COWLES COMMISSION DISCUSSION PAPER: STATISTICS: NO. 323

COMPUTATION OF MAXIMUM LIKELIHOOD ESTIMATES OF THE PARAMETERS OF LINEAR STOCHASTIC DIFFERENCE EQUATIONS IN THE CASE OF SERIALLY CORRELATED DISTURBANCES

Herman Chernoff

1. Resume of the Case of Independent Disturbances. Suppose that our model is given by the structural equations

1) $$\alpha_{ux} x_t' = u_t' \qquad t = 1, 2, \ldots, T$$

where y_t = vector of observations on the endogenous variables in year t

z_t = " " " " predetermined " " "

$x_t = (y_t, z_t)$

The subscripts of a matrix indicate its transformation capabilities and simultaneously the vectors and matrices which may be multiplied by it, e.g.

$$u_t B_{ux} C_{xu} A_{ux} x_t'$$

makes sense. It is to be noticed that u_t and y_t are compatible; that is, they have the same number of components, also $A_{ux} = [A_{uy} \quad A_{uz}]$ In the past the case has been treated where u_t is normally distributed with mean 0, and u_t, u_s are independent for $t \neq s$ and $E(u_t u_t') = \Sigma_{uu}$, $t = 1, 2, \ldots, T$. No conditions were imposed on Σ_{uu} but linear homogenous restrictions were imposed on the coefficients of α_{ux}

It was shown that in this case the log of the likelihood function is given, except for an additive constant, by

2) $$L = T \log |\det \alpha_{uy}| - \frac{T}{2} \log \det \Sigma_{uu} - \frac{1}{2} \sum_{t=1}^{T} \{ x_t \alpha_{ux}' \Sigma_{uu}^{-1} \alpha_{ux} x_t' \}$$

It is in turn seen that L is maximized for $\hat{\alpha}_{ux} = A_{ux}$, $\hat{\Sigma}_{uu} = S_{uu}$ where

3) $$S_{uu} = A_{ux} M_{xx} A_{ux}' \qquad M_{xx} = \frac{1}{T} \sum_{t=1}^{T} x_t' x_t$$

and A_{ux} maximizes

4) $$L^* = \log |\det A_{uy}| - \frac{1}{2} \log \det S_{uu}$$

subject to the above mentioned linear homogeneous restrictions. The method of obtaining A_{ux} is computationally tied in with the expansion of

$L^*(A_{ux} + h D_{ux})$

5) $$L^*(A_{ux} + h D_{ux}) = \log|\det A_{uy}| - \frac{1}{2}\log\det S_{uu} + h\,\mathrm{tr}(A_{uy}'^{-1} D_{uy}') - h\,\mathrm{tr}(S_{uu}^{-1} A_{ux} M_{xx} D_{ux}') + -\frac{h^2}{2}\mathrm{tr}\{A_{uy}'^{-1} D_{uy}' A_{uy}'^{-1} D_{uy}'\} - \frac{h^2}{2}\mathrm{tr}\{S_{uu}^{-1} D_{ux} M_{xx} D_{ux}' + - S_{uu}^{-1} D_{ux} M_{xx} A_{ux}' S_{uu}^{-1} A_{ux} M_{xx} D_{ux}' + - S_{uu}^{-1} A_{ux} M_{xx} D_{ux}' S_{uu}^{-1} A_{ux} M_{xx} D_{ux}'\} + O(h^3)$$

The validity of this expansion depends on the fact that $S_{uu} = A_{ux} M_{xx} A_{ux}'$ and in no way on the fact that these are the maximizing values or that there are restrictions on A_{ux}. Computational facility was obtained in calculating A_{ux} by considering $a_p = \mathrm{vec}_p\, A_{ux}$ = vector obtained by laying out the rows of A_{ux}. The restrictions can then be considered as

6) $$a_p \Phi_{rp}' = 0$$

Thus, there exists a matrix $\bar{\Phi}_{qp}$ orthogonal to Φ_{rp}, $\begin{bmatrix} \bar{\Phi}_{qp} \\ \Phi_{rp} \end{bmatrix}$ is square and a_p can be expressed by

7) $$a_p = \bar{a}_q \bar{\Phi}_{qp}$$

where $\bar{a}_q$ are the independent variables involved in A_{ux}. Once $\bar{\Phi}_{qp}$ is given $\bar{a}_q$ is uniquely determined by a_p. For any vector b_p there is a unique partition

8) $$b_p = \bar{b}_q \bar{\Phi}_{qp} + \bar{b}_r \bar{\Phi}_{rp}$$

The following lemmas were very useful.

9) $\operatorname{tr}\{B_{ux} C'_{ux}\} = b_p c'_p$

10) $\operatorname{tr}\{B_{uu} C_{ux} F_{xx} G'_{ux}\} = c_p \{B_{uu}' \otimes F_{xx}\} g'_p$

where $B_{uu} = \|b_{ij}\|$, $B_{uu} \otimes F_{xx} = \begin{bmatrix} b_{11}F_{xx} & b_{12}F_{xx} & \cdots & b_{1K}F_{xx} \\ \vdots & & & \vdots \\ b_{K1}F_{xx} & \cdots & & b_{KK}F_{xx} \end{bmatrix}$

Now the $\bar{a}_q$ are actually dependent in that there is no unique maximum without normalizing on one element in each row. Once this normalization is carried out, and the conditions sufficient for identification are valid, there is a unique maximum. Let us normalize $\bar{a}_q = (\bar{a}_{q'}, \bar{a}_{q^{-1}})$ where $\bar{a}_{q'}$ is fixed. Thus $\Phi_{qp} = \begin{bmatrix} \Phi_{q'p} \\ \Phi_{q^{-1}p} \end{bmatrix}$ where $\Phi_{q'p}$ and $\Phi_{q^{-1}p}$ are orthogonal and

11) $$a_p = \bar{a}_{q'} \bar{\Phi}_{q'p} + \bar{a}_{q^{-1}} \bar{\Phi}_{q^{-1}p}$$

(This method essentially permits us to treat the case where the original linear restrictions on α_{ux} were not necessarily homogeneous.) Thus it is not necessary to consider general D_{ux} in the above expression. It suffices to consider D_{ux} such that

12) $$d_p = \bar{d}_{q^{-1}} \bar{\Phi}_{q^{-1}p}$$

From the above lemmas we have

13) $$\begin{aligned} l^*(A_{ux} + h D_{ux}) = {} & \log|\det A_{uy}| - \tfrac{1}{2} \log \det S_{uu} + \\ & + h\{\operatorname{vec}_p(A'^{-1}_{uy} \, I_{yx}) \bar{\Phi}'_{q^{-1}p} - a_p(S^{-1}_{uu} \otimes M_{xx}) \bar{\Phi}'_{q^{-1}p}\} \bar{d}'_{q^{-1}} \\ & + \frac{h^2}{2}\{\bar{d}_{q^{-1}} L_{q^{-1}q^{-1}} \bar{d}'_{q^{-1}}\} + O(h^3) \end{aligned}$$

where $L_{q^{-1}q^{-1}}$ can be calculated in a straightforward fashion using the lemmas liberally. This is so because $\bar{d}_{q^{-1}}$ terms occur only quadratically in the second order terms. $\frac{1}{T} L^{-1}_{q^{-1}q^{-1}}$ gives the asymptotic covariance matrix of the estimates if it is calculated at A_{ux}.

2. Dependent Disturbances. Suppose that our disturbances are independent. In fact, let us suppose that they satisfy a simple Markoff process.

14) $$u'_t + \beta_{ru} u'_{t-1} = v'_t$$

where v_t and v_s are independent for $t \neq s$ and $E(v_t' v_t) = \Theta_{vv}$, $t = 1, 2, \ldots, T$. (T_{vv} will be used as an estimate of Θ_{vv} and should not be confused with T = size of sample.) With no restrictions on B_{uv} we have

$$15) \quad \alpha_{ux} x_t' + \beta_{vu} \alpha_{ux} x_{t-1}' = v_t'$$
$$\alpha_{uw} w_t' = v_t'$$

where $w_t = (x_t , x_{t-1})$, $\alpha_{uw} = [\alpha_{ux} \quad \beta_{vu}\alpha_{ux}]$

From (2) we have

$$16) \quad L = T \log|\det \alpha_{uy}| - \frac{T}{2} \log \det \Theta_{vv} - \frac{1}{2}\sum_{t=1}^{T} (w_t \alpha_{uw}' \Theta_{vv}^{-1} \alpha_{uw} w_t')$$

which as in (2) must give

$$17) \quad T_{vv} = A_{uw} M_{ww} A_{uw}' , \quad M_{ww} = \frac{1}{T}\sum_{t=1}^{T} w_t' w_t = \begin{bmatrix} M_{xx} & M_{xx_{-1}} \\ M_{x_{-1}x} & M_{x_{-1}x_{-1}} \end{bmatrix}$$

and where

A_{uw} maximizes

$$18) \quad L^* = \log|\det A_{uy}| - \frac{1}{2} \log \det T_{vv}$$
$$= \log|\det A_{uy}| - \frac{1}{2} \log \det [A_{ux} \quad B_{vu}A_{ux}] M_{ww} [A_{ux} \quad B_{vu}A_{ux}]'$$

Maximizing with respect to B_{vu} we see that we must minimize $\log \det T_{vv}$

In Rubin's thesis there is a lemma which establishes that the value of B_{vu} which minimizes $\det T_{vv}$ is

$$19 \quad B_{vu} = - \{A_{ux} M_{xx_{-1}} A_{ux}'\} \{A_{ux} M_{x_{-1}x_{-1}} A_{ux}'\}^{-1}$$

This may also be established by expanding $\log \det T_{vv}$ with respect to B_{vu}.
First let

$$20) \quad \begin{aligned} P_{0,0} &= A_{ux} M_{xx} A'_{ux} \\ P_{0,-1} &= A_{ux} M_{xx_{-1}} A'_{ux} = P'_{1,0} \\ P_{-1,-1} &= A_{ux} M_{x_{-1}x_{-1}} A'_{ux} \end{aligned}$$

$$21) \quad \log \det \left[A_{ux} \quad (B_{vu} + hE_{vu}) A_{ux} \right] M_{wv} \left[A_{ux} \quad (B_{vu} + hE_{vu}) A_{ux} \right]'$$

$$= \log \det \left[T_{vv} + h (E_{vu} P_{-1,0} + P_{0,-1} E'_{vu} + E_{vu} P_{-1,-1} B'_{vu} + B_{vu} P_{-1,-1} E'_{vu}) \right] + O(h^2)$$

$$= \log \det T_{vv} + h \operatorname{tr} \left\{ T_{vv}^{-1} (E_{vu} P_{-1,0} + P_{0,-1} E'_{vu} + E_{vu} P_{-1,-1} B'_{vu} + B_{vu} P_{-1,-1} E'_{vu}) \right\} + O(h^2)$$

$$= \log \det T_{vv} + 2h \operatorname{tr} \left\{ T_{vv}^{-1} (P_{0,-1} E'_{vu} + B_{vu} P_{-1,-1} E'_{vu}) \right\} + O(h^2)$$

by symmetry. For B_{vu} to maximize $\log \det T_{vv}$ the coefficient of E'_{vu} must vanish

$$T_{vv}^{-1} (P_{0,-1} + B_{vu} P_{-1,-1}) = 0$$

$$22) \quad B_{vu} = - P_{0,-1} P_{-1,-1}^{-1}$$

$$T_{vv} = P_{0,0} - P_{0,-1} P_{-1,-1}^{-1} P_{-1,0}$$

Now our problem reduces to maximizing

$$L^* = \log|\det A_{uy}| - \frac{1}{2}\log\det T_{vv}$$

as a function of the "unrestricted" parameters of A_{ux}. Though expansion (5) is applicable we will find it convenient to redo the part with T_{vv}. We shall use the following expansions:

23) $$\log\det(T_{vv}+hE_{vv}) = \log\det T_{vv} + h\,\mathrm{tr}\{T_{vv}^{-1}E'_{vv}\} - \frac{h^2}{2}\mathrm{tr}\{T_{vv}^{-1}E'_{vv}T_{vv}^{-1}E'_{vv}\} + O(h^3)$$

24) $$(C_{uu}+hF_{uu})^{-1} = C_{uu}^{-1} - hC_{uu}^{-1}F_{uu}C_{uu}^{-1} + h^2C_{uu}^{-1}F_{uu}C_{uu}^{-1}F_{uu}C_{uu}^{-1} + O(h^3)$$

and if C_{uu} is symmetric

$$\mathrm{tr}\{C_{uu}F_{uu}\} = \mathrm{tr}\{C_{uu}F'_{uu}\}$$

$$\mathrm{tr}\{C_{uu}F_{uu}C_{uu}F_{uu}\} = \mathrm{tr}\{C_{uu}F'_{uu}C_{uu}F'_{uu}\}$$

$$\mathrm{tr}\{C_{uu}F_{uu}C_{uu}F'_{uu}\} = \mathrm{tr}\{C_{uu}F'_{uu}C_{uu}F_{uu}\}$$

Consider the effect of a change of hD_{yx} on $P_{0,0}$, $P_{0,-1}$, $P_{-1,0}$, $P_{-1,-1}$, T_{vv}

25) $$P_{\cdot y}\{(A_{yx}+hD_{yx})\} = P_{\cdot y} + hQ_{\cdot y} + hQ^*_{\cdot y} + h^2R_{\cdot y} + O(h^3)$$

where

$$Q^*_{\cdot y} = D_{yx}M_{x_\cdot x_y}A'_{yx}$$

$$Q_{\cdot y} = A_{yx}M_{x_\cdot x_y}D'_{yx}$$

$$R_{\cdot y} = D_{yx}M_{x_\cdot x_y}D'_{yx}$$

$$Q^*_{\cdot y} = Q'_{\cdot y}$$

26) $$T_{vv}\{(A_{ux}+hD_{ux})\} = P_{00}\{(A_{ux}+hD_{ux})\} - P_{0-1}\{(A_{ux}+hD_{ux})\}\, P^{-1}_{-1-1}\{(A_{ux}+hD_{ux})\} \cdot P'_{-10}\{(A_{ux}+hD_{ux})\}$$

$$=(_{00}+hQ_{00}+hQ^*_{00}+h^2R_{00}) - \{(P_{0-1}+hQ_{0-1}+hQ^*_{0,-1}+h^2R_{0-1})\cdot$$

$$\cdot(P^{-1}_{-1-1} - P^{-1}_{-1-1}(hQ_{-1-1}+hQ^*_{-1-1}+h^2R_{-1-1})P^{-1}_{-1-1} + P^{-1}_{-1-1}(hQ_{-1-1}+hQ^*_{-1-1})P^{-1}_{-1-1}(hQ_{-1-1}+hQ^*_{-1-1})P^{-1}_{-1-1}$$

$$\cdot(P_{0-1}+hQ_{0-1}+hQ^*_{0-1}+h^2R_{0-1})\} + O(h^3)$$

$$T_{vv}\{(A+hD)\} - T_{vv} = h\{Q_{00}+Q^*_{00} - (Q_{0-1}+Q^*_{0-1})P^{-1}_{-1-1}P'_{0-1} - P_{0-1}P^{-1}_{-1-1}(Q_{0-1}+Q^*_{0-1}) + P_{0-1}P^{-1}_{-1-1}(Q_{-1-1}+Q^*_{-1-1})P^{-1}_{-1-1}P_{-1,0}\}$$

$$+h^2\{R_{00} - R_{0-1}P^{-1}_{-1-1}P'_{0-1} - P_{0-1}P^{-1}_{-1-1}R'_{0-1} + P_{0-1}P^{-1}_{-1-1}R_{-1-1}P^{-1}_{-1-1}P'_{0-1} +$$

$$-(Q_{0-1}+Q^*_{0-1})P^{-1}_{-1-1}(Q_{-1-1}+Q^*_{-1-1})P^{-1}_{-1-1}P'_{0-1} - [(Q_{0-1}+Q^*_{0-1})P^{-1}_{-1-1}(Q_{-1-1}+Q^*_{-1-1})P^{-1}_{-1-1}P'_{0-1}]'$$

$$-P_{0-1}P^{-1}_{-1-1}(Q_{-1-1}+Q^*_{-1-1})P^{-1}_{-1-1}(Q_{-1-1}+Q^*_{-1-1})P^{-1}_{-1-1}P'_{0-1} - (Q_{0-1}+Q^*_{0-1})P^{-1}_{-1-1}(Q_{0-1}+Q^*_{0-1})\}$$

$$+O(h^3)$$

$$= hG_{vv} + h^2H_{vv} + O(h^3)$$

Thus

$$\log\det T_{vv}\{(A+hD\} = \log\det T_{vv} + h\,tr(T^{-1}_{vv}G_{vv}) + h^2 tr\{T^{-1}_{vv}H_{vv}\} + -\frac{h^2}{2}tr\{T^{-1}_{vv}G_{vv}T^{-1}_{vv}G_{vv}\} + O(h^3)$$

$$\therefore L^*(A_{ux}+hD_{ux}) = \log|\det A_{uy}| + h\,tr\{A'^{-1}_{uy}D'_{uy}\} - \frac{h^2}{2}tr\{A'^{-1}_{uy}D'_{uy}A'^{-1}_{uy}D'_{uy}\}$$

$$-\frac{1}{2}\log\det T_{vv} - h\,tr\{T^{-1}_{vv}(Q_{00}-(Q_{0-1}+Q^*_{0-1})P^{-1}_{-1-1}P'_{0-1}+P_{0-1}P^{-1}_{-1-1}Q_{-1-1}P^{-1}_{-1-1}P'_{0-1}$$

$$+\frac{h^2}{2}tr\{T^{-1}_{vv}G_{vv}T^{-1}_{vv}G_{vv}\} - \frac{h^2}{2}tr\{T^{-1}_{vv}H_{vv}\} + O(h^3)$$

$$\operatorname{tr}\{T_{vv}^{-1}G_{vv}T_{vv}^{-1}G_{vv}\} = \operatorname{tr}\{T_{vv}^{-1}\Big(Q_{00} + Q_{00}^{*} - (Q_{0-1} + Q_{0-1}^{*})P_{-1-1}^{-1}P_{0-1}^{-1} - P_{0-1}P_{-1-1}^{-1}(Q_{0-1} + Q_{0-1}^{*})' +$$

$$+ P_{0-1}P_{-1-1}^{-1}(Q_{-1-1} + Q_{-1-1}^{*})P_{-1-1}P_{0-1}'\Big)T_{vv}^{-1}\Big(Q_{00} - (Q_{0-1} + Q_{0-1}^{*})P_{-1-1}^{-1}P_{0-1}' +$$

$$+ P_{0-1}P_{-1-1}^{-1}Q_{-1-1}P_{-1-1}^{-1}P_{0-1}'\Big)\}$$

$$\operatorname{tr}\{T_{vv}^{-1}H_{vv}\} = \operatorname{tr}\{T_{vv}^{-1}\Big(R_{00} - 2R_{0-1}P_{-1-1}^{-1}P_{0-1}' + P_{0-1}P_{-1-1}^{-1}R_{-1-1}P_{-1-1}^{-1}P_{0-1}' - 2P_{0-1}P_{-1-1}^{-1}(Q_{-1-1} + Q_{-1-1}^{*})P_{-1-1}^{-1}Q_{-1-1}P_{-1-1}^{-1}P_{0-1}'$$

$$-2(Q_{0-1} + Q_{0-1}^{*})P_{-1-1}^{-1}(Q_{-1-1} + Q_{-1-1}^{*})P_{-1-1}^{-1}P_{0-1}' - Q_{0-1}P_{-1-1}^{-1}Q_{0-1}' +$$

$$- Q_{0-1}^{*}P_{-1-1}^{-1}Q_{0-1}^{*\prime} - 2Q_{0-1}^{*}P_{-1-1}^{-1}Q_{0-1}'\Big)\}$$

An evaluation of the $L_{q^{*}q^{*}}$ matrix here is considerably more difficult than in the case of serially uncorrelated disturbances for there are 45 terms here to evaluate compared to 4 in the other case. All the terms require about the same amount of work and some savings is obtained in that the work for some terms is used in others. A typical term is of the form

$$\operatorname{tr}\{T_{vv}^{-1}D_{ux}M_{xx}A_{ux}'B_{uu}A_{ux}M_{xx}D_{ux}'\}$$

In the case where Φ_{qk} consists of diagonal blocks;

$$\Phi_{qk} = \begin{bmatrix} \Phi_{q,x} & 0 & 0 \\ 0 & \Phi_{q,x} & 0 \\ 0 & 0 & \ddots \end{bmatrix} \qquad \Phi_{q,x} = \| \Phi_{ik}^{x} \|$$

$$\mathrm{tr}\{\ \} = \sum T^{ij} \bar{d}^j_u \bar{\Phi}^j_{uk} m_{kl} \Phi^n_{vl} \bar{a}^n_v b_{mn} \bar{a}^n_v \Phi^n_{vs} m_{st} \Phi^i_{vt} \bar{d}^i_v$$
$$= \sum d^j_u h^{nj}_u T^{ij} b_{mn} h^{ni}_v d^i_v$$

where

$$\| h^{nj}_u \| = h_{n q_j} = \bar{a}_{q n} V_{q n q_j}$$
$$V_{q_i q_j} = \Phi_{q_i x} M_{xx} \Phi'_{q_j x} = \| V^{ij}_{uv} \|$$

The portion the above term contributes to $L_{q_i q_j} = \| L^{ij}_{uv} \|$ is $\sum_{m,n} h^{ni}_u T^{ij} b_{mn} h^{nj}_v$. It is to be noted that since $d_p = \bar{d}_q \Phi_{qp}$ it is only necessary to consider L_{qq}.

There are a few redeeming features. The linear terms involve only about 2 to 3 times as much labor as in the serially uncorrelated case. Thus a convergence method like the P_h method would be quite convenient. Also a convergence method using, instead of L_{qq} for the gradient, an expression which is asymptotically equivalent to L_{qq} would probably be quite cheap and efficient. In this case L_{qq} need be computed only once.

The questions of identification and consistency of the estimates arise. They should be treated in detail. In the meantime it can be said that if the equations $A_{ux} x'_t = u'_t$ are identified without considering the lagged variables appearing, then the equations $A_{ux} x'_t + B_{uv} A_{ux} x'_{t-1} = v'_t$ are identified and furthermore the estimates are then consistent. This latter statement follows from an argument in which use is made of the fact that the estimates of the equations would be consistent with no restrictions on $A_{ux_{-1}}$ and that putting extra restrictions on A_{uw} does not affect consistency.