8

FULL INFORMATION ESTIMATION OF DYNAMIC SIMULTANEOUS EQUATIONS MODELS WITH AUTOREGRESSIVE ERRORS

by

Phoebus J. Dhrymes

Discussion Paper Number 8
April, 1971

Full Information Estimation of Dynamic Simultaneous Equations Models with Autoregressive Errors*

Phoebus J. Dhrymes

1. Introduction.

In a previous paper [2] a method was proposed for estimating the parameters of dynamic simultaneous equations models by single equation methods. Its objective was to offer a practicable alternative to the single equation estimator proposed by Fair [5] and one whose properties are easily established.

In the present paper we present a full information estimator whose computation requirements are not excessive. We also show its relation to the maximum likelihood estimator; this relationship easily establishes that if we iterate and the iteration converges, then we obtain the maximum likelihood estimator. Earlier discussions of the problem, notably the important initial paper by Sargan [7] have left the impression of enormous complexity and pointed out the problem of multiple maxima. Subsequently, including a recent implementation by Hendry [6], the objective had been to obtain computer algorithms for the solution of a highly nonlinear set of equations. If the earlier papers had, perhaps, overemphasized the degree of complexity of the problem the present paper will probably

*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.

tend in the other direction. It would rather emphasize the simplicity of the problem and its formal similarities with the scalar model

$$y_t = \alpha x_t + \lambda y_{t-1} + u_t, \qquad u_t = \rho u_{t-1} + \varepsilon_t \ .$$

Indeed, the full information estimator to be developed below is the natural extension of the two step (Aitken) estimator of the parameters of the model above whose properties were extensively examined in [3].

When we examine the single equation analog of the estimator developed here we should be led not to the estimator proposed in [2] but rather to a slight modification thereof which is efficient relative to that in [2].

2. Notation, Assumptions, Conventions and Preliminaries.

The standard structural dynamic simultaneous equations model may be written as

$$y_{t\cdot} = y_{t\cdot} B + y_{t-1\cdot} B^* + w_{t\cdot} C + u_{t\cdot} \ , \qquad t = 1,2,\ldots,T \tag{1}$$

where $y_{t\cdot}$ is the <u>row</u> vector of observations on the m current endogeneous variables of the system at time t; similarly $w_{t\cdot}$ is the observation vector on the s exogenous variables and $u_{t\cdot}$ the m-component vector of errors which are assumed to be generated by a first order autoregression.

Remark 1: As the argument proceeds it will become apparent that the approach will readily accommodate additional lags in the endogenous variables and higher order autoregression in the error process.

The following conventions are observed.

C.1. Identities have been substituted out of the system.

C.2. The system is identified by exclusion restrictions, so that some elements of B, B^*, C are known a priori to be zero.

C.3. $T > \max(2m + 2s, K)$ K being the number of unknown elements in B, B^*, C.

We may write the error vector as

$$(2) \qquad u_{t\cdot} = u_{t-1\cdot} R + \varepsilon_{t\cdot} , \quad R = (r_{ij}) \qquad i,j = 1,2,\ldots,m$$

where

(A.1) $\{\varepsilon'_{t\cdot} : t = 0, \pm 1, \pm 2, \ldots\}$ is a sequence of mutually independent identically distributed (i.i.d.) random vectors obeying

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

and Σ is positive definite.

(A.2) The matrix R is stable, i.e., its roots are less than unity in modulus.

(A.3) $I - B$ is nonsingular so that the reduced form is uniquely defined.

(A.4) The matrix $B^*(I - B)^{-1}$ is stable.

(A.5) $\underset{T\to\infty}{\text{plim}} \frac{1}{T} Q'Q$ exists as a nonstochastic, nonsingular matrix, where

$$Q = (W, W_{-1}, Y_{-1}, Y_{-2}), \quad Y = (y_{ti}), \quad W = (w_{tj}), \qquad i = 1,2,\ldots,m, \quad j = 1,2,\ldots,s .$$

(A.6) The exogenous variables are nonstochastic and bounded.

Remark 2: Assumption (A.6) is not required for the derivation of the estimator or for its consistency. It would be convenient, in dealing with asymptotic distribution aspects. However, this problem is not considered in the discussion below.

We now establish the covariance properties of the error process which make quite clear the similarities (and differences) between the vector and scalar autoregressions.

Lemma 1: The error process in (2) is covariance stationary and its second moment matrix is given by

$$\text{(3)} \qquad \text{Cov}(u'_{t\cdot}) = \Omega, \qquad \text{all} \quad t, \qquad \Omega = \sum_{i=0}^{\infty} R'^{i}\, \Sigma\, R^{i} .$$

Proof: Introducing the lag operator L, we may write, in view of (A.2),

$$\text{(4)} \qquad u'_{t\cdot} = \sum_{i=0}^{\infty} R'^{i}\, \varepsilon'_{t-i} .$$

The proof follows immediately because of (A.1).

Lemma 2: The relation between Ω and Σ is given, explicitly, by

$$\Sigma = \Omega - R'\Omega R \ . \tag{5}$$

Proof: From

$$u'_{t\cdot} = R'u'_{t-1\cdot} + \varepsilon'_{t\cdot} \tag{6}$$

we have, in virtue of Lemma 1,

$$\mathrm{Cov}(u'_{t\cdot}) = \Omega = R'\Omega R + \Sigma \ . \tag{7}$$

Lemma 3: If we put

$$u* = (u_{1\cdot}, u_{2\cdot}, \ldots, u_{T\cdot}) \tag{8}$$

then

$$\mathrm{Cov}(u*') = \begin{bmatrix} \Omega & \Omega R & \Omega R^2 & \ldots & \Omega R^{T-1} \\ R'\Omega & \Omega & \Omega R & \ldots & \Omega R^{T-2} \\ R'^2\Omega & R'\Omega & \Omega & \ldots & \Omega R^{T-3} \\ \vdots & & & & \\ R'^{T-1} & R'^{T-2}\Omega & \ldots & \ldots & \Omega \end{bmatrix} = \Phi \ . \tag{9}$$

Proof: Follows immediately from the representation in (4).

Lemma 4: The inverse of Φ is given by

$$(10)\qquad \Phi^{-1} = \begin{bmatrix} \Omega^{-1}+R\Sigma^{-1}R' & -R\Sigma^{-1} & 0 & \cdots & & & 0 \\ -\Sigma^{-1}R' & \Sigma^{-1}+R\Sigma^{-1}R' & -R\Sigma^{-1} & & & & \vdots \\ 0 & -\Sigma^{-1}R' & \Sigma^{-1}+R\Sigma^{-1}R' & -R\Sigma^{-1} & & & 0 \\ \vdots & & & & & & 0 \\ \vdots & \cdots & & & & \Sigma^{-1}+R\Sigma R' & -R\Sigma^{-1} \\ 0 & & & & 0 & -\Sigma^{-1}R' & \Sigma^{-1} \end{bmatrix}$$

Proof: This is easily verified by multiplication.

Lemma 5: There exists an upper triangular matrix ψ such that

$$(11)\qquad \Phi^{-1} = \psi\psi' .$$

Proof: Take

$$(12)\qquad \psi = \begin{bmatrix} \Omega^{-1/2} & -R\Sigma^{-1/2} & \cdots & & 0 \\ 0 & \Sigma^{-1/2} & -R\Sigma^{-1/2} & & \vdots \\ 0 & 0 & \Sigma^{-1/2} & & -R\Sigma^{-1/2} \\ \vdots & & & \ddots & \\ 0 & & & 0 & \Sigma^{-1/2} \end{bmatrix}$$

where $\Omega^{1/2}$, $\Sigma^{1/2}$ are the "square roots" of the positive definite matrices Ω, Σ respectively. Verify by multiplication.

Lemma 6: The determinant of Φ^{-1} is given by

$$|\Phi^{-1}| = |\Omega^{-1}| \; |\Sigma^{-1}|^{T-1} . \tag{13}$$

Proof: The result follows immediately from the decomposition of Lemma 5.

Lemma 7: The log likelihood function of the observations is, neglecting the problem of initial conditions,

$$L(A,R,\Sigma;Y,W) = -\frac{Tm}{2}\ln(2\pi) - \frac{1}{2}\ln|\Omega| - \frac{T-1}{2}\ln|\Sigma| + \frac{T}{2}\ln|(I-B)'(I-B)| - \frac{1}{2}z^*\Phi^{-1}z^{*\prime} \tag{14}$$

where

$$z_{t\cdot} = (y_{t\cdot}, y_{t-1\cdot}, w_{t\cdot}), \quad z^* = (z_{1\cdot}A, z_{2\cdot}A, \ldots, z_{T\cdot}A), \quad A = \begin{pmatrix} I-B \\ -B^* \\ -C \end{pmatrix} . \tag{15}$$

Proof: Immediate from Lemmas 3, 6 and the fact that the Jacobian of the transformation - neglecting the problem of initial conditions - from u^* to $y^* = (y_{1\cdot}, y_{2\cdot}, \ldots, y_{T\cdot})$ is $\{|I-B|^2\}^{T/2}$.

Now, observe that in view of Lemma 5 we have

$$(16)\quad z^*\Phi^{-1}z^{*\prime} = z_{1\cdot}A\Omega^{-1}A'z_{1\cdot}' + \sum_{t=2}^{T} (z_{t\cdot}A - z_{t-1\cdot}AR)\Sigma^{-1}(z_{t\cdot}A - z_{t-1\cdot}AR)'$$

$$= z_{1\cdot}A\Omega^{-1}A'z_{1\cdot}' + \mathrm{tr}(ZA - Z_{-1}AR)\Sigma^{-1}(ZA - Z_{-1}AR)'$$

where

$$(17)\quad Z = (y_{t\cdot}, y_{t-1\cdot}, w_{t\cdot}),\ Z_{-1} = (y_{t-1\cdot}, y_{t-2\cdot}, w_{t-1\cdot}), \quad t = 2,3,\ldots,T .$$

Since upon division by the sample size (T) the terms $\ln|\Omega|$, $z_{1\cdot}A\Omega^{-1}A'z_{1\cdot}'$ vanish, at least in probability, as $T \to \infty$ we shall neglect them in subsequent discussions.

3. Full Information Dynamic Autoregressive (FIDA) Estimators.

In dealing with FIDA estimators we shall essentially give the natural generalization of the three stage least squares (3SLS) procedure and minimize, with respect to A,

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

given prior estimates of Σ and R; in the above Z_{-1} is as defined in (17) and $\tilde{Z}$ will be defined momentarily. If we iterate the procedure, i.e., if given the new estimate of A we reestimate Σ, R and then minimize the trace above with respect to A and the procedure converges, then the resulting estimator will have the same asymptotic distribution as the maximum likelihood estimator. However, the estimates yielded by the two procedures will differ generally, for finite samples. This is perfectly analogous to the result obtained when the structural error vectors are i.i.d. normal variables and we compare FIML and iterated 3SLS procedures. See for example [4].

If we project $y_{t\cdot}$ on the space spanned by $y_{t-1\cdot}, y_{t-2\cdot}, w_{t\cdot}, w_{t-1\cdot}$ we obtain the relation

$$(18) \qquad Y = \tilde{Y} + \tilde{V}, \ \tilde{Y} = Q(Q'Q)^{-1}Q'Y, \ Q = (W, W_{-1}, Y_{-1}, Y_{-2}) \ .$$

This verifies that

$$(19) \qquad \tilde{V}'\tilde{Y} = 0 \ .$$

Define

$$(20) \qquad \tilde{Z} = (\tilde{Y}, Y_{-1}, W)$$

and note that

$$(21) \qquad Z'Z = \tilde{Z}'\tilde{Z} + \begin{bmatrix} \tilde{V}'\tilde{V} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} .$$

Consequently

$$(22) \qquad \operatorname{tr} \Sigma^{-1}(ZA - Z_{-1}AR)'(ZA - Z_{-1}AR) = \operatorname{tr} \Sigma^{-1}(\tilde{Z}A - Z_{-1}AR)'(\tilde{Z} - Z_{-1}AR)$$
$$+ \operatorname{tr} \Sigma^{-1}(I - B)'\tilde{V}'\tilde{V}(I - B) \ .$$

Since

$$(23) \qquad \underset{T\to\infty}{\text{plim}} \ \text{tr}\ \Sigma^{-1}(I - B)' \ \frac{\tilde{V}'\tilde{V}}{T} \ (I - B) = m$$

we see that we are neglecting an asymptotically negligible component when we deal with

$$\text{tr}\ \Sigma^{-1}(\tilde{Z}A - Z_{-1}AR)'(\tilde{Z}A - Z_{-1}AR) .$$

If we now impose the a priori restrictions on the system in (1) we may write the ith equation as

$$(24) \qquad y_{\cdot i} = Y_i\beta_{\cdot i} + Y_i^*\beta_{\cdot i}^* + W_i\gamma_{\cdot i} + u_{\cdot i} \qquad i = 1,2,\ldots,m$$

where Y_i^* is the submatrix of Y_{-1} corresponding to lagged endogenous variables not known to be excluded from the ith equation. It is now convenient to write (1) in one column. Moreover, for reasons that will become apparent below, it will be convenient to write the (column) vector of all the unknown structural parameters as

$$(25) \qquad \delta^* = (\gamma_{\cdot 1}', \gamma_{\cdot 2}', \ldots, \gamma_{\cdot m}', \beta_{\cdot 1}^{*\prime}, \beta_{\cdot 2}^{*\prime}, \ldots, \beta_{\cdot m}^{*\prime}, \beta_{\cdot 1}', \beta_{\cdot 2}', \ldots, \beta_{\cdot m}')'$$

instead of the customary

$$(26) \qquad \delta = (\delta_{\cdot 1}', \delta_{\cdot 2}', \ldots, \delta_{\cdot m}')', \qquad \delta_{\cdot i} = (\beta_{\cdot i}', \beta_{\cdot i}^{*\prime}, \gamma_{\cdot i}')' .$$

To this effect define the matrices

$$(27)\qquad \begin{aligned} W^* &= \mathrm{diag}(W_1,W_2,\ldots,W_m), & Y^{**} &= \mathrm{diag}(Y_1^*,Y_2^*,\ldots,Y_m^*) \\ Y^* &= \mathrm{diag}(Y_1,Y_2,\ldots,Y_m), & \tilde{Y}^* &= \mathrm{diag}(\tilde{Y}_1,\tilde{Y}_2,\ldots,\tilde{Y}_m) \\ \tilde{Z}^* &= (W^*,Y^{**},\tilde{Y}^*), & Z_{-1}^* &= (W_{-1}^*,Y_{-1}^{**},Y_{-1}^*) \end{aligned}$$

where $\tilde{Y}_i$ is the appropriate submatrix of $\tilde{Y}$ in (18) and a -1 subscript indicates that the time subscripts of the corresponding matrix have been reduced by one. A simple rearrangement yields

$$\begin{aligned}(28)\qquad & \mathrm{tr}\,\Sigma^{-1}(\tilde{Z}A - Z_{-1}AR)'\,(\tilde{Z}A - Z_{-1}AR) \\ &= \{\tilde{y} - (R'\otimes I_{T-1})y_{-1} - [\tilde{Z}^* - (R'\otimes I_{T-1})Z_{-1}^*]\delta^*\}'(\Sigma^{-1}\otimes I_{T-1}) \\ &\quad \{\tilde{y} - (R'\otimes I_{T-1})y_{-1} - [\tilde{Z}^* - R\otimes I_{T-1}]\delta^*\} .\end{aligned}$$

Dropping, for convenience, the subscript, $T-1$, of the identity matrix in (28), we obtain the first order conditions as

$$\begin{aligned}(29)\qquad & [\tilde{Z}^{*\prime} - Z_{-1}^{*\prime}(R\otimes I)](\Sigma^{-1}\otimes I)[\tilde{Z}^* - (R'\otimes I)Z_{-1}^*]\delta \\ &= [\tilde{Z}^{*\prime} - Z_{-1}^{*\prime}(R\otimes I)](\Sigma^{-1}\otimes I)[\tilde{y} - (R'\otimes I)y_{-1}] .\end{aligned}$$

In (29)

$$y = (y_{\cdot 1}', y_{\cdot 2}', y_{\cdot 3}', \ldots, y_{\cdot m}')' , \quad \tilde{y} = (\tilde{y}_{\cdot 1}', \tilde{y}_{\cdot 2}', \ldots, \tilde{y}_{\cdot m}')' ;$$

the quantities $y_{\cdot i}$, $\tilde{y}_{\cdot i}$ are, respectively, the ith columns of Y and $\tilde{Y}$.

Suppose now we have initial constent estimates of R and Σ say $\tilde{R}_0$, $\tilde{\Sigma}_0$. We may define the FIDA estimator as

$$(30)\quad \tilde{\delta}^*_{(1)} = \{[\tilde{Z}^{*\prime} - Z^{*\prime}_{-1}(R_0 \otimes I)][\tilde{\Sigma}_0^{-1} \otimes I][\tilde{Z}^* - (\tilde{R}_0' \otimes I)Z^*_{-1}]\}^{-1}$$

$$[\tilde{Z}^{*\prime} - Z^{*\prime}_{-1}(\tilde{R}_0 \otimes I)][\tilde{\Sigma}_0^{-1} \otimes I][y - (\tilde{R}_0' \otimes I)y_{-1}] \ .$$

The substitution of y for $\tilde{y}$ in (30) is justified in view of the orthogonality of $\tilde{V}$ with $\tilde{Y}$, Y_{-1}, Y_{-2}, W, W_{-1}. We may now prove:

Lemma 8: The estimator in (30) is consistent.

Proof: Since we may write

$$(31)\quad y = Z^*\delta^* + u \ , \ y_{-1} = Z^*_{-1}\delta^* + u_{-1}$$

we have

$$(32)\quad y - (\tilde{R}_0' \otimes I)y_{-1} = [Z^* - (\tilde{R}_0' \otimes I)Z^*_{-1}]\delta^* + u - (\tilde{R}_0' \otimes I)u_{-1} \ .$$

Consequently

$$(33)\quad \tilde{\delta}^*_{(1)} = \delta^* + \{[\tilde{Z}^{*\prime} - Z^{*\prime}_{-1}(\tilde{R}_0 \otimes I)](\tilde{\Sigma}_0^{-1} \otimes I)[\tilde{Z}^* - (\tilde{R}_0' \otimes I)Z^*_{-1}]\}^{-1}$$

$$[\tilde{Z}^{*\prime} - Z^{*\prime}_{-1}(\tilde{R}_0 \otimes I)][\tilde{\Sigma}_0^{-1} \otimes I] \cdot [u - (\tilde{R}_0' \otimes I)u_{-1}] \ .$$

It may be verified that the conditions given earlier are sufficient to guarantee the nonsingularity of

$$(34)\quad \operatorname*{plim}_{T\to\infty} \frac{1}{T} [\tilde{Z}^{*\prime} - Z^{*\prime}_{-1}(\tilde{R}_0 \otimes I)](\tilde{\Sigma}_0^{-1} \otimes I)[\tilde{Z}^{*} - (\tilde{R}_0' \otimes I)Z^{*}_{-1}] = \operatorname*{plim}_{T\to\infty} \tilde{M}$$

$\tilde{M}$ being simply a convenient representation of the matrix in the left member. Consequently, we need only establish that

$$(35)\quad \operatorname*{plim}_{T\to\infty} \frac{1}{T} [\tilde{Z}^{*\prime} - Z^{*\prime}_{-1}(\tilde{R}_0 \otimes I)[(\tilde{\Sigma}_0^{-1} \otimes I)[u - (\tilde{R}_0' \otimes I)u_{-1}] = 0 \ .$$

To this effect we observe that

$$(36)\qquad u = (R' \otimes I)u_{-1} + \varepsilon, \qquad\qquad \varepsilon = (\varepsilon'_{\cdot 1}, \varepsilon'_{\cdot 2}, \ldots, \varepsilon'_{\cdot m})'$$

the $\varepsilon_{\cdot i}$ being the vectors of observations on the ith component of ε_t. in (2) and R being the "true" parameter matrix. Thus

$$(37)\qquad u - (\tilde{R}_0' \otimes I)u_{-1} = \varepsilon - [(\tilde{R}_0' - R') \otimes I]u_{-1}$$

and it is apparent that we can show consistency if we can show

$$(38)\quad \operatorname*{plim}_{T\to\infty} \frac{1}{T} [\tilde{Z}^{*\prime} - Z^{*\prime}_{-1}(\tilde{R}_0 \otimes I)](\tilde{\Sigma}_0^{-1} \otimes I)[(\tilde{R}_0' - R') \otimes I]u_{-1} = 0 \ .$$

Let us deal first with the component

$$\frac{1}{T} \tilde{Z}^{*\prime}(\tilde{\Sigma}_0^{-1} \otimes I)[(\tilde{R}_0' - R') \otimes I]u_{-1} = \frac{1}{T} \tilde{Z}^{*\prime}[\tilde{\Sigma}_0^{-1}(\tilde{R}_0' - R') \otimes I]u_{-1} \ .$$

From (26) we see that

$$(39)\quad \frac{1}{T}\tilde{Z}^{*\prime}[\tilde{\Sigma}_0^{-1}(\tilde{R}_0' - R')\otimes I]u_{-1} = \frac{1}{T}\begin{Bmatrix} W^{*\prime}[\tilde{\Sigma}_0^{-1}(\tilde{R}_0' - R')\otimes I]u_{-1} \\ Y^{**\prime}[\tilde{\Sigma}_0^{-1}(\tilde{R}_0' - R')\otimes I]u_{-1} \\ \tilde{Y}^{*\prime}[\tilde{\Sigma}_0^{-1}(\tilde{R}_0' - R')\otimes I]u_{-1} \end{Bmatrix}$$

In view of assumptions (A.1), (A.5) and (A.6) it is trivial that the first component vanishes. The ith subvector of the second and third components consist of

$$\frac{1}{T}\sum_{j=1}^{m} [\tilde{\Sigma}_0^{-1}(\tilde{R}_0' - R')]_{ij}\, Y_i^{*\prime}u_{-1\cdot j},\ \frac{1}{T}\sum_{j=1}^{m} [\tilde{\Sigma}_0^{-1}(\tilde{R}_0' - R')]_{ij}\, \tilde{Y}_i'u_{-1\cdot j}$$

where $u_{-1\cdot j}$ is the jth subvector of u_{-1} and $[\tilde{\Sigma}_0^{-1}(\tilde{R}_0' - R')]_{ij}$ is the (i,j) element of $\tilde{\Sigma}_0^{-1}(\tilde{R}_0' - R')$; we note that

$$(40)\quad \tilde{Y}_i = Q(Q'Q)^{-1}Q'Y_i,\quad \frac{\tilde{Y}_i'u_{-1\cdot j}}{T} = \frac{Y_i'Q}{T}\left(\frac{Q'Q}{T}\right)^{-1}\frac{Q'u_{-1\cdot j}}{T}$$

and it is clear from the definition of Q in (A.5) and assumptions (A.1), (A.2) and (A.4) that

$$\operatorname*{plim}_{T\to\infty} \frac{1}{T}\tilde{Y}_i'u_{-1.j}$$

is a vector with finite nonstochastic elements. Since Y_i^* is a sub-matrix of Y_{-1} the same considerations imply that

$$\operatorname*{plim}_{T\to\infty} \frac{1}{T} Y_i^{*\prime} u_{-1\cdot j}$$

is also a vector with finite nonstochastic elements. In view of the consistencey of $\tilde{R}_0$ as an estimator of R it follows

(41) $$\operatorname*{plim}_{T\to\infty} \frac{1}{T} \tilde{Z}^{*\prime}[\tilde{\Sigma}_0^{-1}(\tilde{R}_0' - R') \otimes I]u_{-1} = 0 .$$

Since Z^*_{-1} contains only appropriate submatrices of Q the argument above also establishes that

(42) $$\operatorname*{plim}_{T\to\infty} \frac{1}{T} Z^{*\prime}_{-1}[\tilde{R}_0 \, \tilde{\Sigma}^{-1}(\tilde{R}_0' - R') \otimes I]u_{-1} = 0 .$$

Consequently, we have

(43) $$\operatorname*{plim}_{T\to\infty} \tilde{\delta}^*_{(1)} = \delta^* .$$

q.e.d.

Corollary 1: The consistency property of $\tilde{\delta}_0$ does not require that $\tilde{\Sigma}_0$ be a consistent estimator of Σ. It only requires that its probability limit be nonsingular and that the probability limit in (34) be similarly nonsingular. It does require, however, that $\tilde{R}_0$ be a consistent estimator of R. Moreover, all subsequent iterates, say $\tilde{\delta}^*_{(k)}$, are consistent as well.

Proof: Obvious from the proof of Lemma 8.

The preceeding suggest

Lemma 9: Provided $\tilde{\Sigma}_0$ is a consistent estimator of Σ, the asymptotic distrubition of $\sqrt{T}\ (\tilde{\delta}_0^* - \delta^*)$ depends on the properties of the asymptotic distribution of the elements of $\sqrt{T}\ (\tilde{R}_0' - R')$.

Proof: We make use of the following result, which may be found, e.g., in [1, Ch. 3]. If ξ_n, ζ_n are two sequences of random variables and if

$$(44) \qquad \underset{n\to\infty}{\text{plim}}\ (\xi_n - \zeta_n) = 0$$

then, provided ζ_n converges in distribution to the random variable ζ, so does ξ_n.

We note

$$(45) \qquad \sqrt{T}\ (\tilde{\delta}^*_{(1)} - \delta^*) = \tilde{M}^{-1}\frac{1}{\sqrt{T}}\ [\tilde{Z}^{*\prime} - Z^{*\prime}_{-1}(\tilde{R}_0 \otimes I)](\tilde{\Sigma}_0^{-1} \otimes I)$$

$$\{\varepsilon - [(\tilde{R}_0' - R') \otimes I]u_{-1}\}\ .$$

Since as $T \to \infty$, $\tilde{M}^{-1}$ converges to a well defined nonstochastic matrix the asymptotic distribution of $\sqrt{T}\ (\tilde{\delta}_0^* - \delta^*)$ (provided it is well defined) is determined by that of the vector in the right member of (45). Let $\tilde{\Sigma}_1$ be any other consistent estimator of Σ. If $\tilde{M}$ is defined with $\tilde{\Sigma}_1$ instead of $\tilde{\Sigma}_0$ its probability limit is not altered. Now let the vector in the right member of (45) be defined with respect to $\tilde{\Sigma}_1$ and consider the difference of the two. We have

$$(46) \quad \operatorname*{plim}_{T\to\infty} \frac{1}{\sqrt{T}} [\tilde{Z}^{*\prime} - Z^{*\prime}_{-1}(\tilde{R}_0 \otimes I)][(\tilde{\Sigma}_0^{-1} - \tilde{\Sigma}_1^{-1}) \otimes I]\{\tilde{R}_0' - R') \otimes I]u_{-1}\} = 0 .$$

The validity of (46) may be seen as follows. Since $\tilde{\Sigma}_0^{-1} - \tilde{\Sigma}_1^{-1}$ is a consistent estimator of the zero matrix, the result would follow immediately if the vector

$$\frac{1}{\sqrt{T}} [\tilde{Z}^{*\prime} - Z^{*\prime}_{-1}(\tilde{R}_0 \otimes I)][D \otimes I]\{\varepsilon - [(\tilde{R}_0' - R') \otimes I]u_{-1}\}$$

converges in distribution with T, for some nonstochastic well defined finite dimensional matrix D. But this must be so if $\sqrt{T}\,(\tilde{\delta}_0^* - \delta^*)$ has a well defined asymptotic distribution - which we assume. This proves the first part of the Lemma. The proof of the second part is immediate from the expression in (38) and the discussion following.

To complete this particular phase of the estimation problem it is necessary to produce an initial estimator of R and Σ. Observing that

$$(47) \qquad U = U_{-1}R + E, \quad U = (u_{ti}), \quad E = (\varepsilon_{ti}), \qquad i = 1,2,\dots,m$$

we see that a natural estimator of R is

$$(48) \qquad \tilde{R} = (U_{-1}'U_{-1})^{-1} U_{-1}'U .$$

Unfortunately, observations on U and U_{-1} are not directly available. On the other hand from each structural equation

$$y_{\cdot i} = Y_i\beta_{\cdot i} + Y_i^*\beta_{\cdot i}^* + W_i\gamma_{\cdot i} + u_{\cdot i}, \qquad i = 1,2,\ldots,m \tag{49}$$

it is possible, say, by instrumental variables (I.V.) methods, to obtain consistent estimators for the columns of B, B^* and C. Let $\tilde{A}_0$ be the matrix of structural parameters so estimated. Consequently, we may take

$$\tilde{U} = Z\tilde{A}_0, \quad \tilde{U}_{-1} = Z_{-1}\tilde{A}_0 \tag{50}$$

and define the initial estimator

$$\tilde{R}_0 = (\tilde{A}_0'Z_{-1}'Z_{-1}\tilde{A}_0)^{-1}\tilde{A}_0'Z_{-1}Z\tilde{A}_0 . \tag{51}$$

From the residuals of the regression we may also define

$$\tilde{\Sigma}_0 = \frac{1}{T}\tilde{E}'\tilde{E}, \qquad \tilde{E} = [I - \tilde{U}_{-1}(\tilde{U}_{-1}'\tilde{U}_{-1})^{-1}\tilde{U}_{-1}']\tilde{U} \tag{52}$$

and thus we complete the initial phase of estimation.

Remark 3: The FIDA estimator described above is a very close analog to the 3SLS procedure[1]. Conceptually, it involves three steps

i. Project Y on W, W_{-1}, Y_{-1}, Y_{-2} and obtain the residual matrix $\tilde{V}$.

ii. By single equation (I.V.) methods obtain consistent estimates of the structural parameters and hence of Σ and R as in (52) and (51) respectively.

[1]Perhaps, for the reasons given below, it should be termed 3SLS dynamic autoregressive, 3SLSDA.

iii. Obtain the FIDA estimator of the structural parameters after "purging" Y of its "stochastic" component, $\tilde{V}$, which is correlated with the error of the equations with which we operate.

The difference between FIDA and 3SLS lies in the second step, where estimation cannot be carried out by least squares methods after substitution in the ith equation of $Y_i - \tilde{V}_i$ for Y_i.

Remark 4: The asymptotic distribution aspects of the FIDA estimator above are, apart from complications induced by simultaneity, entirely analogous to those of the two step Aitken estimator in the context of the scalar model

$$(53) \qquad y_t = \alpha x_t + \lambda y_{t-1} + u_t, \quad u_t = \rho u_{t-1} + \varepsilon_t \ .$$

This asymptotic distribution was studied extensively in [3] and was compared with that of the maximum likelihood estimator. Monte Carlo results (also in [3]) for certain empirically interesting parametric configurations show the two step estimator to perform quite well relative to the maximum likelihood estimator - in fact almost as well.

Comparisons of asymptotic distributions show the efficiency of the two step estimator to depend <u>only</u> on the difference $s_0 - s_1$, where s_0 is the asymptotic variance of the I.V. estimator of λ and s_1 is the asymptotic variance of the maximum likelihood estimator of λ. This, therefore, suggests that if after we obtain the two step estimator of α and λ we iterate the procedure the <u>efficiency of</u>

successive iterates is improved. This is so since if s_{01} is the asymptotic variance of the (initial) two step estimator of λ then

$$s_0 - s_{01} \geq 0 .$$

Indeed if we iterate until convergence is obtained we would obtain a solution to the equations defining the maximum likelihood estimator.

Motivated by these considerations we therefore examine the properties of iterated FIDA estimators.

4. Maximum Likelihood and Iterated FIDA Estimators.

The $(k + 1)^{st}$ iterate of the iterated FIDA estimator obeys the following conditions

$$(54) \quad \tilde{\delta}^*_{(k+1)} = \{[\tilde{Z}^{*\prime} - Z^{*\prime}_{-1}(\tilde{R}_{(k)} \otimes I)][\tilde{\Sigma}^{-1}_{(k)} \otimes I][\tilde{Z}^* - (\tilde{R}'_{(k)} \otimes I)Z^*_{-1}]\}^{-1}$$

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

$$(55) \quad \tilde{R}_{(k)} = (\tilde{A}'_{(k)}Z'_{-1}Z_{-1}\tilde{A}_k)^{-1}\,\tilde{A}'_{(k)}Z'_{-1}Z\tilde{A}_{(k)}, \quad \tilde{\Sigma}_{(k)} = \frac{1}{T}\,\tilde{E}'_{(k)}\tilde{E}_{(k)} .$$

The converging iterate of the process above (CIFIDA) is therefore a fix-point of the equation in (54). We shall not investigate here the conditions under which convergence is obtained. Rather, we shall show that asymptotically the maximum likelihood estimator satisfies an equivalent set of equations and consequently that the asymptotic distributions of CIFIDA and FIML estimators coincide. Unfortunately, the asymptotic distribution of the FIML estimator in this context has not been explicitly established. One suspects, however, as in the scalar cases investigated in [3] that by using an appropriate central limit theorem for m-dependent variables one can establish asymptotic normality with covariance matrix obtained from the expected value of the Hessian of the likelihood function, after we have concentrated the latter with respect to all other parameters except δ^*.

The likelihood function of the observations has been obtained in Lemma 7. Neglecting terms like $-\frac{1}{2}\ln|\Omega| -\frac{1}{2} z_{1.}A\Omega^{-1}A'z_{1.}'$ which are asymptotically insignificant we may write the likelihood function conveniently as[2]

$$
(56) \qquad L(A,R,\Sigma;Y,W) = -\frac{Tm}{2} - \frac{T}{2}\ln\left|\frac{\tilde{V}'\tilde{V}}{T}\right| - \frac{T}{2}\ln|\Sigma| + \frac{T}{2}\ln\left|(I-B)'\frac{\tilde{V}'\tilde{V}}{T}(I-B)\right| - \frac{T}{2}\operatorname{tr}\Sigma^{-1}\left[\frac{(ZA-Z_{-1}AR)'(ZA-Z_{-1}AR)}{T}\right]
$$

The term $\ln\left|\frac{\tilde{V}'\tilde{V}}{T}\right|$ depends only on the data - as is obvious from (18) - and its introduction does not affect the maximization process.

Differentiating with respect to Σ^{-1} we obtain

$$
(57) \qquad \Sigma = \frac{1}{T}(ZA - Z_{-1}AR)'(ZA - Z_{-1}AR) .
$$

Differentiating with respect to R we find

$$
(58) \qquad R = (A'Z_{-1}'Z_{-1}'A)^{-1} A'Z_{-1}'ZA .
$$

[2] If one wished one could be more careful whether one should use T, $T-1$ or $T-2$. Asymptotically, however, such matters are inconsequential.

Now, in view of (27) we can write the trace in (56) as

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

The derivative of this expression with respect to δ^* has already been established, implicitly, in (28) - subject to appropriate modifications.

If we put

$$S = (I - B)' \frac{\tilde{V}'\tilde{V}}{T} (I - B) \tag{59}$$

we have that

$$\frac{\partial \ln |S|}{\partial \beta} = 2[\tilde{V}^{*\prime}(S^{-1} \otimes I)V^*\beta - \tilde{V}^{*\prime}(S^{-1} \otimes I)\tilde{v}], \tag{60}$$

where $\beta = (\beta_{\cdot 1}', \beta_{\cdot 2}', \ldots, \beta_{\cdot m}')'$, $\tilde{V}^* = \mathrm{diag}(\tilde{V}_1, \tilde{V}_2, \ldots, \tilde{V}_m)$, $\tilde{v} = (\tilde{v}_{\cdot 1}', \tilde{v}_{\cdot 2}', \ldots, \tilde{v}_{\cdot m}')'$, $\tilde{v}_{\cdot i}$ are the columns of $\tilde{V}$ and $\tilde{V}_i$ is the residual matrix corresponding to Y_i. The latter is the matrix of observations on the current endogenous variables appearing as explanatory variables in the ith structural equation.

Hence, the first order conditions with respect to δ^* yield

$$(61) \quad \{[Z^{*\prime} - Z^{*\prime}_{-1}(R \otimes I)][\Sigma^{-1} \otimes I][Z^* - (R' \otimes I)Z^*_{-1}] - Z^{**\prime}(S^{-1} \otimes I)Z^{**}\}\delta$$

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

where we have put

$$(62) \qquad Z^{**} = (0,0,\tilde{V}^*)$$

and have noted that

$$(63) \qquad \tilde{V}^{*\prime}(S^{-1} \otimes I)\ \tilde{v} = \tilde{V}^{*\prime}(S^{-1} \otimes I)y\ .$$

In view of the orthogonality of $\tilde{V}^*$ and Z^*_{-1} we may write (54) - omitting the subscripts pertaining to the iterative aspects of the computation - as

$$(64) \quad \{[Z^{*\prime} - Z^{*\prime}_{-1}(R \otimes I)][\Sigma^{-1} \otimes I][Z^* - (R' \otimes I)Z^*_{-1}] - Z^{**\prime}(\Sigma^{-1} \otimes I)Z^{**}\}\delta^*$$

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

A comparison with (61) shows that the FIDA estimators differ from FIML only in the way in which the residual components $\tilde{V}_i$ are treated; FIDA operates with quantities of the form

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

while the FIML estimator operates with the quantities

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

the $\tilde{\sigma}^{ij}$, $\tilde{s}^{ij}$ being computed by using a consistent estimator of δ^*.

Thus, the CIFIDA and FIML estimators differ only in the way in which they "purge" the relevant stochastic component from the moment matrices of the current endogenous variables.

We note, however, that

$$\text{(65)} \qquad \operatorname*{plim}_{T\to\infty} \tilde{S} = \Sigma$$

where

$$\text{(66)} \qquad \tilde{S} = (I - \tilde{B}_{ML})' \, \frac{\tilde{V}'\tilde{V}}{T} \, (I - \tilde{B}_{ML}) \, .$$

Consequently, for large samples CIFIDA and the FIML estimators are identical and hence they would have the same asymptotic distribution. To see this, one has only to compare equations (54) and (55) defining the CIFIDA estimator and equations (57), (58), (61) defining the FIML estimator. Needless to say the last three equations can only be solved by iteration.

Remark 5: In view of the asymptotic equivalence of the CIFIDA and FIML estimators it would seem to be computationally simpler to use the schemes in (54) and (55) rather than the one implied by the FIML estimator in (57), (58) and (61). The latter require recomputation at each iteration of the inessential matrix $\tilde{S}$ which has the same probability limit as $\tilde{\Sigma}$. This aspect entails no asymptotic gain and considerable additional computational burden. Whether there is small sample advantage in the recomputation of $\tilde{S}$ is an open question.

Remark 6: It is possible that the convergence of the iteration in (54) and (55) may be established by an argument similar to that employed by Sargan [7].

The preceding has established

Theorem: Consider the dynamic structural econometric model in (1) and (2) together with conventions C.1-C.3 and assumptions A.1-A.6. Then the FIDA estimator as given in (29) is consistent, as is any subsequent iterate.

Moreover, provided convergence holds, the convergent iterated FIDA estimator obtained as a fix-point of (54) is asymptotically equivalent to the FIML estimator and consequently the two have the same asymptotic distribution.

5. Conclusion.

In this paper we have defined a sequence of estimators for the parameters of a dynamic structural model with autoregressive errors as in (1) and (2). The sequence is defined by an iteration procedure given in (54) and (55).

All iterates are consistent. It is conjectured that the ith iterate is (weakly) efficient relative to the jth one, $j < i$. Moreover, if the iteration converges then we obtain the maximum likelihood estimator, in the sense that asymptotically the convergent iterate is equivalent to the FIML estimator. The difference between the FIML and CIFIDA estimators lies in the way in which they "purge" the second moment matrices of the current endogenous variables of their appropriate "stochastic" component. The FIDA estimator $\tilde{\delta}^*_{(1)}$ (first iterate) is an almost exact analog of the usual 3SLS estimator.

The significance of the preceding discussion is twofold. First, it elucidates the nature of the FIML estimator and second, it provides a relatively convenient computational procedure for obtaining an estimator which is asymptotically equivalent to FIML.

Finally, it is simple to see that the estimator presented in [2] differs from the FIDA estimator - when specialized to the conditions of [2] - essentially in the way in which one treats the lagged endogenous variables. It is then obvious what modifications are needed in [2] in order to obtain the simple equation analog of FIDA.

REFERENCES

1. Dhrymes, P. J.: Econometrics: Statistical Foundations and Applications, New York, Harper and Row, 1970.
2. ___________, R. Berner and D. Cummins: "Limited Information Estimation of Simultaneous Equations Models with Lagged Endogenous Variables and Autoregressive Errors" Discussion Paper No. 183, Department of Economics. University of Pennsylvania, 1970.
3. ___________: Distributed Lags: Problems of Formulation and Estimation, San Francisco, Holden-Day, 1971.
4. ___________: "Small Sample and Asymptotic Relations between Maximum Likelihood and Three Stage Least Squares Estimators" Discussion Paper No. 199 , Department of Economics, University of Pennsylvania, 1971.
5. 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.
6. Hendry, D.E.: "Maximum Likelihood Estimation of Systems of Simultaneous Regression Equations with Errors Generated by a Vector Autoregressive Process" International Economic Review, vol. 12 (1971), pp.
7. Sargan, J. D.: "The Maximum Likelihood Estimation of Economic Relationships with Autoregressive Residuals" Econometrica, vol. 29 (1961), pp. 414-426.
8. ___________: "Wages and Prices in the United Kingdom: A Study in Econometric Methodology" Colston Papers, London, Butterworth's Scientific Publications, 1964.