiply (44) by T times the transpose of (43) and take limits to obtain

$$(45) \qquad \Phi = C_3 + \lim_{T \to \infty} T\ D(\Sigma \otimes I_T)D' \qquad .$$

which establishes the desired result.

COROLLARY 2. 3SLS is strictly efficient relative to RRFIIV (SLS), in the sense that for at least one element of $\hat{\delta}$ the asymptotic 3SLS variance is strictly less than the corresponding SLS variance, unless

$$(46) \qquad \lim_{T \to \infty} T\ D(\Sigma \otimes I_T)D' = 0 .$$

<u>Proof</u>. Obvious from the Lemma.

LEMMA 5. RRFIIV (SLS) has the same asymptotic distribution as 3SLS if

i. $K = \Sigma$

ii. All equations of the system are just identified.

Proof. Obvious from (43) since, under either i. or ii. $D = 0$

Remark 9: Since 3SLS and FIML have the same asymptotic distribution, Lemmas 4, 5 and Corollary 2 hold with respect to FIML estimators as well.

The relation of RRFIIV (SLS) to 2SLS and limited information maximum likelihood (LIML) estimators is established by

LEMMA 6: RRFIIV (SLS) estimators are asymptotically

i. equivalent to 2SLS if the equations of the system are just identified, or if K^{-1} is a diagonal matrix.

ii. efficient relative to 2SLS if $K = \Sigma$, or (by continuity) if $K^{-1} = \Sigma^{-1} + \varepsilon R$, for suitably small ε, R being some $m \times m$ matrix such that $\Sigma^{-1} + \varepsilon R$ is nonsingular.

iii. inefficient relative to 2SLS if Σ is a diagonal matrix.

Proof. To prove i. we observe first that

$$(47) \qquad \sqrt{T}\,(\hat{\delta}_{2SLS} - \delta^0) \sim N(0, C_2)$$

where

$$(48) \qquad C_2 = \lim_{T\to\infty} \left[\left(\frac{\bar{Z}^{*\prime}\bar{Z}^*}{T} \right)^{-1} \frac{\bar{Z}^{*\prime}(\Sigma \otimes I_T)\bar{Z}^*}{T} \left(\frac{\bar{Z}^{*\prime}\bar{Z}^*}{T} \right)^{-1} \right].$$

When K^{-1} is a diagonal matrix, then

$$(49) \quad [\bar{Z}^{*\prime -1}(K^{-1} \otimes I_T)\bar{Z}^{*}\,]^{-1}\bar{Z}^{*\prime}(K^{-1} \otimes I_T) = (\bar{Z}^{*\prime}\bar{Z}^{*})^{-1}\bar{Z}^{*\prime}$$

and it is easily verified that

$$(50) \qquad \Phi = C_2 \ .$$

When the system is just identified 2SLS coincides with 3SLS and thus the first part of i. follows from Lemma 4.

To prove ii. we note that if $K = \Sigma$ then $\Phi = C_3$; the second part of ii. follows easily by continuity and the arguments given in [1, Ch. 4].

To prove iii. we note that when Σ is diagonal, then

$$(51) \qquad C_2 = C_3 \ .$$

The result then follows from Lemma 4.

Remark 10: Since LIML has the same asymptotic distribution as 2SLS Lemma 6 applies to LIML estimators as well. It is possible to give more precise conditions under which 2SLS is inefficient or efficient relative to RRFIIV (SLS) estimators in terms of the relation between the characteristic roots of K^{-1} and those of Σ. It does not seem worthwhile to do so, however, since our inability to rank RRFIIV (SLS) relative to 2SLS is well established by Lemma 6.

We have therefore established

THEOREM 2. RRFIIV (SLS) estimators are inefficient relative to 3SLS or FIML in the sense that the difference of the covariance matrices of their respective asymptotic distributions - in the order stated - is a positive semidefinite matrix. Thus, FIML and 3SLS dominate RRFIIV (SLS).

It is not possible to rank, uniformly, RRFIIV (SLS) and 2SLS or LIML estimators in terms of relative asymptotic efficiency. They are equivalent if K^{-1} is a diagonal matrix, 2SLS (or LIML) dominates if Σ is a diagonal matrix, and RRFIIV (SLS) dominates if $K = \Sigma$.

4. Conclusion.

In this paper we have studied extensively the properties of the SLS estimator proposed by Brown [2]. We have given it an iterated instrumental variables estimator interpretation and have shown that if we begin the iteration with a consistent estimate then all subsequent iterates are consistent. Moreover we have established that SLS is (asymptotically) dominated by 3SLS (or FIML) estimators and that it neither dominates nor is dominated by 2SLS (or LIML) estimators. Which of the two situations holds depends on the parametric configuration of the problem.

The results we have obtained provide considerable guidance in the design of Monte Carlo experiments which would explore the small sample properties of SLS (RRFIIV) relative to other full or limited information estimators.

REFERENCES

1. Bellman, R: "Introduction to Matrix Analysis, New York, McGraw-Hill, 1960.
2. Brown, T.M: "Simultaneous Least Squares: A Distribution-free Method of Equation System Structure Estimation," International Economic Review, vol. 1 (1960) pp. 173-191.
3. Cramer, H: Mathematical Methods of Statistics, Princeton, Princeton University Press, 1946.
4. Dhrymes, P.J: Econometrics: Statistical Foundations and Applications, New York, Harper and Row, 1970.
5. __________: Distributed Lags: Problems of Formulation and Estimation, San Francisco, Holden-Day (in press).
6. Nakamura, M: "A Note on the Consistency of Simultaneous Least Squares Estimation," International Economic Review, vol. 1 (1960) pp. 192-197.

APPENDIX

In this appendix we derive an expression for the normal equations of SLS and obtain the probability limit of the matrix of second derivatives of the minimand of SLS. The minimand is given by - see equation (13) in the text -

$$(A.1)\qquad tr(Y - X\Pi)'(Y - X\Pi) = tr\, K^{-1}A'Z'ZA = \sum_{i=1}^{m}\sum_{j=1}^{m} k^{ij}\, m_{ji}$$

where

$$(A.2)\qquad M = (m_{ij}) = A'Z'ZA, \quad K^{-1} = (k^{ij}), \qquad i,j = 1,2,\ldots,m .$$

If we differentiate (A.1) with respect to the first element of δ, as given in (16), we would have, symbolically,

$$(A.3)\qquad \frac{\partial}{\partial\gamma_{11}}\, tr\, K^{-1}A'Z'ZA = \sum_{i=1}^{m}\sum_{j=1}^{m} k^{ij}\frac{\partial}{\partial\gamma_{11}} m_{ji} + \sum_{i=1}^{m}\sum_{j=1}^{m}\left(\frac{\partial}{\partial\gamma_{11}} k^{ij}\right) m_{ji}$$

$$= tr\, K^{-1}\left(\frac{\partial}{\partial\gamma_{11}} M\right) + tr\left(\frac{\partial K^{-1}}{\partial\gamma_{11}}\right) M .$$

If we repeat this operation until all elements of δ have been exhausted, and write the result in the form of a column vector, we shall denote this by

$$\text{(A.4)}\qquad \frac{\partial}{\partial\delta}\ \text{tr}\ K^{-1}M = \text{tr}\ K^{-1}\left(\frac{\partial}{\partial\delta}M\right) + \text{tr}\left(\frac{\partial K^{-1}}{\partial\delta}\right)M\ .$$

Although the notation is grossly inaccurate the symbolism is useful for our purposes.

Let us concentrate on the first component and observe that

$$\text{(A.5)}\qquad m_{ji} = (y_{\cdot j} - Y_j\beta_{\cdot j} - X_j\gamma_{\cdot j})'\ (y_{\cdot i} - Y_i\beta_{\cdot i} - X_i\gamma_{\cdot i})\ .$$

Consequently, $\gamma_{\cdot 1}$ appears only in the elements m_{1j} and m_{j1}, $j = 1,2,\ldots,m$. Thus

$$\text{(A.6)}\qquad \text{tr}\ K^{-1}\left(\frac{\partial M}{\partial\gamma_{\cdot 1}}\right) = \sum_{j=1}^{m} k^{1j}\ \frac{\partial m_{j1}}{\partial\gamma_{\cdot 1}} + \sum_{j=1}^{m} \frac{\partial m_{1j}}{\partial\gamma_{\cdot 1}}\ k^{j1}$$

$$= -2\ X_1' \sum_{j=1}^{m} k^{1j}[y_{\cdot j} - Y_j\beta_{\cdot j} - X_j\gamma_{\cdot j}]\ .$$

Similarly differentiating with respect to $\beta_{\cdot 1}$ we have

$$\text{(A.7)}\qquad \text{tr}\ K^{-1}\left(\frac{\partial M}{\partial\beta_{\cdot 1}}\right) = \sum_{j=1}^{m} k^{1j}\ \frac{\partial m_{j1}}{\partial\beta_{\cdot 1}} + \Sigma\ \frac{\partial m_{1j}}{\partial\beta_{\cdot 1}}\ k^{j1}$$

$$= -2\ Y_1'\ \Sigma\ k^{1j}[y_{\cdot j} - Y_j\beta_{\cdot j} - X_j\gamma_{\cdot j}]\ .$$

Evidently we obtain a similar equation when we differentiate with respect to $\gamma_{\cdot i}$, $\beta_{\cdot i}$, $i = 2,3,\ldots,m$. Consequently

$$\text{(A.8)} \qquad \operatorname{tr} K^{-1}\left(\frac{\partial M}{\partial \gamma}\right) = -2\, X^{*\prime}(K^{-1} \otimes I_T)(y - Z^*\delta)$$

$$\text{(A.9)} \qquad \operatorname{tr} K^{-1}\left(\frac{\partial M}{\partial \beta}\right) = -2\, Y^{*\prime}(K^{-1} \otimes I_T)(y - Z^*\delta)$$

and finally

$$\text{(A.10)} \qquad \operatorname{tr} K^{-1}\left(\frac{\partial M}{\partial \delta}\right) = 2\, Z^{*\prime}(K^{-1} \otimes I_T)[Z^*\delta - y] \; .$$

Turning now to the second term in the right member of (A.4) we observe that the derivative is null with respect to all the elements in γ. We then may confine ourselves to derivatives with respect to elements of β, i.e., elements of the matrix B. We have

PROPOSITION 1. Let A be an $m \times m$ nonsingular matrix. Then

$$\text{(A.11)} \qquad \frac{\partial A^{-1}}{\partial a_{ij}} = -A^{-1} e_i' e_j A^{-1}, \qquad\qquad e_i = (0,\ldots,0,1,0,\ldots,0\quad)$$

e_i being an m-element row vector whose elements are all zero save the ith, which is unity.

Proof. Obvious from the identity

$$AA^{-1} = I \; .$$

An immediate cosequence of Proposition 1 is

PROPOSITION 2.

$$\frac{\partial}{\partial b_{ij}} K^{-1} = K^{-1}[(I - B')e_i'e_j + e_j'e_i(I - B)]K^{-1} .$$

Proof. Note that

$$K = I - B - B' + B'B \tag{A.12}$$

and apply Proposition 1.

PROPOSITION 3.

$$\operatorname{tr}\left(\frac{\partial}{\partial b_{ij}} K^{-1}\right)M = 2\, b^{*\prime}_{\cdot j} q_{i\cdot}$$

where $b^{*}_{\cdot j}$ is the jth column of $(I - B')^{-1}$ and $q_{i\cdot}$ is the ith row of

$$Q = (I - B')^{-1\prime} M(I - B)^{-1} = (Y - X\Pi)'(Y - X\Pi) . \tag{A.13}$$

Proof. Applying Proposition 2, we have

$$\operatorname{tr}(I - B)^{-1}(I - B')^{-1}[(I - B')e_i'e_j + e_j'e_i(I - B)](I - B)^{-1}(I - B')^{-1}M \tag{A.14}$$
$$= \operatorname{tr}[e_i'e_j(I - B)^{-1} + (I - B')^{-1}e_j'e_i]\,Q .$$

But $(I - B')^{-1}e_j'e_i$ <u>consists of a matrix all of whose columns are null</u>

except the ith, which is given by $b^*_{\cdot j}$; since

$$e_i' e_j (I - B)^{-1} = [(I - B')^{-1} e_j' e_i]'$$

$e_i' e_j (I - B)^{-1}$ consists of a matrix all of whose rows are null except the ith which is given by $b^{*\prime}_{\cdot j}$. Consequently,

$$\text{(A.15)} \qquad \mathrm{tr}\left(\frac{\partial}{\partial b_{ij}} K^{-1}\right) M = 2\, b^{*\prime}_{\cdot j} q_i . \qquad \text{q.e.d.}$$

Finally, we have

PROPOSITION 4.

$$\frac{d}{d\delta} \;\mathrm{tr}\, K^{-1}M = 2[Z^{*\prime}(K^{-1} \otimes I_T)Z^*\delta - Z^{*\prime}(K^{-1} \otimes I_T)y + \begin{pmatrix} 0 \\ V^{*\prime} \end{pmatrix} (I_m \otimes V)b^*]$$

where

$$\text{(A.16)} \qquad V^* = \mathrm{diag}(V_1, V_2, \ldots, V_m),\; V_i = Y_i - X\Pi_i,$$

$$V = Y - X\Pi,\; b^* = (b^{*\prime}_{\cdot 1}, b^{*\prime}_{\cdot 2}, \ldots, b^{*\prime}_{\cdot m})' .$$

Proof. Immediate from Proposition 2 and 3 if we take into account the restrictions on the elements of B.

The expression above may be simplified considerably if we note that the ith subvector of $V^{*\prime}(I_m \otimes V)b^*$ is given by

$$\text{(A.17)} \qquad V_i' V b^*_{\cdot i} = V_i' \; ZA(I - B)^{-1}\, b^*_{\cdot i} = V_i' \; ZA\; k^*_{\cdot i}$$

where $k^*_{\cdot i}$ is the ith column of K^{-1}. Hence

$$\text{(A.18)} \quad V^{*\prime}(I_m \otimes V)b^* = \begin{pmatrix} V_1' \; ZA \; k^*_{\cdot 1} \\ V_2' \; ZA \; k^*_{\cdot 2} \\ \vdots \\ V_m' \; ZA \; k^*_{\cdot m} \end{pmatrix} = V^{*\prime}(K^{-1} \otimes I_T)(y - Z^*\delta) .$$

Consequently, we may also write

$$\text{(A.19)} \quad \frac{d}{d\delta} \operatorname{tr} K^{-1}M = 2\Big[Z^{*\prime}(K^{-1} \otimes I_T)Z^*\delta - Z^{*\prime}(K^{-1} \otimes I_T)y + \begin{pmatrix} 0 \\ V^{*\prime} \end{pmatrix} (K^{-1} \otimes I_T)y - \begin{pmatrix} 0 \\ V^{*\prime} \end{pmatrix} (K^{-1} \otimes I_T)Z^*\delta\Big] .$$

Notice that even though the notation does not make it explicit V^* depends on δ.

The matrix of second derivatives with respect to δ may be written down, in a somewhat inaccurate notation, as

$$2\Big[Z^{*\prime}(K^{-1} \otimes I_T)Z^* - \begin{pmatrix} 0 \\ V^{*\prime} \end{pmatrix} (K^{-1} \otimes I_T)Z^* - Z^{*\prime}\left(\frac{d}{d\delta} K^{-1} \otimes I_T\right)(y - Z^*\delta) + \begin{pmatrix} 0 \\ V^{*\prime} \end{pmatrix} \left(\frac{d}{d\delta} K^{-1} \otimes I_T\right)(y - Z^*\delta) + \begin{pmatrix} 0 \\ \frac{dV^*}{d\delta} \end{pmatrix}' (K^{-1} \otimes I_T)(y - Z^*\delta)\Big] .$$

The expression $Z^{*\prime} \; \frac{d}{d\delta} K^{-1} \otimes I_T \; (y - Z\delta^*)$, for example, though notationally atrocious indicates the matrix of derivatives of $Z^{*\prime}(K^{-1} \otimes I_T)(y - Z\delta^*)$ with respect to δ only to the extent that such parameters enter in K^{-1}.

It is then entirely obvious that if we evaluate the matrix above at $\delta = \delta^0$, divide by T and take probability limits the last three terms will vanish and we shall obtain

$$\text{(A.20)} \qquad \underset{T\to\infty}{\text{plim}} \frac{1}{T} [\bar{Z}^{*\prime}(K^{-1} \otimes I_T)Z^*] = \lim_{T\to\infty} \frac{1}{T} [\bar{Z}^{*\prime}(K^{-1} \otimes I_T)\bar{Z}^*] = \Psi \; .$$