A COMPARISON OF SOME LIMITED INFORMATION ESTIMATORS FOR DYNAMIC SIMULTANEOUS EQUATIONS MODELS WITH AUTOCORRELATED ERRORS

By

Phoebus J. Dhrymes, R. Berner and D. Cummins

Discussion Paper Number 25
October 1972

Preliminary Report on Research in Progress
Not to be quoted without permission of the author.

A Comparison of Some Limited Information Estimators for Dynamic Simultaneous Equations Models with Autocorrelated Errors[†]

by

Phoebus J. Dhrymes, R. Berner and D. Cummins

0. Introduction.

While most econometric models extant contain lagged endogenous among their predetermined variables, relatively little attention has been paid to the complications entailed by the, possibly, autoregressive character of structural errors. Indeed, most such models are estimated by some variant of two stage least squares (2SLS). The problem has been examined first by Sargan [15] and later by Amemiya [1] and Fair [9]. Sargan operates in a very general context and Amemiya assumes in his models the absence of lagged endogenous variables. The present paper has been stimulated by Fair's contribution which, while an interesting addition to the literature of such models leaves a number of questions unanswered.

In a previous version [8] we considered a simple two step estimator whose asymptotic distribution was invariant to iteration. Here we shall examine systematically the problem of limited information estimation in the context of a model containing first order lags in the endogenous variables and an autogregressive error process of the first order.

[†] The research on which this paper is based was, in part, supported by NSF Grant GS2289 at the University of Pennsylvania (Dhrymes), the National Defense Education Act, Title IV (Berner) and the S. S. Huebner Foundation (Cummins)

Extension to higher order lags in the dependent variables and higher order autoregressions is rather routine.

We shall show that one of the procedures suggested by Fair is similar, but not identical to the limited information variant of the estimators discussed in Dhrymes [4]. We shall also show how a modified method of scoring approach yields estimators which are asymptotically equivalent to the converging iterate of the two stage least squares estimator (C2SLSA) as discussed in [4].

Finally, having produced the asymptotic distributions of the extimators above, we shall obtain a ranking of the various alternatives based on the covariance matrices of the appropriate asymptotic distributions.

1. Formulation of the Problem

Consider the standard linear dynamic structural econometric model for which we have a sample of size T. It may be written, compactly, as

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

where Y is $T \times m$, X is $T \times G$ and U is $T \times m$ being, respectively, the matrices of current endogenous variables, predetermined variables and error terms of the system. The matrices B, C are $m \times m$, $G \times m$ and comprise, subject to certain identifiability restrictions, the unknown structural parameters of the problem.

Following the practice in Amemiya [1] and Fair [9] we write the error specification as

$$U = U_{-1} R + E \tag{2}$$

where

(3) $\quad R = \mathrm{diag}(\rho_1, \rho_2, \ldots, \rho_m), \ |\rho_i| < 1 \qquad i = 1,2,\ldots,m$

It is assumed that the rows of E, i.e., $\varepsilon_{t\cdot} = (\varepsilon_{t1}, \varepsilon_{t2}, \ldots, \varepsilon_{tm})$, $t = 1,2,\ldots,T$ constitute a random sample, or more generally $\{\varepsilon'_{t\cdot} = t = 0, \pm 1, \pm 2, \ldots\}$ is a sequence of independent identically distributed (i.i.d.) random variables. In addition, it is assumed that

(4) $$E(\varepsilon'_{t\cdot}) = 0, \quad \mathrm{Cov}(\varepsilon'_{t\cdot}) = \Sigma$$

Σ being a general (unrestricted) positive definite matrix, and that the equations of the system are identifiable.

Remark 1. The fact that R is assumed to be a diagonal matrix in no way implies that the error terms of the system are uncorrelated or mutually independent across equations. One may, quite easily, show that

(5) $$\Omega = R \Omega R + \Sigma, \qquad \Omega = \mathrm{Cov}(u'_{t\cdot})$$

Consequently, the u_{ti}, $i = 1,2,\ldots,m$ are not generally uncorrelated (or in the case of normality independent).

Let us now focus on one of the equations of the system, say the first.

We may write

(6) $$y_{\cdot 1} = Y_1 \beta_{\cdot 1} + X_1 \gamma_{\cdot 1} + u_{\cdot 1}$$

where $y_{\cdot 1}$, $u_{\cdot 1}$ are, respectively, the first columns of Y and U; $\beta_{\cdot 1}$, $\gamma_{\cdot 1}$ are the first columns of B, C after elements known to be zero have been suppressed; Y_1, X_1 are, thus, appropriate submatrices of Y, X.

For simplicity we assume that the system contains only first order lags[1].

Consequently, partition

$$X_1 = ({}_1Y^*_{-1}, W_1) \tag{7}$$

where ${}_1Y^*_{-1}$ is an appropriate submatrix of Y_{-1} and W_1 is an appropriate submatrix of W, a $T \times s$ matrix containing all the exogenous variables of the system.

To conclude this section it seems appropriate to restate Fair's proposal in the context of our notation. This will elucidate at least one of the alternatives he has dealt with and indicate the similarities and differences of his procedure relative to the estimators to be discussed below.

While a number of suggestions are put forth, perhaps the most clearly enunciated of Fair's procedures is the following:

i. regress the variables in Y_1 on a set of variables that, at least <u>includes</u>[2] $Y_{1,-1}$, the variables in Y_1 lagged one period the variables in ${}_1Y^*_{-1}$ as well their one period lags and the variables in W_1 as well as their one period lags and obtain the "predicted" matrix $\hat{Y}_1$.

---

[1] Higher order lags can be easily accommodated at the cost of some analytical complications. No essentially new problem, however, is introduced.

[2] This is the aspect of the procedure that makes it ill specified.

ii. Regress $(y_{t1} - \rho_1 y_{t-1,1})$ on the variables contained in the matrices $(\hat{Y}_1 - \rho_1 Y_{1,-1})$, $({}_1Y^*_{-1} - \rho_1\, {}_1Y^*_{-2})$, $(W_1 - \rho_1 W_{1,-1})$ repeatedly, with ρ varying over the interval, say, $[-.99,.99]$ and select that regression which minimizes the sum of squared residuals. From that regression one, thus, obtains estimators $\delta_{\cdot 1}$, ρ_1, say $\tilde{\delta}_{\cdot 1}$, $\tilde{\rho}_1$.

A difficulty with this procedure is that the list of regressors is not specified and no attempt is made to determine whether the properties of the structural estimators are affected by the choice of regressors at the first stage. Moreover, if we follow the procedure not merely for one equation of the system but for all we would expect to have a situation in which a variable in Y_1 is "explained" (at the first stage) by different specifications depending on which structural equation is considered at the moment.

2. A Two Step Alternative

Consider again, under the standard assumptions, the model in (1), whose first equation is as exhibited in (9). Let W be the $T \times s$ matrix of exogenous variables, in the entire system, of which W_1 is an appropriate submatrix. Using W_1 and as many other columns of W (and W_{-1}, if necessary) as there are columns in Y_1 and ${}_1Y^*_{-1}$ estimate[3] $\delta_{\cdot 1}$ by instrumental variables. More precisely, let P_1 be the matrix of instruments so selected. The first stage estimator of $\delta_{\cdot 1}$ is then

$$(10) \qquad \tilde{\delta}_{\cdot 1} = (P_1' Z_1)^{-1} P_1' y_{\cdot 1}$$

[3] It is assumed that the model contains enough exogenous variables for this to be feasible.

Under the standard assumptions - a matter we shall examine at greater length below - the estimator in (10) is consistent.

Do this for every equation and thus estimate consistently the parameters of B and C not known a priori to be zero.

Partition conformably,

$$(11) \qquad X = (Y_{-1}, X), \quad \Pi = \begin{pmatrix} \Pi_0 \\ \Pi_1 \end{pmatrix} \quad \Pi = C(I - B)^{-1}$$

and note that the t^{th} row of the reduced form is given by

$$(12) \qquad y_{t\cdot} = y_{t-1\cdot}\,\Pi_0 + w_{t\cdot}\,\Pi_1 + v_{t\cdot}, \quad v_{t\cdot} = u_{t\cdot}\,(I - B)^{-1}$$

$u_{t\cdot}$ being the t^{th} row of U, $y_{t\cdot}$ the t^{th} row of Y and $w_{t\cdot}$ the t^{th} row of W.

Since $\tilde{B}$, $\tilde{C}$ have been estimated consistently we can obtain "predictions" of $y_{t\cdot}$, recursively, from

$$(13) \qquad \tilde{y}_{t\cdot} = \tilde{y}_{t-1\cdot}\,\tilde{\Pi}_0 + w_{t\cdot}\,\tilde{\Pi}_1$$

given some fixed initial condition, the simplest of which is $\tilde{y}_{0\cdot} = 0$

Remark 2. The initial condition $\tilde{y}_{0\cdot} = 0$ is of course quite arbitrary and the particular estimates are likely, in small samples, to be sensitive to this requirement i.e., the numbers one obtains in any given application are likely to vary as we alter the specification. The properties of the estimators, however, are unaffected by such practices if the model is stable, as one customarily assumes. For alternative ways of handling this aspect see Pesaran [14]. Such alternatives, proposed in the context of single equation models, entail specific assumptions on the sequence of exogenous variables $\{w_{t\cdot} : t = 0, \pm 1, \pm 2 \ldots\}$

In this connection, it should be pointed out that Monte Carlo studies, in a single equation context, indicate that small sample "bias" and "mean square errors" are not materially affected by whether we "estimate" initial conditions or we set them equal to zero; see for example Morrison [12]. On the other hand such "estimates" of initial conditions are not consistent.

From (13) we thus obtain the matrix $\tilde{Y}$, of "predictions" and lagging every element once we obtain $\tilde{Y}_{-1}$.

In addition, from the first stage estimators we have

$$\tilde{u}_{\cdot 1} = y_{\cdot 1} - Z_1 \tilde{\delta}_{\cdot 1} \tag{13}$$

and from these residuals we obtain, by the usual methods, $\tilde{\rho}_1$. Define the matrices

$$\tilde{V}_1^{-1} = \begin{vmatrix} 1 & -\tilde{\rho}_1 & 0 & 0 \cdots 0 \\ -\tilde{\rho}_1 & 1 + \tilde{\rho}_1^2 & -\tilde{\rho}_1 & 0 \cdots 0 \\ 0 & -\tilde{\rho}_1 & 1 + \tilde{\rho}_1^2 & \cdots 0 \\ \cdots & -\tilde{\rho}_1 & 1 + \tilde{\rho}_1^2 & -\tilde{\rho}_1 \\ 0 \cdots & & -\tilde{\rho}_1 & 1 \end{vmatrix} \qquad \tilde{Z}_1 = (\tilde{Y}_1, \tilde{Y}^*_{-1}, W_1) \tag{14}$$

and obtain the instrumental variables estimator

$$\tilde{\delta}_{\cdot 1} = (\tilde{Z}_1' \tilde{V}_1^{-1} Z_1)^{-1} \tilde{Z}_1' \tilde{V}_1^{-1} y_{\cdot 1} \tag{15}$$

Evidently, this procedure may be applied to every equation yielding, in the obvious notation

$$\tilde{\delta}_{\cdot i} = (\tilde{Z}_i' \tilde{V}_i^{-1} Z_i)^{-1} \tilde{Z}_i' \tilde{V}^{-1} y_{\cdot i} \qquad i = 1,2,\ldots,m \tag{16}$$

Remark 3. It may be thought that the estimator above is equivalent to the limited information maximum likelihood estimator when the structural errors are jointly normal. This conjecture may be prompted by the similarity of the procedure above to the interpretation given by Amemiya and Fuller [2] to the distributed lag estimator proposed by Hannan [9]. This, however, is not so. Perhaps the simplest way of noting the differences and similarities is to revert to the single equation model. If we write

$$y_t = \alpha w_t + \lambda y_{t-1} + u_t, \quad u_t = \rho u_{t-1} + \varepsilon_t \tag{17}$$

the estimation procedure we suggest in the preceding amounts of the following: Estimate α, λ by instrumental variables, obtain the residuals and estimate ρ. Compute

$$\tilde{y}_{t-1} = \tilde{\alpha} \sum_{i=0}^{t-2} \tilde{\lambda}^i w_{t-1-i} \tag{18}$$

Define

$$\tilde{X} = (\tilde{y}_{-1}, w) \tag{19}$$

in the obvious notation, and obtain

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\lambda} \end{pmatrix} = (\tilde{X}' \tilde{V}^{-1} X)^{-1} \tilde{X}' \tilde{V}^{-1} y \tag{20}$$

If

$$(21) \qquad \operatorname*{plim}_{T \to \infty} \frac{(\tilde{X}' \tilde{V}^{-1} X)}{T} = Q$$

exists as a nonsingular nonstochastic matrix

and if

$$(22) \qquad \operatorname*{plim}_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} u_t w_{t+\tau} = 0 \quad \text{for all} \quad \tau$$

then the asymptotic distribution of (20) is given by

$$(23) \qquad \sqrt{T} \left[\begin{pmatrix} \hat{\alpha} \\ \hat{\alpha} \end{pmatrix} - \begin{pmatrix} \alpha \\ \lambda \end{pmatrix} \right] \sim N(0, \Phi_{IIV}), \; \Phi_{IIV} = \sigma^2 \operatorname*{plim}_{T\to\infty} \left(\frac{\overline{X}' V^{-1} \overline{X}}{T} \right)^{-1}$$

where

$$(24) \qquad \overline{X} = (\overline{y}_{-1}, w) \; \overline{y}_{-1} = (\overline{y}_1, \overline{y}_2, \ldots, \overline{y}_{T-1}) \; \overline{y}_t = \alpha \sum_{t=0}^{\infty} \lambda^i w_{t-i}$$

On the other hand the maximum likelihood (ML) (or in case of nonnormality, the minimum chi square) estimator of α and λ has the distribution (see e.g. [6, ch. 7]

$$(25) \qquad \sqrt{T} \left[\begin{pmatrix} \hat{\alpha} \\ \hat{\lambda} \end{pmatrix}_{ML} - \begin{pmatrix} \alpha \\ \lambda \end{pmatrix} \right] \sim N(0, \Phi_{ML}) \; .$$

where

$$(26) \qquad \Phi_{ML}^{-1} = \sigma^2 \left[\lim_{T \to \infty} \frac{\overline{X}' V^{-1} \overline{X}}{T} + S, \right] \quad S = \begin{bmatrix} 0 & 0 \\ 0 & \dfrac{(\rho - \lambda)^2}{(1-\lambda^2)\,(1-\lambda\rho)^2} \end{bmatrix}$$

It is thus easy to see that

$$(27) \qquad \Phi_{IIV} - \Phi_{ML} \geq 0$$

in the sense that the difference is positive semidefinite.

In the Amemiya and Fuller problem we have

$$(28 \qquad y_t = \alpha \sum_{i=0}^{\infty} \lambda^i w_{t-i} + u_t, \quad u_t = \rho u_{t-1} + \varepsilon_t \, .$$

Here λ is a parameter that enters only the mean of the dependent variable, but not its variance. For this "reason" the estimators of λ and ρ (by ML methods) are asymptotically independent. "Hence", the asymptotic properties of the ML estimators of α and λ are independent of the properties of the consistent estimator of ρ. If we solve the difference equation in (17) we find

$$(29) \qquad y_t = \alpha \sum_{i=0}^{\infty} \lambda^i w_{t-i} + \sum_{i=0}^{\infty} \lambda^i u_{t-i}$$

Thus, in (29), we have a second order autoregressive error process and λ is both a mean and a variance parameter. "Consequently", the estimators of λ and ρ are not independent even asymptotically. Thus, how well λ is estimated will depend on how well we estimate ρ in the "first" stage since we treat these two parameters asymmetrically.

The point of using the iterated instrumental variables (IIV) approach instead of the direct maximum likelihood is that the IIV estimator does not depend, in its asymptotic distribution, on anything but the consistency of the "first stage" estimator of ρ.

3. Properties of the Estimator

Consider again the model as exhibited in (12).

We assume

(A.1) The matrix Π_0 is stable, in the sense that its roots are less than unity, in modulus.

(A.2) The sequence $\{\varepsilon'_{t\cdot} : t = 0, \pm 1, \pm 2, \ldots\}$ is one of mutually independent identically distributed random variables (i.i.d.) with mean zero and covariance matrix Σ, and the matrix R, in (3), is stable.

(A.3) The exogenous variables are (uniformly) bounded nonstochastic[4] and

$$\lim_{T\to\infty} \frac{1}{T} (\bar{Z}'_1 V^{-1} \bar{Z}_i) = \Phi_{ii}^{-1} \qquad i = 1,2,\ldots,m$$

exist as nonsingular matrices, where

$$\bar{Z}_i = (\bar{Y}_i, {}_i\bar{Y}^*_{-1}, W_i)$$

and $\bar{Y}_i$, ${}_i\bar{Y}^*_{-1}$ are obtained from the systematic part of the final form

$$y'_t = (I - \Pi'_0 L)^{-1} \Pi'_1 w'_{t\cdot} + (I - \Pi'_0 L)^{-1} v'_{t\cdot} \tag{30}$$

[4] This assumption may be relaxed at the cost of some analytical complications, as will be indicated below.

Remark 4. The assumption in (A.3) entails some restrictions, when it so happens that the columns of $\bar{Z}_i$ are linearly dependent (at least asymptotically). This will occur, e.g., when the system contains only two exogenous variables the two being a sine and a cosine[5]. When this is so $\bar{Y}_i$, ${}_i\bar{Y}^*_{-1}$ contain only linear combinations of sines and cosines thus rendering (A.3) invalid. The same is true when the only exogenous variables of the system are polynomials of various degrees in "time". In most applications, however, these limitations are innocuous since the various columns of $\bar{Y}_i$, ${}_i\bar{Y}^*_{-1}$ are (infinite) linear combinations of all the predetermined variables <u>and</u> all their lags. In general we would not expect such linear dependencies to materialize, and (A.3) rules out explicitly the case where the t^{th} row of $\bar{Y}_i$, ${}_i\bar{Y}^*_{-1}$ consist of linear combinations of the t^{th} observations on the exogenous variables of the system.

If we define

$$(30)\qquad \begin{aligned} \tilde{Z}^* &= \mathrm{diag}(\tilde{Z}_1,\tilde{Z}_2,\dots,\tilde{Z}_m), \quad \delta = (\delta'_{\cdot 1},\delta'_{\cdot 2},\dots,\delta'_{\cdot m})' \\ \tilde{V}^{-1} &= \mathrm{diag}(V_1^{-1},V_2^{-1},\dots,V_m^{-1}), \quad y = (y'_{\cdot 1},y'_{\cdot 2},\dots,y'_{\cdot m})' \end{aligned}$$

then the IIV estimator of the system as a whole is given by

$$(31)\qquad \tilde{\delta} = (\tilde{Z}'^* \tilde{V}^{-1} Z^*)^{-1} \tilde{Z}'^* \tilde{V}^{-1} y$$

We have

LEMMA 1. Given assumptions (A.1) through (A.3) and the conventions of the model as exhibited in (1), the estimator in (31) is consistent.

[5] This limitation was pointed out to us by a reader of the earlier version of this paper.

Proof. After substitution for y we obtain

(32) $$\tilde{\delta} - \delta = (\tilde{Z}'* \tilde{V}^{-1} Z)^{-1} \tilde{Z}'* \tilde{V}^{-1} u , \quad u = (u'_{\cdot 1}, u'_{\cdot 2}, \ldots, u'_{\cdot m})' .$$

It will suffice to show that

(33) $$\operatorname*{plim}_{T \to \infty} \frac{\tilde{Z}'* \tilde{V}^{-1} u}{T} = 0 .$$

This is so, since

(34) $$\operatorname*{plim}_{T \to \infty} \frac{\tilde{Z}' \tilde{V}^{-1} Z}{T} = \Phi^*, \quad \Phi^* = \operatorname{diag}(\Phi^*_{11}, \Phi^*_{22}, \ldots, \Phi^*_{mm})$$

and the Φ^*_{ii} are nonsingular by (A.3).

Consider the i^{th} subvector of (33). We observe that

(35) $$\operatorname*{plim}_{T \to \infty} \frac{\tilde{Z}'_i \tilde{V}^{-1}_i u_{\cdot i}}{T} = \operatorname*{plim}_{T \to \infty} \frac{\bar{Z}'_i V^{-1}_i u_{\cdot i}}{T} = 0$$

The last equality follows easily from (A.2) and the fact that the exogenous variables of the system are bounded non-stochastic.

LEMMA 2. The asymptotic distribution of the estimator in (31) is given by

(37) $$\Phi^*_{IIV} = \lim_{T \to \infty} \frac{(\bar{Z}*' V^{-1} Z*)^{-1}}{T} \; \frac{\bar{Z}*' M*' (\Sigma \otimes I_T) M* \bar{Z}*}{T} \; \frac{(\bar{Z}*' V^{-1} Z*)^{-1}}{T}$$

$$\bar{Z}* = \operatorname{diag}(\bar{Z}_1, \bar{Z}_2, \ldots, \bar{Z}_m), \quad M'_i M_i = V^{-1}_i$$

$$M* = \operatorname{diag}(M_1, M_2, \ldots, M_m)$$

Proof. First, we observe that

$$\underset{T\to\infty}{\text{plim}} \frac{\tilde{Z}_i'(\tilde{V}_i^{-1} - V_i^{-1})u_{\cdot i}}{\sqrt{T}}$$

(38)

$$= \text{plim}\ \sqrt{T}(\tilde{\rho}_i - \rho_i)\ \frac{1}{T}\tilde{Z}_i' \begin{bmatrix} 0 & -1 & 0 & \cdots & & 0 \\ -1 & \tilde{\rho}_i + \rho_i & -1 & & & \\ 0 & -1 & \tilde{\rho}_i + \rho_i & -1 & 0 & \\ 0 & & \vdots & & & \\ \vdots & & -1 & \tilde{\rho}_i + \rho_i & -1 & \\ 0 & 0 & \cdots & 0 & -1 & 0 \end{bmatrix} u_{\cdot i} = 0$$

since $\sqrt{T}\,(\tilde{\rho}_i - \rho_i)$ has a well defined asymptotic distribution as $T \to \infty$. Next, by similar arguments, we see that

(39)
$$\underset{T\to\infty}{\text{plim}} \frac{(\tilde{Z}_i - \bar{Z}_i)'\ V_i^{-1} u_{\cdot i}}{\sqrt{T}} = 0$$

Thus, we need be concerned only with

(40)
$$\frac{\bar{Z}^{*\prime}\ V^{-1}u}{\sqrt{T}} = \frac{\bar{Z}^{*\prime}\ M^{*\prime}\ \varepsilon^*}{\sqrt{T}}$$

where $\varepsilon^* = (\varepsilon^{*\prime}_{\cdot 1}, \varepsilon^{*\prime}_{\cdot 2}, \ldots, \varepsilon^{*\prime}_{\cdot m})'$
$\varepsilon_{\cdot i} = (\varepsilon_{1i}, \varepsilon_{2i}, \ldots, \varepsilon_{Ti})'$ and $\varepsilon^*_{\cdot i} = \varepsilon_{\cdot i}$ with the exception that
$\varepsilon^*_{1i} = \sqrt{1 - \rho_i^2}\, u_{t1}$, $i = 1,2,\ldots,m$.

Let ${}_i\bar{z}^*_{s\cdot}$ be the s^{th} row of $M_i\bar{Z}_i$ and note that

$$(41) \qquad \frac{Z^{*\prime} V^{-1} u}{\sqrt{T}} = \frac{1}{\sqrt{T}} \begin{bmatrix} {}_1\bar{z}^{*\prime}_{1\cdot} \\ \\ {}_m\bar{z}^{*\prime}_{1\cdot} \end{bmatrix} \varepsilon^{*\prime}_{1\cdot} + \frac{1}{\sqrt{T}} \sum_{s=2}^{T} \begin{vmatrix} {}_1\bar{z}^{*\prime}_{s\cdot} \\ {}_2\bar{z}^{*\prime}_{s\cdot} \\ {}_m\bar{z}^{*\prime}_{s\cdot} \end{vmatrix} \varepsilon^{\prime}_{s\cdot}$$

where $\varepsilon_{s\cdot} = (\varepsilon_{s1}, \varepsilon_{s2}, \ldots, \varepsilon_{sm})$.

But the first term in the right member of (41) vanishes in probability with T; morevoer in view of (A.1), (A.2) and (A.3) the Lindeberg-Feller Theorem [3, ch. 3] applies and thus we conclude

$$(42) \qquad \frac{\bar{Z}' V^{-1} u}{\sqrt{T}} \sim N\left(0, \lim_{T\to\infty} \frac{\bar{Z}^{*\prime} M^{*\prime} (\Sigma \otimes I_T) M^* \bar{Z}^*}{T}\right)$$

It is then immediate from (32) that

$$(43) \qquad \sqrt{T}\,(\tilde{\delta} - \delta) \sim N(0, \Phi^*_{IIV}) \qquad \text{q.e.d.}$$

COROLLARY 1. Iterating the estimator in (31) will not alter its asymptotic properties.

Proof. Obvious from Lemmas 1, 2.

Remark 5. The motivation for the two step alternative considered here may be seen from (38). It is evident that the asymptotic distribution in (43) does not depend on the distribution of the "first step" estimators of the autoregressive parameters ρ_i, $i = 1,2,\ldots,m$.

This is so since the vector multiplying $\sqrt{T}\,(\tilde{\rho}_i - \rho_i)$ in the right member of (38) vanishes asymptotically. It is in order to effectuate this property that we have defined both $\tilde{Y}_i$ and ${}_i\tilde{Y}^*_{-1}$ in terms of the final form of the system.

The estimator we shall consider in the next section will lack this simplicity since its distribution will be shown to depend on the properties of the quantities $\sqrt{T}\,(\tilde{\rho}_i - \rho_i)$ as obtained from the "first stage".

4. Two and Three Stage Least Squares Variants in the Autoregressive Errors Case

In this discussion it is useful, but not necessary, in motivating the estimators to assume, in addition to (A.2), that

$$\varepsilon'_{t\cdot} \sim N(0,\Sigma) \tag{44}$$

It may be shown, see e.g., [4], that the log likelihood function of the observations may be written as

$$L(A,R,\Sigma;\,Y,W) = -\frac{Tm}{2}\ln(2\pi) - \frac{1}{2}\ln|\Omega| - \frac{1}{2}z_{1\cdot}A\Omega^{-1}A'z'_{1\cdot} - \frac{T}{2}\ln\left|\frac{\tilde{V}'\tilde{V}}{T}\right|$$
$$- \frac{T-1}{2}\ln|\Sigma| + \frac{T}{2}\ln\left|(I-B)'\frac{\tilde{V}'\tilde{V}}{T}(I-B)\right| - \frac{1}{2}\operatorname{tr}\{\Sigma^{-1}(ZA - Z_{-1}AR)'(ZA - Z_{1}AR)\} \tag{45}$$

In (45) the symbols have a slightly different meaning than was the case earlier; in particular, we have put

$$Y = YB + Y_{-1}C_0 + WC_1 + U, \quad ZA = U \tag{46}$$

where, of course,

(47) $$Z = (Y, Y_{-1}, W), \qquad A = (I - B', -C_0' - C_1')'$$

and the observation corresponding to $t = 2$ is eliminated from the definition of Z in (47). This observation appears separately as the row $z_{1\cdot}$. Further, it is clear that the terms $-\frac{1}{2}[\ln|\Omega| + z_{1\cdot} A\Omega^{-1}A' z_{1\cdot}']$ play no role in obtaining the (asymptotic) properties of the maximum likelihood (ML) estimators; this is so since the relative weight of such terms vanishes as $T \to \infty$.

It is shown in [4] that the three stage least squares (3SLS) analogue in the present model is obtained by minimizing, with respect to the unknown elements of A,

(48) $$\Lambda = \mathrm{tr}\, \Sigma^{-1} (\tilde{Z}A - Z_{-1}AR)' (\tilde{Z}A - Z_{-1}AR)$$

subject to a prior estimate of Ω and R. In the preceding

(49) $$\tilde{Z} = (\tilde{Y}, Y_{-1}, W),$$

and $\tilde{Y}$ is the projection of Y on the space spanned by the columns of $Y_{-1}, Y_{-2}, W, W_{-1}$.

If $\tilde{A}$ is a prior estimate of A, then the prior estimates of Σ and R are given by

(50) $$\tilde{\Sigma} = \frac{1}{T}\tilde{E}'\tilde{E}, \quad \tilde{R} = (\tilde{U}_{-1}'\tilde{U}_{-1})^{-1}\tilde{U}_{-1}'\tilde{U}, \quad \tilde{U} = Z\tilde{A}, \quad \tilde{E} = \tilde{U} - \tilde{U}_{-1}\tilde{R}$$

It is, further, shown in [4] that if we iterate this procedure (with respect to $\tilde{A}$, $\tilde{\Sigma}$ and $\tilde{R}$) until it converges then the resulting esimators have the same asymptotic distribution as the ML estimators of

A, Σ, R. While this is not shown directly, what is shown is that these two sets of estimators satisfy, asymptotically the same set of equations.

We specialize these results to the case under consideration, i.e., to the case where R is diagonal.

Before proceeding to this task it will be useful to introduce a more convenient notation to overcome the confusion arising from the many subscripts one has to employ in such situations. Thus, introduce the selection matrices S_{i1}, S_{i2}, S_{i3} such that

(51) $$YS_{i1} = Y_i, \quad Y_{-1}S_{i2} = {}_iY^*_{-1} \quad WS_{i3} = W_i, \qquad i = 1,2,\ldots,m.$$

Putting

(52) $$S_i = \begin{bmatrix} S_{i1} & 0 & 0 \\ 0 & S_{i2} & 0 \\ 0 & 0 & S_{i3} \end{bmatrix}, \qquad S = \mathrm{diag}(S_1, S_2, \ldots, S_m)$$

we have that

(53) $$Z_i = ZS_i \qquad Z^* = (I_m \otimes Z)S, \qquad Z^*_{-1} = (I_m \otimes Z_{-1})S$$

$$y = Z^*\delta + u$$

At any rate minimizing Λ of (48) with respect to the unknown elements of A we find

(54) $$[\tilde{Z}^* - (\tilde{R}' \otimes I)Z^*_{-1}]' (\tilde{\Sigma}^{-1} \otimes I) [\tilde{Z}^* - (\tilde{R}' \otimes I)Z^*_{-1}]\hat{\delta}$$

$$= [\tilde{Z}^* - (\tilde{R}' \otimes I)Z^*_{-1}]' (\tilde{\Sigma}^{-1} \otimes I) [y - (\tilde{R}' \otimes I)y_{-1}]$$

where y is a defined in (30) - except that observations for $t = 2$ are ommitted - and y_{-1} is obtained by reducing the time subscript of the elements of y by one.

It is clear that the two stage least squares (2SLS) variant of this estimator is obtained by neglecting the stochastic dependence of the system's errors across equations, which means, operationally, that we set, in (54), $\tilde{\Sigma}^{-1} = I$. Thus, substituting for y in (54) we find

$$(55) \qquad \hat{\delta} - \delta = \{[\tilde{Z}^* - (\tilde{R} \otimes I)Z^*_{-1}]' \; [\tilde{Z}^* - (\tilde{R}' \otimes I)Z^*_{-1}]\}^{-1} \; [\tilde{Z}^* - (\tilde{R}' \otimes I)Z^*_{-1}]' \; [u - (\tilde{R}' \otimes I)u_{-1}]$$

Remark 6. The relations in (54) and (55) show that it is not necessary to define the $\tilde{Y}$ component of $\tilde{Z}$ as the projection of Y on the space spanned by the columns of Y_{-1}, Y_{-2}, W, W_{-1}. This is necessary only in order to obtain orthogonality between the matrix of residuals and Z^*_{-1} and thus justify the transition from (54) to (55). It is interesting that an (asymptotically) equivalent procedure can be derived that makes repeated use of instrumented variables estimators. To be precise, if $\tilde{Y}$ is not obtained by projection then it is necessary that it be obtained from the relationship

$$(56) \qquad \tilde{Y} = Y_{-1}\tilde{F}_1 + Y_{-2}\tilde{F}_2 + W\tilde{F}_3 + W_{-1}\tilde{F}_4$$

where

$$\tilde{F}_1 = \tilde{C}_0(I - \tilde{B})^{-1} + (I - \tilde{B})\tilde{R}(I - \tilde{B})^{-1}, \quad \tilde{F}_2 = \tilde{C}_0\tilde{R}(I - \tilde{B})^{-1}$$

$$\tilde{F}_3 = \tilde{C}_1(I - \tilde{B})^{-1}, \quad \tilde{F}_4 = \tilde{C}_1\tilde{R}(I - \tilde{B})^{-1}$$

The alternative, but asymptotically equivalent procedure is as follows: Estimate, by instrumental variables methods the vectors $\beta_{\cdot i}$, $\gamma_{\cdot i}$, $i = 1,2,\ldots,m$ using the relations in (6) and thus obtain the estimates $\tilde{B}$, $\tilde{C}_0$, $\tilde{C}_1$. Using these estimates we can compute the residual matrix $\tilde{U}$ and obtain an estimate of R by computing

$$(57) \qquad \tilde{\rho}_i = (\Sigma\, \tilde{u}_{t-1,i}\, \tilde{u}_{ti})/(\Sigma\, \tilde{u}^2_{t-1,i}), \qquad i = 1,2,\ldots,m$$

This is all that is required to obtain the matrices $\tilde{F}_1$, $\tilde{F}_2$, $\tilde{F}_3$, $\tilde{F}_4$ and thus $\tilde{Y}$. The estimator that is (asymptotically) equivalent to the one in (55) is

$$(58) \qquad \hat{\delta} = \{[\tilde{Z}* - (\tilde{R}' \otimes I)Z*_{-1}]' \, [Z* - (\tilde{R}' \otimes I)Z*_{-1}]\}^{-1} \, [\tilde{Z}* - (\tilde{R}' \otimes I)Z*_{-1}]' \, [y - (\tilde{R}' \otimes I)y_{-1}]$$

Note that is <u>not necessary to recompute the</u> $\tilde{F}_i$, $i = 1,2,3,4$ since <u>iteration would rely only on</u> (57) <u>and</u> (58) <u>and would not involve</u> (56).

<u>Remark 7</u>. It is shown in [4] that the asymptotic properties of the estimator in (54) do not depend on any of the properties of $\tilde{\Sigma}$ <u>beyond consistency</u>: they do, however, depend rather crucially on the estimator of R. Provided Σ and R are initially estimated consistently, all subsequent iterates are shown to be consistent as well.

<u>Remark 8</u>. The procedure discussed in connection with (58) is the exact analogue, in the autoregressive errors case, of the LIIV estimator given in Dhrymes [7]; where the estimator is modified to take into account

the covariance matrix Σ, i.e., when it is based on (54) rather than on (55) then it is the exact analogue of the FIIV estimator, also given in [7]. One difference ought to be noted, however: while in the absence of autocorrelated errors the LIIV and FIIV extimators are asymptotically equivalent to 2SLS and 3SLS estimators respectively, this is not so here. The estimator in (58) (and its fullinformation analogue) would have to be iterated until convergence is obtained. <u>Only the convergent iterate is asymptotically equivalent to the autoregressive two and three stage least square estimators</u>.

To establish the asymptotic properties of the estimator in (54) and (55) we put

$$(59) \qquad \tilde{G}(\tilde{\Sigma}) = \frac{1}{T} [\tilde{Z}^* - (\tilde{R}' \otimes I)Z^*_{-1}]' \; [\tilde{\Sigma}^{-1} \otimes I] \; [\tilde{Z}^* - (\tilde{R}' \otimes I)Z^*]$$

and by slightly expanding (A.3) to read, in the relevant part,

$$(A.3)' \quad \underset{T\to\infty}{\text{plim}} \frac{1}{T} [\bar{Z}^* - (R' \otimes I)Z^*_{-1}]' \; (\Sigma^{-1} \otimes I) \; [\bar{Z}^* - (R' \otimes I]Z^*_{-1}]$$

exists as a nonstochastic nonsingular matrix we conclude that

$$(60) \qquad G(\Sigma) = \underset{T\to\infty}{\text{plim}} \frac{1}{T} [\tilde{Z}^* - (\tilde{R}' \otimes I)Z^*_{-1}] \; (\tilde{\Sigma}^{-1} \otimes I) \; [\tilde{Z}^* - (\tilde{R}' \otimes I)Z^*_{-1}]$$

exists as a nonstochastic nonsingular matrix.

It is then easy to show that the asymptotic distribution of the estimator in (54) is given by

$$(61) \quad \sqrt{T}\,(\hat{\delta} - \delta)_{3SLSA} \sim [G(\Sigma)]^{-1} \frac{1}{\sqrt{T}} [\bar{Z}^* - (R' \otimes I)Z^*_{-1}]' \, (\Sigma^{-1} \otimes I) \{\varepsilon - [(\tilde{R} - R)' \otimes I]u_{-1}\}$$

where, we remind the reader,

(62) $\bar{Z}^* = (I_m \otimes \bar{Z})S, \quad \bar{Z} = (\bar{Y}, Y_{-1}, W), \quad \bar{Y} = Y_{-1}F_1 + Y_{-2}F_2 + WF_3 + W_{-1}F_4$

the F_i being the probability limits of the quantities $\tilde{F}_i$, $i = 1,2,3,4$ defined in (57)

We now proceed to find the distribution of the limited information estimator as exhibited in (55). Clearly, this is given by

$$\sqrt{T}\,(\hat{\delta} - \delta)_{2SISA} \sim [G(I)^{-1} \frac{1}{\sqrt{T}} [\bar{Z}^* - (R' \otimes I)Z^*_{-1}]' \{\varepsilon - [(\tilde{R} - R) \otimes I]u_{-1}\} . \tag{63}$$

Let $\bar{z}^{(i)}_{t\cdot}$ be the t^{th} row of $\bar{Z}_i = \bar{Z}S_i$, $i = 1,2,\ldots,m$; $z^{(i)}_{t\cdot}$ is, obviously, the t^{th} row of $Z_i = ZS_i$, $i = 1,2,\ldots,m$. We remind the reader that, in this context, $t = 3,4,\ldots,T$. Let

$$\bar{z}^{*(i)}_{t\cdot} = \bar{z}^{(i)}_{t\cdot} - \rho_i z^{(i)}_{t-1\cdot}, \quad z^*_t = \mathrm{diag}(\bar{z}^{*(1)}_{t\cdot}{}', \bar{z}^{*(2)}_{t\cdot}{}', \ldots, \bar{z}^{*(m)}_{t\cdot}{}') \tag{64}$$

Let

$$\operatorname*{plim}_{T\to\infty} \frac{1}{T} [\bar{Z}^* - (R' \otimes I)Z^*_{-1}]u_{-1} = \begin{bmatrix} \xi_{11} \\ \xi_{22} \\ \cdot \\ \cdot \\ \cdot \\ \xi_{mm} \end{bmatrix} \tag{65}$$

ξ_{ii} being the probability limit of $\frac{1}{T}(\bar{Z}_i - \rho_i\, {}_iZ_{-1})'u_{\cdot i,-1}$; similarly,

$$\operatorname*{plim}_{T\to\infty} \frac{1}{T} [Z^* - (R' \otimes I)Z^*_{-1}]'u_{-1} = \begin{bmatrix} \zeta_{11} \\ \zeta_{22} \\ \cdot \\ \cdot \\ \cdot \\ \zeta_{mm} \end{bmatrix} \tag{66}$$

(67) $$\xi = \text{diag}(\xi_{11}, \xi_{22}, \ldots, \xi_{mm}), \qquad \zeta = \text{diag}(\zeta_{11}, \zeta_{22}, \ldots, \zeta_{mm}) .$$

and define ξ^* to be the matrix in which ξ_{ii} is replaced by

$\frac{1 - \rho_i^2}{\sigma_{ii}} \xi_{ii}$. Finally, letting

(68) $$u_t^* = \text{diag}(u_{t1}, u_{t2}, \ldots, u_{tm})$$

We see that we may rewrite (63) as

(69) $$\sqrt{T}(\tilde{\delta} - \delta)_{2SLSA} \sim G^{-1}\left[\frac{1}{\sqrt{T}} \sum_{t=3}^{T} z_t^* \varepsilon_{t\cdot}' - \frac{1}{\sqrt{T}} \xi^* \sum_{t=3}^{T} u_{t-1}^* \varepsilon_{t\cdot}' + \xi^* \zeta' \sqrt{T}(\tilde{\delta} - \delta)\right]$$

where $\tilde{\delta}$ is the initial estimator of the structural parameters on the basis of which we estimate the autoregressive parameters ρ_i, $i = 1, 2, \ldots, m$ and $G(I)$ was written, for simplicity, as G. The third term in the right member of (69), thus, shows quite clearly how the estimator of the autoregressive parameters affects, in this context, even the asymptotic properties of the system's (other) structural parameter estimators.

To complete the problem we must obtain an explicit expression for the ξ_{ii} and ζ_{ii}. To this effect define

(70) $$\zeta_i = \left[\frac{(\Pi_0' - \rho_i I)(I - \rho_i \Pi_0')^{-1}(I - B)^{-1}(I - \rho_i R)^{-1} + (I - B')^{-1}(I - \rho_i R)^{-1}}{(1 - \rho_i^2)(I - \rho_i \Pi_0')^{-1}(I - B)^{-1}(I - \rho_i R)^{-1}}\right] \sigma_{\cdot i}$$

where $\sigma_{\cdot i}$ is the i^{th} column Σ.

We note, then, that

(71) $$\zeta_{ii} = S_i' \zeta_{i\cdot} \qquad \xi_{ii} = \zeta_{ii}$$

Instead of obtaining the distribution of $\sqrt{T}\,(\hat{\delta} - \delta)_{2SLSA}$ for arbitrary estimator $\tilde{\delta}$, we obtain the distribution of the converging iterate of such estimator.

Asymptotically, this estimator behaves as

(72) $$(G - \xi^*\zeta')\,\sqrt{T}\,(\hat{\delta} - \delta)_{C2SLSA} \sim \frac{1}{\sqrt{T}} \left[\sum_{t=3}^{T} (z_t^* - \xi^* u_{t-1}^*)\varepsilon_{t\cdot}' \right]$$

Unfortunately, the right member of (72) is not the sum of independent random vectors. Thus, the Lindeberg-Feller theorem invoked in Section 3 is not applicable here. The random vectors in (72) are dependent; because of the special character of the autoregression from which this dependence arises, it is possible to use the results of Mann and Wald [12] to convert the problem to one involving n-dependent variables. We illustrate this for the case

(73) $$\frac{1}{\sqrt{T}} \xi^* \sum_{t=3}^{T} u_{t-1\cdot}^* \varepsilon_{t\cdot}' = \frac{1}{\sqrt{T}} \xi^* \sum_{t=3}^{T} u_{t-1}^{*n} \varepsilon_{t\cdot}' + \frac{1}{\sqrt{T}} \xi^* R^n \sum_{t=3}^{T} u_{t-n-1}^* \varepsilon_{t\cdot}'$$

It is clear that the second term in the right member of (73) can be made arbitrarily small in probability by appropriate[6] choice of n, since its expectation is identically zero. Thus, the asymptotic behavior of the left member of (73) is determined, essentially, by the first term in

[6] To see this, note that for an arbitrary (conformable) vector of constants α, $\mathrm{Var}[1/\sqrt{T})\,\alpha'\xi^* R^n \Sigma_{t=3}^{T} u^*_{t-n-1}\varepsilon_{t\cdot}']$ can be made arbitrarily small by proper choice of n; apply, then, Chebyshev's inequality.

the right member. Evidently, we may apply the same reasoning to both terms of the right member of (72). This is perfectly feasible since the (stochastic) components of z_t^* may be expressed in terms of the final form; the latter, however, represents a rational lag in the exogenous variables and a rational lag in the ε-process.

Having used this truncation argument in (72) we do, in effect, find the asymptotic distribution of a random vector which differs from $\sqrt{T}\,(\hat{\delta} - \delta)_{C2SLSA}$ by an arbitrarily small quantity with probability as close to unity as desired. This is essentially the meaning of Lemma 1 in Mann and Wald [12].

Now, there are many variants of central limit theorems for n-dependent variables. If the process is assumed to be Gaussian then, in view of (A.3), the conditions of Hoeffding-Robbins theorem [11] or [6] will hold[7] and thus we conclude that, asymptotically,

$$(74) \qquad \frac{1}{\sqrt{T}} \sum_{t=3}^{T} (z_t^* - \xi^* u_{t-1}^*)\varepsilon_{t\cdot}' \sim N(0,\ A_1 + A_2 - A_3)$$

where

$$(75) \qquad A_1 = \underset{T\to\infty}{\text{plim}} \frac{1}{T} [\bar{Z}^* - (R' \otimes I)Z_{-1}^*]'(\Sigma \otimes I)\ [\bar{Z}^* - (R' \otimes I)Z_{-1}^*]$$

$$(76) \qquad A_2 = \left[r_{ij}^2 \frac{(1-\rho_i^2)(1-\rho_j^2)}{1-\rho_i\rho_j} \zeta_{ij}\zeta_{jj}' \right], \quad r_{ij}^2 = \frac{\sigma_{jj}^2}{\sigma_{ii}\sigma_{jj}}, \quad i,j = 1,2,\ldots,m$$

---

[7] The Hoeffding-Robbins theorem is usually stated for scalar random variables; its application to vector random variables is possible by the following result, see Dhrymes [3, ch. 3]. Let $\{x_T : T = 1,2,\ldots\}$ be a sequence of random vectors and α an arbitrary real conformable vector. If $\{z_T : z_T = \alpha' x_T,\ T = 1,2,\ldots\}$ converges in distribution to $N(\alpha'\mu, \alpha'\Sigma\alpha)$ for all real α, then $\{x_T : T = 1,2,\ldots\}$ converges in distribution to a $N(\mu,\Sigma)$ random vector.

(77) $$A_3 = \left[\sigma_{ij}(\xi^*_{ii}\zeta'_{ji} + \zeta_{ij}\xi^{*\prime}_{jj})\right]$$

and

(78) $$\zeta_{ji} = \operatorname*{plim}_{T\to\infty} \frac{1}{T} (\overline{Z}_j - \rho_j\, {}_jZ_{-1})'u_{\cdot i,-1} = S'_j\zeta_i$$

the ζ_i being as defined in (70)

It follows then, immediately that

(79) $$\sqrt{T}\,(\hat{\delta} - \delta)_{C2SILSA} \sim N(0,\ \Phi_{C2SILSA})$$

where

(80) $$\Phi_{C2SILSA} = (G - \xi^*\zeta')^{-1}\,(A_1 + A_2 - A_3)(G - \zeta\,\xi^{*\prime})^{-1}$$

We have therefore proved

LEMMA 3. Consider the linear structural econometric model as given in (1), (2) subject to the assumptions (A.1), (A.2), (A.3) and the additional requirements that the sequence $\{\varepsilon'_{t\cdot} : t = 0, \pm 1, \pm 2, \ldots\}$ has finite sixth order absolute moments and that

(81) $$G = \operatorname*{plim}_{T\to\infty} \frac{1}{T} [\tilde{Z}^* - (\tilde{R}' \otimes I)Z^*_{-1}]'[\tilde{Z}^* - (\tilde{R}' \otimes I)Z^*_{-1}]$$

exists as a nonsingular matrix with nonstochastic elements. Then the converging iterate of the limited information dynamic autoregressive (LIDA) estimator as exhibited in (72) has the asymptotic distribution

(82) $$\sqrt{T}\,(\hat{\delta} - \delta)_{C2SILSA} \sim N(0,\ \Phi_{C2SILSA})$$

where

$$(83) \qquad \Phi_{C2SLSA} = (G - \xi^*\zeta')^{-1} (A_1 + A_2 - A_3) (G - \zeta\xi^{*\prime})^{-1}$$

the A_i, $i = 1,2,3$ being as defined in (75), (76) and (77) respectively.

COROLLARY 2. The k^{th} (nonconvergent) iterate of the LIDA estimator does not have the same asymptotic distribution as the $(k+1)^{st}$ (nonconvergent) iterate or the convergent iterate $\sqrt{T}\,(\hat{\delta} - \delta)_{C2SLSA}$.

Proof. Obvious from (69)

Remark 9. It is interesting that we can give a two step estimator that is asymptotically equivalent to the C2SLSA estimator, by using a modified form of the method of scoring [3, ch. 3]. It ought to be noted that this is a general procedure applicable to all problems that obtain estimators by extremising some function. To this effect, note that the C2SLSA estimator is obtained by minimizing the function

$$(84) \qquad \Lambda = \mathrm{tr}(\tilde{Z}A - Z_{-1}AR)'(\tilde{Z}A - Z_{-1}AR)$$

with respect to the unknown elements in A and R. If we put

$$(85) \qquad \delta^* = (\delta, \rho^{*\prime})' \quad \rho^* = (\rho_1, \rho_2, \ldots, \rho_m)'$$

then the C2SLSA estimator is a consistent root of $\frac{\partial\Lambda}{\partial\delta^*} = 0$.

Now, let $\tilde{\delta}^*$ be any consistent estimator of δ^* and consider the estimator

$$(86) \qquad \hat{\delta}^* = \tilde{\delta}^* - \left[\frac{\partial^2\Lambda}{\partial\delta^*\partial\delta^*} (\tilde{\delta}^*) \right]^{-1} \frac{\partial\Lambda}{\partial\delta^*} (\tilde{\delta}^*)$$

Clearly, this estimator can always be obtained given the initial estimator $\tilde{\delta}*$ and is easily recognized as a modified form of the first iterate involved in the application of the method of scoring.

Now expand

$$(87) \qquad \frac{\partial\Lambda}{\partial\delta*}(\delta*) = \frac{\partial\Lambda}{\partial\delta*}(\delta_o^*) + \frac{\partial^2\Lambda}{\partial\delta*\partial\delta*}(\delta**)(\delta* - \delta_o^*)$$

where $|\delta** - \delta_o^*| < |\tilde{\delta}* - \delta_o^*|$, δ_o^* being the true value of the vector δ_0^*. Substituting (87) in (86) we obtain

$$(88) \qquad \sqrt{T}(\hat{\delta}* - \delta_0^*) = \sqrt{T}(\tilde{\delta}* - \delta_0^*) - \left[\frac{1}{T}\frac{\partial^2\Lambda}{\partial\delta*\partial\delta*}(\tilde{\delta}*)\right]^{-1}\left[\frac{1}{\sqrt{T}}\frac{\partial\Lambda}{\partial\delta*}(\delta_0^*) + \frac{1}{T}\frac{\partial^2\Lambda}{\partial\delta*\partial\delta*}(\delta**)\sqrt{T}(\tilde{\delta}* - \delta_0^*)\right]$$

Given the usual regularity conditions and the fact that, if $\text{plim}\ \tilde{\delta}* = \delta_0^*$, then $\text{plim}\ \delta** = \delta_0^*$ as well, we conclude that, asymptotically,

$$(89) \qquad \sqrt{T}(\hat{\delta}* - \delta_0^*) \sim -M^{-1}\frac{\partial S}{\partial\delta*}(\delta_0^*)$$

where

$$(90) \qquad M = \underset{T\to\infty}{\text{plim}}\ \frac{1}{T}\frac{\partial^2\Lambda}{\partial\delta*\partial\delta*}(\delta_0^*)$$

But the asymptotic distribution of the right member of (89) is that of the consistent root of the equation $\frac{\partial\Lambda}{\partial\delta*} = 0$. Since the C2SLSA estimator gives a consistent solution of this equation the (asymptotic) equivalence claimed above is demonstrated.[8] Finally, it should be pointed out that, in

---

[8] A certain uniqueness condition is clearly involved in this claim.

our derivation in (82) we have given the marginal asymptotic distribution of the δ component of δ^*, as the latter is defined in (85).

5. Description and Comparison of Alternative Estimators

There are a number of limited information estimators for problems similar to the one dealt with in this paper. However, no distribution theory exists except in cases which are far more restrictive than the ones considered here. Thus, e.g., Amemiya [1] deals with a model not containing lagged endogenous variables. We could, if we wished, specialize our results and compare with what he calls MS2SLS estimators. In Fair [9] a problem entirely similar to ours is consider but the distribution theory of his proposed estimators is not adequately developed. His search and iterated estimators are similar to the C2SLSA estimator developed in this model, although it is not clear what effect, if any, the selection of regressors in the first stage might have on the properties of the resulting estimators.

It will be helpful before we undertake a comparison to give an outline of the various procedures considered here.

1. Fair's procedure was outlined at the end of Section 1 and need not be repeated here.

2. Converging iterate two stage least squares autoregressive (C2SLSA). First obtain the reduced model

$$Y = Y_{-1}F_1 + Y_{-2}F_2 + WF_3 + W_{-1}F_4 + E$$

and regress Y on Y_{-1}, Y_{-2}, W, W_{-1} thus obtaining the projection $\tilde{Y}$, of Y, on the space spanned by the columns of Y_{-1}, Y_{-2}, W, W_{-1}.

Obtain by instrumental variables (or other) methods an initial consistent estimate of C_0, C_1, B and using these obtain an initial consistent estimate of ρ_i, $i = 1,2,\ldots,m$ as indicated in equation (57). Using these estimates (of the ρ_i) estimate δ as in (54) (putting $\tilde{\Sigma} = I$). From the new estimates of δ obtain the residual matrix $\tilde{U}$ and recompute the ρ_i as in (57). Using these recompute estimates of δ as in (54) and continue until convergence is obtained.[9]

3. The instrumental variables version of the C2SLSA estimator. This is essentially the same as under 2 except that $\tilde{Y}$ is not the projection of Y on the space spanned by the columns of Y_{-1}, Y_{-2}, W, W_{-1}. Rather, given the initial consistent estimates of C_0, C_1, B we compute $\tilde{Y}$ as

$$\tilde{Y} = Y_{-1}\tilde{F}_1 + Y_{-2}\tilde{F}_2 + W\tilde{F}_3 + W_{-1}\tilde{F}_4$$

the $\tilde{F}_i$, $i = 1,2,3,4$ being as defined just below equation (56). This version obtains estimators for δ and ρ_i, $i = 1,2,\ldots,m$ by iterating, until convergence, equations (57) and (58). Here the initial consistent estimate of C_0, C_1, B serves both in defining $\tilde{Y}$ as well as in obtaining the initial estimates of the ρ_i.

4. The modified method of scoring version of C2SLSA estimators. Here, from an initial consistent estimate of δ and ρ^*, say $\tilde{\delta}^*$, we evaluate $\frac{\partial \Lambda}{\partial \delta^*}$ $\frac{\partial^2 \Lambda}{\partial \delta^* \partial \delta^*}$ at the point $\tilde{\delta}^*$ and obtain the desired estimator as in equation (86). In the present case Λ is defined in (84) and the $\tilde{Y}$ component of $\tilde{Z}$ is the projection of Y on Y_{-1}, Y_{-2}, W, W_{-1}.

---

[9] It should be noted that no theorem has been produced giving the precise conditions under which convergence will obtain.

5. When the ρ_i, $i = 1,2,\ldots,m$ are known we may operate with the reduced model

$$y_{\cdot i} - \rho_i y_{\cdot i,-1} = (Z_i - \rho_i\, {}_iZ_{-1})\delta_{\cdot i} + \varepsilon_{\cdot i} \qquad i = 1,2,\ldots,m$$

treating $y_{\cdot i} - \rho_i y_{\cdot i,-1}$ and $Z_i - \rho_i\, {}_iZ_{-1}$ as known data. Here, operating analogously with the 2SLS procedure means that we use the estimator

$$[(\tilde{Z}_i - \rho_i\, {}_iZ_{-1})'(\tilde{Z}_i - \rho_i\, {}_iZ_{-1})]^{-1} (\tilde{Z}_i - \rho_i\, {}_iZ_{-1})' (y_{\cdot i} - \rho_i y_{\cdot i,-1}) \quad i=1,2,\ldots,m$$

the $\tilde{Y}_i$ component of $\tilde{Z}_i$ being obtained from the projection of Y on the space spanned by the columns of Y_{-1}, W.

6. The iterated (two-step) instrumental variables estimator developed in Section 2 is obtained as follows: From an initial (instrumental variables) estimator of C_0, C_1, B compute $\tilde{Y}$ from the final form as indicated in (13), and the $\tilde{\rho}_i$ as indicated in (57). The estimator is then as defined in (31) where the Y component of Z^* is obtained from the final form, i.e., for the i^{th} structural equation we use the matrix of instruments $\tilde{Z}_i' \tilde{V}_i^{-1}$, the $\tilde{Y}_i$ component of $\tilde{Z}_i$ being an appropriate submatrix of $\tilde{Y}$, as defined immediately above.

Fair has not explicitly derived the asymptotic distribution of his estimator and thus direct comparison is not possible in this case. On the other hand we may extend one of Fair's suggestions so that it becomes the C2SLSA estimator.

The procedures under 2., 3., and 4. are asymptotically equivalent, provided in the first two we employ an iterative scheme and the iteration converges.

Thus, effectively, the comparison is to be carried out with respect to C2SLSA (item 2), the obvious modification of 2SLS when the autocorrelation parameters are known (item 5) and the iterated (two-step) instrumental variables estimator (item 6).

Inspection of the matrices Φ^*_{IIV} and Φ_{C2SLSA} reveals that comparison in terms of the estimators of all the parameters of the model is rather cumbersome. Thus, we confine ourselves to a comparison of the marginal asymptotic distribution of the parameters of, say the i^{th} structural equation.

Extracting the appropriate submatrix from (83) we find

$$\sqrt{T}\,(\hat{\delta}_{\cdot i} - \delta_{\cdot i})_{C2SLSA} \sim N(0, \Phi_{ii}) \tag{91}$$

$$\Phi_{ii} = \sigma_{ii}\,(G_{ii} - \xi^*_{ii}\zeta'_{ii})^{-1} \tag{92}$$

where G_{ii} is the i^{th} block diagonal element of G.

Examining now the iterated instrumental variables estimator of Section 3 we see that the (marginal) asymptotic distribution of $\sqrt{T}\,(\tilde{\delta}_{\cdot i} - \delta_{\cdot i})_{IIV}$ is normal with covariance matrix[10]

$$\Phi^*_{ii} = \underset{T\to\infty}{\text{plim}} \left[\frac{(\bar{Z}_i - \rho_i\, {}_i\bar{Z}_{-1})'\,(\bar{Z}_i - \rho_i\, {}_i\bar{Z}_{-1})}{T} \right]^{-1} \tag{93}$$

Here, however, $\bar{Z}_i = (\bar{Y}_i, {}_i\bar{Y}^*_{-1}, W_i)$ and $\bar{Y}_i$, ${}_i\bar{Y}^*_{-1}$ are obtained from

[10] Strictly speaking this is not the expression given in Section 3; it differs from the former only in excluding from $\bar{Y}_i$, W_i observations corresponding to $t = 2$ and from ${}_i\bar{Y}^*_{-1}$ observations corresponding to $t = 1$. Clearly, in a limiting context this is of no consequence.

the systematic part of the final form.

We observe that G_{ii} is defined by

$$(94) \qquad G_{ii} = \operatorname*{plim}_{T\to\infty} \frac{1}{T} (\bar{Z}_i - \rho_i \, {}_iZ_{-1})' (\bar{Z}_i - \rho_i \, {}_iZ_{-1})$$

where now

$$(95) \qquad \bar{Z}_i = (\bar{Y}_i, {}_iY^*_{-1}, W_i), \qquad {}_iZ_{-1} = (Y_{i,-1}, {}_iY^*_{-2}, W_{i,-1})$$

and the rows of $\bar{Y}_i$ as given in (95) consist of subvectors of

$$(96) \qquad \bar{y}_{t\cdot} = y_{t-1\cdot}\Pi_0 + w_{t\cdot}\Pi_1 + u_{t-1\cdot} R(I - B)^{-1}$$

By contrast, $\bar{Z}_i$ as it appears in (95) differs from the quantity defined in (93); the difference is twofold. First the $\bar{Y}_i$ component of the former consists of subvectors of

$$(97) \qquad \bar{y}'_{t\cdot} = (I - \Pi'_0 L)^{-1} \Pi'_1 w'_{t\cdot}, \qquad t = 1,2,\ldots$$

and second, the Y^*_i component of the former consists of subvectors of

$$(98) \qquad \bar{y}'_{t-1\cdot} = (I - \Pi'_0 L)^{-1} \Pi'_1 w'_{t-1\cdot}$$

In both cases the parameters Π_0, Π_1 have been estimated consistently by instrumental variable or other methods. Consequently

$$(99) \qquad G_{ii} = \Phi^{*-1}_{ii} + S'_i Q \left[\operatorname*{plim}_{T\to\infty} \frac{1}{T} \sum_{t=3}^{T} \begin{bmatrix} u'_{t-1\cdot} \\ v^{*\prime}_{t-1\cdot} \\ v^{*\prime}_{t-2\cdot} \end{bmatrix} [u_{t-1\cdot}, v^*_{t-1\cdot}, v^*_{t-2\cdot}] \right] Q' S_i$$

where

(100)
$$Q = \begin{bmatrix} (I - B')^{-1} R' & (\Pi_0' - \rho_i\ I) & 0 \\ 0 & I & -\rho_i I \\ 0 & 0 & 0 \end{bmatrix}$$

and $v^*_{t-1\cdot}$ is the row vector of final form disturbances for the system as a whole.

We observe (see for example the arguments given in Dhrymes [5]) that the sum whose probability limit it taken in (99) converges to its limit with probability one as well; thus, the probability limit in question may be determined as the expectation of the matrix under the summation sign. Denote this probability limit by N. Comparing with (66) and using the Schwarz inequality for integrals we conclude

(101)
$$S_i'\zeta_i\zeta_i'S_i \leq S_i'\ QNQ'\ S_i\sigma_{ii}$$

in the sense that the difference of the two matrices is negative semi-definite. Thus,

(102)
$$G_{ii} - \xi^*_{ii}\zeta'_{ii} - \Phi^{*-1}_{ii} = S_i'\ QNQ'\ S_i - \frac{1 - \rho_i^2}{\sigma_{ii}} S'\ \zeta_i\ \zeta_i'\ S_i$$

which is positive semidefinite by (101). Consequently

(103)
$$\Phi^*_{ii} - (G_i - \xi^*_{ii}\ \zeta'_{ii})^{-1} \geq 0$$

which shows the relative (marginal) asymptotic efficiency of the C2SLSA estimator relative to the IIV estimator examined earlier.

Finally, it is easy to show that if ρ_i is known then operating with the reduced model

$$(104) \qquad y_{\cdot i} - \rho_i y_{\cdot i,-1} = (Z_i - \rho_i \, {}_iZ_{-1})\delta_{\cdot i} + \varepsilon_{\cdot i}$$

and applying the usual 2SLS procedure the asymptotic distribution of the resulting estimator is given by

$$(105) \qquad \sqrt{T}\,(\hat{\delta}_{\cdot i} - \delta_{\cdot i})_{\rho_i=\rho_{io}} \sim N(0, \sigma_{ii}G_{ii}^{-1})$$

Since

$$(106) \qquad G_{ii} \geq G_{ii} - \xi^*_{ii}\zeta'_{ii} \geq \Phi^{*-1}_{ii}$$

we conclude that the relative (marginal) asymptotic efficiency rankings are as follows: the estimator in (105) is asymptotically efficient relative to the C2SLSA and the latter is asymptotically efficient relative to the IIV estimator.

The substance of our results may be summarized in

THEOREM. Consider the linear structural econometric model exhibited in (46), its error process obeying (2), and subject to assumptions (A.1), (A.2), (A.3), (A.3)' and the usual identifiability conditions. Then

i. The iterated (two step) instrumental variables estimator in (31) is consistent and its asymptotic distribution is given in (36). Moreover, iterating this estimator further will not alter its asymptotic properties.

ii. It the autocorrelation parameters ρ_i, $i = 1,2,\ldots,m$ are known then applying the usual 2SLS procedure to the reduced model in (104) yields an estimator which is consistent and has the asymptotic distribution given in (105)

iii. If an initial consistent estimator of the structural parameters is used to obtain the estimator in (55) then all successive iterates are consistent. Successive iterates have, generally, different asymptotic distributions. The asymptotic distribution of the converging iterate (C2SLSA), - if one exists - is given by (82).

iv. The instrumental variables version of the convergent iterate two stage least squares autoregressive (C2SLSA) estimator and the modified method of scoring version (item 3 and 4) are, asymptotically, equivalent to C2SLSA.

v. Employing the criterion of marginal (relative asymptotic efficiency the estimator in ii. dominates thos in i and iii.; the converging iterate of iii. dominates the estimator in i.

Remark 10. It is conjectured that the $(k+1)^{st}$ iterate of the estimator in ii. is, asymptotically, efficient relative to the k^{th} iterate.

REFERENCES

1. Amemiya, T: "Specification Analysis in the Estimation of Parameters of a Simultaneous Equations Model with Autoregressive Residuals", Econometrica, vol. 34 (1966) pp. 283-306.

2. Amemiya, T. and W. Fuller, "A Comparative Study of Alternative Estimators in a Distributed Lag Model" Econometrica, vol. 35 (1967) pp. 509-529.

3. Dhrymes, P.J.: Econometrics: Statistical Foundation and Applications, Harper and Row, New York, 1970.

4. __________: "Full Information Estimation of Dynamic Simultaneous Equations Models with Autoregressive Errors" Discussion Paper No. 203, University of Pennsylvania, March, 1971

5. __________: "On the Strong Consistency of Estimators for Certain Distributed Lag Models with Autocorrelated Errors", International Economic Review, vol. 12 (1971) pp. 329-343.

6. __________: Distributed Lags: Problems of Estimation and Formulation, Holden-Day, San Francisco, 1971.

7. __________: "A Simplified Structural Estimator for Large-Scale Econometric Models", Australian Journal of Statistics, vol. 13 (1971) pp. 168-175.

8. __________, R. Berner and D. Cummins: "Limited Information Estimation of Simultaneous Equations Models with Lagged Endogenous Variables and Autocorrelated Errors", Discussion Paper No. 183, University of Pennsylvania, October 1970.

9. Fair, R.C.: "The Estimation of Simultaneous Equations Models with Lagged Endogenous Variables and First Order Serially Correlated Errors", Econometrica, vol. 38 (1970), pp. 507-516

10. Hannan, E.J.: "The Estimation of Relationships Involving Distributed Lags", Econometrica, vol. 33 (1965), pp. 409-418.

11. Hoeffding, W. and H. Robbins: "The Central Limit Theorem for Dependent Variables" Duke Mathematical Journal, vol. 15 (1948), pp. 773-780.

12. Mann, H.B. and A. Wald: "On the Statistical Treatment of Linear Stochastic Difference Equations", Econometrica, vol. 11 (1943), pp. 173-220.

13. Morrison, J.L.: "Small Sample Properties of Selected Distributed Lag Estimators: A Monte Carlo Experiment" International Economic Review, vol. 11 (1970), pp. 13-23.

14. Pesaran, M.H.: "The Small Sample Problem of Truncation Remainders in the Estimation of Distributed Lag Models with Auto-Correlated Errors", Mimeographed, undated.

15. Sargan, J.D.: "The Maximum Likelihood Estimation of Economic Relationships with Autoregressive Residuals", Econometrica vol. 29 (1961), pp. 414-426.