A list of Working Papers on
the last pages

No. 132, 1984

A GENERAL FIML ESTIMATOR FOR A CERTAIN CLASS OF MODELS THAT ARE NONLINEAR IN THE VARIABLES

by
Erik Mellander

Paper presented at the European Meeting of the Econometric Society in Pisa, Fall 1983.

September, 1984

ABSTRACT

Most econometric multi-equation models estimated are assumed to be linear in both the variables and the parameters. One reason is that, in general, methods of linear algebra cannot be applied to nonlinear systems.

In this paper a certain class of nonlinear models is defined, however, the members of which can be formulated in matrix terms. Particular interest is focused upon nonlinearities in the variables.

An algorithm for full information maximum likelihood (FIML) is described, including the linear model as a special case. Neither the likelihood function presented, nor its first order derivates are overly complicated relative to the usual (linear) FIML case. The latter makes the suggested approach particularly attractive compared to "derivative-free" methods when dealing with systems containing many parameters. It is also shown how the efficiency in the actual computations can be greatly increased by exploiting certain properties of the involved matrices.

1 Introduction

Econometric models are usually assumed to be linear in both the variables and the parameters. One of the main reasons is that methods of linear algebra mostly cannot be applied to nonlinear systems, making it impossible to express them in a way which is both general and explicit.

Sometimes nonlinear models may be treated as if they are linear, however. This paper provides an example of such a case. It shows how the FIML estimator by means of linear algebra can be extended to a certain class of models that are nonlinear in the variables. These models can be formalized according to

$$By_t *_1 (z_t'p) + Cz_t *_2 (z_t'q) = u_t \quad t = 1,\dots,T \qquad (1.1)$$

where y_t and z_t are m and n component vectors of observations at time t on the endogeneous and predetermined variables respectively. The matrices B and C and the vectors p and q contain the unknown coefficients. The sign "*" may denote either multiplication or division. For the sake of convenience it is assumed that

$$u_t \sim NID(0,\Sigma) \text{ for all } t. \qquad (1.2)$$

In the simple case where the scalars $(z_t'p)$ and $(z_t'q)$ are equal to one for all t,(1.1) represents the usual system of simultaneous equations. It should be noted that according to (1.1) the same scalar(s) should enter all the equations. This requirement considerably facilitates the calculations but at the same time restricts the applicability of the algorithm.

An example of the type of systems to which the proposed method may be applied is the system of expenditure shares corresponding to the translog indirect utility function (see Christensen, Jorgenson and Lau (1975))

$$w_{it} = \frac{\alpha_i + \sum_{j=1}^{n} \gamma_{ij} \log(P_{jt}/M_t)}{-1 + \sum_{i=1}^{n} \sum_{j=1}^{n} \gamma_{ij} \log(P_{jt}/M_t)} - u_{it} \qquad (1.3)$$

$i=1,\ldots,n.$

where w_{it} is the budget share of the i:th commodity at time t, M_t is total expenditure and the parameters are restricted according to

$$\sum_{i=1}^{n} \alpha_i = -1.0, \quad \gamma_{ij} = \gamma_{ji}.$$

Comparing (1.3) with (1.1) the former is seen to correspond to the special case $B=-I$, $(z_t'p) = 1$ for all t and $*_2$ denoting division.[1]

By regarding the elements of B,C,p and q as functions of a set of unrestricted parameters, i.e.

$$B = B(\theta), \; C = C(\theta), \; p = p(\theta), \; q = q(\theta) \qquad (1.4)$$

where

$$\theta' = (\theta_1,\ldots,\theta_\ell). \qquad (1.5)$$

nonlinear constraints on the parameters can also be incorporated in (1.1). This approach has been used by e.g. Jansson and Mellander (1983) and will also be followed here.

The likelihood function corresponding to (1.1) is given in section 4.1. In the section following its gradient vector is derived. Thus, it is possible to maximize the likelihood function by means of e.g. some quasi-Newton method[2]. It is shown that neither the likelihood function nor the derivatives become overly complicated, compared to the usual FIML case. However, the rather large matrices involved pose a storage problem, especially concerning the calculation of the gradient. A solution to that problem is given in section 4.3 where it is shown how considerable gains can be made in computational efficiency.

For easy reference in section 4, brief reviews of some matrix operations and of matrix differential calculus are presented in the next two sections.

2 Matrix operators and operations

As the reader is assumed to be familiar with e.g. the properties of Kronecker products, the presentation here is very brief. No proofs are given. More extensive treatments can be found in Pollock (1979, ch. 4) and Magnus and Neudecker (1979).

A (m,n) matrix X may be regarded as an array of column vectors $x_{.j}$, j=1,...,n or as an array of row vectors $x_{i.}$, i=1,...,m. Thus X may be vectorized either by stacking its columns on top of each other

$$X^c = \begin{matrix} x_{.1} \\ \vdots \\ x_{.n} \end{matrix} \qquad (2.1)$$

or by putting its rows after one another

$$X^r = (x_{1.}, \ldots, x_{m.}). \quad (2.2)$$

In particular, if y is a column vector then
$y^c = y'^c = y$ and $y^r = y'^r = y'$.[3]

The two operators are related by the identity

$$X^r = X'^{c\prime}. \quad (2.3)$$

Let $A=(a_{ij})$ be an (r,m) and B an (s,n) matrix. The Kronecker product is then defined as the (rs,mn) matrix

$$A \otimes B = (a_{ij}B). \quad (2.4)$$

Some of the properties of Kronecker products are (the matrices involved are assumed to be conformable)

$$(A \otimes B)(C \otimes D) = AC \otimes BD \quad (2.5)$$

$$(A \otimes B)' = A' \otimes B' \quad (2.6)$$

$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}. \quad (2.7)$$

The following two relations give the connection between the operators "c" and "r" and the Kronecker product

$$(AXB')^c = (B \otimes A)X^c \quad (2.8)$$

$$(AXB')^r = X^r(A' \otimes B'). \quad (2.9)$$

The commutation matrix K_{mn} is the (mn,mn) matrix[4] which transforms the vectorized (m,n) matrix X according to

$$K_{mn}X^c = X'^c \tag{2.10}$$

$$X^r K_{mn} = X'^r \ . \tag{2.11}$$

Following Magnus and Neudecker the (m^2,m^2) matrix K_{mm} will be written K_m for notational brevity.

Further, if A is (r,m) and B is (s,n) then

$$K_{rs}(B \otimes A)X^c = (A \otimes B)K_{mn}X^c. \tag{2.12}$$

The last rule can be established by means of (2.8) and (2.10).

3 Matrix differentiation

Following Pollock, matrix differentiation is treated within the framework of vector differentiation. No proofs are given in the text. Two are however indicated in notes.

According to the basic definition, the partial derivative of the (m,1) vector y = y(x) with respect to the (n,1) vector x is the (m,n) matrix[5]

$$\frac{\partial y}{\partial x} = \frac{\partial y_i}{\partial x_j} \ . \tag{3.1}$$

A standard example of (3.1) is

$$\frac{\partial Ax}{\partial x} = A. \tag{3.2}$$

Special cases of (3.2) are the derivative of a scalar with respect to a (column) vector

$$\frac{\partial y_i}{\partial x} = \left[\frac{\partial y_i}{\partial x_1},\dots,\frac{\partial y_i}{\partial x_n}\right] \tag{3.3}$$

and of a (column) vector with respect to a scalar

$$\frac{\partial y}{\partial x_j} = \left[\frac{\partial y_1}{\partial x_j},\dots,\frac{\partial y_m}{\partial x_j}\right]' . \tag{3.4}$$

Regarding matrices, for the (r,s) and (m,n) matrices Y and X, let the function Y=Y(X) be written $Y^c=Y^c(X^c)$. Then the derivative of Y^c with respect to X^c is the (rs,mn) matrix

$$\frac{\partial Y^c}{\partial X^c} = \left[\frac{\partial y_{.i}}{\partial x_{.j}}\right] \quad i=1,\dots,s\ ,\ j = 1,\dots,n \tag{3.5}$$

Notice that the typical "element" in this case is a (r,m) matrix.

Further, using (2.8) and (3.1)

$$\frac{\partial (AXB')^c}{\partial X^c} = \frac{\partial (B\otimes A)X^c}{\partial X^c} = B\otimes A \tag{3.6}$$

and, by (2.10)

$$\frac{\partial X'^c}{\partial X^c} = \frac{\partial K_{mn}X^c}{\partial X^c} = K_{mn} \tag{3.7}$$

The two fundamental rules of matrix differential calculus that will be used in the next section are the chain rule and the product rule. The former states that if Y=Y(U) and U=U(X) then

$$\frac{\partial Y^c}{\partial X^c} = \frac{\partial Y^c}{\partial U^c} \frac{\partial U^c}{\partial X^c} \quad . \tag{3.8}$$

According to the product rule, if Y=UVW is a matrix wherein U=U(X), V=V(X) and W=W(X) then

$$\frac{\partial (UVW)^c}{\partial X^c} = \left((VW)'\otimes I\right) \frac{\partial U^c}{\partial X^c} + (W'\otimes U) \frac{\partial V^c}{\partial X^c} + (I\otimes UV) \frac{\partial W^c}{\partial X^c} \tag{3.9}$$

This rule stems from the fact that $(UVW)^c$ by use of (2.8) can be rewritten in any of the alternative forms provided by the identity

$$\begin{aligned} (UVW)^c &= \left((VW)'\otimes I\right)U^c \\ &= (W'\otimes U)V^c \\ &= (I\otimes UV)W^c, \end{aligned}$$

(cf. Pollock (1979, pp. 77-79)).

One useful result which can be obtained with the help of the product rule is[6]

$$\frac{\partial Z^{-1c}}{\partial Z^c} = - (Z^{-1\prime}\otimes Z^{-1}). \tag{3.10}$$

Finally, the following result concerning the differentiation of determinants is needed[7]

$$\frac{\partial |Z|}{\partial Z^c} = |Z| Z^{-1r}. \tag{3.11}$$

4 Full-information maximum likelihood estimation of the nonlinear model

To simplify the notation, rewrite (1.1) according to

$$Ax_t = u_t \qquad t=1,\dots,T \tag{4.1}$$

where the n vector x_t and the (m,m+n) matrix A are partitioned as

$$x_t' = (y_t' *_1 (z_t'p) | z_t' *_2 (z_t'q)), \quad A=(B|C)$$

Further, let

$$\upsilon = m + n. \tag{4.2}$$

The stochastic assumption and the general specification of the parameters have been given in (1.2) and (1.4)-(1.5).

In all the following sections both $*_1$ and $*_2$ will initially be assumed to denote division. At the end of the respective sections it will then be shown how the derivations can be adjusted to accomodate for the other possible cases.

4.1 The likelihood function

If both $*_1$ and $*_2$ denote division the vector x_t becomes

$$x_t' = (y_t'(z_t'p)^{-1} | z_t'(z_t'q)^{-1}) \tag{4.3}$$

Letting the vectors y_t' and z_t' constitute the t:th rows of the matrices Y and Z, the set of all T realizations of the relationship (4.1), given (4.3), can be compiled in the matrix equation

$$AV' = BY'D_1^{-1} + CZ'D_2^{-1} = U' \tag{4.4}$$

where the (T,υ) matrix V is partitioned as

$$V = (D_1^{-1}Y|D_2^{-1}Z)$$

and

$$D_1 = \left(\delta_{ij}(z_{i.}p)\right) , \quad i=1,\ldots,T \tag{4.5}$$

$$D_2 = \left(\delta_{ij}(z_{i.}q)\right) , \quad i=1,\ldots,T \tag{4.6}$$

δ_{ij} being the Kronecker delta and $z_{i.}$ the i:th row of the matrix Z.

Hence, although in general it is not possible to express nonlinear systems in linear form, (4.4) shows how this can be done in a simple way in this particular case. A minor difficulty is that the D matrices are rather inefficient means of storing the scalars $(z_{i.}p)$ and $(z_{i.}q)$. This issue will however be deferred until the last section.

The log-likelihood function for the system (4.4) is[8]

$$L(\Theta,\Sigma) = k + \sum_{t=1}^{T} \log|J_t| - \tfrac{1}{2}T\log|\Sigma| -$$

$$\tfrac{1}{2}tr(AV'\Sigma^{-1}VA') \tag{4.7}$$

$$k = -\tfrac{1}{2}mT\log(2\Pi)$$

where J_t is the Jacobian matrix $\partial u_t/\partial y_t$.

Application of (3.2) to (4.1) shows that if $*_1$ denotes division then

$$\sum_{t=1}^{T} \log|J_t| = \sum_{t=1}^{T} \log|\partial u_t/\partial y_t|$$

$$= \sum_{t=1}^{T} \log|(z_t'p)^{-1}B|$$

$$= T\log|B| - m\log|D_1|. \tag{4.8}$$

If Σ is unrestricted then (4.7) can be maximized analytically with respect to this matrix

$$\frac{\partial L}{\partial \Sigma} = 0 \text{ which implies } \hat{\Sigma}(\Theta) = T^{-1}AV'VA' \tag{4.9}$$

(cf. Eisenpress and Greenstadt (1966)).

By substitution of (4.9) in (4.7) the log-likelihood function can be made a function of Θ only

$$L^*(\Theta) = k' + T\log|B| - m\log|D_1| - \frac{1}{2} T\log|\hat{\Sigma}| \tag{4.10}$$

$$k' = -\frac{1}{2} mT(\log(2\Pi) + 1) \quad .$$

If instead both $*_1$ and $*_2$ denote multiplication (4.4) is changed to

$$A\mathring{V}' = BY'D_1 + CZ'D_2 = \mathring{U}' \quad , \tag{4.4'}$$

$$\mathring{V} = (D_1Y|D_2Z)$$

and $\mathring{V}$ has to be substituted for V in all formulas above. In the same way if e.g. $*_1$ denotes multiplication and $*_2$ denotes division

$$A\bar{V}' = BY'D_1 + CZ'D_2^{-1} = \bar{U}', \tag{4.4''}$$

$$\bar{V} = (D_1Y|D_2^{-1}Z).$$

Further, with $*_1$ denoting multiplication the sign in front of $m\log|D_1|$ in (4.10) should be changed.

4.2 Calculation of first order derivatives

To calculate the first order derivatives of the concentrated log-likelihood function is a rather laborious task. Hence, often "derivative-free" algorithms, operating from function values only, are used to maximize L*. One such procedure is the conjugate gradient method of Powell (1964). These routines usually are easy to work with. However, with many parameters to estimate, the computing time can become very long compared to the case where the derivatives are analytically calculated. Another disadvantage concerns the estimated standard errors, which can be obtained from the Hessian matrix. The estimates seem to be accurate when the gradient is analytically computed, but can be rather unreliable when it is numerically approximated[9]. Thus, besides being of theoretical interest, the calculations below are also justified on practical grounds.

By (3.3) $\partial L^*/\partial\Theta$ is a row vector. Using the chain rule (3.8) and the result (3.11) for determinants it can be expressed as

$$\frac{\partial L^*}{\partial\Theta} = T\frac{\partial\log|B|}{\partial B^c}\frac{\partial B^c}{\partial\Theta} - m\frac{\partial\log|D_1|}{\partial D_1^c}\frac{\partial D_1^c}{\partial\Theta} - \frac{1}{2}T\frac{\partial\log|\hat{\Sigma}|}{\partial\hat{\Sigma}^c}\frac{\partial\hat{\Sigma}^c}{\partial\Theta}$$

$$= TB^{-1r}\frac{\partial B^c}{\partial\Theta} - mD_1^{-1r}\frac{\partial D_1^c}{\partial\Theta} - \frac{1}{2}T\hat{\Sigma}^{-1r}\frac{\partial\hat{\Sigma}^c}{\partial\Theta} \tag{4.11}$$

(notice that $\theta^c=\theta$).

The factor $\partial\hat{\Sigma}^c/\partial\theta$ in the last term of (4.11) can be developed in the following way.

All the matrices making up the product matrix $\hat{\Sigma}^c$ are functions of θ. Utilizing (2.8) $\hat{\Sigma}^c$ can be written in any of the alternative forms

$$\hat{\Sigma}^c = T^{-1}(AV'V\otimes I_m)A^c \qquad (4.12a)$$

$$= T^{-1}(A\otimes A)(V'V)^c \qquad (4.12b)$$

$$= T^{-1}(I_m\otimes AV'V)A'^c . \qquad (4.12c)$$

Further, by use of (2.12) and (2.10) the following result is obtained

$$T^{-1}(AV'V\otimes I_m)A^c = T^{-1}K_m(I_m\otimes AV'V)A'^c . \qquad (4.13)$$

Hence, application of first the product and the chain rules [(3.9) and (3.8)] and subsequently (3.7) gives

$$\frac{\partial\hat{\Sigma}^c}{\partial\theta} = T^{-1}(I+K_m)(I_m\otimes AV'V)K_{m\upsilon}\frac{\partial A^c}{\partial\theta} + T^{-1}(A\otimes A)\frac{\partial(V'V)^c}{\partial V^c}\frac{\partial V^c}{\partial\theta} . \qquad (4.14)$$

[Orders of identity matrices will be suppressed when they are obvious as in $(I+K_m)$]

The second term of (4.14) can be further elaborated. Making use of the product rule again the factor $\partial(V'V)^c/\partial V^c$ becomes

$$\frac{\partial (V'V)^c}{\partial V^c} = (I_\upsilon \otimes V') + (V' \otimes I_\upsilon) K_{T\upsilon} \ . \tag{4.15}$$

Some manipulating by means of (2.5) and (2.12) shows that

$$T^{-1}(A \otimes A)\frac{\partial (V'V)^c}{\partial V^c} = T^{-1}(I+K_m)(A \otimes AV') \ . \tag{4.16}$$

Since

$$V^c = (D_1^{-1}Y|D_2^{-1}Z)^c = \begin{matrix} (D_1^{-1}Y)^c \\ (D_2^{-1}Z)^c \end{matrix} \tag{4.17}$$

the derivative $\partial V^c/\partial\Theta$ can be evaluated by differentiating in turn $(D_1^{-1}Y)^c$ and $(D_2^{-1}Z)^c$. The two derivatives being completely analogous it is sufficient to show the calculation of $\partial(D_1{}^{-1}Y)^c/\partial\Theta$. Application of (3.8) and then (3.6) and (3.10) gives

$$\frac{\partial (D_1^{-1}Y)^c}{\partial\Theta} = -(Y'D_1^{-1} \otimes D_1^{-1}) \frac{\partial D_1^c}{\partial\Theta}. \tag{4.18}$$

Exploiting the chain rule the (T^2,ℓ) matrix $\partial D_1^c/\partial\Theta$ can be expressed as

$$\frac{\partial D_1^c}{\partial\Theta} = \left(\delta_{T(j-1)+i} \left(z_{i.} \frac{\partial p}{\partial\Theta_k}\right)\right) \tag{4.19}$$

where $T(j-1)+i$ is the row index, k the column index and

$$\delta_{T(j-1)+i} = \begin{cases} 1 \text{ if } i=j \\ 0 \text{ otherwise} \end{cases} , \tag{4.20}$$

$i,j = 1,\dots,T \quad k = 1,\dots,\ell$. Thus,

$$\frac{\partial D_1^c}{\partial \Theta} = N \frac{\partial p}{\partial \Theta} \qquad (4.21)$$

where the (T^2, m) matrix N can be written

$$N = \begin{matrix} e_{.1} z_{1.} \\ \vdots \\ e_{.T} z_{T.} \end{matrix} \qquad (4.22)$$

$e_{.j}$ being the j:th column of the identity matrix of order T.

Using the same technique with respect to $\partial(D_2^{-1}Z)^c/\partial\Theta$ and collecting the results the derivative $\partial V^c/\partial\Theta$ becomes

$$\frac{\partial V^c}{\partial \Theta} = - \begin{matrix} (Y'D_1^{-1} \otimes D_1^{-1})N(\partial p/\partial\Theta) \\ (Z'D_2^{-1} \otimes D_2^{-1})N(\partial q/\partial\Theta) \end{matrix} . \qquad (4.23)$$

As

$$(A\otimes AV') = (B\otimes AV' | C\otimes AV') \qquad (4.24)$$

the last term of (4.14) can now be rewritten according to

$$T^{-1}(A\otimes A) \frac{\partial (V'V)^c}{\partial V^c} \frac{\partial V^c}{\partial \Theta} = T^{-1}(I+K_m) \times$$

$$\Big((S'\otimes AV'D_1^{-1})N(\partial p/\partial\Theta) + \qquad (4.25)$$

$$\big((AV'-S')\otimes AV'D_2^{-1}\big)N(\partial q/\partial\Theta)\Big)$$

where

$$S = D_1^{-1}YB' \qquad (4.26)$$

When substituting in (4.11) notice that $\hat{\Sigma}^{-1}$ is symmetric. Hence, utilizing (2.9) the last term of (4.11) becomes

$$\frac{1}{2} T\hat{\Sigma}^{-1r} \frac{\partial\hat{\Sigma}^c}{\partial\Theta} = (\hat{\Sigma}^{-1}AV'V)^r K_{m\upsilon} \frac{\partial A^c}{\partial\Theta} - (S\hat{\Sigma}^{-1}AV'D_1^{-1})^r N \frac{\partial p}{\partial\Theta} - \left((VA'-S)\hat{\Sigma}^{-1}AV'D_2^{-1}\right)^r N \frac{\partial q}{\partial\Theta} \qquad (4.27)$$

Observing that B is contained in A and making use of (2.11) the derivative $\partial L^*/\partial\Theta$ can be given its final form

$$\frac{\partial L^*}{\partial\Theta} = \left(T(B^{-1}{}'|0)' - V'W\right)^r \frac{\partial A^c}{\partial\Theta} + (SW'D_1^{-1}-mD_1^{-1})^r N \frac{\partial p}{\partial\Theta} + \left((VA'-S)W'D_2^{-1}\right)^r N \frac{\partial q}{\partial\Theta} \qquad (4.11a)$$

where

$$W = VA'\hat{\Sigma}^{-1} \qquad (4.28)$$

and the zero matrix is of dimension (n,m). With $D_1 = D_2 = I_T$ the first term is the gradient in the usual (linear) FIML case.

Apart from the matrices $N(\partial p/\partial\Theta)$ and $N(\partial q/\partial\Theta)$, (4.11a) does not involve very much extra calculations compared to the linear case. With the vectors and the matrices in the last two terms being of dimension $(1,T^2)$ and (T^2,ℓ) respectively, storage space requirements seem very strong, though. For instance, with T=30 and ℓ=10 there would be almost 20 000 elements to store! Fortunately, there are ways to reduce these requirements considerably, as will be shown in the next section.

The calculations do not change very much if one or both of $*_1$ and $*_2$ denote multiplication instead of division. In (4.11) the sign of $mD_1^{-1r}(\partial D_1^c/\partial\Theta)$ has to be changed. The formulas (4.12) - (4.16) all remain valid if $\mathring{V}$ or $\bar{V}$ are substituted for V.

If $*_1$ and $*_2$ both denote multiplication, i.e. when (4.4') is valid, (4.18) becomes

$$\frac{\partial(D_1Y)^c}{\partial\Theta} = (Y'\otimes I_T)\,\frac{\partial D_1^c}{\partial\Theta} \tag{4.18'}$$

Consequently (4.23) will change to

$$\frac{\partial\mathring{V}^c}{\partial\Theta} = \begin{matrix}(Y'\otimes I_T)N(\partial p/\partial\Theta)\\(Z'\otimes I_T)N(\partial q/\partial\Theta)\end{matrix} \tag{4.23'}$$

Substitution of (4.23') in (4.25) and subsequently (4.11) gives, with suitable change in notation

$$\begin{aligned}\frac{\partial L^*}{\partial\Theta} = &\left(T(B^{-1\prime}|0)' - \mathring{V}'\mathring{W}\right)^r \frac{\partial A^c}{\partial\Theta} - \\ &(D_1^{-1}\mathring{S}\mathring{W}' - mD_1^{-1})^r\, N\,\frac{\partial p}{\partial\Theta} - \\ &\left(D_2^{-1}(\mathring{V}A' - \mathring{S})\mathring{W}'\right)^r N\,\frac{\partial q}{\partial\Theta}\end{aligned} \tag{4.11a'}$$

where

$$\mathring{S} = D_1YB' \tag{4.26'}$$

$$\mathring{W} = \mathring{V}A'\hat{\Sigma}^{-1}\ . \tag{4.28'}$$

With $*_1$ denoting multiplication and $*_2$ division (i.e. (4.4")) the derivative becomes

$$\frac{\partial L^*}{\partial \Theta} = \left(T(B^{-1}{}'|0) - \bar{V}'\bar{W}\right)^r \frac{\partial A^c}{\partial \Theta} +$$

$$(D_1^{-1}\bar{S}\bar{W}' - mD_1^{-1})^r \; N \; \frac{\partial p}{\partial \Theta} \qquad (4.11a")$$

$$\left((\bar{V}A' - \bar{S})\bar{W}'D_2^{-1}\right)^r \; N \; \frac{\partial q}{\partial \Theta}$$

where

$$\bar{S} = \mathring{S} \qquad (4.26")$$

and

$$\bar{W} = \bar{V}A'\hat{\Sigma}^{-1} \qquad (4.28")$$

4.3 Increasing computational efficiency

When transforming the formulas presented above into computer algorithms it becomes necessary to consider computational efficiency. This concept relates to both computing time and storage space requirements. Here particularly the latter aspect will be delat with. It will be shown how storage space requirements can be substantially reduced by simply exploiting the diagonality of the matrices D_1 and D_2. The decrease is thus, in principle, obtained by avoiding storage of a lot of zero elements. In this way gains will also be made in computing time, too.

First, considering the effects of pre- and postmultiplication by diagonal matrices, it is obvious that instead of setting up D_1 and D_1, one may calculate the vectors

$$d^1 = Zp \tag{4.29}$$

and

$$d^2 = Zq. \tag{4.30}$$

In the case e.g. of (4.4), multiplying the elements of the i:th column of BY' by the inverse of the i:th element of d^1 is equivalent to postmultiplying the same matrix by D_1^{-1}.

Looking at (4.22) it can be seen that any row-vectorized (T,T) matrix which is postmultiplied by N can be treated as if it were a diagonal matrix. Thus, only the diagonal elements of $SW'D_1^{-1}$ and $(VA'-S)W'D_2^{-1}$ are needed.

Further, by "compressing" the N matrix so as to retain only its nonzero elements the dimension of the matrix $\partial D_1^c/\partial\theta$ can be reduced from (T^2,ℓ) to (T,ℓ) according to

$$\frac{\partial D_1^c}{\partial\theta}^* = Z\,\frac{\partial p}{\partial\theta} \tag{4.31}$$

where the k:th column is the derivative with respect to θ_k. Application of the chain rule to (4.29) shows that (4.31) can alternatively be written

$$\frac{\partial D_1^c}{\partial\theta}^* = \frac{\partial d^1}{\partial\theta} \tag{4.31a}$$

Corresponding operations can be performed on $\partial D_2^c/\partial\theta$.

Thus, when calculating $\partial L^*/\partial\Theta$ one can proceed as follows, differentiating with respect to one parameter at a time:

i. Set up the matrix W. (Many of the calculations needed to achieve W are the same as those needed to get the likelihood value L*).

ii. Compute the column vectors $\partial d^1/\partial\Theta_k$ and $\partial d^2/\partial\Theta_k$. Assign that part of the derivative which corresponds to the last two terms of (4.11a), to the k:th element of the gradient vector g, according to

$$g_k \leftarrow \sum_{i=1}^{T}\left((w_{i.}s'_{i.}) - m\right)(\partial d^1/\partial\Theta_k)_i/d_i^1 + \sum_{i=1}^{T}\left(w_{i.}(u'_{i.} - s'_{i.})\right)(\partial d^2/\partial\Theta_k)_i/d_i^2 \qquad (4.32)$$

$k = 1,\ldots,\ell$

where d_i^1 and $(\partial d^1/\partial\Theta_k)_i$ are the i:th elements of d^1 and $\partial d^1/\partial\Theta_k$. The vectors $w_{i.}$, $s_{i.}$ and $u_{i.}$ constitute the i:th rows of the matrices W, S and U(= AV'), respectively.

Note that the product matrices WS' and W(U'-S') need not be calculated. With U already computed (to get L*) it is sufficient to set up W and S and then form the required inner products from the corresponding rows.

iii. Compute V'W, set up the partitioned matrix containing B^{-1} and assign

$$g_k \leftarrow g_k + \left(T(B^{-1\prime}|0)' - V'W\right)^r \frac{\partial A^c}{\partial\Theta_{k.}} \qquad (4.33)$$

Compared to direct calculation of the derivatives according to (4.11a) this procedure offers an appreciable increase in computational efficiency. Besides gaining computing time, it considerably reduces the requirements for additional storage space relative to the linear case. The extra storage requirement is confined to the matrices S, $(\partial D_1^c/\partial\Theta)^*$ and $(\partial D_2^c/\partial\Theta)^*$. With respect to the latter two it is only necessary to reserve space for one of their respective columns.

Now consider the case where both $*_1$ and $*_2$ denote multiplication, i.e. (4.11a'). If the matrices $\dot{V}$, $\dot{W}$ and $\dot{S}$ are used instead of V, W and S the only change with respect to (4.32) and (4.33) is that the former should be multiplied by -1. This very nice property is due to the above mentioned fact that, with regards to the last two terms of $\partial L^*/\partial\Theta$, only the diagonal elements of the involved row-vectorized matrices are essential.

For similar reasons, when calculating (4.11a") the first term of (4.32) will, after appropriate substitutions, get a negative sign .

NOTES

[1] A common characteristic feature of all systems of share equations like (1.3) is that their residual variance - covariance matrices are singular. As that problem has been dealt with elsewhere, see e.g. Barten (1969), it will not be taken up in this paper. In other words, Σ in (1.2) is assumed to have full rank.

[2] These methods differ from the Newton-Raphson procedures in that they do not require analytical second order derivatives, and hence are much simpler to use. Also, unlike Newton-Raphson they have been found to work well even if the initial values are far from the optimum. See e.g. Fletcher and Powell (1963) and Fletcher (1970).

The advantages of using quasi-Newton algorithms instead of numerically approximating both first and second order derivatives are shortly discussed in the beginning of section 4.2.

[3] The use of the indices "c" and "r" is due to Pollock. Often vec X is used instead of X^c. There is no counterpart to X^r, however, which makes Pollock's notation preferable.

[4] The name commutation matrix and the denotation K_{mn} were introduced by Magnus and Neudecker. The same matrix was defined in Pollock who called it the tensor commutator and denoted it by an encircled T. The former notation is used here because it is simpler and shows the dimension of the matrix.

[5] It may be noted that (3.1) differs from conventional practice which arrays the partial derivatives in an (n,m) matrix. (See e.g. Neudecker (1969).)

[6] Differentiate $(ZZ^{-1})^c$ according to (3.9)

$$\partial(ZZ^{-1})^c = (Z^{-1\prime} \otimes I) + (I \otimes Z)\, \frac{\partial Z^{-1c}}{\partial Z^c} = 0 \quad .$$

The result must be equal to the zero matrix since $ZZ^{-1} = I$. Rearranging and applying (2.7) gives (3.10).

[7] This result comes from the identity

$$Z^{-1} = |Z|^{-1} \text{ adj } Z$$

which can be equivalently expressed as

$$I|Z| = (\text{adj } Z)Z.$$

Differentiation gives

$$\frac{\partial(I|Z|)^c}{\partial Z^c} = I \otimes \text{adj } Z \quad .$$

Rewriting adj Z according to the basic identity and noting that (3.3) is applicable gives (3.11).

[8] Cf. Goldfeld and Quandt (1972, pp.233-234).

[9] $\text{Var}(\Theta) = -T^{-1}\text{plim}\left(\frac{\partial(\partial L^*/\partial\Theta)'}{\partial\Theta}\right)$

Only asymptotic properties of the estimated standard errors are known. It seems likely, however, that better approximations of the second order derivates should in general result in more accurate estimates of the variances. That is the argument underlying the statement in the text.

Belsley (1980) concludes that the Hessians produced by quasi-Newton methods are very close to the analytic values, even in the case of strong nonlinearities.

In general it is difficult to approximate the analytic Hessian equally well when the gradient is numerically calculated. In this case the estimated standard errors are also sometimes sensitive to the scaling of the function value L*.

REFERENCES

Barten, A.P., 1969, "Maximum Likelihood Estimation of a Complete System of Demand Equations", European Economic Review, Fall, pp. 7-73.

Belsley, D.A., 1980, "On the Efficient Computation of the Nonlinear Full-information Maximum Likelihood Estimator", Journal of Econometrics Vol. 14, No. 2.

Christensen, L.R., D.W. Jorgenson and L.J. Lau, 1975, "Transcendental Logarithmic Utility Functions", The American Economic Review, Vol. LXV, No. 3.

Eisenpress, H. and J. Greenstadt, 1966, "The Estimation of Nonlinear Econometric Estimators", Econometrica, 34, Oct., pp. 851-861.

Fletcher, R. and M.J.D. Powell, 1963, "A rapidly Convergent Descent Method for Minimization". The Computer Journal,No. 6, pp. 163-168.

Fletcher, R., 1970, "A New Approach to Variable Metric Algorithms", The Computer Journal, Vol. 13, No. 3 pp. 317-322.

Goldfeld, M.S. and R.E Quandt, 1972, Nonlinear Methods in Econometrics, Amsterdam, North Holland Publishing Co.

Jansson, L. and E. Mellander, 1983, "Optimization under Nonlinear Constraints", Economics Letters 11, pp. 343-346.

Magnus, J.R. and H. Neudecker, 1979, "The Commutation Matrix: Some Properties and Applications", The Annals of Statistics, Vol. 7, No. 2, pp. 381-394.

Neudecker, H., 1969, "Some Theorems on Matrix Differentiation with Special Reference to Kronecker Products", Journal of the American Statistical Association, pp. 953-963.

Pollock, D.S.G., 1979, The Algebra of Econometrics, Chichester, John Wiley and Sons.

Powell, M.J.D., 1964, "An Efficient Method for Finding the Minimum of a Function of Several Variables Without Calculating Derivatives, The Computer Journal, Vol. 7, pp. 155-162.

Erik Mellander
The Industrial Institute for
Economic and Social Research
Grevgatan 34, 5
S-114 53 Stockholm
Sweden
Tel: (08) 783 84 13

WORKING PAPERS (Missing numbers indicate publication elsewhere)

1976

1. Corporate and Personal Taxation and the Growing Firm
by Ulf Jakobsson

8. Estimation and Analysis with a WDI Production Function
by Göran Eriksson, Ulf Jakobsson and Leif Jansson

1977

15. Pay Differentials between Government and Private Sector Employees in Sweden
by Siv Gustafsson

1980

25. On Unexplained Price Differences
by Bo Axell

34. Imperfect Information Equilibrium, Existence, Configuration and Stability
by Bo Axell

1981

36. Energi, stabilitet och tillväxt i svensk ekonomi (Energy, Stability and Growth in the Swedish Economy)
by Bengt-Christer Ysander

38. Utiliy in Local Government Budgeting
by Bengt-Christer Ysander

40. Wage Earners Funds and Rational Expectations
by Bo Axell

42. The Structure of the ISAC Model
by Leif Jansson, Tomas Nordström and Bengt-Christer Ysander

43. An Econometric Model of Local Government and Budgeting
by Bengt-Christer Ysander

44. Local Authorities, Economic Stability and the Efficiency of Fiscal Policy
by Tomas Nordström and Bengt-Christer Ysander

45. Growth, Exit and Entry of Firms
by Göran Eriksson

56. Central Control of the Local Government Sector in Sweden
by Richard Murray

59. Longitudinal Lessons from the Panel Study of Income Dynamics
by Greg J. Duncan and James N. Morgan

1982

61. Var står den nationalekonomiska centralteorin idag?
av Bo Axell

63. General Search Market Equilibrium
by James W. Albrecht and Bo Axell
General Equilibrium without an Auctioneer
by James W. Albrecht, Bo Axell and Harald Lang

64. The Structure and Working of the ISAC Model
by Leif Jansson, Thomas Nordström and Bengt-Christer Ysander

65. Comparative Advantage and Development Policy Twenty Years Later
by Anne O. Krueger

67. Computable Multi-Country Models of Production and Trade
by James M. Henderson

69. Relative Competitiveness of Foreign Subsidiary Operations of a Multinational Company 1962-77
by Anders Grufman

71. Technology, Pricing and Investment in Telecommunications
by Tomas Pousette

72. The Micro Initialization of MOSES
by James W Albrecht and Thomas Lindberg

75. The MOSES Manual
by Fredrik Bergholm

76. Differential Patterns of Unemployment in Sweden
by Linda Leighton and Siv Gustafsson

77. Household Market and a Nonmarket Activities (HUS)
- A Pilot Study
by Anders Klevmarken

102. Collecting Data for Micro Analysis.
Experiences from the Hus-Pilot Study
by N. Anders Klevmarken

103. The Swedish Micro-to-Macro Model
- Idea, Design and Application
by Gunnar Eliasson

105. Estimation of Wage Gains and Welfare Gains from
Self-Selection Models
by Anders Björklund and Robert Moffitt

106. Public Policy Evaluation in Sweden
by Bengt-Christer Ysander

108. Entry, Industry Growth and the Microdynamics
of Industry Supply
by John C. Hause and Gunnar Du Rietz

112. Schumpeterian Competition in Alternative Technological
Regimes
by Sidney Winter

113. The Industrial Finance Systems; Europe, U.S. and Japan
by Tad Rybczynski

117. Micro Heterogeneity of Firms and the Stability of Industrial
Growth
by Gunnar Eliasson

118. The Initialization Process - The Moses Manual, Part 2
by Fredrik Bergholm

<u>1984</u>

119. Forestalling the Demise of Empirical Economics: The Role
of Microdata in Labor Economics Research
by Frank Stafford

120. Unemployment and Mental Health - Some Evidence from
Panel Data
by Anders Björklund

121. Comparative Keynesian Dynamics
by Richard H. Day

122. The Firm and Financial Markets in the Swedish
Micro-to-Macro Model (MOSES)
- Theory, Model and Verification
by Gunnar Eliasson

123. What Can Be Learned from the Chrysler Experience?
by Frank Stafford

125. Estimates of Bounds for Export Price Elasticities
in an Underidentified Simultaneous Two Equation System
by Eva Christina Horwitz

126. An Easy Fixup of the Flaw in Lucas' "Neutrality of Money"
by Harald Lang

127. Comment on Feldstein and Poterba's "Unemployment
Insurance and Reservation Wages"
by Harald Lang

128. Information Criterion and Estimation of Misspecified
Qualitative Choice Models
by David Brownstone

129. Organisation och styrsystem i ett modernt storföretag
av Gunnar Eliasson

130. Fertility and the Opportunities of Children: Evidence on
Quantity-Quality Interactions in Sweden
by Bertil Holmlund

131. CONRAD - A Maximum Likelihood Program for Estimation
of Simultaneous Equations Models that are Non-Linear
in the Parameters
by Leif Jansson and Erik Mellander

132. A General FIML Estimator for a Certain Class of Models
that are Nonlinear in the Variables
by Erik Mellander