Econometric Theory, 4, 1988, 159–170. Printed in the United States of America.

MAXIMUM LIKELIHOOD ESTIMATION OF THE PARAMETERS OF A SYSTEM OF SIMULTANEOUS REGRESSION EQUATIONS

JAMES DURBIN
London School of Economics and Political Science

1. INTRODUCTION

Procedures for computing the full information maximum likelihood (FIML) estimates of the parameters of a system of simultaneous regression equations have been described by Koopmans, Rubin, and Leipnik [5], Chernoff and Divinsky [2], Brown [1], and Eisenpress [4]. However, all of these methods are rather complicated since they are based on estimating equations that are expressed in an inconvenient form. In this paper, a transformation of the maximum likelihood (ML) equations is developed which not only leads to simpler computations but which also simplifies the study of the properties of the estimates. The equations are obtained in a form which is capable of solution by a modified Newton–Raphson iterative procedure. The form obtained also shows up very clearly the relation between the maximum likelihood estimates and those obtained by the three-stage least squares method of Zellner and Theil [9].

2. MAXIMUM LIKELIHOOD EQUATIONS

The problem we shall study is the estimation of the parameters of the model

$$YA + XB = U, \tag{1}$$

where Y is a $T \times p$ matrix of T observations of each of p jointly-dependent variables and X is a $T \times q$ matrix of observations of q predetermined variables. A and B are matrices of regression coefficients and U is a matrix of disturbances. We make the following assumptions:

a. A is nonsingular.
b. Each equation of the system is identified by virtue of the fact that certain elements of A and B are known to be equal to zero. (This assumption is relaxed to permit consideration of underidentified equations in Section 7.) In addition,

Editor's Note: This paper was originally presented at the 1963 European Meetings of the Econometric Society in Copenhagen. Since that time the paper has become well known by word of mouth but few have had the opportunity to read it until now.

 0266-4666/88 $5.00 + .00

we follow the normalizing convention according to which every element in the leading diagonal of A is unity.

c. The elements of X are either fixed constants or are lagged values of the dependent variables. In the latter case, values of the dependent variables occurring prior to the T time periods under study are assumed to be fixed constants. The columns of X are supposed to be linearly independent.

d. The rows of U are independently and normally distributed with vector mean zero and nonsingular variance matrix V. (This assumption is relaxed to permit the inclusion of identities in Section 6.)

Based upon these assumptions, the density is proportional to $|V^{-1}|^{1/2T} \times \exp(-\frac{1}{2}\mathrm{tr}\ U'UV^{-1})$ and the Jacobian of the transformation form U to Y is $|A|^T$; consequently,

$$\log L = \text{constant} + T\log|A| + \frac{1}{2}T\log|V^{-1}| - \frac{1}{2}\mathrm{tr}(YA + XB)'(YA + XB)V^{-1}, \tag{2}$$

where $|\cdot|$ denotes the determinant of a matrix and tr denotes its trace. The maximum likelihood estimates are obtained by equating to zero the derivatives of this with respect to the elements of A, B, and V^{-1}. Let the matrix of derivatives of log L with respect to the elements of a matrix H be denoted by $\partial\log L/\partial H$. We then have

$$\frac{\partial\log L}{\partial A} = TA'^{-1} - Y'(YA + XB)V^{-1}, \tag{3}$$

$$\frac{\partial\log L}{\partial B} = -X'(YA + XB)V^{-1}, \tag{4}$$

$$\frac{\partial\log L}{\partial V^{-1}} = \frac{1}{2}TV - \frac{1}{2}(YA + XB)'(YA + XB). \tag{5}$$

These results are obtained by the application of the following rules, which may be verified directly, for the differentiation of functions of matrices:

$$\frac{\partial\log|\alpha|}{\partial\alpha} = \alpha'^{-1}, \tag{6a}$$

$$\frac{\partial\mathrm{tr}\ \beta\gamma}{\partial\beta} = \gamma', \tag{6b}$$

$$\frac{\partial\mathrm{tr}\ \beta'\delta\beta\epsilon}{\partial\beta} = 2\delta\beta\epsilon, \tag{6c}$$

where α is a square matrix, β and γ are matrices such that $\beta\gamma$ is square, and δ and ϵ are symmetric matrices.

When equating the elements of (3)–(5) to zero, the restrictions on A and

B arising from the fact that some of their elements are known to be unity or zero have to be allowed for. It is only the derivatives with respect to unknown elements which are equated to zero. The estimating equations therefore take the form

$$TA'^{-1} + Y'(YA + XB)V^{-1} = 0, \qquad (*) \tag{7}$$

$$X'(YA + XB)V^{-1} = 0, \qquad (**) \tag{8}$$

$$TV - (YA + XB)'(YA + XB) = 0, \tag{9}$$

where (*) indicates that only elements corresponding to unknown elements of A are taken and (**) that only elements corresponding to unknown elements of B are taken. No restrictions apply to (9) since throughout this paper V is taken to be unrestricted. The "0" on the right-hand sides of (7)–(9) indicates a matrix of zero elements of appropriate order in each case.

The difficulties in solving these equations arise partly from the presence of the term A'^{-1} in (7) and partly from the restrictions in (7) and (8). We shall aim at removing each of these sources of difficulty in turn.

Equation (9) can be rewritten as

$$TV - A'Y'(YA + XB) = B'X'(YA + XB).$$

Premultiplying by A'^{-1} and postmultiplying by V^{-1}, we obtain

$$TA'^{-1} - Y'(YA + XB)V^{-1} = A'^{-1}B'X'(YA + XB)V^{-1}. \tag{10}$$

Since (9) is unrestricted so is (10), i.e., every element on the left-hand side of (10) equals the corresponding element on the right-hand side. Consequently, (7) and (10) together imply

$$A'^{-1}B'X'(YA + XB)V^{-1} = 0. \qquad (*) \tag{11}$$

Like (7) this is a restricted matrix equation the interpretation of which is that elements of the left-hand side of (11) corresponding to unknown elements of A are equated to zero, but that elements corresponding to zero or unit elements of A need not be equal to zero.

The basic proposal of this paper is that (11) should replace (7) in the estimation procedure, i.e., that the maximum likelihood estimates should be obtained from (8), (9), and (11) rather than from (7)–(9). The advantage of doing so will emerge when, in the next section, the equations are expressed in an unrestricted form.

3. THE EQUATIONS IN UNRESTRICTED FORM

Let W denote the matrix $[Y{:}X]$ and $\overline{W}$ the matrix $[\overline{Y}{:}X]$ obtained by replacing Y by the set of reduced-form regressions $\overline{Y} = -XBA^{-1}$, i.e., $\overline{W} =$

$-XBA^{-1}:X]$. Denote the matrix $\begin{bmatrix} A \\ B \end{bmatrix}$ by C. Then (8) and (11) can be written together as

$$\overline{W}'WCV^{-1} = 0, \qquad (***) \tag{12}$$

where (***) indicates that only elements of the left-hand side of (12) corresponding to unknown elements of C are equated to zero.

Let c_j denote the jth column of C $(j = 1, \ldots, p)$. The jth element of c_j is known to equal unity since all elements of the leading diagonal of A are unity; other elements of c_j are zero by virtue of the prior restrictions on A and B. Let m_j be the number of unknown elements in c_j and let the $m_j \times 1$ vector of these unknown elements be denoted by $-\delta_j$. Thus, $-\delta_j$ is simply the vector of unknown coefficients in the jth equation of the original model. Let the columns of W corresponding to unknown elements of c_j be arranged as a $T \times m_j$ matrix Z_j. Finally, let y_j denote the jth column of Y. Thus, in the product Wc_j, y_j has a coefficient of unity and the columns of Z_j have coefficients equal to the elements of $-\delta_j$; all other columns of W have zero coefficients. We therefore have $Wc_j = y_j - Z_j\delta_j$. Consequently,

$$WC = [y_1 - Z_1\delta_1, y_2 - Z_2\delta_2, \ldots, y_p - Z_p\delta_p]. \tag{13}$$

The advantage of this representation, which is the one used by Zellner and Theil [9], is that the restrictions on A and B are allowed for automatically in the sense that all elements of $\delta_1, \ldots, \delta_p$ are unknown coefficients to be estimated.

Substituting into (12), we have

$$\overline{W}' \sum_{k=1}^{p} (y_k - Z_k\delta_k)[v^{k1}, \ldots, v^{kp}] = 0, \qquad (***) \tag{14}$$

where v^{kj} is the kjth element of V^{-1}. To take account of the restrictions (***), we have to pick out elements of the left-hand side of (14) corresponding to unknown elements of C and equate them to zero. The jth column of the left-hand side of (14) is $\overline{W}' \sum_{k=1}^{p} (y_k - Z_k\delta_k)v^{kj}$. We require the elements of this corresponding to the m_j unknown elements of c_j. Let $\bar{Z}_j$ be the $T \times m_j$ matrix formed by the columns of $\overline{W}$ corresponding to unknown elements of c_j, i.e., $\bar{Z}_j$ is obtained from $\overline{W}$ by the same process as Z_j is obtained from W. Thus, $\bar{Z}_j$ is the same as Z_j except that every dependent variable in Z_j is replaced by its reduced-form regression. The m_j elements of $\overline{W}' \sum_{k=1}^{p} (y_k - Z_k\delta_k)v^{kj}$ corresponding to unknown elements of c_j are therefore the elements of the $m_j \times 1$ vector $\bar{Z}_j' \sum_{k=1}^{p} (y_k - Z_k\delta_k)v^{kj}$. Consequently, (14) is equivalent to

$$\bar{Z}_j' \sum_{k=1}^{p} (y_k - Z_k\delta_k)v^{kj} = 0, \qquad (j = 1, \ldots, p). \tag{15}$$

The set of equations (15) is identical with the set specified by (8) and (11). The advantage of the form (15) is that it is an unrestricted set of equations, the restrictions on A and B having been taken care of by the notation employed.

We now write (15) in the more compact form

$$\bar{Z}'G(y - Z\delta) = 0, \tag{16}$$

i.e.,

$$\bar{Z}'GZ\delta = \bar{Z}'Gy, \tag{17}$$

where Z and $\bar{Z}$ are the $pT \times m$ matrices

$$\begin{bmatrix} Z_1 & 0 & \cdots & 0 \\ 0 & Z_2 & \cdots & \vdots \\ \vdots & & & 0 \\ 0 & \cdots & 0 & Z_p \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \bar{Z}_1 & 0 & \cdots & 0 \\ 0 & \bar{Z}_2 & \cdots & \vdots \\ \vdots & & & 0 \\ 0 & \cdots & 0 & \bar{Z}_p \end{bmatrix},$$

$m = \sum_{j=1}^{p} m_j$ being the total number of regression coefficients being estimated; G is the $pT \times pT$ matrix

$$\begin{bmatrix} v^{11}I & v^{12}I & \cdots & v^{1p}I \\ \vdots & \ddots & & \vdots \\ v^{1p}I & \cdots & & v^{pp}I \end{bmatrix},$$

where I is the $T \times T$ unit matrix, i.e., G is the Kronecker product of V^{-1} and I. Finally, δ and y are the $m \times 1$ and $pT \times 1$ vectors $[\delta_1', \ldots, \delta_p']'$ and $[y_1', \ldots, y_p']'$.

From now on we use a circumflex to indicate a maximum likelihood estimate; thus, $\hat{\delta}$ is the maximum likelihood estimate of δ, $\hat{A}$ of A, $\hat{G}$ of G, etc. In order to avoid an over-elaborate notation, however, the maximum likelihood estimate of $\overline{W} = [-XBA^{-1}:X]$ is denoted not by $\hat{\overline{W}}$ but by $\hat{W}$; it is given by $\hat{W} = [-X\hat{B}\hat{A}^{-1}:X]$. Similarly, the estimate of $\bar{Z}$, which is an arrangement of the columns of $\overline{W}$, is denoted by $\hat{Z}$, which is the same arrangement of the columns of $\hat{W}$. In this notation, the maximum likelihood estimate of δ is the solution $\hat{\delta}$ of the equations

$$\hat{Z}'\hat{G}Z\hat{\delta} = \hat{Z}'\hat{G}y, \tag{18}$$

where $\hat{G}$ is obtained from $\hat{V} = T^{-1}(Y\hat{A} + X\hat{B})'(Y\hat{A} + X\hat{B})$, the ijth element of which is $\hat{v}_{ij} = T^{-1}(y_i - Z_i\hat{\delta}_i)'(y_j - Z_j\hat{\delta}_j)$.

The form of (18) shows up very clearly the relation of Zellner and Theil's three-stage least squares estimates [9] to the full information maximum likelihood estimates $\hat{\delta}$. The Zellner–Theil estimates $\tilde{\delta}$ are, in fact, the solution of the equations

$$\tilde{Z}'\tilde{G}Z\tilde{\delta} = \tilde{Z}'\tilde{G}y \tag{19}$$

(Zellner and Theil [9], Eq. 2.16) where $\tilde{G}$ is the same as $\hat{G}$ except that the two-stage least squares estimator of V is used in place of the maximum likelihood estimator $\hat{V}$, and $\tilde{Z}$ is the same as $\hat{Z}$ except that the unrestricted reduced-form regressions $X(X'X)^{-1}X'Y$ are used in place of the regressions $X\hat{B}\hat{A}^{-1}$ based on the maximum likelihood estimators $\hat{A}$ and $\hat{B}$; these, of course, take account of the restrictions on A and B whereas the unrestricted regressions $X(X'X)^{-1}X'Y$ do not. Comparing (18) and (19), we see that the magnitudes of the difference between $\tilde{\delta}$ and $\hat{\delta}$ is of second-order only, the consequence of which is that the three-stage least squares estimates are asymptotically efficient. This has recently been demonstrated independently by a number of different methods (Madansky [6], Rothenberg and Leenders [7], and Sargan [8]).

4. ITERATIVE SOLUTION

We now discuss the solution of the equations (18) by iteration and for simplicity of presentation we shall assume that the moment matrix of predetermined variables $T^{-1}X'X$ converges in probability to a finite positive-definite matrix. In practice, this is a rather mild restriction since, even if at the outset it is violated, a transformation of variables will usually enable it to be satisfied. For example, suppose one column of X is composed of the values $x_t = t$, then $T^{-1}\Sigma_{t=1}^{T} x_t^2 \to \infty$. However, if x_t is replaced by $x_t^* = T^{-1}x_t$ with a consequential change of the coefficient of this variable in each equation in which it occurs, then the model is essentially unaltered but the condition can now be satisfied since $T^{-1}\Sigma_{t=1}^{T} x_t^{*2} \to 1/6$. One might also remark that the iterative method to be described is likely to work well under a much greater range of circumstances than that indicated by the assumption even after transformations of the kind mentioned.

We shall derive the iterative procedure by applying a form of the Newton–Raphson method to the solution of (18). Let $\delta_{(1)}$ be an initial estimate of δ whose elements differ from the corresponding elements of δ by quantities $O(T^{-1/2})$;[1] for example, two-stage least squares could be used to provide such an estimate. Let $\delta_{(2)} = \delta_{(1)} + d\delta$ be the second approximation. Substituting $\delta_{(2)}$ for $\hat{\delta}$ in (18) and taking the first two terms only in the Taylor expansion, we have

$$\bar{Z}'_{(1)} G_{(1)} (Z\delta_{(1)} - y) + \bar{Z}'_{(1)} G_{(1)} Z d\delta + (d\bar{Z}' G_{(1)} + \bar{Z}'_{(1)} dG)(y - Z\delta_{(1)}) = 0,$$

where $\bar{Z}_{(1)}, G_{(1)}$ are the values of $\bar{Z}, G$ calculated by taking $\delta = \delta_{(1)}$, and $d\bar{Z}, dG$ are the increments in $\bar{Z}$ and G due to the change from $\delta_{(1)}$ to $\delta_{(2)}$. Because the elements of $d\bar{Z}' G_{(1)}$ and $\bar{Z}'_{(1)} dG$ are small compared with those of $\bar{Z}'_{(1)} G_{(1)}$ we have, to the first order of approximation,

$$d\delta = (\bar{Z}'_{(1)} G_{(1)} Z)^{-1} \bar{Z}'_{(1)} G_{(1)} (y - Z\delta_{(1)}),$$

i.e.,

$$\delta_{(2)} = (\bar{Z}'_{(1)} G_{(1)} Z)^{-1} \bar{Z}'_{(1)} G_{(1)} y. \tag{20}$$

By repeated application of this result, we obtain the following general formulae for the $(r + 1)$th-round estimates in terms of the rth

$$\delta_{(r+1)} = \delta_{(r)} + (\bar{Z}'_{(r)} G_{(r)} Z)^{-1} \bar{Z}'_{(r)} G_{(r)} (y - Z\delta_{(r)}), \tag{21}$$

i.e.,

$$\delta_{(r+1)} = (\bar{Z}'_{(r)} G_{(r)} Z)^{-1} \bar{Z}'_{(r)} G_{(r)} y, \qquad r = 1,2,\ldots. \tag{22}$$

The simplest way of carrying out the iteration is to continue applying (22) until the difference between successive sets of values is sufficiently small. In practice, however, it is not necessary to compute the matrix $(\bar{Z}'_{(r)} G_{(r)} Z)^{-1}$ at each step of the iteration. One may instead use the same value of this matrix for a number of successive steps, n say, giving the iteration

$$\delta_{(nr+s+1)} = \delta_{(nr+s)} + (\bar{Z}'_{(nr+1)} G_{(nr+1)} Z)^{-1} \bar{Z}'_{(nr+s)} G_{(nr+s)} (y - \delta_{(nr+s)}),$$
$$r = 0,1,2,\ldots, \qquad s = 1,2,\ldots,n. \tag{23}$$

The iterations (22) and (23) are variants of the Newton–Raphson method which previous investigations have found to have good asymptotic convergence properties for the present problem (Koopmans, Rubin, and Leipnik [5], Section 4.35; Chernoff and Divinsky [2], Section 7; and Crockett and Chernoff [3]). In practice, it will generally be advisable to incorporate modifications designed to speed up convergence and to inhibit oscillations. Examples of the type of modification required are given in Section 8. Asymptotically, the rate of convergence is very fast as can be seen from the following heuristic argument which indicates that $\delta_{(r)}$ differs from $\hat{\delta}$ by quantities $O(T^{-r/2})$.

The original model (1) can be written in the form $y = Z\delta + u$, where u is the $pT \times 1$ vector composed of the columns of U placed under one another. Substituting in (18), we have

$$\hat{\delta} = \delta + (\hat{Z}'\hat{G}Z)^{-1} \hat{Z}'\hat{G}u, \tag{24}$$

while substituting in (20), gives

$$\delta_{(2)} = \delta + (\bar{Z}'_{(1)} G_{(1)} Z)^{-1} \bar{Z}'_{(1)} G_{(1)} u.$$

Thus,

$$\delta_{(2)} - \hat{\delta} = H_1 u,$$

where $H_1 = (\bar{Z}'_{(1)} G_{(1)} Z)^{-1} \bar{Z}'_{(1)} G_{(1)} - (\hat{Z}'\hat{G}Z)^{-1}\hat{Z}'\hat{G}$ and is therefore the difference between two matrices the elements of which differ only because those of one are functions of $\delta_{(1)}$ and those of the other are functions of $\hat{\delta}$. Now the elements of both $\delta_{(1)}$ and $\hat{\delta}$ differ from those of δ by quantities $O(T^{-1/2})$ and they therefore differ from each other by quantities of the same order. By the mean-value theorem, each element of H_1 has the form $a'(\delta_{(1)} - \hat{\delta})$ where a is a vector of derivatives of the corresponding element of $(\hat{Z}'\hat{G}Z)^{-1}\hat{Z}'\hat{G}$ with respect to elements of $\hat{\delta}$ and evaluated at a point intermediate between $\delta_{(1)}$ and $\hat{\delta}$. Thus, the ratio of each element of H_1 to the corresponding element of $(\hat{Z}'\hat{G}Z)^{-1}\hat{Z}\hat{G}$ is a quantity $O(T^{-1/2})$. But the elements of $(\hat{Z}'\hat{G}Z)^{-1}\hat{Z}\hat{G}u$ are $O(T^{-1/2})$. The elements of $H_1 u$ are therefore $O(T^{-1})$.

Similarly,

$$\begin{aligned}\delta_{(3)} &= \delta + (\bar{Z}'_{(2)} G_{(2)} Z)^{-1} \bar{Z}'_{(2)} G_{(2)} u,\\ &= \hat{\delta} + H_2 u,\end{aligned}$$

where $H_2 = (\bar{Z}'_{(2)} G_{(2)} Z)^{-1} \bar{Z}'_{(2)} G_{(2)} - (\hat{Z}'\hat{G}Z)^{-1}\hat{Z}\hat{G}$. Because the elements of $\delta_{(2)} - \hat{\delta}$ are $O(T^{-1})$, the ratio of elements of H_2 to corresponding elements of $(\hat{Z}'\hat{G}Z)^{-1}\hat{Z}\hat{G}$ are $O(T^{-1})$. Thus, the elements of $H_2 u$ are $O(T^{-3/2})$. Continuing in this way, it follows that the elements of $\delta_{(r)} - \hat{\delta}$ are $O(T^{-r/2})$ for $r = 1,2,\ldots.$

5. VARIANCE MATRIX OF THE ESTIMATES

The approach of the last section furnishes us with a simple heuristic derivation of the asymptotic variance matrix of the estimates. From (24), we have

$$\begin{aligned}\hat{\delta} - \delta &= (\hat{Z}'\hat{G}Z)^{-1}\hat{Z}'\hat{G}u,\\ &= (\bar{Z}'GZ)^{-1}\bar{Z}'Gu + K_1 u,\end{aligned}$$

where the elements of $K_1 = (\hat{Z}'\hat{G}Z)^{-1}\hat{Z}'\hat{G} - (\bar{Z}'GZ)^{-1}\bar{Z}'G$ are $O(T^{-1/2})$ compared with those of $(\bar{Z}'GZ)^{-1}\bar{Z}'G$. Now

$$\bar{Z}'GZ = \begin{bmatrix} \bar{Z}'_1 & 0 & \cdots & 0 \\ 0 & \bar{Z}'_2 & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & & \cdots & \bar{Z}'_p \end{bmatrix} \begin{bmatrix} v^{11}I & \cdots & v^{1p}I \\ \vdots & \ddots & \vdots \\ v^{1p}I & \cdots & v^{pp}I \end{bmatrix} \begin{bmatrix} Z_1 & 0 & \cdots & 0 \\ 0 & Z_2 & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & & \cdots & Z_p \end{bmatrix},$$

a typical submatrix of which is $v^{ij}\bar{Z}_i'Z_j = v^{ij}\bar{Z}_i'\bar{Z}_j + v^{ij}\bar{Z}_i'(Z_j - \bar{Z}_j)$. Each column of $Z_j - \bar{Z}_j$ consists either of zeros or of independent normal variables with zero means and constant variance. Because $T^{-1}X'X$ converges to a finite positive-definite matrix so does $T^{-1}\bar{Z}_i'\bar{Z}_i$. Consequently, the elements of $\bar{Z}_i'(Z_j - \bar{Z}_j)$ are $O(T^{-1/2})$ compared with those of $\bar{Z}_i'\bar{Z}$. Thus, the elements of $\bar{Z}'GZ$ differ from those of $\bar{Z}'G\bar{Z}$ by terms of relative order $O(T^{-1/2})$. Consequently, $(\bar{Z}'GZ)^{-1}\bar{Z}'G = (\bar{Z}'G\bar{Z})^{-1}\bar{Z}'G + K_2u$ where K_2 is a matrix whose elements are $O(T^{-1/2})$ compared with those of $(\bar{Z}'G\bar{Z})^{-1}\bar{Z}'G$. Thus,

$$\hat{\delta} - \delta = (\bar{Z}'G\bar{Z})^{-1}\bar{Z}'Gu + (K_1 + K_2)u.$$

Taking the leading term only, we have for the variance matrix of $\hat{\delta}$,

$$\begin{aligned} V(\hat{\delta}) &= E(\hat{\delta} - \delta)(\hat{\delta} - \delta)', \\ &= E[(\bar{Z}'G\bar{Z})^{-1}\bar{Z}'Guu'G\bar{Z}(\bar{Z}'G\bar{Z})^{-1}], \end{aligned}$$

i.e.,

$$V(\hat{\delta}) = (\bar{Z}'G\bar{Z})^{-1} \tag{25}$$

to the first order since $E(uu') = G$. This result is the same as that obtained by Rothenberg and Leenders ([7], Formula 4.27) by inverting the matrix of expected values of the second derivatives of the concentrated log-likelihood. It is asymptotically equivalent to the expression for the variance matrix of three-stage estimates obtained by Zellner and Theil [9]. The estimated large-sample variance is $\hat{V}(\hat{\delta}) = (\hat{Z}\hat{G}\hat{Z})^{-1}$.

6. TREATMENT OF IDENTITIES

The results obtained so far have been derived on the assumption that the variance matrix V is nonsingular. However, models constructed for econometric applications frequently contain identities in which all coefficients are known and the disturbances are identically zero. In principle, it is usually a simple matter to substitute for some of the dependent variables from the identities into the remaining equations to give a new model, equivalent to the old one, with fewer equations and with disturbances possessing a nonsingular variance matrix. Nevertheless, as with three-stage least squares it can be shown that this elimination procedure is strictly speaking unnecessary since the estimation methods developed in Sections 2–4 can be applied as they stand to the equations of the original model which possess stochastic disturbances. The only point at which the identities are required is in the estimation of the reduced-form vectors of $\bar{Z}$.

We justify this claim by an argument similar to that used by Rothenberg and Leenders [7]. Suppose that of the p equations in the original model p_1

are stochastic and p_2 are identities. The model can then be written in the partitioned form

$$Y_1A_1 + Y_2A_2 + XB_1 = U, \tag{26}$$

$$Y_1A_3 + Y_2A_4 + XB_2 = 0, \tag{27}$$

where (26) represents the stochastic equations with unknown coefficients and (27) the identities with known coefficients. The rows of U are supposed to be independent normal vectors with zero means and nonsingular variance matrix V as before.

Suppose that we had substituted for Y_2 from (27) into (26). The coefficient matrix of Y_1 in the new equations would be $A_1 - A_2A_4^{-1}A_3$, where we assume A_4 to be nonsingular. The Jacobian of the transformation from U to Y_1 would then be $|A_1 - A_2A_4^{-1}A_3|$. Taking determinants of both sides of the matrix identity

$$\begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix}\begin{bmatrix} I & 0 \\ -A_4^{-1}A_3 & I \end{bmatrix} = \begin{bmatrix} A_1 - A_2A_4^{-1}A_3 & A_2 \\ 0 & A_4 \end{bmatrix}$$

gives $|A_1 - A_2A_4^{-1}A_3| = |A|\,|A_4|^{-1}$. Because $|A_4|$ is known, we therefore have for the log-likelihood

$$\log L = \text{constant} + T\log|A| + \frac{1}{2}T\log|V^{-1}|$$

$$- \frac{1}{2}\operatorname{tr}(Y_1A_1 + Y_2A_2 + XB_1')'(Y_1A_1 + Y_2A_2 + XB_2)V^{-1}. \tag{28}$$

Differentiating with respect to the unknown elements of the matrices A, B, and V^{-1} we have

$$\frac{\partial \log L}{\partial A} = TA'^{-1} - Y'(Y_1A_1 + Y_2A_2 + XB_1)V^{-1}, \qquad (*) \tag{29}$$

$$\frac{\partial \log L}{\partial B} = -X'(Y_1A_1 + Y_2A_2 + XB_1)V^{-1}, \qquad (**) \tag{30}$$

$$\frac{\partial \log L}{\partial V^{-1}} = TV^{-1} - (Y_1A_1 + Y_2A_2 + XB_1)'(Y_1A_1 + Y_2A_2 + XB_2),$$

where (*) and (**) indicate that we are only taking elements corresponding to unknown elements of A and B. Equating the elements of $\partial\log L/\partial V^{-1}$ to zero and augmenting, we have

$$T\begin{bmatrix} V & 0 \\ 0 & 0 \end{bmatrix} = (YA + XB)'(YA + XB),$$

$$= A'Y'(YA + XB) + B'X'(YA + XB).$$

Pre-multiplying by A'^{-1} and post-multiplying by $\begin{bmatrix} V^{-1} & 0 \\ 0 & 0 \end{bmatrix}$, we have

$$T\bar{A}'^{-1} - Y'(Y_1A_1 + Y_2A_2 + XB_1)V^{-1}$$
$$= A'^{-1}B'X'(Y_1A_1 + Y_2A_2 + XB_1)V^{-1},$$

where $\bar{A}'^{-1}$ indicates the first p_1 columns of A'^{-1}. Substituting in (29) gives

$$\frac{\partial \log L}{\partial A} = A'^{-1}B'X'(Y_1A_1 + Y_2A_2 + XB_1)V^{-1}, \qquad (*). \tag{31}$$

The implication of these results is that the basic estimating equations (18) remain valid when applied to the stochastic equations of a system containing identities, on the understanding that $\hat{Z}$ is calculated from the reduced-form values $-X\hat{B}\hat{A}^{-1}$ obtained from the entire system.

7. JUST-IDENTIFIED AND UNDER-IDENTIFIED EQUATIONS

In this section, we consider the special features which arise when the system contains just-identified and under-identified equations. For simplicity let us suppose that the system contains no identities. The complete p-equation model $YA + XB = U$ can then be written in the form

$$Y[A_1 : A_2 : A_3] + X[B_1 : B_2 : B_3] = [u_1 : u_2 : u_3], \tag{32}$$

where $[A_1B_1]$, $[A_2B_2]$, and $[A_3B_3]$ are the coefficient matrices of p_1 over-identified equations, p_2 just-identified equations, and p_3 under-identified equations, respectively. The reduced form can be partitioned correspondingly, i.e.,

$$Y[I_1 : I_2 : I_3] = X[\Gamma_1 : \Gamma_2 : \Gamma_3] + [u_1^* : u_2^* : u_3^*], \tag{33}$$

where $[I_1 : I_2 : I_3]$ denotes the $p \times p$ unit matrix partitioned into p_1, p_2, and p_3 columns while $[\Gamma_1 : \Gamma_2 : \Gamma_3]$ and $[u_1^* : u_2^* : u_3^*]$ are the partitioned forms of the coefficient matrix $\Gamma = -BA^{-1}$ and the matrix of reduced-form disturbances $U^* = UA^{-1}$.

Because the coefficients of the under-identified equations cannot be estimated, and since the coefficients of the just-identified equations are easily recoverable from the reduced form, it is advantageous to replace both sets of equations by their reduced forms and consequently to apply the estimation procedure to the equivalent model

$$Y[A_1 : I_2 : I_3] = X[B_1 : \Gamma_2 : \Gamma_3] + [U_1 : U_2^* : U_3^*] \tag{34}$$

instead of the original model (32). The coefficient matrices A_2 and B_2 of the just-identified equations are then obtained from the complete reduced-form matrix Γ, which is estimated when fitting model (34), by virtue of the relation

$$B_2 = -\Gamma A_2. \tag{35}$$

The advantage of following this procedure, instead of fitting the model in which the just-identified equations occur in their original form, is that the more reduced-form equations there are in the model fitted by the iterative procedure described in Section 4, the simpler are the calculations.

NOTES

1. We say that a quantity $g(T)$ is $O(T^{-s})$ if $T^s g(T)$ is either nonstochastic and bounded or is a random variable with bounded mean square.

REFERENCES

1. Brown, T.M. Simplified full information maximum likelihood and comparative structural estimates. *Econometrica* 27 (1959): 638-653.
2. Chernoff, J. & N. Divinsky. The computation of maximum likelihood estimates of linear structural equations. In W.C. Hood & T.C. Koopmans (eds.), *Studies in Econometric Method*, Chapter 10, New York: Wiley, 1953.
3. Crockett, J.B. & H. Chernoff. Gradient methods of maximization. *Pacific Journal of Mathematics* 5 (1955).
4. Eisenpress, H. Note on the computation of full information maximum likelihood estimates of coefficients of a simultaneous system. *Econometrica* 30 (1962): 343-348.
5. Koopmans, T.C., H. Rubin, & R.B. Leipnik. Measuring the equation systems of dynamic economics. In T.C. Koopmans (ed.), *Statistical Inference in Dynamic Economic Models*, Chapter 2, New York: Wiley, 1950.
6. Madansky, A. On the efficiency of three-stage least squares estimation. RAND Corporation Memorandum RM-3557-PR, March (1963).
7. Rothenberg, T. & T.C. Leenders. Efficient estimation of simultaneous equation systems. *Econometrica*, 32 (1964), 57-76.
8. Sargan, J.D. Three-stage least squares and full maximum likelihood estimates. *Econometrica*, 32 (1964), 77-81..
9. Zellner, A. & H. Theil. Three-stage least squares: Simultaneous estimation of simultaneous equations. *Econometrica* 30 (1962): 54-78.