1973

Conversion of Simultaneous Least Squares to a Malinvaud Minimum Distance Estimator: MDSLS

T. Merritt Brown

RESEARCH REPORT 7305

CONVERSION OF SIMULTANEOUS LEAST SQUARES TO A MALINVAUD MINIMUM DISTANCE ESTIMATOR: MDSLS

by

T. Merritt Brown

February, 1973

1. INTRODUCTION

This article explores the theory, and then develops the computational techniques, for converting Simultaneous Least Squares (SLS) [2] to a Malinvaud Minimum Distance Estimator (MDSLS). It begins with a brief review of the linear theory which leads by analogy to Professor Malinvaud's definition of a minimum distance estimator [10, p. 325]. His definition broadens to cover the situation where the parameters enter the estimation problem in a nonlinear form. Such nonlinearity can occur when a complete linear model of several simultaneous equations is being estimated as a simultaneous operation; or when a single equation, or model of several equations, is initially structurally nonlinear in parameters and variables [10, Chap. 9], [3]

One would expect the computational program to become more burdensome because of the added generality in theoretical foundations. Also since Malinvaud states at various points in [10] that his estimator is equivalent to the maximum likelihood estimator if the disturbances are normal, one would expect MDSLS to be at least as burdensome as the computations for Full Information Maximum Likelihood (FML), or QFML when the disturbance properties are unknown [9, pp. 134-9]. It is demonstrated below that MDSLS and QFML do indeed converge to the same estimates. Bur surprisingly the computations for MDSLS appear to be shorter and easier to handle than the outlines previously presented for SLS and FML in [1], [2], [5], [6], [8], [9], [11].

2. MALINVAUD'S DEFINITION OF A MINIMUM DISTANCE ESTIMATOR

On pages 325-6 of [10] Malinvaud defines a minimum distance estimator (MDE) as that $\hat{y}$ which minimizes the quadratic form

(1) $(x-y)' Q^{-1} (x-y)$, where x is an N dimensional vector with covariance matrix Q and Ex = y. $\hat{y}$ = est y. y does not necessarily belong to a linear subspace, as it does in the general linear model. He points out that "no general theory of this

method of estimation can be given," and that "There is no known optimal property valid for $\hat{y}$.

Only one result can be stated, and it is obvious; if x is normally distributed, $\hat{y}$ is the maximum likelihood estimator."

But Malinvaud is looking for a procedure with less restrictive assumptions than normality, and which is not confined to asymptotically optimal properties.

Malinvaud's definition is presumably drawn by analogy from the theory of estimation of the general linear model.

(2) $\underset{Tx1}{Y} = \underset{Txk}{X}\ \underset{kx1}{\alpha} + \underset{Tx1}{\mu}$, with T observations in a sample of observed data for Y and X. Assume that

(3) $E\ \mu\mu' = V$ is a full covariance matrix implying heteroscedasticity and all possible orders of autocorrelation in μ. If μ is normally distributed, then its density or likelihood function is given by

$$(4) \qquad L(\mu) = (2\Pi)^{-\frac{1}{2}T} (\det V)\exp - \frac{1}{2}\, \mu' V^{-1} \mu\ .$$

Then the maximum likelihood estimator $\hat{\alpha}$ of α will be those estimates which max L or min $\hat{\mu}' V^{-1} \hat{\mu}$, and is consequently an MDE.

If μ is not assumed normal we find the best linear unbiased estimator a of α and u of μ using the Aitken Generalized Least Squares (GLS) approach. In this process we assume μ to exist in a non-Euclidian or non-orthogonal basis or space R(TxT) such that $R\mu = \mathcal{E}$. Then $\mathcal{E}$ is assumed to exist in an Euclidian basis, such that it is homoscedastic and not autocorrelated. Thus

$$(5) \qquad E\ \mathcal{E}\mathcal{E}' = \sigma_{\mathcal{E}}^{2}\ I\ ;\ (E\mathcal{E}'\mathcal{E})^{\frac{1}{2}} = E\|\mathcal{E}\| = \sqrt{T}\ \sigma_{\mathcal{E}} \text{ an Euclidian distance.}$$

The corresponding non-Euclidian or generalized distance for μ associated with $\mathcal{E}$ can be taken as

$$(6) \qquad (\mu' R' R \mu)^{\frac{1}{2}} = \|\mathcal{E}\| = \sigma_{\mathcal{E}} (\mu' V^{-1} \mu)^{\frac{1}{2}}\ .$$

$\mu' V^{-1} \mu$ can properly define a distance function or metric if V^{-1} is a positive definite matrix. This is always so for V a covariance matrix.

Let u = est μ and e = est $\mathcal{E}$. Then the GLS procedure involves minimizing

(7) $u' V^{-1} \mu = e'e/\sigma_{\mathcal{E}}^2$, and hence is also a minimum distance estimator according to Malinvaud's definition above.

Let us now refer to the generalized non-Euclidian distance function or metric used by Malinvaud as

(8) $MD = u' V^{-1} u$, in the context of (2) above,

$$= \sum_{t,s=1}^{T} v^{ts} u_t u_s .$$

What the mathematical analysis tells us here is that for an optimal estimation of a = est α, all squares and cross products of the estimated disturbances or residuals must be put on a kind of equal footing, before minimization proceeds. For our estimation to be proper, we must have E uu' = V, as well as E a = α. A straight minimization of $u'u$ would not achieve this, for it would attempt to make u_t^2 as small as u_p^2, when in fact according to the probability structure of the system μ_t possibly should in general be larger than μ_p. Thus if in V we have v_{tt} and v_{ts} large, while v_{pp} and v_{pq} are small, we should expect u_t^2 and $u_t u_s$ to be large, and u_p^2 and $u_p u_q$ to be small. Then our minimization procedure, if it is faithfully to reflect the probability structure V of the system, should encourage largeness in u_t^2 and $u_t u_s$ relative to smallness in u_p^2 and $u_p u_q$. This is precisely what the minimum distance estimator MD = min achieves for us.

For in V^{-1} we have v^{tt} and v^{ts} tending to be small, when v_{tt} and v_{ts} are large, while v^{pp} and v^{pq} tend to be large when v_{pp} and v_{pq} are small. Thus u_t^2 and $u_t u_s$ are given small weights in the minimization process, while u_p^2 and $u_p u_q$ are given large weights, so that the former pair will not be excessively diminished while the

latter pairs will be strongly reduced. Through this process we obtain optimal estimates of α and μ jointly, and through repeated samples of data could make an optimal estimate of $V = E\mu\mu'$.

The insight provided by the mathematical analysis of the ML and GLS methods contains an irresistible logic. The analysis is rigorous when EY is explained in a linear subspace. But when EY is explained by a nonlinear function of parameters and variables, we can as yet only use the insight provided by the linear case. Malinvaud's MDE applies this insight. From this vantage, Malinvaud's minimum distance estimation formula generalized to include nonlinear subspaces for EY seems eminently satisfactory.

3. A COMPLETE LINEAR MODEL OF AN ECONOMIC SYSTEM

Let such a model be represented by

$$(9) \qquad \underset{h\times h\ h\times T}{BY} + \underset{h\times k\ k\times T}{CZ} + \underset{h\times T}{U} = 0 = AX + U$$

Y is an hxT matrix of T observations on h endogenous variables, Z represents k predetermined variables, and U is the matrix of estimated disturbances.

$A(h \times (h+k)) = B,C$; $X((h+k) \times T) = \begin{bmatrix} Y \\ Z \end{bmatrix}$; b_{ii} of $B = -1$, for $i=1,\ldots,h$, representing normalization. h-g of the equations in (9) are identities with known parameters and zero disturbances, so that only g of the equations are stochastic. True or population parameters corresponding to B, C, A and U are β, Γ, $\mathcal{A}$ and $\mathcal{U}$. The population covariance matrix of disturbances is Σ, and

$$(10) \qquad S = \frac{UU'}{T} = \text{est } \Sigma = \frac{AXX'A'}{T} .$$

We make the following assumptions about the disturbances $\mathcal{U}$ in this model. The disturbance for each equation has zero mean and constant variance (homoscedastic). The contemporaneous covariance of disturbances for pairs of equations are constant and may be non-zero. There is no auto- or serial-correlation among the disturbances. Thus

(11) $E\,\mu_{it} = 0\ ;\ E\,\mu_{it}^2 = \sigma_i^2\ ;\ E\,\mu_{it}\mu_{jt} = \sigma_{ij}\ ;$

$E\,\mu_{it}\mu_{it-r} = 0\ ;\ E\,\mu_{it}\mu_{jt-r} = 0\ ;\ i,j=1,\ldots,g\ ;\ t=1,\ldots,T\ ;\ r=0,1,\ldots,T\ .$

The parameters which are to be estimated for this model are the true values corresponding to A, U and S. The behavioural parameters of A which require estimation are set out as a vector by two operations. Let the rows of A be represented by a_i and the columns by a^j $(i=1,\ldots,h\ ;\ j=1,\ldots,h+k)$. Then

(12) $$A = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_h \end{bmatrix} = [a^1\ a^2\ \ldots\ a^{h+k}]$$

vec A $= (a_1, a_2, \ldots, a_h)$

vec*A = vec A with all zeros and constants deleted = a', a row vector containing n elements $(a_1, a_2, \ldots, a_n)$.

For any econometric method requiring a non-singular Σ matrix it is necessary for us to remove the identities from (9). This is because each identity produces a row and column of zeros in Σ. Removal is carried out by substituting the identities into the stochastic equations, a process which preserves the information in the identities and the endogenous character of all variables explained by them. The following model demonstrates this process.

(13) $$C = b_{13}Y + c_{11}L + c_{10} + u_1$$

$$I = b_{23}Y + c_{23}Y_{-1} + c_{20} + u_2$$

$$Y = C + I + G$$

The endogenous variables are $C = Y_1$ = consumer demand, $I = Y_2$ = investment demand, $Y = Y_3$ = GNP; the predetermined variables are $L = Z_1$ = liquid assets of households, $G = Z_2$ = government spending, Y_{-1} = GNP lagged one period = Z_3, and $Z_o \equiv 1$.

Vec*A $= a' = (b_{13},\ c_{11},\ c_{10},\ b_{23},\ c_{23},\ c_{20})$.

The model is over identified.

Substituting the identity into (13) produces the reduced model.

(14) $$(b_{13}-1)C + b_{13}I + b_{13}G + c_{11}L + c_{10}Z_o + u_1 = 0$$

$$b_{23}C + (b_{23}-1)I + b_{23}G + c_{23}Y_{-1} + c_{20}Z_o + u_2 = 0$$

We do not normalize the reduced model, which is represented by

(15) $B_r Y_r + C_r Z + U = 0 = A_r X_r + U$. Note that U is unchanged by the substitution, and that

(16) $$B_r = \begin{bmatrix} (b_{13}-1), & b_{13} \\ b_{23}, & (b_{23}-1) \end{bmatrix} ; \; C_r = \begin{bmatrix} c_{11}, & b_{13}, & 0, & c_{10} \\ 0, & b_{23}, & c_{23}, & c_{20} \end{bmatrix}$$

If we let $a_r' = \text{vec}^* A_r$, we have each $a_{ri} = a_{ri}(a)$, a function of the unrestricted parameters a of the original model (9).

The explanation of the endogenous variables and data Y_r (henceforth Y without subscript) by the reduced model separates into a systematic component Y_s and a random component U_s, as we solve (15) for Y.

(17) $$Y = - B_r^{-1} C_r Z - B_r^{-1} U = Y_s + U_s .$$

Here U_s represents the total residuals of the complete model in simultaneous solution, and $Y_s = E\,Y = - B_r^{-1} C_r Z = F_r Z$. We define (17) as the structural reduced form (SRF) of the model, with SRF residuals U_s and covariance matrix S_s .

(18) $$S_s = B_r^{-1} \frac{UU'}{T} B_r^{-1\prime} = B_r^{-1} S B_r^{-1\prime} = \text{est } \beta_r^{-1} \Sigma \beta_r^{-1\prime} = \text{est } \Sigma_s .$$

4. SLS AND ITS CONVERSION TO A MALINVAUD MINIMUM ESTIMATOR

The principle of SLS is to find that parameter structure A and a which will cause the vector of total residuals $u_s = \text{vec}' U_s$ to have a minimum Euclidian distance. Our distance function D is the square of this distance

(19) $$D = u_s' u_s = u_{s11}^2 + \ldots + u_{s1T}^2 + \ldots\ldots + u_{sg1}^2 + \ldots + u_{sgT}^2 = \text{tr } TS_s = \text{tr } S_s^* = \text{tr } U_s U_s' \quad [2] .$$

In terms of (1) and Sec. 2 above, EY is explained in a nonlinear subspace of the parameters a to be estimated. Thus $EY = -B_r^{-1}C_r Z$ is a nonlinear function of a. Also $\text{cov}(\text{vec } Y) = \text{cov } y = \text{cov}(y - Ey) = \text{cov } \mu_s = E\,\mu_s\mu_s' = \Sigma_s \otimes I_T \neq I_{gT}$. Hence using SLS to estimate (9) and (15) is analogous to using LS in the general linear model, when GLS should be used for increased efficiency. Consequently applying the principles of Sec. 2 above, we have for a generalized Malinvaud distance (squared)

(20) $MD = u_s'(\Sigma_s \otimes I_T)^{-1}u_s = \min$, as an improved basis for estimation of β, Γ and Σ.

Conversion of MD to Matrix Form for Convenient Computation

A useful symbolism in what follows is the following. Let A and B be matrices of the same dimension, mxn. Define $A \circ B = [a_{ij}b_{ij}]$ $i=1,\ldots,m$; $j=1,\ldots,n$ = simple direct product of A and B. Define operator s in front of a matrix to mean to sum up all elements of the matrix. Thus $sC = \sum_{i,j} c_{ij}$. Then

(21) $s(A \circ B) = a_{11}b_{11}+\ldots+a_{ij}b_{ij}+\ldots+a_{mn}b_{mn} = \text{vec } A \text{ vec}' B = \text{tr } AB'$

From (20) we have

(22) $MD = \text{tr } \Sigma_s^{-1} \otimes I_T u_s u_s' = s\left\{(\Sigma_s^{-1} \otimes I_T) \circ u_s u_s'\right\}$

$$= s \begin{bmatrix} \sigma_s^{11} I_T & \cdots & \sigma_s^{1g} I_T \\ \vdots & & \vdots \\ \sigma_s^{g1} I_T & \cdots & \sigma_s^{gg} I_T \end{bmatrix} \circ \begin{pmatrix} u_{s11} \\ \vdots \\ u_{s1T} \\ \vdots \\ u_{sg1} \\ \vdots \\ u_{sgT} \end{pmatrix} (u_{s11}\ldots u_{s1T}\ldots u_{sg1}\ldots u_{sgT})$$

$$= \sigma_s^{11}(u_{s11}^2+\ldots+u_{s1T}^2) + \cdots\cdots + \sigma_s^{1g}(u_{s11}u_{sg1}+\ldots+u_{s1T}u_{sgT})$$

$$\vdots \qquad\qquad\qquad\qquad\qquad\qquad \vdots$$

$$\sigma^{g1}(u_{sg1}u_{s11}+\ldots+u_{sgT}u_{s1T}) + \cdots\cdots + \sigma_s^{gg}(u_{sg1}^2+\ldots+u_{sgT}^2)$$

$$= \sum_{t=1}^{T} \sum_{i,j=1}^{g} \sigma_s^{ij} u_{sit} u_{sjt} = \sum_{t=1}^{T} u_s^{t\prime} \Sigma_s^{-1} u_s^t$$

(23) $MD = \text{tr } U_s' \Sigma_s^{-1} U_s = \text{tr}(\Sigma_s^{-1} U_s) U_s'$ (since tr $AB' = $ tr $A'B$) $= \text{tr } \Sigma_s^{-1} S_s^* = s\Sigma_s^{-1} \circ S_s^*$.

This result supports Professor Dhrymes' conjecture in [6], p. 203, (15).

We note from (23) that while SLS is based on min tr S_s^*, MDSLS (or MD for short) involves min weighted sum of S_s^*, with each element of S_s^* weighted by its corresponding element in Σ_s^{-1}. This is in conformity with the theory of Section 2, found in optimal estimation of the general linear model.

Iteration Formula for Min MD

The iteration formula proposed is the one used in [9], [8], [1], [2], [5]. It is essentially the Newton method using the vector of steepest descent, $-\frac{\partial MD}{\partial a} = -MD_a$, adjusted by the inverse "curvature" matrix $\left[\frac{\partial^2 MD}{\partial a \partial a'}\right]^{-1} = MD_{aa}^{-1}$, to approach a geodesic line on the surface $MD = MD(a)$. [1, pp. 641-2].

(24) $$a^{m+1} = a^m - h_m \left[\frac{\partial^2 MD}{\partial a \partial a'}\right]_m^{-1} \left(\frac{\partial MD}{\partial a}\right)_m ; \; h_o = .5, \; h_1 = .75, \; h_2 = 1.0, \; h_3 = 1.0, \text{ etc.}$$

Calculation of First Order Partials, MD_a

From (23)

(25) $$\frac{\partial MD}{\partial a_i} = s \, \Sigma_s^{-1} \circ \frac{\partial S_s^*}{\partial a_i} ; \; S_s^* = (Y - F_r Z)(Y - F_r Z)'$$

(26) $$\frac{\partial S_s^*}{\partial a_i} = -\frac{\partial F_r}{\partial a_i} Z(Y - F_r Z)' + (Y - F_r Z)(-Z' \frac{\partial F_r'}{\partial a_i})$$

$$= \left\{\frac{\partial F_r}{\partial a_i}(ZZ'F_r' - ZY')\right\} + \left\{\frac{\partial F_r}{\partial a_i}(ZZ'F' - ZY')\right\}' = \left\{\frac{\partial F_r}{\partial a_i}(ZZ'F_r' - ZY')\right\} + \left\{ '' \right\}',$$

where ('') means to repeat the entire preceding matrix. Let

(27) $(ZZ'F_r' - ZY') = Z(Z'F_r' - Y') = E$. Then

(28) $$\frac{\partial S_s^*}{\partial a_i} = \frac{\partial F_r}{\partial a_i} E + (")' .$$

(29) $$F_r = - B_r^{-1} C_r \; ; \; dF_r = B_r^{-1}(dB_r)B_r^{-1}C_r - B_r^{-1}dC_r .$$

(30) $$\frac{\partial F_r}{\partial a_i} = - B_r^{-1}\left(\frac{\partial B_r}{\partial a_i} F_r + \frac{\partial C_r}{\partial a_i}\right) \; ; \; [2.\ p.\ 179].$$

Note that in a linear model, $\frac{\partial B_r}{\partial a_i}$ and $\frac{\partial C_r}{\partial a_i}$ are constant matrices which are easy to calculate and store. Thus, returning to (16),

(31) $$\frac{\partial B_r}{\partial b_{13}} = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \; ; \; \frac{\partial C_r}{\partial b_{13}} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

It is now easy to put together the components of $\frac{\partial MD}{\partial a_i}$. We combine matrices like (31) into (30), (30) and (27) into (28), (28) into (25).

Calculation of Second Order Partials, MD_{aa}

(32) $$\frac{\partial^2 MD}{\partial a_i \partial a_j} = s\, \Sigma_s^{-1} \circ \frac{\partial^2 S_s^*}{\partial a_i \partial a_j} .$$ From (28)

(33) $$\frac{\partial^2 S_s^*}{\partial a_i \partial a_j} = \left(\frac{\partial^2 F_r}{\partial a_i \partial a_j} E + \frac{\partial F_r}{\partial a_i}\frac{\partial E}{\partial a_j}\right) + (")'$$

(34) $$\frac{\partial E}{\partial a_j} = ZZ'\frac{\partial F_r'}{\partial a_j} .$$ Then from (30)

(35) $$\frac{\partial^2 F_r}{\partial a_i \partial a_j} = B_r^{-1}\frac{\partial B_r}{\partial a_j} B_r^{-1}\left(\frac{\partial B_r}{\partial a_i} F_r + \frac{\partial C_r}{\partial a_i}\right) - B_r^{-1}\frac{\partial B_r}{\partial a_i}\frac{\partial F_r}{\partial a_j}$$

$$= - B_r^{-1}\left(\frac{\partial B_r}{\partial a_j}\frac{\partial F_r}{\partial a_i}\right) - B_r^{-1}\frac{\partial B_r}{\partial a_i}\frac{\partial F_r}{\partial a_j}$$

$$(36)\quad \frac{\partial^2 F_r}{\partial a_i \partial a_j} = -\,B_r^{-1}\left(\frac{\partial B_r}{\partial a_i}\frac{\partial F_r}{\partial a_j} + \frac{\partial B_r}{\partial a_j}\frac{\partial F_r}{\partial a_i}\right) \qquad \text{(Cf. Carter [5], p. 4).}$$

Combining (36), (34), (30) and (27) into (33), we are then able to complete (32), and hence MD_{aa}.

The above formulas (25)-(36) represent all of the basic calculations needed for the iteration program (24). They are all simple matrix operations, and have the further advantage of iterating directly to a and A, without having to go through a_r or vec F_r first. These calculations seem to be easier and more direct than previous calculation programs outlined for SLS or FML.

Estimation of Σ_s

Up to this point we have assumed that Σ_s is a known constant. In practice this covariance structure is usually not known, and hence must be estimated. A consistent prior estimate could be derived from the unrestricted reduced form (URF) as

$$(37)\quad \hat{\Sigma}_{su} = \frac{U_{su}U_{su}'}{T} = \frac{1}{T}\,Y(I - Z'(ZZ')^{-1}Z)Y' .$$

However a more accurate estimate of this matrix, given a small sample of data, would be derived from 2SLS estimates of the structure of the model, B_r^o, C_r^o, F_r^o, followed by use of the SRF, giving

$$(38)\quad \hat{\Sigma}_s^o = \frac{(Y - F_r^o Z)(Y - F_r^o Z)'}{T} = S_s^o$$

$\hat{\Sigma}_s$ can then be updated after each iteration of MD, so that

$$(39)\quad \Sigma_s^m = \frac{(Y - F_r^m Z)(Y - F_r^m)'}{T} = S_s^m$$ would be used as a constant in the iteration which provides $a^{(m+1)}$, A^{m+1}, B_r^{m+1}, C_r^{m+1} and F_r^{m+1} .

Value of Min MD as Check on Convergence[1]

At the conclusion of each iteration we have from (23)

(40) $MD = tr\left(S_s^m\right)^{-1} S_s^{m+1}\, T$. Near convergence $S_s^m \to S_s^{m+1}$, so that

$MD \to tr\left(S_s^{m+1}\right)^{-1} S_s^{m+1} T = gT$.

A knowledge of this minimum value (gT) of MD can be useful as a check on progress and proximity of convergence during computations.

5. COMPARISON OF MDSLS TO CLOSELY RELATED METHODS

(a) SLS. As observed above, and in [6], SLS does not make sufficient use of the probability structure of the model. It works on min trace of S_s, while MD minimizes the sum of all of S_s, after appropriately weighting each element with the probability structure.

Of further interest is the feature that while SLS for small samples of data is not invariant to changes in units and scale of the endogenous variables, MD is invariant to such changes. This checks using the approach of [4].

(b) LWV. The method of Least Weighted Variance was discovered by Professors Ronald J. and Thomas H. Wonnacott [12, pp. 365 ff] through geometrical analysis on single equation estimation. It involves minimizing the sum of the squares of the SRF residuals related to the endogenous variables in a single equation of a complete model, with each residual appropriately weighted. The weight applied to each residual is the reciprocal of the variance of the URF residual corresponding to the SRF residual.

If the concept of LWV is carried over to a complete model and translated into a full systems estimator, it becomes a close variant of MDSLS. In terms

---

[1] I am indebted to Dr. R. A. L. Carter for this suggestion.

of (23) above it becomes as follows. If M is a square matrix let DM be the diagonal matrix formed when all but the diagonal elements of M are converted to zeros. Thus

$$(41) \qquad DM = \begin{bmatrix} m_{11} & 0 \ldots\ldots 0 \\ 0 & m_{22} \ldots . 0 \\ \vdots & \\ 0 & 0 \ldots\ldots m_{nn} \end{bmatrix}$$

Then the LWV method is

$$(42) \qquad LWV(a) = s \left(D\hat{\Sigma}_{su} \right)^{-1} \circ S_s^{\cdot} = \min.$$

$$(43) \qquad \frac{\partial LWV}{\partial a_i} = s \left(D\hat{\Sigma}_{su} \right)^{-1} \circ \frac{\partial S_s^{\cdot}}{\partial a_i} \; ; \; \frac{\partial^2 LWV}{\partial a_i \partial a_j} = s \left(D\hat{\Sigma}_{su} \right)^{-1} \circ \frac{\partial^2 S_s^{\cdot}}{\partial a_i \partial a_j}$$

It might be found desirable to modify LWV to use

$$(44) \qquad LWV = s \left(DS_s^{m-1} \right)^{-1} \circ S_s^{m\cdot} \text{ , thereby improving the weighting matrix with}$$

each iteration.

(c) QFML

This method ([9], [8], [1], [2]) involves minimizing the determinant

(45) $d(a) = |S_s|$. This would appear to be a method quite different from MD as summarized in (23). If however we make a monotonic transformation of d to $\ln(d) = LD(a)$, the difference begins to close.

$$(46) \qquad LD = \ln|S_s| \; ; \; dLD = \frac{d|S_s|}{|S_s|} = \text{vec } S_s^{-1} \text{ vec}' dS_s' = s \; S_s^{-1} \circ d \; S_s = \min .$$

$$(47) \qquad \frac{\partial LD}{\partial a_i} = s \; S_s^{-1} \circ \frac{\partial S_s}{\partial a_i} = \frac{1}{T} \; s \; S_s^{-1} \circ \frac{\partial S_s^{\cdot}}{\partial a_i}$$

Thus rather surprisingly the first order partial derivatives of LD are quite

close to those of MD. Recalling (25), (39) and (40), these differ in a factor of $\frac{1}{T}$, and in the use of the current value of $\left(S_s^m\right)^{-1}$ in (47) compared to the value of $\left(S_s^{m-1}\right)^{-1}$ from the previous iteration in (39). They become effectively the same, as we near convergence.

The second order partials appear at first sight to differ more substantially.

$$(48)\quad \frac{\partial^2 LD}{\partial a_i \partial a_j} = -\frac{1}{T}\, s\, S_s^{-1} \frac{\partial S_s}{\partial a_j} S_s^{-1} \circ \frac{\partial S_s^{\cdot}}{\partial a_i} + \frac{1}{T}\, s\, S_s^{-1} \circ \frac{\partial^2 S_s^{\cdot}}{\partial a_i \partial a_j}$$

In the iteration formula (24) the factor $\frac{1}{T}$ in (47) and (48) would disappear. The second term in (48) is quite similar to the (32), (39) procedure, again with the difference between the use of $\left(S_s^m\right)^{-1}$ in (48) and $\left(S_s^{m-1}\right)^{-1}$ in (32), (39). As in (47), this latter difference will vanish as we near convergence. But the first term in (48) appears to be more formidable. However we note from (47) that, as we near convergence, $\frac{\partial LD}{\partial a_i} \to 0$, and so $\frac{\partial S_s}{\partial a_i} \to 0$. Hence in (48) both $\frac{\partial S_s}{\partial a_j}$ and $\frac{\partial S_s^{\cdot}}{\partial a_i} \to 0$ with convergence, so that the first term in (48) gradually disappears.

We conclude that the values $\bar{a}$ which bring LD to its lowest minimum do the same for MD. For $LD_a \to MD_a \Big/ T$ and $LD_{aa} \to MD_{aa} \Big/ T$, as $a^m \to \bar{a}$. Hence $\bar{a}(QFML) = \bar{a}(MDSLS)$.

6. CONCLUSIONS

By analogy from the mathematical analysis of the general linear equation, it appears that on theoretical grounds MDSLS is to be preferred to SLS. This result is also confirmed by the finding that maximum likelihood applied to a complete linear model is equivalent to MDSLS, although the optimal properties

of FML are only known for the asymptotic situation.

When iterations are made from small samples, we are still not sure where we stand. To invert $S_s = B_r^{-1} A_r XX' A_r' B_r^{-1'}$ we must have S_s and B_r of rank g, with X of rank at least g, so that we must have $T > g$, and preferably $T > g + k$, in order for MDSLS to proceed. This favours the removal of identities for SLS, where on other grounds it is not required. Also by analogy with the single equation theory, S presumably requires a degrees of freedem (df) correction such as $S = \frac{UU'}{T-p}$, to provide an unbaised estimation of Σ. Again by analogy, p could equal the number of parameters to be estimated in the minimization. There are n parameters in α and $\frac{g(g+1)}{2}$ in Σ. Hence the df in MD could be $T - n - \frac{1}{2}g(g+1)$, while in SLS df could be T-n.

Since sample error dispersion widens as df or effective sample size falls, SLS might have a practical advantage over MD and FML in small sample operations, with given T.

From a computational point of view the procedure outlined above for MD appears to be simpler than previous programs for SLS and FML. Hence SLS could be revised to use (23), (25) and (32) above with $\Sigma_s = I$. For FML we would use the same program as for MD.

Finally the MD procedure provides a convenient format for computing LWV for a full system.

REFERENCES

[1] BROWN, T. MERRITT, "Simplified Full Maximum Likelihood and Comparative Structural Estimates," Econometrica, Vol. 27, 1959, p. 638.

[2] _____, "Simultaneous Least Squares: A Distribution Free Method of Equation System Estimation," International Economic Review, Vol. 1, 1960, p. 173.

[3] _____, "Structure Estimation for Nonlinear Systems of Simultaneous Equations," International Economic Review, Vol. 4, 1963, p. 117.

[4] _____, "Simultaneous Least Squares and Invariance Under Changes of Units of Measurement," International Economic Review, Vol. 8, 1967, p. 97.

[5] CARTER, R. A. L., "A Note on the Computation of Full Information Maximum Likelihood Parameter Estimates," mimeograph, 1968.

[6] DHRYMES, PHOEBUS, J., "Asymptotic Properties of Simultaneous Least Squares Estimators," International Economic Review, Vol. 13, 1972, p. 201.

[7] GOLDBERGER, ARTHUR, S. and INGRAM OLKIN, "A Minimum-Distance Interpretation of Limited Information Estimation," Econometrica, Vol. 39, 1971, p. 635.

[8] HOOD, Wm. C., and TJALLING, C. KOOPMANS, eds., Studies in Econometric Methods, John Wiley and Sons, New York, 1953.

[9] KOOPMANS, TJALLING, C., ed., Statistical Inference in Dynamic Economic Models, John Wiley and Sons, New York, 1950.

[10] MALINVAUD, E., Statistical Methods of Econometrics, (second revised ed.), North-Holland Publishing Co., Amsterdam, 1970.

[11] ROTHENBERG, T. J. and C. T. LEENDERS, "Efficient Estimation of Simultaneous Equation Systems, Econometrica, Vol. 32, 1964, p. 57.

[12] WONNACOTT, RONALD, J., and THOMAS H. WONNACOTT, Econometrics, John Wiley and Sons, New York, 1970.

[13] ZELLNER, ARNOLD, "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias," Journal of the American Statistical Association, Vol. 57, 1962, p. 348.