ON THE COMPUTATION OF FULL-INFORMATION MAXIMUM LIKELIHOOD ESTIMATES FOR NON-LINEAR EQUATIONS SYSTEMS

Gregory C. Chow

Econometric Research Program
Research Memorandum No. 142
June 1972

The research described in this paper was supported by NSF Grant GS 32003X.

Econometric Research Program
PRINCETON UNIVERSITY
207 Dickinson Hall
Princeton, New Jersey

ON THE COMPUTATION OF FULL-INFORMATION MAXIMUM LIKELIHOOD ESTIMATES FOR NON-LINEAR EQUATION SYSTEMS

Gregory C. Chow*

1. Introduction

In this paper, I will generalize the modified Newton method previously applied in Chow (1968) to the computation of full-information maximum likelihood estimates of parameters of a system of linear structural equations to the case of a system of non-linear structural equations. The success of that method for linear systems[1] has stimulated my present attempt to generalize it for non-linear systems.

The subject of maximum likelihood estimation of non-linear simultaneous equation systems has been studied by Eisenpress and Greenstadt (1966). There are three main differences between their approach and ours. First, their basic formulation is more general, assuming that all parameters in the system may appear in every equation,[2] whereas we assume as the basic setup that there is a distinct set of parameters belonging to each equation. Second, partly because of the first, we are able to obtain simpler and more explicit expressions for the derivatives of likelihood function required in the calculations. Third, and also partly because of the first, we can conveniently deal with linear restrictions on the parameters in the same equation or in different equations, an important problem in econometrics. A fourth feature of this paper, and in fact a feature which has partly motivated it, is the contrast of the linear with the non-linear case. As it will be shown, there are many similarities in the computations of both. This demonstration can enhance our understanding of the nature of the estimation equations. Two additional features of this paper are the treatments of identities in the system and of residuals which may follow an autoregressive scheme.

We will derive in section 2 the estimation equations for non-linear systems, under the assumptions that each structural equation contains a distinct set of parameters, that the parameters are not subject to any linear restrictions, and that the (additive) residuals are serially uncorrelated. Section 3 treats the special case when some equations are linear, and contrasts this case with the non-linear case. Section 4 deals with identities and linear restrictions on the parameters. Section 5 is concerned with the problem of autoregressive residuals.

2. Derivation of the Estimation Equations

Let the t^{th} observation on the g^{th} structural equation be

$$(2.1) \qquad \Phi_g(y_{1t},\ldots,y_{Nt} ; \beta_g) = u_{gt} \qquad \begin{array}{c}(g=1,\ldots,G)\\ (t=1,\ldots,T)\end{array}$$

where y_{it} include G dependent variables followed by predetermined variables, β_g is a row vector of n_g unknown parameters in the g^{th} equation, and u_{gt} are normally distributed with means zero and covariances $Eu_{gt}u_{hs} = \delta_{ts}\,\sigma_{gh}$, δ_{ts} being the kronecker delta.

The logarithm of the concentrated likelihood function, after the unknown σ_{gh} have been eliminated by

(2.2) $$\sigma_{gh} = s_{gh} = \frac{1}{T}\sum_t u_{gt}u_{ht} ,$$

is known to be[3]

(2.3) $$L = \text{const.} - \frac{T}{2}\log|S| + \sum_{t=1}^{T}\log|\tilde{B}_t|$$

where

(2.4) $$S = (s_{gh}) = (\frac{1}{T}\,u_g'u_h)$$

with u_g' denoting the row vector $(u_{g1}\ldots u_{gT})$ whose elements are treated as functions defined by (2.1), and

(2.5) $$\tilde{B}_t = (\tilde{\beta}_{gh,t}) = (\frac{\partial u_{gt}}{\partial y_{ht}}) \qquad (g,h=1,\ldots,G).$$

To maximize the likelihood function, one differentiates (2.3) with respect to the column vector β_i',

(2.6) $$\frac{\partial L}{\partial \beta_i'} = -\frac{T}{2}\sum_{g,h}\frac{\partial \log|S|}{\partial s_{gh}}\cdot\frac{\partial s_{gh}}{\partial \beta_i'} + \sum_t\sum_{g,h}\frac{\partial \log|\tilde{B}_t|}{\partial \tilde{\beta}_{gh,t}}\cdot\frac{\partial \tilde{\beta}_{gh,t}}{\partial \beta_i'} .$$

Note that

$$(2.7) \qquad \frac{\partial \log|S|}{\partial s_{gh}} = s^{hg} \equiv h\text{-}g \text{ element of } S^{-1}$$

$$(2.8) \qquad \frac{\partial s_{gh}}{\partial \beta_i'} = \begin{cases} 0 & \text{for } i \neq g \text{ and } i \neq h \\ \frac{1}{T}\left(\frac{\partial u_{i1}}{\partial \beta_i'} \cdots \frac{\partial u_{iT}}{\partial \beta_i'}\right) u_h \equiv \frac{1}{T} \tilde{Y}_i' u_h & \text{for } i = g \neq h \end{cases}$$

where we have denoted by $\tilde{Y}_i'$ the $n_i \times T$ matrix of the derivatives of the T elements of u_i with respect to the n_i elements of β_i'. Also

$$(2.9) \qquad \frac{\partial \log|\tilde{B}_t|}{\partial \tilde{\beta}_{gh,t}} = \tilde{\beta}_t^{hg} \equiv h\text{-}g \text{ element of } \tilde{B}_t^{-1}$$

$$(2.10) \qquad \frac{\partial \tilde{\beta}_{gh,t}}{\partial \beta_i'} = \begin{cases} 0 & \text{for } i \neq g \\ \frac{\partial^2 u_{it}}{\partial \beta_i' \, \partial y_{ht}} & \text{for } i = g \end{cases} .$$

Using (2.7), (2.8), (2.9), and (2.10), we rewrite (2.6) as

$$(2.11) \qquad \frac{\partial L}{\partial \beta_i'} = - \sum_h s^{hi} \tilde{Y}_i' u_h + \sum_t \sum_h \tilde{\beta}_t^{hi} \cdot \frac{\partial^2 u_{it}}{\partial \beta_i' \, \partial y_{ht}} = 0 \qquad (i=1,\ldots,G).$$

These equations will be solved for the unknowns $\beta_1, \ldots, \beta_G$, to obtain the maximum likelihood estimates.

If we let α denote the vector consisting of all the unknowns $\beta_1, \ldots, \beta_G$, and f denote all the derivatives (2.11) for $i=1,\ldots,G$, then Newton's method for solving

$$(2.12) \qquad f(\alpha) = 0$$

amounts to iterating by the formula

$$(2.13) \qquad \alpha^{r+1} - \alpha^r = - [F(\alpha^r)]^{-1} f(\alpha^r)$$

where α^r is the value of α in the r^{th} iteration, and F is the matrix of partial derivatives of the elements of f with respect to the element of α.

To obtain the elements of F, we differentiate (2.11) with respect to the row vector β_j $(j=1,\ldots,G)$. This will be done separately for the two components of (2.11). For the first component,

$$(2.14) \qquad \frac{\partial}{\partial \beta_j} \left(- \sum_h s^{hi} \tilde{Y}_i' u_h \right)$$

$$= - \sum_h s^{hi} \tilde{Y}_i' \cdot \frac{\partial u_h}{\partial \beta_j} - \sum_h \tilde{Y}_i' u_h \cdot \frac{\partial s^{hi}}{\partial \beta_j} - \sum_h s^{hi} \sum_t u_{ht} \frac{\partial u_{it}}{\partial \beta_i' \partial \beta_j} .$$

The only term in (2.14) that requires further evaluation is the row vector

$$(2.15) \qquad \frac{\partial s^{hi}}{\partial \beta_j} = \sum_{m,n} \frac{\partial s^{hi}}{\partial s_{mn}} \cdot \frac{\partial s_{mn}}{\partial \beta_j}$$

which, on account of (2.8) and of the fact

$$(2.16) \qquad \frac{\partial s^{hi}}{\partial s_{mn}} = - s^{hm} s^{ni} \quad ,$$

can be written as

$$(2.17) \qquad \frac{\partial s^{hi}}{\partial \beta_j} = \frac{1}{T} \sum_n \frac{\partial s^{hi}}{\partial s_{jn}} \cdot u_n' \cdot \tilde{Y}_j + \frac{1}{T} \sum_{m \neq j} \frac{\partial s^{hi}}{\partial s_{mj}} \cdot u_m' \tilde{Y}_j$$

$$= - \frac{1}{T} \sum_n (s^{hj} s^{ni} + s^{hn} s^{ji}) u_n' \tilde{Y}_j \quad .$$

Using (2.17) and (2.8), we rewrite (2.14) finally as

$$(2.18) \qquad \frac{\partial}{\partial \beta_j}\left(- \sum_h s^{hi} \tilde{Y}_i' u_h \right) = - s^{ji} \tilde{Y}_i' \tilde{Y}_j + \frac{1}{T} \tilde{Y}_i' \sum_{h,n} u_h (s^{hj} s^{ni} + s^{hn} s^{ji}) u_n' \tilde{Y}_j$$

$$- \sum_t \frac{\partial^2 u_{it}}{\partial \beta_i' \partial \beta_j} \cdot \sum_h s^{hi} u_{ht} \quad .$$

The differentiation of the second component of (2.11) can be similarly performed.

$$(2.19) \qquad \frac{\partial}{\partial \beta_j} \left(\sum_t \sum_h \tilde{\beta}_t^{hi} \cdot \frac{\partial^2 u_{it}}{\partial \beta_i' \partial y_{ht}} \right)$$

$$= \sum_t \sum_h \left(\tilde{\beta}_t^{hi} \cdot \frac{\partial^3 u_{it}}{\partial \beta_i' \partial \beta_j \, \partial y_{ht}} + \frac{\partial^2 u_{it}}{\partial \beta_i' \partial y_{ht}} \cdot \frac{\partial \tilde{\beta}_t^{hi}}{\partial \beta_j} \right) \quad .$$

Analogous to (2.15), we have, on account of (2.10),

$$(2.20)\qquad \frac{\partial \tilde{\beta}^{hi}_t}{\partial \beta_j} = \sum_{m,n} \frac{\partial \tilde{\beta}^{hi}_t}{\partial \tilde{\beta}_{mn,t}} \cdot \frac{\partial \tilde{\beta}_{mn,t}}{\partial \beta_j} = \sum_n (-\tilde{\beta}^{hj}_t \, \tilde{\beta}^{ni}_t) \frac{\partial^2 u_{jt}}{\partial \beta_j \partial y_{nt}} ,$$

so that (2.19) becomes

$$(2.21)\qquad \frac{\partial}{\partial \beta_j} \left(\sum_t \sum_h \tilde{\beta}^{hi}_t \cdot \frac{\partial^2 u_{it}}{\partial \beta_i' \partial y_{ht}} \right)$$

$$= \sum_t \sum_h \left(\tilde{\beta}^{hi}_t \cdot \frac{\partial^3 u_{it}}{\partial \beta_i' \partial \beta_j \;\partial y_{ht}} - \frac{\partial^2 u_{it}}{\partial \beta_i' \partial y_{ht}} \sum_n [\tilde{\beta}^{hj}_t \; \tilde{\beta}^{ni}_t] \frac{\partial^2 u_{jt}}{\partial \beta_j \; \partial y_{nt}} \right) .$$

To summarize, Newton's method iterates by formula (2.13), where the i^{th} subvector of $f(\alpha)$ is given by (2.11) and where the i-j submatrix of $F(\alpha)$ is given by the sum of (2.18) and (2.21). Note that

$$(2.22)\qquad \frac{\partial^2 u_{it}}{\partial \beta_i' \; \partial \beta_j} = 0 \qquad \text{for } i \neq j \quad .$$

Hence, for $i \neq j$, the third term of (2.18) and the first term of (2.21) are zero.

The computations can be performed as follows. Given the data y_{it} $(i=1,\ldots,G;\ t=1,\ldots,T)$, and z_{it} $(i=1,\ldots,K;\ t=1,\ldots,T)$, and given the values for β_{ig} $(i=1,\ldots,n_g;\ g=1,\ldots,G)$ in the r^{th} iteration,[4] compute

1. u_{gt} as given by (2.1), $g=1,\ldots,G$, $t=1,\ldots,T$;

2. s_{gh} as given by (2.4) and s^{hg} by (2.7), $g, h=1,\ldots,G$;

3. $\tilde{\beta}_{gh,t} = \frac{\partial u_{gt}}{\partial y_{ht}}$ and $\tilde{\beta}^{hg}_{t}$ by (2.9), $g, h=1,\ldots,G$, $t=1,\ldots,T$;

4. $\tilde{y}_{it} = \frac{\partial u_{it}}{\partial \beta_i'}$, the t^{th} column of $\tilde{Y}_i'$, $i=1,\ldots,g$, $t=1,\ldots,T$;

5. $\frac{\partial^2 u_{it}}{\partial \beta_i' \partial y_{ht}}$, a column vector, $i, h=1,\ldots,G$, $t=1,\ldots,T$;

6. $f(\alpha)$ by (2.11) and the preceding results ;

7. $\frac{\partial^2 u_{it}}{\partial \beta_i' \partial \beta_i}$, an $n_i \times n_i$ matrix, $i=1,\ldots,G$, $t=1,\ldots,T$;

8. $\frac{\partial^3 u_{it}}{\partial \beta_i' \partial \beta_i \partial y_{ht}}$, an $n_i \times n_i$ matrix, $i, h=1,\ldots,G$, $t=1,\ldots,T$;

9. $F(\alpha)$ as the sum of (2.18) and (2.21) ;

10. $\alpha^{r+1} - \alpha^r$ by the right-side of (2.13), or a multiple h thereof, where h depends on the value of the likelihood at the new estimate as explained in Chow (1968).

The above calculations can be accomplished by evaluating the first derivatives of steps 3 and 4 , the second derivatives of steps

5 and 7 , and the third derivatives of step 8. Since analytical derivatives can be obtained in the TROLL system,[5] we will use the method in TROLL to evaluate the above derivatives analytically.

3. Estimating Equations When a Structural Equation is Linear

In order to appreciate the nature and the difficulty of estimating the parameters of non-linear equations, it is useful to present the estimating equations when any structural equation is linear. It is also of practical importance to do so, since linear structural equations are often encountered in practice, and one would wish to exploit the linearity to simplify computations. When a system consists of both linear and non-linear equations, I would suggest first setting up the estimating equations as derived from differentiating the likelihood function with respect to the parameters of the linear equations. These are now derived, following closely the development of Section 2.

If the g^{th} structural equation is linear, we can write the T observations on this equation as

$$y_g - Y_g \beta_g' = u_g \tag{3.1}$$

where y_g is a column vector of the T observations on the g^{th} dependent variable, Y_g consists of n_g columns of observations on the other variables appearing in equation g, with the dependent variables listed first, to be followed by the predetermined variables, so that the column vector β'_g of coefficients will consist first of coefficients of the dependent variables, and u_g is as defined in section 2.

Following the derivation of section 2, we will find that equations (2.2), (2.3) and (2.4) remain valid here, but the g^{th} row of equation (2.5) will be reduced to

$$(3.5) \qquad \left(\frac{\partial u_{gt}}{\partial y_{1t}} \cdots \frac{\partial u_{gt}}{\partial y_{Gt}}\right) = (\beta_{g1} \cdots \beta_{gG})$$

where some of the coefficients β_{gj} may be specified to be zero. Concerning the derivations from (2.6) on, one notes the following changes. (2.8) will become

$$(3.8) \qquad \frac{\partial s_{gh}}{\partial \beta'_i} = \begin{cases} 0 & \text{for } i \neq g \text{ and } i \neq h \\ \frac{1}{T} Y'_i u_h & \text{for } i = g \neq h \end{cases}$$

where Y_i consists of n_i columns, each of T observations on the variables (other than y_i) appearing in equation i, as defined for (3.1); (2.10) will become

$$(3.10) \qquad \frac{\partial \tilde{\beta}_{gh,t}}{\partial \beta_{ik}} = \begin{cases} 0 & \text{for } i \neq g \\ 0 & \text{for } i = g;\ k \neq h \\ 1 & \text{for } i = g;\ k = h \end{cases} .$$

The end result is that the g^{th} subvector of $f(\alpha) = 0$ will become a special case of (2.11), namely,

$$(3.11) \qquad \frac{\partial L}{\partial \beta_g'} = - \sum_h s^{hg} Y_g' u_h + \sum_t \tilde{\beta}_t^{(g)g} = 0 ,$$

where $\tilde{\beta}_t^{(g)g}$ is a column vector consisting of elements of the g^{th} column of $\tilde{\beta}_t^{-1}$ which correspond to only the unknown coefficients of the dependent variables in equation (g). Note also that the derivatives (3.10) with respect to the coefficients of the predetermined variables are zero, implying that the second term on the right-hand side of (3.11) vanishes for these coefficients.

Concerning the matrix $F(\alpha)$, we follow the development from equation (2.14) on. Denoting the i-j submatrix of F by F_{ij}, we note that F_{gj} will be changed. It has two components originating from (3.11). The first component, previously given by (2.18), will become, for $i=g$,

$$(3.18) \quad \frac{\partial}{\partial \beta_j}\left(-\sum_h s^{hg} Y_g' u_h\right) = - s^{jg} Y_g' \tilde{Y}_j + \frac{1}{T} Y_g' \sum_{h,n} u_h (s^{hj} s^{ng} + s^{hn} s^{jg}) u_n' \tilde{Y}_j$$

since the derivatives of the elements of Y_g with respect to β_j are zero. The second component, previously given by (2.21), will become, for $i=g$,

$$(3.21) \qquad \frac{\partial}{\partial \beta_j}\left(\sum_t \tilde{\beta}_t^{(g)g}\right) = \sum_t \sum_n \left(-\tilde{\beta}_t^{(g)j} \, \tilde{\beta}_t^{ng}\right) \frac{\partial^2 u_{jt}}{\partial \beta_j \, \partial y_{nt}} ,$$

where $\tilde{\beta}_t^{(g)j}$ is a column vector consisting of those elements in the j^{th} column of $\tilde{B}_t^{-1}$ which correspond to the unknown coefficients of the dependent variables in equation (g) .

If equation (j) is also linear, $\tilde{Y}_j$ in (3.18) will become Y_j and (3.21) will further be reduced to the $n_g \times n_j$ matrix

$$(3.21a) \quad \frac{\partial}{\partial\beta_j}\left(\sum_t \tilde{\beta}_t^{(g)g}\right) = -\sum_t \tilde{\beta}_t^{(g)j} \sum_n \tilde{\beta}_t^{ng} \frac{\partial^2 u_{jt}}{\partial\beta_j \, \partial y_{nt}} = -\sum_t \tilde{\beta}_t^{(g)j} \tilde{\beta}_t^{(j)g'} ,$$

where $\tilde{\beta}_t^{(j)g}$ is a column vector consisting of those elements in the g^{th} column of $\tilde{\beta}_t^{-1}$ which correspond to the unknown coefficients of the dependent variables in equation (j).

To pinpoint the computational simplifications when the g^{th} equation is linear, let us review the ten steps of section 2. Steps 1 and 2 are the same. In step 3, $\tilde{\beta}_{gh,t} = \beta_{gh}$, and in step 4, $\tilde{Y}_g = Y_g$. In step 5, $\partial^2 u_{gt}/\partial\beta_{gk}\partial y_{ht}$ equals 1 for $k=h$, and equals zero otherwise. In step 6, $f(\alpha)$ is computed by (3.11). The matrices in steps 7 and 8 are zero for $i=g$. The g^{th} row of submatrices of $F(\alpha)$ in step 9 is given by the sum of (3.18) and (3.21). Finally, step 10 is the same as before. Thus, the main simplifications result from avoiding the first derivatives of steps 3 and 4, the second derivatives of steps 5 and 7, and the third derivatives of step 8.

4. Treatment of Identities and Linear Restrictions

When an equation is an identity, it contains no unknown parameters and no random residual. Let the first M equations in the system be stochastic, and the remaining $G-M=M'$ equations be identities. We will show that the only modification to the likelihood function (2.3) required in this situation is to interpret the matrix S as an $M\times M$ covariance matrix of the residuals. Therefore, all the computations specified by equations (2.11), (2.18) and (2.21) will remain valid, with the understanding that all terms involving the elements of S and S^{-1} will sum only to M.

To demonstrate the effect of identities on the likelihood function, let the $M\times 1$ vector $u_{\cdot t}$ have the normal density

$$\text{const}\cdot|\Sigma|^{-\frac{1}{2}}\exp\{-\tfrac{1}{2}u'_{\cdot t}\Sigma^{-1}u_{\cdot t}\} \tag{4.1}$$

and let the system consist of two sets of equations

$$\begin{aligned}\Phi_1(y_{1t},y_{2t}) &= u_{\cdot t}\\ \Phi_2(y_{1t},y_{2t}) &= 0\end{aligned} \tag{4.2}$$

where y_{1t} is $M\times 1$, y_{2t} and Φ_2 are both $M'\times 1$, and the identities Φ_2 are known functions. To obtain the density of y_{1t}, one needs the Jacobian

$$(4.3)\qquad \left|\frac{\partial u'_{\cdot t}}{\partial y_{1t}}\right| = \left|\frac{\partial \Phi'_1}{\partial y_{1t}} + \frac{\partial \Phi'_1}{\partial y_{2t}} \cdot \frac{\partial y'_{2t}}{\partial y_{1t}}\right| .$$

From $\Phi_2 = 0$, we have

$$(4.4)\qquad \frac{\partial y'_{2t}}{\partial y_{1t}} = -\left(\frac{\partial \Phi'_2}{\partial y_{2t}}\right)^{-1}\left(\frac{\partial \Phi'_2}{\partial y_{1t}}\right) .$$

Hence, the Jacobian becomes

$$(4.5)\qquad \left| \frac{\partial \Phi'_1}{\partial y_{1t}} - \frac{\partial \Phi'_1}{\partial y_{2t}} \cdot \left(\frac{\partial \Phi'_2}{\partial y_{2t}}\right)^{-1} \cdot \frac{\partial \Phi'_2}{\partial y_{1t}} \right| .$$

If we had ignored the fact that Φ_2 is a set of identities, the Jacobian required would be

$$(4.6)\qquad |\tilde{B}_t| = \begin{vmatrix} \frac{\partial \Phi'_1}{\partial y_{1t}} & \frac{\partial \Phi'_1}{\partial y_{2t}} \\ \frac{\partial \Phi'_2}{\partial y_{1t}} & \frac{\partial \Phi'_2}{\partial y_{2t}} \end{vmatrix} = \left| \frac{\partial \Phi'_1}{\partial y_{1t}} - \frac{\partial \Phi'_1}{\partial y_{2t}} \left(\frac{\partial \Phi'_2}{\partial y_{2t}}\right)^{-1} \frac{\partial \Phi'_2}{\partial y_{1t}} \right| \cdot \left| \frac{\partial \Phi'_2}{\partial y_{2t}} \right| .$$

Since (4.5) and (4.6) differ by a multiplicative constant, the density of y_{1t} will simply be

$$(4.7)\qquad \text{const} \cdot |\Sigma|^{-\frac{1}{2}} \cdot |\tilde{B}_t| \cdot \exp\{-\tfrac{1}{2} u'_{\cdot t} \Sigma^{-1} u_{\cdot t}\}$$

where Σ is $M \times M$, $\tilde{B}_t$ is $G \times G$, and $u_{\cdot t}$ is interpreted as the function ϕ_1 of (4.2). Straight-forward manipulations of (4.7) will yield the likelihood function (2.3), with S interpreted as an $M \times M$ matrix.

If there are linear restrictions on the unknown parameters (these parameters may belong to different structural equations), one has to modify the vector $f(\alpha)$ and the matrix $F(\alpha)$ in equations (2.12) and (2.13).[6] The modifications required can be seen by considering the restriction

$$\alpha_i = c\alpha_j + d\alpha_k \ . \tag{4.8}$$

The unknown α_i will be eliminated, since it is a known linear function of two of the remaining unknowns α_j and α_k. The likelihood function L will be replaced by a new function L^* of a new set of variables α^* (having one fewer element than α), by substituting the right-side of (4.8) for α_i in L. By (4.8) and the chain rule of differentiation, the new $f^*(\alpha^*) = 0$ will contain the following equations

$$\begin{aligned} \frac{\delta L^*}{\delta \alpha_j} &= \frac{\delta L}{\delta \alpha_j} + \frac{\delta L}{\delta \alpha_i} \cdot c = 0 \ , \\ \frac{\delta L^*}{\delta \alpha_k} &= \frac{\delta L}{\delta \alpha_k} + \frac{\delta L}{\delta \alpha_i} \cdot d = 0 \ , \end{aligned} \tag{4.9}$$

where it is understood that the argument α_i of any derivative of L is replaced by the right side of (4.8) -- likewise for equations

(4.10) and (4.11) below. If $f(\alpha)$ has n elements, say, then $f^*(\alpha^*)$ and $f(\alpha)$ are related by the equation

$$f^*(\alpha^*) = M\, f(\alpha)\,, \tag{4.10}$$

where M is an $(n-1)\times n$ matrix which is constructed from the $n\times n$ identity matrix by (1) elminating its i^{th} row, (2) replacing the zero in the i^{th} position of the j^{th} row by c, and (3) replacing the zero in the i^{th} position of the k^{th} row by d.

By differentiating the elements of $f^*(\alpha^*)$ with respect to the remaining $n-1$ variables, one can obtain the new matrix $F^*(\alpha^*)$ of second partial derivatives:

$$F^*(\alpha^*) = M\, F(\alpha)\, M'\,. \tag{4.11}$$

Equations (4.10) and (4.11) can then be used to modify equation (2.13) in order to perform iterations by Newton's method. If there is a second linear restriction, then another matrix, say N, can be used to multiply f^* and F^* in the same way M was used in equations (4.10) and (4.11) to multiply f and F. This process can be repeated for any number of linear restrictions. Setting a coefficient equal to a constant c amounts to setting it equal to c times the dummy variable 1 in the list of predetermined variables; similarly, non-homogeneous linear restrictions can be treated by using this dummy variable.

5. Autoregressive Residuals

If the residuals are assumed to follow an auto-regressive scheme, e.g.,

$$(5.1) \qquad u_{\cdot t} = R_1 u_{\cdot t-1} + R_2 u_{\cdot t-2} + e_{\cdot t}$$

where $e_{\cdot t}$ is normal with zero mean and $E\, e_{\cdot t} e'_{\cdot s} = \delta_{ts} \Sigma$, and if the vector of M stochastic equations is written as

$$(5.2) \qquad \phi_{\cdot t} = \begin{bmatrix} \phi_1(y_{\cdot t}, & \beta_1) \\ \vdots & \\ \phi_M(y_{\cdot t}, & \beta_M) \end{bmatrix} = u_{\cdot t}$$

then

$$(5.3) \qquad \phi_{\cdot t} - R_1 \phi_{\cdot t-1} - R_2 \phi_{\cdot t-2} = e_{\cdot t}$$

will be a system with serially uncorrelated residuals. An obvious method of obtaining maximum likelihood estimates of the parameters $\beta_1, \ldots, \beta_M$, R_1 and R_2 is simply to redefine our structural equations according to (5.3) and apply the procedure previously set forth.

By this method, the g^{th} new stochastic equation will be

$$(5.4) \quad \phi_g(y_{\cdot t}, \beta_g) - \sum_k r_{1,gk}\, \phi_k(y_{\cdot t-1}, \beta_k) - \sum_j r_{2,gj}\, \phi_j(y_{\cdot t-2}, \beta_j) = e_{gt}$$

where $r_{1,gk}$ are the elements of the g^{th} row of R_1 . Comparing

(5.4) with (2.1), one finds two new sets of parameters introduced in the g^{th} equation. The first consists of $r_{1,gk}$ and $r_{2,gj}$. The second consists of the β_k and β_j that are associated with the first in equation (5.4). The computations will be performed by the same ten steps as listed in section 2, with the understanding that e_{gt} will replace u_{gt} and that the parameters of equation (g) are now expanded. The derivatives involving the new parameters are as follows.

Note first that the derivatives with respect to y_{ht} as required in steps 3, 5, and 8 will remain exactly the same as before, since only the leading term of (5.4) involves y_{ht}. In other words, if we let α_g denote a vector of the parameters in equation (g), which is composed of all the $r_{\tau,gk}$ and β_k appearing in the function e_{gt}, the derivatives in steps 3, 5, and 8 will be:

3. $\tilde{\beta}_{gh,t} = \frac{\partial e_{gt}}{\partial y_{ht}} = \frac{\partial u_{gt}}{\partial y_{gt}}$, same as before.

5. $\frac{\partial^2 e_{it}}{\partial \beta_i' \partial y_{ht}} = \frac{\partial^2 u_{it}}{\partial \beta_i' \partial y_{ht}}$, same as before,

second derivatives with respect to other elements of α_i are zero.

8. $\frac{\partial^3 e_{it}}{\partial \beta_i' \partial \beta_i \partial y_{ht}} = \frac{\partial^3 u_{it}}{\partial \beta_i' \partial \beta_i \partial y_{ht}}$, same as before,

third derivatives with respect to other elements of α_i are zero.

In step 4, the new derivatives are

4a. $$\frac{\partial e_{it}}{\partial r_{\tau,ik}} = \Phi_k(y_{\cdot t-\tau}, \beta_k) = u_{k,t-\tau} \qquad \tau=1,2$$

4b. $$\frac{\partial e_{it}}{\partial \beta_k} = \delta_{ik} \frac{\partial u_{i,t}}{\partial \beta_k} - r_{1,ik} \frac{\partial u_{k,t-1}}{\partial \beta_k} - r_{2,ik} \frac{\partial u_{k,t-2}}{\partial \beta_k} \ .$$

If R_1 and R_2 are diagonal, as it may often occur in practice, α_i is composed only of $r_{1,ii}$, $r_{2,ii}$, and β_i. In step 7, the new derivatives are

7a(1). $$\frac{\partial^2 e_{it}}{\partial r_{\tau,ik}\, \partial r_{s,ij}} = 0 \qquad \text{for all } \tau, s, k\ j$$

7a(2). $$\frac{\partial^2 e_{it}}{\partial r_{\tau,ik}\, \partial \beta_j} = \begin{cases} 0 & \text{for } j \neq k \\ \dfrac{\partial u_{k,t-\tau}}{\partial \beta_k} & \text{for } j = k \end{cases}$$

7b. $$\frac{\partial^2 e_{it}}{\partial \beta_k'\, \partial \beta_j} = \begin{cases} 0 & \text{for } j \neq k \\ \delta_{ik} \dfrac{\partial u_{i,t}}{\partial \beta_k'\, \partial \beta_k} - r_{1,ik} \dfrac{\partial u_{k,t-1}}{\partial \beta_k'\, \partial \beta_k} - r_{2,ik} \dfrac{\partial u_{k,t-2}}{\partial \beta_k'\, \partial \beta_k} & \text{for } j=k \ . \end{cases}$$

The derivatives 4a, 4b, 7a and 7b are thus easily obtained from calculations already required for the problem of section 2 when the residuals are assumed to be serially uncorrelated. However, one should not forget that section 2 was based on the assumption of disjoint parameter set in each equation. In-so-far as β_k serves as parameters in several e_{it}, linear restrictions would have to be imposed on these parameters by the procedure of section 4. Of course, these restrictions are not required if R_1 and R_2 are diagonal.

If many elements of R_1 and R_2 are to be estimated, it may be advantageous to employ a two-part iteration method similar to the one suggested by Chow and Fair (1970) for linear systems. By this method, each iteration would consist of two parts. The first is to estimate the β's, taking the r's as given. This can be accomplished by following the same ten steps as just described, ignoring the derivatives in 4a and 7a, and treating the r's as given. The second is to estimate the r's, taking the β's as given. This can be accomplished by taking steps 1 (with e_{gt} replacing u_{gt}, of course), 2, 4a, 6, 7a(1), 9 and 10. The other derivatives in steps 4, 5, 7, and 8 are obviously zero. Although the derivatives $\tilde{\beta}_{gh,t}$ in step 3 are not zero, they will be multiplied by zeros in equation (2.11) for step 6 and in equation (2.21) for step 9. Note that if no restrictions are imposed on the elements of R_1 and R_2, the above can be accomplished by performing ordinary least squares on each equation of (5.3), with ϕ_{gt} as the dependent variable and $\phi_{\cdot t-1}$ and $\phi_{\cdot t-2}$ as explanatory variables. As it is well-known, these least squares estimates are maximum-likelihood estimates for the multivariate regression (5.3) when $\phi_{.t}$, $\phi_{.t-1}$ and $\phi_{.t-2}$ are observable data.[7]

Although the two-part iteration method may provide computational advantages when many elements of R_1 and R_2 are to be estimated, one would still need to find the inverse of $-F(\alpha)$, involving all elements of α as evaluated at the last iteration, to serve as an approximate covariance matrix of the estimates. The same matrix can be used to test linear hypotheses on the parameters, as usual.

REFERENCES

Anderson, T.W., Introduction to Multivariate Statistical Analysis (New York: John Wiley and Sons, Inc., 1958).

Chapman, D.R., and Fair, R.C., "Full-Information Maximum Likelihood Program: USER'S GUIDE," Econometric Research Program Research Memorandum No. 137, Princeton University (April 1972).

Chow, G.C., "Two Methods of Computing Full-Information Maximum Likelihood Estimates in Simultaneous Stochastic Equations," International Economic Review, Vol. 9, No. 1 (February, 1968), pp. 100-112.

Chow, G.C., and Fair, R.C., "Maximum Likelihood Estimation of Linear Equation Systems with Auto-Regressive Residuals," Econometric Research Program Research Memorandum No. 118, Princeton University (November 1970).

Eisenpress, H., and Greenstadt, J., "The Estimation of Non-linear Econometric Systems," Econometrica, Vol. 34, No. 4 (October, 1966), pp. 851-861.

Eisner, M., and Pindyck, R.S., A Generalized Approach to Estimation for the TROLL/1 System, Mimeo, April, 1971.

Maling, W., Preliminary Version of the TROLL/1 Reference Manual, Mimeo, May, 1971.

FOOTNOTES

* This research was carried out while the author was consultant to the NBER Computer Research Center; it was also partly supported by the National Science Foundation through Grant NSF GS 32003X. The author is indebted to Ray C. Fair, Edwin Kuh, and Richard E. Quandt for helpful comments and discussions.

1. This method and the associated method of Chow and Fair (1970) for linear systems with autoregressive residuals have been successfully applied to estimate 82 parameters (including 49 autoregressive coefficients) in a model of 7 stochastic equations constructed by Ray Fair, and to estimate some 48 parameters in a model of 9 stochastic equations with serially uncorrelated residuals constructed by Robert Pindyck of M.I.T.. The latter example involves inverting a 48×48 matrix in each iteration. There is no reason why the estimation of a much larger model, say two or three times larger, could not be attempted. Interested users of the method may consult Chapman and Fair (1972). I am indebted to Ray Fair for supplying the above information.

2. See equation (3-2) of Eisenpress and Greenstadt (1966), p.852.

3. See, for example, equation (3-13) of Eisenpress and Greenstadt (1966), p. 854, and section 4 of the present paper.

4. To obtain an initial estimate to start the iterations, one may apply the method to each equation separately since each equation is a special case of a system once we choose one variable as the dependent variable. If the equation is linear in the parameters, the method of least squares may be used. If not, one can use a set of zeros as the starting value for the iterations on the separate equation. Experience with the program described in Chapman and Fair (1972) has shown that the method often converges for linear systems by using zeros as the initial estimate.

5. See Eisner and Pindyck (1971) and Maling (1971).

6. This treatment of linear restrictions is adopted from Chow and Fair (1970).

7. See, for example, Anderson (1958), p. 181.