7

SMALL SAMPLE AND ASYMPTOTIC RELATIONS BETWEEN MAXIMUM LIKELIHOOD AND THREE STAGE LEAST SQUARES ESTIMATORS

by

Phoebus J. Dhrymes

Discussion Paper No. 7

Small Sample and Asymptotic Relations Between Maximum Likelihood and Three Stage Least Squares Estimators*

Phoebus J. Dhrymes

1. Introduction.

The relationship between full information maximum likelihood (FIML) and three stage least squares (3SLS) estimators has received considerable attention in the literature. In particular, Madansky [5], Sargan [7] and Rothenberg and Leenders [6] have shown that, asymptotically, the two procedures are equivalent in the sense of having the same asymptotic distribution under a given set of assumptions. Chow [1] examines such relations by comparing the minimands of the two procedures but his derivation of the 3SLS estimator appears to be in error. The present author [2], [3] has shown that the maximand of the FIML procedure can be decomposed into two components, one of which converges to zero in probability upon division by the sample size. It is then shown that 3SLS may be viewed as maximizing the first component given a prior consistent estimate of the convariance matrix of the system's structural errors.

---

*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.

Hence, it was argued, iterating 3SLS until convergence is obtained will not yield the FIML estimator, because we disregard the second component. On the other hand, this approach makes clear the asymptotic equivalence of the two procedures. An objection may be raised - an unfounded one as it turns out - that in such arguments one ought to be dealing with the estimators themselves, not with their associated maximands.

In this paper we establish the nature of the small sample difference between 3SLS and FIML estimators, which makes it absolutely transparent why the two estimators have the same asymptotic distribution. The difference of the two estimators reduces essentially to the manner in which the explanatory current endogenous variables are "purged" of their stochastic component. We also point out how the Chow [1] derivation of the 3SLS estimator is in error and indeed we obtain a "linearized" FIML estimator similar in motivation to the one given in Chow [2]. This "linearized" estimator requires little additional computation over what is required for 3SLS and has the same asymptotic distribution as FIML and 3SLS. Indeed, 3SLS may be viewed as a kind of "linearized" FIML.

2. Specifications, Assumptions and Notation.

The standard simultaneous equations model may be written in the notation and conventions of [4] as

$$Y = YB + XC + U \tag{1}$$

where

$$Y = (y_{\cdot 1}, y_{\cdot 2}, \ldots, y_{\cdot m}),\ X = (x_{\cdot 1}, x_{\cdot 2}, \ldots, x_{\cdot G}),\ U = (u_{\cdot 1}, u_{\cdot 2}, \ldots, u_{\cdot m}) \tag{2}$$

the $y_{.i}$, $x_{.j}$, $u_{.i}$ being, respectively the T element (column) vectors of observations on the $i^{\underline{th}}$ jointly dependent variable, $j^{\underline{th}}$ predetermined variable and structural error of the $i^{\underline{th}}$ equation. It is assumed that all identities have been substituted out and that all equations obey the rank condition for identifiability. If lagged endogenous are included among the predetermined variables the system is, in addition, assumed to be stable. Moreover, the second order moment matrix of the current endogenous and predetermined variables is assumed to have a well defined nonsingular probability limit.

The $t^{\underline{th}}$ observation on the system in (1) may be written as

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

where

$$y_{t.} = (y_{t1}, y_{t2}, \ldots, y_{tm}), \; x_{t.} = (x_{t1}, x_{t2}, \ldots, x_{tG}), \; u_{t.} = (u_{t1}, u_{t2}, \ldots, u_{tm})$$

it being implied that the system contains m jointly dependent and G predetermined variables.

Concerning the error structure we assume that the vectors $\{u'_{t.} : t = 1,2,\ldots\}$ are mutually independent identically distributed and moreover

$$(4) \qquad u'_{t.} \sim N(0, \Sigma)$$

Σ being a positive definite matrix. No restrictions are imposed on Σ. The identifiability conditions exclude some variables from each equation so that we may write, for example,

(5) $$y_{.i} = Y_i\beta_{.i} + X_i\gamma_{.i} + u_{.i}, \qquad i = 1,2,\ldots,m$$

the vectors $\beta_{.i}$, $\gamma_{.i}$ containing, respectively, m_i and G_i elements not known a priori to be zero.

3. Chow's Derivation.

In his interesting paper, [1], Chow claims that 3SLS is obtained by minimizing

(6) $$|\Sigma^*| = |A'\tilde{Z}'\tilde{Z}A|$$

where

(7) $$Z = (Y,X), \quad A = [I - B', C']' \quad \tilde{Z} = (\tilde{Y},X) \quad \tilde{Y} = X(X'X)^{-1}X'Y \ .$$

Unfortunately, however, this is false and moreover if one does follow such procedure one would obtain either a degenerate estimator or an inconsistent one. This may be shown quite easily as follows: From equation (7.10) in [1] we easily see that the estimator obtained by minimizing (6) obeys

(8) $$[\tilde{Z}^{*\prime}(\Sigma^{*-1} \otimes I_T)\tilde{Z}^*]\delta = \tilde{Z}^{*\prime}(\Sigma^{*-1} \otimes I_T)\tilde{y}$$

where

$$\tilde{Z}^* = \text{diag}(\tilde{Z}_1,\tilde{Z}_2,\ldots,\tilde{Z}_m) \quad \tilde{y} = (\tilde{y}'_{.1},\tilde{y}'_{.2},\ldots,\tilde{y}'_{.m})', \quad \tilde{Z}_i = (\tilde{Y}_i,X_i)$$

(9) $$\delta = (\delta'_{.1},\delta'_{.2},\ldots,\delta'_{.m})' \quad \delta_{.i} = (\beta'_{.i},\gamma'_{.i})'$$

and $\tilde{Y}_i$ is an appropriate submatrix of $\tilde{Y}$ in (7). But in (8) the matrix Σ^{*-1} contains the unknown parameters $\delta_{\cdot i}$ and in the derivation of 3SLS one utilizes a prior consistent estimate for them, typically the two stage least squares (2SLS) one. If we do so - indeed if we substitute <u>any</u> consistent estimator for $\delta_{\cdot i}$ - we see that

$$(10) \qquad \underset{T\to\infty}{\text{plim}} \frac{1}{T} \tilde{\Sigma}^* = \frac{\tilde{A}'\tilde{Z}'\tilde{Z}\tilde{A}}{T} = \lim_{T\to\infty} \frac{1}{T} [C'X'XC - C'X'XC] = 0.$$

Consequently, this estimator is not well defined in the limit. On the other hand if (8) is solved iteratively until convergence is obtained then either the resulting estimator is inconsistent, or else the same indeterminacy noted in (10) will prevail. Finally, if any of the equations of the system obey the rank conditions for just identifiability and for $\delta_{\cdot i}$, in Σ^{*-1}, of (8) we substitute its 2SLS estimator, say $\tilde{\delta}_{\cdot i}$, an appropriate row and column of $\tilde{\Sigma}^*$ will become zero. This is so since

$$(11) \quad \sigma^*_{ij} = (\tilde{y}_{\cdot i} - \tilde{Z}_i\delta_{\cdot i})'(\tilde{y}_{\cdot j} - \tilde{Z}_j\delta_{\cdot j})$$

$$= [X(X'X)^{-1}X'(y_{\cdot i} - Z_i\delta_{\cdot i})]'[X(X'X)^{-1}X'(y_{\cdot j}-Z_j\delta_{\cdot j})] \quad i,j = 1,2,\ldots,m.$$

If for $\delta_{\cdot i}$ we substitute its 2SLS estimator then (11) becomes

$$(12) \qquad \tilde{\sigma}^*_{ij} = \tilde{u}'_{\cdot i}X(X'X)^{-1}X'\tilde{u}_{\cdot j}, \qquad \tilde{u}_{\cdot i} = y_{\cdot i} - Z_i\tilde{\delta}_{\cdot i} \; .$$

If the $i^{\underline{th}}$ equation is just identified then it may be shown, [4, p. 198], that the 2SLS residuals are orthogonal to the predetermined variables of the system. Consequently

$$(13) \qquad \tilde{\sigma}^*_{ij} = 0 \qquad j = 1,2,\ldots,m$$

and the inverse $\tilde{\Sigma}^{*-1}$ does not exist.

In fact, 3SLS estimators are obtained by minimizing

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

while 2SLS estimators are obtained by minimizing

$$\text{tr } A'\tilde{Z}'\tilde{Z} A$$

as the reader may readily verify.

4. The Relation Between FIML and 3SLS estimators.

Given the assumptions in (4) we can write the likelihood function of the sample as

$$(14) \quad L(A, \Sigma; Y,X) = -\frac{mT}{2}\ln(2\pi) - \frac{T}{2}\ln|\Sigma| + \frac{T}{2}\ln|(I - B)'(I - B)| - \frac{T}{2}\text{tr}\{\Sigma^{-1} A' \frac{Z'Z}{T} A\} .$$

Partially maximizing the function with respect to the elements of Σ - which are assumed to be unrestricted - we find

$$(15) \qquad \Sigma(A) = \frac{A'Z'Z A}{T}$$

and concentrating the likelihood function we have

$$(16) \qquad L(A;Y,X) = -\frac{mT}{2}[\ln(2\pi)+1] - \frac{T}{2}\ln\left|A'\frac{Z'Z}{T}A\right| + \frac{T}{2}\ln|(I-B)'(I-B)| \; .$$

Defining

$$(17) \qquad N = X(X'X)^{-1}X' , \qquad \tilde{V} = (I-N)Y$$

we can rewrite (16) more suggestively as

$$(18) \qquad L(A;Y,X) = -\frac{mT}{2}[\ln(2\pi)+1] - \frac{T}{2}\ln\left|\frac{\tilde{V}'\tilde{V}}{T}\right| - \frac{T}{2}\ln\left|\frac{A'Z'Z'A}{T}\right| + \frac{T}{2}\ln\left|(I-B)'\frac{\tilde{V}'\tilde{V}}{T}(I-B)\right| \; .$$

Differentiating successively with respect to the columns of A, after a priori restrictions have been imposed, we find

$$(19) \qquad [Z^{*\prime}(\Sigma^{-1}\otimes I_T)Z^* - V^{*\prime}(S^{-1}\otimes I_T)V^*]\delta = Z^{*\prime}(\Sigma^{-1}\otimes I_T)y - V^{*\prime}(S^{-1}\otimes I_T)v$$

where

$$(20) \qquad V^* = \operatorname{diag}(\tilde{V}_1^*,\tilde{V}_2^*,\ldots,\tilde{V}_m^*) \quad v = (\tilde{v}'_{\cdot 1},\tilde{v}'_{\cdot 2},\ldots,\tilde{v}'_{\cdot m})', \quad \tilde{V}_i^* = (\tilde{V}_i,0) \; .$$

$\tilde{V}_i$ being a submatrix of $\tilde{V}$ and bearing the same relationship to the latter as Y_i bears to Y; the $\tilde{v}_{\cdot i}$ are the columns of $\tilde{V}$.

In addition

$$S = (I - B)' \frac{\tilde{V}'\tilde{V}}{T} (I - B) \tag{21}$$

Since Σ in (19) is as defined in (15) and S as defined in (21) it follows that (19) is a highly nonlinear function of δ and can only be solved by iteration. On the other hand the preceding discussion affords a particularly simple method of "linearizing" the FIML estimator.

Let $\tilde{\Sigma}$, $\tilde{S}$ be the matrices resulting when for the unknown parameter δ we substitute its 2SLS estimator $\tilde{\delta}$ in Σ and S respectively. The "linearized" FIML estimator of δ is then easily obtained as

$$\hat{\delta} = [Z^{*\prime}(\tilde{\Sigma}^{-1} \otimes I_T)Z^* - V^{*\prime}(\tilde{S}^{-1} \otimes I_T)V^*]^{-1} [Z^{*\prime}(\tilde{\Sigma}^{-1} \otimes I_T)y - V^{*\prime}(\tilde{S}^{-1} \otimes I_T)v] \tag{22}$$

It is now simple to verify that the estimator above has the same asymptotic distribution as the FIML estimator. Since the distribution of the latter is well known we shall confine ourselves to a very brief demonstration. To this effect define

$$\tilde{V}_i = (I - N)Y_i, \quad Y_i^* = (Y_i, 0), \quad Y^* = \mathrm{diag}(Y_1^*, Y_2^*, \ldots, Y_m^*) \tag{23}$$

and note that

$$V^* = [I_m \otimes (I - N)]Y^*, \qquad v = [I_m \otimes (I - N)]y \tag{24}$$

Consequently, (22) may be written as

$$(25) \quad \hat{\delta} = [Z^{*\prime}(\tilde{\Sigma}^{-1} \otimes I_T)Z^* - Y^{*\prime}[\tilde{S}^{-1} \otimes (I - N)]Y^*]^{-1}[Z^{*\prime}(\tilde{\Sigma}^{-1} \otimes I_T) - Y^{*\prime}[\tilde{S}^{-1} \otimes (I - N)]]y \ .$$

Since $y = Z^*\delta + u, \quad u = (u'_{.1}, u'_{.2}, \ldots, u'_{.m})'$ it follows that

$$(26) \quad \hat{\delta} - \delta = [Z^{*\prime}(\tilde{\Sigma}^{-1} \otimes I_T)Z^* - Y^{*\prime}[\tilde{S}^{-1} \otimes (I - N)]Y^*]^{-1}[Z^{*\prime}(\tilde{\Sigma}^{-1} \otimes I_T) - Y^{*\prime}(\tilde{S}^{-1} \otimes (I - N)]u$$

The desired result is then immediate if we note that, asymptotically,

$$(27) \quad \sqrt{T}\,(\hat{\delta} - \delta) \sim \left[\frac{(Z^* - V^*)'(\Sigma^{-1} \otimes I_T)(Z^* - V^*)}{T}\right]^{-1} \frac{1}{\sqrt{T}}\,[(Z^* - V^*)'(\Sigma^{-1} \otimes I_T)]u \ .$$

which yields exactly the asymptotic distribution of the 3SLS estimator; and hence that of the FIML estimator.

REAMRK 1: The expression in (25) would appear to be particularly convenient for computational purposes. Note that, in iterating, we need only recompute the $m \times m$ matrices $\tilde{\Sigma}(\delta)$, $\tilde{S}(\delta)$. Moreover, it is known that under suitable regularity conditions the normal equations of the maximum likelihood procedure have at most one consistent root and this corresponds to the one that gives the global maximum of the likelihood function; thus, it would appear that if we begin the iteration with, say, the 2SLS estimator of δ and the iteration converges we would indeed have found the FIML estimator since we are assured of the consistency of the convergent iterate.

Now, what is the small sample relation between FIML and 3SLS estimators? The answer is readily determined from (25). First we note that the 3SLS estimator may be written as

$$(28) \qquad \hat{\delta}_{3SLS} = [Z^{*\prime}(\tilde{\Sigma}_3^{-1} \otimes I_T)Z^* - Y^{*\prime}[\tilde{\Sigma}_3^{-1} \otimes (I-N)]Y^*]^{-1}[Z^{*\prime}(\tilde{\Sigma}_3^{-1} \otimes I_T) - Y^{*\prime}[\tilde{\Sigma}_3^{-1} \otimes (I-N)]]y$$

where

$$(29) \qquad \tilde{\Sigma}_3 = \frac{\tilde{A}'Z'Z\tilde{A}}{T}$$

and $\tilde{A}$ has been obtained by 2SLS methods. If we iterate 3SLS, and the iteration converges, we shall obtain an estimator obeying (28) but in this case $\tilde{\Sigma}_3$ of (29) will be computed with $\tilde{A}$ as obtained by 3SLS methods.

On the other hand if we iterate (25) and the iteration converges then the FIML estimator will obey (25) but with

$$(30) \qquad \tilde{\Sigma} = \frac{\tilde{A}'_{ML}Z'Z\tilde{A}_{ML}}{T}, \qquad \tilde{S} = (I - \tilde{B}_{ML})' \frac{\tilde{V}'\tilde{V}}{T}(I - \tilde{B}_{ML})$$

In general, a converging iteration of 3SLS will not produce the FIML estimator. Intuitively, the essential difference between the two estimators is that FIML employs the quantities

$$\tilde{\sigma}^{ij}Y_i'Y_j - \tilde{s}^{ij}\tilde{V}_i'\tilde{V}_j, \qquad \tilde{\sigma}^{ij}Y_i'y_{\cdot j} - \tilde{s}^{ij}\tilde{V}_i'\tilde{v}_{\cdot j}$$

while the 3SLS estimator operates with the quantities

$$\tilde{\sigma}^{ij}(Y_i - \tilde{V}_i)'(Y_j - \tilde{V}_j), \qquad \tilde{\sigma}^{ij}(Y_i - \tilde{V}_i)'(y_{\cdot j} - \tilde{v}_{\cdot j})$$

The "reason" why asymptotically the two estimators are equivalent in terms of their asymptotic distribution is that

$$\text{(31)} \qquad \underset{T\to\infty}{\text{plim}}\ \tilde{\Sigma} = \underset{T\to\infty}{\text{plim}}\ \tilde{S} = \Sigma$$

where $\tilde{\Sigma}$, $\tilde{S}$ are defined as in (30) in the case of the FIML estimator, and $\tilde{\Sigma}$ is defined as in (29) in the case of the 3SLS estimator. Notice that $\tilde{S}$ is the estimator of Σ obtained from the residuals of the unrestricted reduced form, as modified in (30). On the other hand, $\tilde{\Sigma}$ in (30) and $\tilde{\Sigma}_3$ in (29) are estimators of Σ obtained from the residuals of the restricted reduced form, the former as induced by the FIML estimator of δ, the latter as induced by the 2SLS estimator of δ. Are there any conditions under which, for every sample size, FIML and 3SLS estimators will coincide? The answer appears to be yes, and the condition is that all equations of the system obey the rank (and order) condition for just identifiability. If in (25) we commence the iteration with $\hat{\delta}_{2SLS}$, then $\tilde{S}$ and $\tilde{\Sigma}$, thus computed, would be identical, since indirect least squares and 2SLS estimated structural parameters will coincide. Consequently, the first iterate would be simply the 3SLS estimate. But under just identifiability conditions 3SLS and 2SLS estimators coincide. Thus, nothing will be gained by further iteration. Indeed under conditions of just identifiability (for all equations of the system) it would appear that 2SLS, 3SLS and FIML estimators are identical for every sample size.

The preceding discussion has therefore established

THEOREM: Consider the model in (3) and (4) together with the conditions customarily assumed for such simultaneous equations models. Then

i. The "linearized" FIML estimator exhibited in (25) has the same asymptotic distribution as the FIML estimator

ii. Iterating 3SLS until convergence is obtained does not yield the FIML estimator

iii. Under conditions of just identifiability for all the equations of the system, FIML and 3SLS estimators coincide for every sample size.

5. Conclusion.

In this paper we have elucidated the small sample relation between 3SLS and FIML estimators. Moreover, we have established that iteration of 3SLS does not yield FIML estimates. An interesting byproduct of this approach is the result that under just identifiability conditions for all the equations of the system FIML and 3SLS estimators coincide for every sample size. Since it is known that 2SLS and 3SLS also coincide under such conditions, it is therefore established that in such a case all commonly employed limited and full information estimators yield identical estimates - apart from roundoff errors.

REFERENCES

1. Chow, G. C: "A Comparison of Simultaneous Estimators for Simultaneous Equations" <u>Econometrica</u>, vol. 32 (1964) pp. 532-553.

2. ____________: "Two Methods of Computing Full Information Maximum Likelihood Estimates in Simultaneous Stochastic Equations" <u>International Economic Review</u>, vol. 9 (1968) pp. 100-112.

3. Dhrymes, P. J: "A Relation Between Three Stage Least Squares and Full Information Maximum Likelihood Estimators," Discussion Paper No. 83 (1968) Department of Economics, University of Pennsylvania.

4. ____________: <u>Econometrics</u>: <u>Statistical Foundations and Applications</u> New York, Harper and Row, 1970.

5. Madansky, A: "On the Efficiency of Three Stage Least Squares Estimation" <u>Econometrica</u>, vol. 32 (1964) pp. 51-56.

6. Rothenberg, T. and C. T. Leenders: "Efficient Estimation of Simultaneous Equation Systems" <u>Econometrica</u>, vol. 32 (1964) pp. 57-76.

7. Sargan, J. D: "Three Stage Least Squares and Full Maximum Likelihood Estimates" <u>Econometrica</u>, vol. 32 (1964) pp. 77-81.