1973

A General Quasi-Asymptotic Formula for the Sampling Error Covariance Matrix of Econometric Estimators

T. Merritt Brown

Research Report 7310

A GENERAL QUASI-ASYMPTOTIC FORMULA FOR THE SAMPLING ERROR COVARIANCE MATRIX OF ECONOMETRIC ESTIMATORS

by

T. Merritt Brown

A General Quasi-Asymptotic Formula for the Sampling Error Covariance Matrix of Econometric Estimators

by T. Merritt Brown

1. INTRODUCTION

The main part of this paper is the development of a near or seemingly asymptotic formula for the error covariance matrix S(a) for econometric methods which are derived by a maximization (max) or minimization (min) procedure. These methods may have a closed form expression or formula for the actual estimates of the parameter vector a′, but often do not. Methods to be studied for which there is no closed form are Simultaneous Least Squares (SLS), Full Information Maximum Likelihood (FML), and Generalized Minimum Distance Simultaneous Least Squares (MDSLS). In the closed form group we study Ordinary Least Squares (LS) and Two Stage Least Squares (2SLS).

While the formulas herein derived are necessarily asymptotic, they contain more sample related information than is found in the corresponding conventional asymptotic formulas. This additional detail may make the formulas a little more analytic for use with small samples.

The author, with a group of colleagues, is currently engaged in a Monte Carlo research project in which we are testing various econometric methods, including these formulas. Interim results indicate that the formulas are better than fully asymptotic formulas for FML and SLS, but not quite as good for 2SLS.

It currently seems safe to assume that, because of ease of computation and general applicability to max and min methods, that the formulas may be useful for hypothesis testing, as an aid in assessing the qualities of different econometric methods, and for the estimation of S(a) for estimates of the parameters of nonlinear systems.

2. CAUSES OR REASONS FOR SAMPLING ERRORS

In this analysis let us abstract from measurement errors in the data, specification errors, and serious collinearities in the variables in the equations. The observed data provide T observations of g endogenous variables in Y (g x T), and of k exogenous variables in Z (k x T). A complete model of an economic system is represented in basic structural form (SF) as

$$BY + CZ + U = 0; \quad \text{or} \quad AX + U = 0 \qquad (1)$$

where $A = [B,C]$, $X = \begin{bmatrix} Y \\ Z \end{bmatrix}$, and U is a g x T matrix of estimated disturbances or unexplained residuals. The corresponding true or population parameters of the model are $[\beta,\Gamma] = A_p$, with true disturbances $U_p = [\mu_{it}]$ $\left(\begin{smallmatrix} i=1,\dots,g \\ t=1,\dots\dots,T \end{smallmatrix}\right)$.

We make the following stochastic assumptions:

$$E\mu_{it} = 0; \quad E\mu_{it}^2 = \sigma_i^2; \quad E\mu_{it}\mu_{jt} = \sigma_{ij}; \quad E\mu_{it}\mu_{is} = 0 \ (t\neq s=1,\dots,T); \quad E\mu_{it}\mu_{js} = 0;$$

$$E\,\frac{U_pU_p'}{T} = \Sigma = E\mu^t\mu^{t\prime} = [\sigma_{ij}]_{i,j=1,\dots,g} = \operatorname*{Plim}_{T\to\infty} \frac{U_pU_p'}{T} \qquad (2)$$

We use the notation that μ^t is the t^{th} column and μ_i is the i^{th} row of U_p. Thus μ^t is drawn from a g-variate probability distribution with covariance matrix Σ, with corresponding sample estimates u^t, and $S = \frac{UU'}{T}$.

We occasionally find it useful to convert matrices into vectors, and for this we use the operator

$$\text{vec } U = (u_1,u_2,\dots,u_g) = u' = \text{est } \mu' \qquad (3)$$

For matrices like A which may contain many zeros and a few constants we use the operator vec*A = vec A with all zeros and constants deleted, and the remainder closed up so that

$$\text{vec*}A = a' = (a_1,\ldots,a_n) = \text{est } \alpha' \tag{4}$$

To review the causes of sampling error in our estimates, we begin by assuming a population of observed data, PX, from which samples of data SX are drawn. It is from these samples that estimates are made. PX is the sample space of SX, and every sample SX is a subset of PX. Thus PX includes every possible observation X_{it}.

We represent the development of the sample space symbolically in Figure 1 below. In this analysis we make use of the structural reduced form (SRF) of the economic model in (1). This is obtained by solving the model for Y, thereby delineating the determinants of Y.

$$Y = -\beta^{-1}\Gamma Z - \beta^{-1}U_p = Y_{sp} + U_{sp} \; ; \; EY = Y_{sp} \tag{5}$$

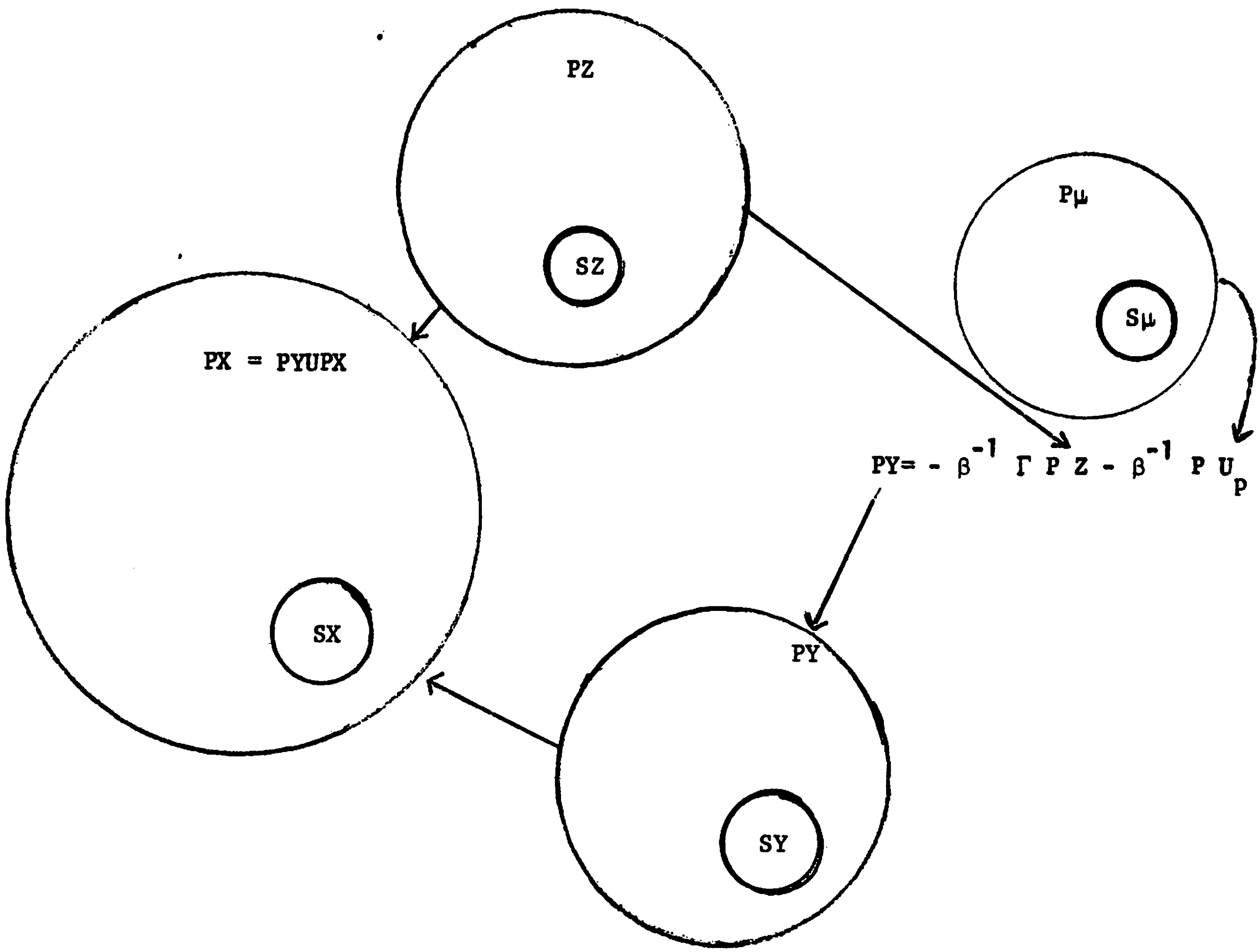
Thus Y_s represents the systematic explanation of Y by the complete model, and U_s represents the total residuals of the complete model in simultaneous solution.

In Figure 1 is shown the formation of PY from PZ, A_p and PU_p, with samples SY derived from SZ, A_p and SU_p. PX is then formed from PZ and PY, while any SX is formed from the related SZ and SY.

We next consider the process of working from the X data, through an estimation method, to the estimated structure. This is represented in Figure 2. We assume that the econometric method used is consistent, so that with the full population PX, and correct specification of structure, our estimating procedure will produce the true structure α and A_p, as well as the related $\sigma' = \text{vec } \Sigma$.

In Figure 2 we indicate that if the whole population or sample space of X is used in the estimation method, the result produced will be the true

Figure 1



structure α, σ. (We only indicate α in the diagram.) Next we indicate the occurrence of three random samples, S_1X, S_2X and S_3X, drawn from PX. S_3X is represented by the dots spread rather uniformly throughout PX. As each of these is fed through the estimation method we get corresponding estimates a^1, s^1; a^2,s^2; and a^3,s^3. We expect these estimates to differ from α,σ since none of the samples or subsets covers the whole population.

Let us represent the sample errors of a^i only by $a^i - \alpha = \delta_i$, $i=1,2,3$. We expect δ_1 to be moderately large, since S_1X though centrally located,

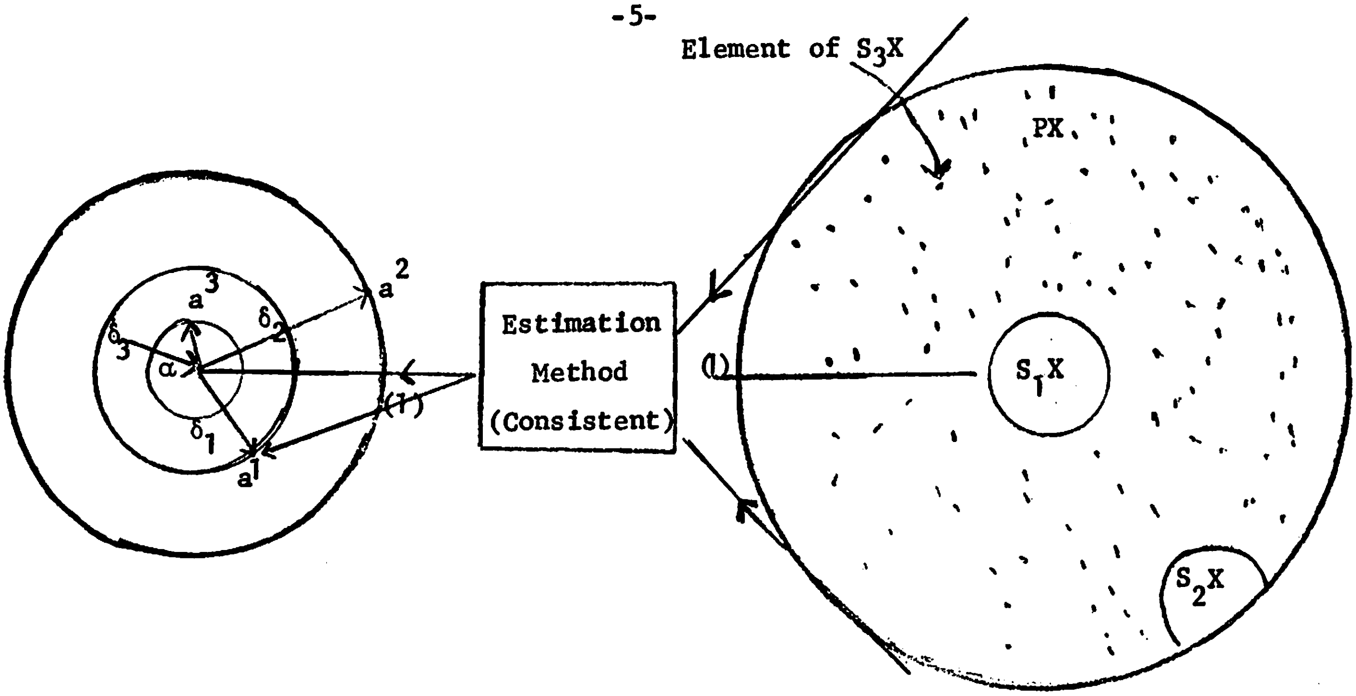


Figure 2

represents a very incomplete coverage of the population. S_2X likewise leaves a very large portion of the population unrepresented, and in addition is located well off center and near the boundary of PX. We expect a very large error δ_2 from it. S_3X on the other hand has points in all regions of PX. Even though it is not a large sample, it is meant (symbolically in Figure 2) to be a sample which is very representative of PX. From it then we may expect an estimate close to α.

The basic assumption which comes out of this intuitive analysis is that the more representative the sample is of the population, the more accurate will be the estimates a' and s'; and conversely, the less representative the sample, the greater the sampling error.

Since most econometric methods for simultaneous equation systems are believed to be biased when using small samples, we shall develop our covariance matrix about expected values E_a rather than true values α. We define a <u>sampling deviation operator</u> d_s in the following:

$$d_s a = a - Ea;\ E\left[d_s a\ d_s a'\right] = \Sigma(a);\ \text{Est}\ \Sigma(a) = S(a) \tag{6}$$

S(a) is thus more properly a <u>sampling deviation</u>, or dispersion matrix, rather than a full sampling error covariance matrix.

The causes of sampling error, which our review above associates with the use of unrepresentative samples, can now be briefly summarized.

<u>Cause 1: Sample Size T</u>. With small T the probability of obtaining a fully representative sample of PX becomes quite low. Indeed when T is very small it may even become impossible to have the elements of SX spaced well enough throughout PX to provide a representative sample.

<u>Cause 2: The Extent of Variation in Z</u>. Referring to Figure 1, if SZ has low variation relative to its potential provided by PZ, then SZ cannot be a representative sample of PZ. We analyze this in relation to estimation formulas as follows. The Z data frequently enter our estimation formulas in the form of raw moments ZZ'. Individual elements of ZZ' are of the form $\sum_{t=1}^{T} Z_{it}^2$ and $\sum_{t=1}^{T} Z_{it}Z_{jt}$, (i,j=1,...,k). The variation in Z_t is defined to be equal to $\sum^{T}(Z_{it} - \bar{Z}_i)^2 = \sum^{T} z_{it}^2$, where $\bar{Z}_i = \frac{1}{T}\sum^{T} Z_{it}$. It follows that $\sum^{T} Z_{it}^2 = \sum^{T} z_{it}^2 + T\,\bar{Z}_i$, so that $\sum^{T} Z_{it}^2$ = variation of Z_i + T (average size of Z_i) = T{(variance of Z_i) + average size of Z_i}. Thus $\sum^{T} Z_{it}^2$ is of the order of sample size T, sample variance of Z_i and average size of Z_i = O(T,V,S). If we let $\frac{1}{T}\sum^{T} Z = (\bar{Z}_1,\ldots,\bar{Z}_k)' = \bar{Z}$, we have the general formula,

$$ZZ' = T\left[\frac{zz'}{T} + \bar{Z}\bar{Z}'\right] = T(CS) \qquad (7)$$

where CS = sum of covariance (C) matrix and size (S) matrix. The larger is ZZ'/T, the larger are C and S, and the greater the probability that SZ is representative of PZ. We could of course have large ZZ'/T with low C and high S, when the sample by chance consists mainly of large elements in Z, and is unrepresentative. But the probability of this is low.

Cause 3: The Dispersion of PUp as Represented by Σ. The higher the "noise" level of the system relative to the (CS) of PZ, the larger the space of PX relative to PZ. With PX enlarged the chances of a given sample SX being representative of PX are reduced.[1] Alternatively one can argue that, for a given SZ fixed in repeated samples, the changes in SU_p for repeated samples can be great, when Σ is large. This will cause SX to vary considerably from sample to sample, and will thereby cause a' to experience wide variability in repeated samples.

To sum up this review, we can observe that when each of the above causes enters our estimation in an unfavourable way--small T; small CS of SZ relative to CS of PZ; and large Σ relative to CS of PZ--we can anticipate reduced probability of a representative sample, and an expansion of the sample space of $\delta = a - \alpha$. These three forces should be separately represented in our subsequent formulas, even though we abstract from bias, and work with $d_s a$ and $\Sigma(a)$. We should find $\Sigma(a)$ an inverse function of T, and of CS of Z, but a direct dunction of Σ. Our formulas should also reflect that with increasing T, accompanied by a presumed increase

[1] I am indebted to Dr. R. J. Wonnacott for this suggestion.

in CS of Z, $\Sigma(a) \rightarrow 0$, irrespective of the relative size of Σ to CS of PZ. For each of the econometric methods studied is consistent.

3. GENERAL FORMULA FOR $\Sigma(a)$ = Cov (a) WHEN a' IS DERIVED BY A MAX OR MIN PROCESS

These estimators include closed form cases, as for LS and 2SLS; and non-closed form cases where the estimates can only be derived by iterating to the max or min of some function f(a,X), as for FML. Suppose that Z can be treated as exogenous and fixed in repeated samples. Then our function can be written as f(a,Y), and both a' and Y will vary from sample to sample.

The estimates are those values of a, say $\bar{a}$ which, for a given sample of data X, cause f(a,Y) to be at its global max or min. At this point

$$\left(\frac{\partial f(a,Y)}{\partial a'}\right)_{\bar{a}} = \left(\frac{\partial f}{\partial a_1} \;----\; \frac{\partial f}{\partial a_n}\right)_{\bar{a}} = f'_{\bar{a}} = 0 \tag{8}$$

Suppose now we think of repeated samples of X, giving us different values of both Y and $\bar{a}$. These have mean values Ea and EY. (Let us henceforth use a as surrogate for $\bar{a}$, our estimates for each sample.) Our sampling deviations for these random variables are

$$d_s a = a - Ea;\; d_s Y = Y - EY;\; \beta EY + \Gamma Z = 0,\; EY = -\beta^{-1}\Gamma Z;\; Y - EY = -\beta^{-1}U_p = U_{sp};$$

$$Y = -\beta^{-1}\Gamma Z - \beta^{-1}U_p = Y_{sp} \text{ (systematic explanation)} + U_{sp} \text{ (random component)} \tag{9}$$

$U_s = -B^{-1}U$ are the residuals of the structural reduced form (SRF) of (1), that is the total residuals of the complete model when solved to simulate the sample of Y. Henceforth we shall drop the subscript s from d, and treat d as the sampling deviation operator.

Let us now apply a Taylor expansion to (8), about Ea and EY, using

$y' = \text{vec } Y = (Y_{11} \text{----} Y_{1T} \text{------} Y_{g1} \text{----} Y_{gT})$.

$$f_a(Ea + dA,\ EY + dY) = f_a(Ea,\ EY) + f_{aa}(Ea,\ EY)da + f_{ay}(Ea, EY)dy + \text{----} = 0 \qquad (10)$$

where

$$f_{aa} = \left[\frac{\partial^2 f}{\partial a \partial a'} \right], \text{ and } f_{ay} = \left[\frac{\partial^2 f}{\partial a \partial y'} \right]$$

We next make the following approximations:

a) Use $a = \bar{a}$ as a substitute for Ea in f_{aa} and f_{ay}.

b) Use Y_s as a substitute for EY in (10).[1]

c) Omit second and higher order terms of the expansion in (10).

With these approximations we can derive a formula for the sampling deviation of a, and hence the error covariance matrix $\Sigma(a)$.

$$\frac{\partial^2 f}{\partial a \partial a'} da + \frac{\partial^2 f}{\partial a \partial y'} dy = 0 \ ; \ da = - f_{aa} f_{ay} \mu_s \qquad (11)$$

Since dada$'$ is a product of six stochastic matrices, we cannot use the ordinary expectations operator on it to derive $\Sigma(a)$. However if we use asymptotic expectations and the Plim operator, an asymptotic formula in five constant matrices emerges.

$$\underset{T \to \infty}{\text{Plim}}\ \bar{E}[dada'] = (\text{Plim } f_{aa})^{-1} (\text{Plim } f_{ay}) \text{Plim} \bar{E}[\mu_s \mu_s'] (\text{Plim } f_{ya}) (\text{Plim } f_{aa})^{-1}$$

---

[1]The LHS term of (10) is zero for all samples. Using our approximations, the first term on RHS of (10) is also found to be zero for each of the methods studied. Experimentation is still needed to discover whether in our final formulas we can use Y as a substitute for Y_s in evaluating f_{aa} and f_{ay} in S(a).

$$\Sigma(a) = \Phi_{\alpha\alpha}^{-1}\, \Phi_{\alpha y}\, \Sigma_s \otimes I_T\, \Phi_{y\alpha}\, \Phi_{\alpha\alpha}^{-1} ;$$

$$S(a) = f_{aa}^{-1}\, f_{ay}\, S_s \otimes I_T\, f_{ya}\, f_{aa}^{-1} = \text{Est } \Sigma(a) \qquad (12)$$

Formula (12) is referred to as quasi-asymptotic since, though clearly asymptotic, it contains more sample information than the fully asymptotic formulas. The omission of higher order terms in the Taylor expansion of (10) could be its Achilles' heel. However the success of (12) in our Monte Carlo research so far is an indication that the errors caused by this omission are not disastrous.

4. INTERPRETATION OF FORMULA (12) FOR S(a)

What does our formula for S(a) tell us about causes which influence the dispersion of our estimates?

S_s. Our theory in Section 2 above gave a prominent role in S(a) to the dispersions of the primary residuals μ, relative to the CS matrix ZZ'/T. Formula (12) brings out this role, but we note that for a complete model it is the total residuals of the model (SRF residuals) which are relevant. Thus it is $S_s = B^{-1}SB^{-1'}$, rather than S, which appears in the formula.

f_{aa}. This matrix is evaluated at the peak or bottom of f, depending on whether the method involves a max or min. The elements of f_{aa} are the rates of change of the slopes of tangents to the surface in the direction of the axes, and hence are an indication of curvature. For a narrow peak at a max the curvature will be high, and the elements of f_{aa} will be negative, and large in absolute value. For a broad humped, nearly flat peak, the absolute values will be low.

The narrow peak will provide us with more precision or resolution in our estimates, with a tendency to small sampling deviations, as indicated in Figure 3.

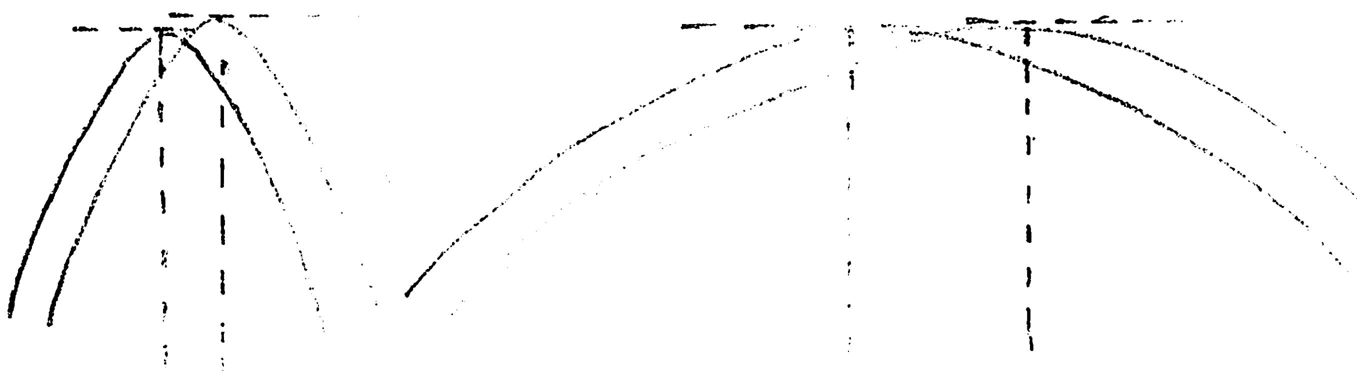


Figure 3

Conversely the broad humped max will give less precision and wider sampling deviations.

In (12) f_{aa} enters twice as an inverse, so that in the narrow peaked system f_{aa}^{-1} will tend to have small values, and hence will contribute to smallness in S(a). The opposite will follow for the broad humped, low curvature max.

On the basis of our earlier theory, we can assume that high curvature in f will be associated with large T, and with large CS matrix ZZ'/T. We can check this as we analyze f_{aa} for specific methods of estimation.

f_{ay}. This matrix of second order partials tells us the rates at which tangent slopes to f = f(a,Y) change from zero at a max or min to some different value as we change the elements of Y to a new sample. This would tend to favour a broad humped max (see Figure 3) to achieve smallness in the effect of f_{ay} on S(a). It thus appears that f_{ay}, at least in part, offsets some of the effect of f_{aa}^{-1}. As we examine specific econometric methods below, we usually find that f_{ay} is O(1) in elements of Y and Z, so that $f_{ay} S_s \otimes I_T f_{ya}$ consists of quadratic and bilinear forms in elements of Y and Z, which are O(T) and O(CS).

Thus f_{ay} seems to play the role of a brake on f_{aa}^{-1}, with two of the latter well overcoming the combined effect of the two f_{ay} matrices. Ultimately f_{ay} is directly associated with $dy = \mu$ in (10), providing an interaction effect between μ and $f_a(a,X)$.

5. APPLICATION OF THE GENERAL COVARIANCE FORMULA TO SPECIFIC ECONOMETRIC METHODS

5.1 S(a) for LS

Here we have a single equation assumed to satisfy Markoff conditions.

$$\underset{T\times 1}{y} = \underset{T\times k}{Z} \ \underset{k\times 1}{a} + \underset{T\times 1}{u} \qquad (13)$$

The function to be minimized is

$$f(a,X) = u'u = y'y - y'Za - a'Z'y + a'Z'Za \qquad (14)$$

$$\frac{\partial F}{\partial a} = f_a = -2Z'y + 2Z'Za \ ; \ \frac{\partial^2 f}{\partial a \partial a'} = f_{aa} = 2Z'Z \qquad (15)$$

$$\frac{\partial^2 f}{\partial a \partial y_1} = -2Z_1' \text{ ----}, \frac{\partial^2 f}{\partial a \partial y_T} = -2Z_T' \ ; \ f_{ay} = -2Z' \qquad (16)$$

Applying (12), we have

$$S(a) = 2^{-1}(Z'Z)^{-1}(-2Z')s^2(u) \otimes I_T(-2Z)2^{-1}(Z'Z)^{-1} = s^2(u)(Z'Z)^{-1} = s^2 \frac{(CS)^{-1}}{T} \qquad (17)$$

Analysis of (12) for LS

In (15)-(17) we find that the curvature matrix f_{aa} is positive, since we are finding a min, and is large if $Z'Z = T(CS)$ is large. The influence of large T and large CS in reducing S(a) is brought in through f_{aa}^{-1}. f_{ay} is not related to T, and is increased by large Z's. The combined effect of $f_{ay} s^2 \otimes I_T f_{ya} = 4\, s^2 Z'Z = s^2 O(T,CS)$. The term $O(T,CS)$ is offset by one of the f_{aa}^{-1} .

5.2 S(a) for 2SLS

Let one of the equations in the complete model represented by (1) be expressed in the form

$$\underset{Tx1}{Y_1} = \underset{Txd}{Y_r}\, b + \underset{Txe}{Z_r}\, c + u_1 \tag{18}$$

where Y_1 and Y_r are endogenous variables included in Y in (1), and Y_r means RHS endogenous variables which are explanatory to Y_1. Z_r are RHS exogenous variables which help explain Y_1, and which are included in Z in (1). There are d+e = n RHS variables, and hence n parameters to be estimated in (18).

The equation which is estimated conceptually in 2SLS as a surrogate for (18) is

$$Y_1 = Y_r^{\bullet} b + Z_r c + u_1^{\bullet} = X_r^{\bullet} a + u_1^{\bullet} \tag{19}$$

where $X_r^{\bullet} = \left[Y_r^{\bullet}, Z_r\right]$; $a' = (b', c')$; and $Y_r^{\bullet} = EY_r$ in the Unrestricted Reduced Form (URF) of (1). The URF of (1) is

$$\underset{Txg}{Y} = \underset{Txk}{Z}\, P + U_u = Z(Z'Z)^{-1} Z'Y + U_u, \text{ with } Y_r^{\bullet} = Z(Z'Z)^{-1} Z'Y_r = HY_r \text{ and}$$

$$u_1^{\bullet} = Y_1 - X_r^{\bullet} a = u_1 + U_{ur} b \tag{20}$$

Note that H is TxT, symmetric and indempotent.

The function to be minimized in 2SLS in order to estimate a is

$$f(a, X^*) = u_1^{\bullet\prime} u_1^{\bullet} = Y_1' Y_1 - Y_1' X_r^{\bullet} a - a' X_r^{\bullet\prime} Y_1 + a' X_r^{\bullet\prime} X_r^{\bullet} a , \tag{21}$$

where $X^* = \left[Y_1, Y_r, Z_r\right]$.

$$f_a = \frac{\partial f}{\partial a} = -\ 2X_r^{\cdot\prime} Y_1 + 2\ X_r^{\cdot\prime} X_r^{\cdot} a \tag{22}$$

$$f_{aa} = \frac{\partial^2 f}{\partial a \partial a'} = 2\ X_r^{\cdot\prime} X_r^{\cdot} = 2\begin{bmatrix} Y_r^{\cdot\prime} Y_r^{\cdot} & , & Y_r^{\cdot\prime} Z_r \\ Z_r' Y_r^{\cdot} & , & Z_r' Z_r \end{bmatrix} = 2\begin{bmatrix} Y_r' H Y_r & , & Y_r' Z_r \\ Z_r' Y_r & , & Z_r' Z_r \end{bmatrix} \tag{23}$$

since $Y_r^{\cdot\prime} Z_r = (Y_r - U_{ur})' Z_r = Y_r' Z_r$.

To calculate f_{ay}, we begin by assembling all endogenous variables in the equation into matrix $Y^* = [Y_1, Y_r]$, with $\text{vec}\ Y^* = y' = (Y_{11}---Y_{1T}, Y_{r11}----Y_{r1T}, ------, Y_{rd1}----Y_{rdT})$.

$$f_a = -2\begin{bmatrix} Y_r' H \\ Z_r' \end{bmatrix} Y_1 + 2\begin{bmatrix} Y_r' H Y_r & , & Y_r' Z_r \\ Z_r' Y_r & , & Z_r' Z_r \end{bmatrix} a = 2\begin{bmatrix} Y_r' H Y_1 & , & Y_r' H Y_r & , & Y_r' Z_r \\ Z_r' Y_1 & , & Z_r' Y_r & , & Z_r' Z_r \end{bmatrix}\begin{pmatrix} -1 \\ b \\ c \end{pmatrix}$$

$$= -\ 2\begin{bmatrix} Y_r' H\ u_1 \\ Z_r' \end{bmatrix} = -\ 2X_r^{\cdot\prime} u_1 \text{ [1]} \tag{24}$$

$$f_{ay} = \underset{nx(n+1)T}{\frac{\partial^2 f}{\partial a \partial y'}} = 2\left[\begin{pmatrix} Y_r' H \\ Z_r' \end{pmatrix}\left(\frac{\partial Y_1}{\partial Y_{1t}}\right),\ \underset{nxd}{0},\ \underset{nxe}{0}\right]\begin{pmatrix} -1 \\ b \\ c \end{pmatrix},$$

$(t=1,\ldots,T)$

$$2\begin{bmatrix} \left(\frac{\partial Y_r'}{\partial Y_{rit}}\right) H Y_1 & , & \left(\frac{\partial Y_r'\ H Y_r}{\partial Y_{rit}}\right) + (\,'' \,)' & , & \frac{\partial Y_r'}{\partial Y_{rit}} Z_r \\ \underset{nx1}{0_{ex1}} & , & \underset{nxd}{Z_r' \frac{\partial Y_r}{\partial Y_{rit}}} & , & \underset{nxe}{0_{exe}} \end{bmatrix}\begin{pmatrix} -1 \\ b \\ c \end{pmatrix} \tag{25}$$

$(i=1---d;t=1-----T)$

---

[1] I am indebted to Dr. R. A. L. Carter for suggesting this final arrangement.

In (25) we introduce the symbolism that (″) is to represent a duplication of the expression or combination of matrices of the immediately preceding brackets of the same form ( ). Similarly for {″} and [″].

According to our general theory in Section 3 leading to (12) we should use $Y_s = \widehat{EY}$ in evaluating (23) and (25). In the asymptotic case however we have Plim $Y^{\bullet}$ = Plim Y_{su} = Plim $\overline{EY}$ = Plim Y_{sp}.[1] If we substitute $Y^{\bullet}$ for Y in (23) and (25) our formulas remain invariant. Monte Carlo tests should however be made to see if the substitution of Y_s for Y in these formulas, in the case of small samples, improves their quality. This is equally true for the methods studied below.

Analysis of (12) for 2SLS

The central matrix of (12) applied to 2SLS is $S_s^* \otimes I_T$, where S_s^* is that portion of S_s for the complete model which corresponds to Y_1 and Y_r, with rows and columns in the order in which these appear in vec $Y^* = y'$ defined above, that is in the order of $Y_1, Y_{r1}, ---, Y_{rd}$.

---

[1] I am indebted to Professor A. L. Nagar for a proof of this surprising result. From the SRF we have

$$Y = -\beta^{-1}\Gamma Z - \beta^{-1}U_p = \Phi Z + U_{sp} = Y_{sp} + U_{sp} \qquad (1)$$

The URF is

$$\underset{g\times T}{Y} = PZ + U_u = YZ'(ZZ')^{-1}Z + U_u = Y_{su} + U_u \qquad (2)$$

Substituting (1) into (2) gives

$$\underset{T\to\infty}{\text{Plim}}\, Y_{su} = \text{Plim}(\Phi Z+U_{sp})Z'(ZZ')^{-1}Z = \Phi Z = \text{Plim}\, Y_{sp} \qquad (3)$$

From (23) we observe that f_{aa} is of $O(T.CS \text{ of } X^{\cdot}_r)$. Hence large T, and large CS of $X^{\cdot}_r$ ultimately caused by Z, will give high curvature to f at min, and a sharp resolution of estimates.

f_{ay} of (25) has the following composition. In the first bracket is column t of $Y^{\cdot\prime}_r$ over column t of $Z^{\prime}_r$, followed by zeros. In the second bracket we have: a column of n zeros, except for $Y^{\cdot}_{1t}$ in the i^{th} position; a dxd matrix of zeros, except for the i^{th} row which is $Y^{\cdot}_{rt}$, + transpose, over an exd matrix of zeros, except for the i^{th} column which is $Z^{\prime}_{rt}$; a dxe matrix of zeros, except for the i^{th} row, which is Z_{rt}, over an exe matrix of zeros. This complete matrix is multiplied by vector $(-1,a')' = p$. f_{ay} thus consists of elements of $X^{*\cdot}$, post multiplied by p. In the composite term $f_{ay} S^*_s \otimes I_T f_{ya}$ we have quadratic and bilinear forms of these elements, weighted by p and S^*_s, and hence the composite term appears to be of $O(T.CS \text{ of } X^{*\cdot})$. It will be offset in terms of order of magnitude O by one of the f^{-1}_{aa}.

We conclude that the total error covariance matrix

$S(a)$ is $O\left(\frac{1}{T}, \frac{1}{CS \text{ of } X^{*\cdot}}, S^*_s\right)$ and $\to 0$ as $T \to \infty$.

Our formula appears to contain more analytic detail relevant to sample variability than does the conventional asymptotic formula for 2SLS, which is

$$S(a) = s^2(u)(X^{\cdot\prime}_r X^{\cdot}_r)^{-1} = 2\ s^2(u) f^{-1}_{aa} \qquad (26)$$

5.3 S(a) for FML

This and the following methods all work with the simultaneous estimations of all parameters of a complete linear model. Any identities in the model are eliminated by substitution into the remaining equations, producing a reduced model

$$B_r \underset{gxT}{Y} + C_r \underset{kxT}{Z} + U = 0\ ;\ A_r X + U = 0 \qquad (27)$$

We now assume that the original model contained g+q endogenous variables, q of which were explained by the identities. The reduced model then contains only g endogenous variables.

FML estimates are derived by finding the values of the unrestricted parameters of the original model ($a' = \text{vec}^*A$) which maximize the likelihood function

$$L(A_r,X) = k + \ln\left|\det B_r\right| - \frac{1}{2}\ln\det S \tag{28}$$

(Cf. Koopmans (1950, p. 213); Brown (1959, p. 640)).

The calculation procedures of Brown (1959) can be further simplified as follows. First we define the operation of <u>Simple Direct Product</u> of two matrices D and E, each mxn, as

$$D\circ E = \left[d_{ij}e_{ij}\right]_{\substack{i=1---m \\ j=1-----n}} = \left[f_{ij}\right] = F = D'\circ E' \tag{29}$$

Then we define the operation of summing all elements of a matrix F with the operator s.

$$sF = \sum_{i,j} f_{ij}, \quad \begin{matrix} i=1---m \\ j=1-----n \end{matrix} \tag{30}$$

Using these operations,

$$dL = \text{tr}\, B_r^{-1\prime} dB_r' - \frac{1}{2}\text{tr}S^{-1}dS = \text{vec}\, B_r^{-1\prime}\, \text{vec}'dB_r - \frac{1}{2}\,\text{vec}\, S^{-1}\text{vec}'dS'$$

$$= s\left[B_r^{-1\prime} \circ dB_r\right] - \frac{1}{2}\, s\left[S^{-1} \circ dS\right] \tag{31}$$

We next define $S^{\bullet} = TS = A_r XX'A_r'$, so that

$$S^{-1} \circ dS = TS^{\bullet -1} \circ d\frac{S^{\bullet}}{T} = S^{\bullet -1} \circ dS^{\bullet} = \text{tr}\, S^{\bullet -1}(dA_r)XX'A_r' + \text{tr}\, S^{\bullet -1}A_r XX'dA_r'$$

$$= 2\text{tr}\, S^{\bullet -1}A_r XX'dA_r' = 2s(S^{\bullet -1}A_r XX' \circ dA_r)\ .$$

$$dL = s\left\{\left[B_r^{-1\prime} I_{g(g+k)} - S^{\cdot -1} A_r XX'\right] \circ dA_r\right\} \quad , \tag{32}$$

where $I_{g(g+k)} = \left[I_g, 0_{gxk}\right]$.

Proceeding directly from (32) we formulate L_a, L_{aa} and L_{ay}, element by element.

$\underline{L_a}$

$$\frac{\partial L}{\partial a_j} = s\left[B_r^{-1\prime} I_{g(g+k)} - S^{\cdot -1} A_r XX'\right] \circ \frac{\partial A_r}{\partial a_j} , \tag{33}$$

where s is now understood to operate on the complete simple direct product which follows it.

$\underline{L_{aa}}$

$$\frac{\partial^2 L}{\partial a_i \partial a_j} = s\left[-B_r^{-1\prime} \frac{\partial B_r'}{\partial a_i} B_r^{-1\prime} I_{g(g+k)} + S^{\cdot -1} \frac{\partial S^{\cdot}}{\partial a_i} S^{\cdot -1} A_r XX' - S^{\cdot -1} \frac{\partial A_r}{\partial a_i} XX'\right] \circ \frac{\partial A_r}{\partial a_j} \tag{34}$$

$$\frac{\partial S^{\cdot}}{\partial a_i} = \left(\frac{\partial A_r}{\partial a_i} XX' A_r'\right) + (\,''\,)' \tag{35}$$

$\underline{L_{ay}}$

Recall that $y' = \text{vec } Y = (Y_{11} --- Y_{1T} ------ Y_{g1} --- Y_{gT})$. The differential of $\frac{\partial L}{\partial a_j}$ for changes in y only is

$$d_y\left(\frac{\partial L}{\partial a_j}\right) = s\left[S^{\cdot -1}(d_y S^{\cdot})S^{\cdot -1} A_r XX' - S^{\cdot -1} A_r d_y XX'\right] \circ \frac{\partial A_r}{\partial a_j} \tag{36}$$

$$d_y S^{\cdot} = A_r (d_y XX') A_r'; \quad d_y XX' = \begin{bmatrix} (dY)Y' + YdY' , & (dY)Z' \\ ZdY' & , 0_{k \times k} \end{bmatrix} \tag{37}$$

$$\frac{\partial XX'}{\partial Y_{it}} = \begin{bmatrix} \frac{\partial Y}{\partial Y_{it}} \cdot Y' + ('')' , & \frac{\partial Y}{\partial Y_{it}} \cdot Z' \\ Z \frac{\partial Y'}{\partial Y_{it}} , & 0_{k \times k} \end{bmatrix} \tag{38}$$

In (38) we note that $\frac{\partial Y}{\partial Y_{it}} \cdot Y'$ is a gxg matrix of zeros, except that its i^{th} row is the t^{th} column of Y. Similarly $\frac{\partial Y}{\partial Y_{it}} \cdot Z'$ is a gxk matrix of zeros, except that its i^{th} row is the t^{th} column of Z.

$$\frac{\partial^2 L}{\partial a_j \partial Y_{it}} = s\left[S^{\cdot -1} A_r \frac{\partial XX'}{\partial Y_{it}} A_r' S^{\cdot -1} A_r XX' - S^{\cdot -1} A_r \frac{\partial XX'}{\partial Y_{it}}\right] \circ \frac{\partial A_r}{\partial a_j} \tag{39}$$

<u>Analysis of (12) for FML</u>

When we examine (34) and (35) we find that $\frac{\partial S^{\cdot}}{\partial a_i}$ is O(T,CS of X), moderated by A_r. Let us assume that $S^{\cdot}$ is small relative to XX′, which implies that A_r is small. Then $\frac{\partial S^{\cdot}}{\partial a_i}$ will tend to be small, $S^{\cdot -1}$ large, $A_r XX'$ small. Thus L_{aa} is O(1) but increases with increase in XX′ relative to $S^{\cdot}$. L is accordingly sharper peaked at max when XX′ is large relative to UU′. It follows that L_{aa}^{-1} is smaller, the larger is XX′ relative to $S^{\cdot}$, that is the variation and size of X relative to variation and size of U.

When we come to analyze L_{ay}, we observe that $\frac{\partial XX'}{\partial Y_{it}}$ is O(X) = O(1), influenced by the size of X. Assuming XX′ large relative to UU′ = $S^{\cdot}$; $A_r \frac{\partial XX'}{\partial Y_{it}}$ is O(1), small. $S^{\cdot}$ = T(CS of U), so that $S^{\cdot -1} = \frac{1}{T}(\text{CS of U})^{-1}$ will be large according to smallness of CS of U, with smallness restored as T increases. Now looking at individual terms in (39), term 1 is $O\left(\frac{1}{T}(\text{CS of U})^{-1}\right)$ x O(1) pre- and post-multiplied by A_r(small) x $O\left(\frac{1}{T}(\text{CS of U})^{-1}\right)$ x A_r O(T(CS of X)) - $O\left(\frac{1}{T}(\text{CS of U})^{-1}\right) A_r$ O(1). The large and small components, based on assuming

XX' large relative to UU', tend to balance out, leaving the whole net result of f_{ay} of $O(\frac{1}{T})$.

The combined effect of $L_{aa}^{-1}L_{ay}$ under our assumptions is for L_{aa}^{-1} to be small and $O(1)$, with L_{ay} of $O(\frac{1}{T})$.

The direct role of Σ or CS of U in S(a) is brought to bear through S_s in the central term of (12).

Finally as T increases, we can expect S(a) to diminish through the joint effects of $\frac{1}{T}$ and a possible slight increase in L_{aa} and hence a slight decrease in L_{aa}^{-1} with samples of X more representative of PX.

The fully asymptotic formula for FML is $S(a) = -\frac{1}{T} L_{aa}^{-1}$. It thus reflects only the influence of sample size, and the relative values of XX' and $S^{\cdot}$. The direct effects of L_{ay} and Σ are ignored, which is quite appropriate for very large samples, but this may reduce sensitivity in the case of small samples.

5.4 S(a) for SLS

The original computational program for this method was presented in Brown (1960). In this method we find the values of a′ which minimize a distance function D. We begin with the structural form (SF) and the SRF for the reduced model, with identities removed by substitution.

$$B_r Y + C_r Z + U = 0 \; ; \; Y = -B_r^{-1} CZ - B_r^{-1} U = FZ + U_s \qquad (40)$$

The principle of SLS is to minimize the Euclidean distance between the vector of observed Y and the vector of the systematic explanation of Y by the complete model. The distance (squared) function is

$$D = \mathrm{tr}\, S_s^{\cdot} = \mathrm{tr}(Y - F_r Z)(Y - F_r Z)' = \mathrm{tr}\, U_s U_s' \qquad (41)$$

The computational procedure appears to be simplified by calculating the individual elements of D_a and D_{aa}, instead of attempting to derive the whole vector and matrix as single formulas.[1] A new and seemingly simpler computational format is now summarized, using this plan.

$\underline{D_a}$ From (41)

$$\frac{\partial D}{\partial a_i} = tr\left[\left\{\frac{\partial F_r}{\partial a_i}(ZZ'F_r' - ZY')\right\} + \left\{''\right\}'\right] = 2\ tr\left(\frac{\partial F_r}{\partial a_i} E\right);\ E = ZZ'F_r' - ZY' \tag{42}$$

$$\frac{\partial F_r}{\partial a_i} = B_r^{-1} \frac{\partial B_r}{\partial a_i} B_r^{-1} C - B_r^{-1} \frac{\partial C_r}{\partial a_i} = - B_r^{-1}\left(\frac{\partial B_r}{\partial a_i} F_r + \frac{\partial C_r}{\partial a_i}\right) \tag{43}$$

The $\frac{\partial B_r}{\partial a_i}$ and $\frac{\partial C_r}{\partial a_i}$ are constant matrices of 1's and zeros.

$\underline{D_{aa}}$

$$\frac{\partial^2 D}{\partial a_i \partial a_j} = 2\ tr\left[\frac{\partial^2 F_r}{\partial a_i \partial a_j} E + \frac{\partial F_r}{\partial a_i}\frac{\partial E}{\partial a_j}\right];\ \frac{\partial E}{\partial a_j} = ZZ'\frac{\partial F_r'}{\partial a_j} \tag{44}$$

$$\frac{\partial^2 D}{\partial a_i \partial a_j} = tr\ \frac{\partial^2 S_s^{\cdot}}{\partial a_i \partial a_j} = 2tr\left[\frac{\partial^2 F_r}{\partial a_i \partial a_j} E + \frac{\partial F_r}{\partial a_i} ZZ' \frac{\partial F_r'}{\partial a_j}\right] \tag{45}$$

From (43) we develop

$$\frac{\partial^2 F_r}{\partial a_i \partial a_j} = - B_r^{-1}\left(\frac{\partial B_r}{\partial a_i}\frac{\partial F_r}{\partial a_j} + \frac{\partial B_r}{\partial a_j}\frac{\partial F_r}{\partial a_i}\right) \tag{46}$$

(See also Carter (1968) and Brown (1973)).

---

[1] The procedure of working with individual elements of D_a and D_{aa} was suggested by Dr. R. A. L. Carter, who found this approach easier to program for the computer.

$\underline{D_{ay}}$

Differentiating (42) with respect to Y_{jt}, we obtain

$$\frac{\partial^2 D}{\partial a_i \partial Y_{jt}} = - \operatorname{tr}\left[\left(\frac{\partial F_r}{\partial a_i} Z \frac{\partial Y'}{\partial Y_{jt}}\right) + (")'\right] = - 2\operatorname{tr}\left(\frac{\partial F_r}{\partial a_i} Z \frac{\partial Y'}{\partial Y_{jt}}\right) \qquad (47)$$

In (47) $Z \frac{\partial Y'}{\partial Y_{jt}}$ is a k x g matrix of zeros, except that the j^{th} column is the t^{th} column of Z.

Analysis of (12) for SLS

The first term of (45) is dominated by $E = - Z U_s'$, which will be small. The second term however is dominated by $ZZ' = T\ O$ (CS of Z). Thus the curvature of the D function at min increases with T and CS of Z, so that D_{aa}^{-1} decreases with increasing T and CS of Z.

From (47) we observe that D_{ay} is dominated by Z. Hence $D_{ay}\ S_s \otimes I_T\ D_{ya}$ will consist of quadratic and bilinear forms of modified rows of Z and will be O(T).

The resulting formula (12) for S(a) gives a central role to the "noise" level of the model, provided by Σ_s; is influenced inversely by CS of Z; and is $O(\frac{1}{T})$.

The above formulation can be compared with the fully asymptotic formula in Dhrymes (1972).

5.5 S(a) for MDSLS

The generalized Minimum Distance Simultaneous Least Squares estimator is developed in Brown (1973). This work was stimulated by discussions with Professor Phoebus Dhrymes when he was working on Dhrymes (1972). I began my approach with Professor Malinvaud's definition (Malinvaud, 1970,

p. 325 ff) of a minimum distance estimator when the parameters enter the estimation problem in a nonlinear form, as they do in SLS. The resulting estimator confirms Dhrymes observation (1972, p. 203) regarding possible conversion of SLS to a full information estimator. In Brown (1973) I also show that it converges to the same estimates as QFML, and makes easy allowance for the handling of identities.

In this method the generalized distance (squared) function to be minimized is

$$MD = \mathrm{tr}\ U_s' \Sigma_s^{-1} U_s = \mathrm{tr}\ \Sigma_s^{-1} U_s U_s' = s\ \Sigma_s^{-1} \circ S_s^{\cdot} \tag{48}$$

Thus, instead of minimizing tr $S_s^{\cdot}$ as in SLS, we are minimizing a weighted sum of $S_s^{\cdot}$, with each element weighted by the corresponding element in the inverse of the covariance matrix of SRF disturbances.

$\underline{MD_{aa}}$

$$\frac{\partial MD}{\partial a_i} = s\ \Sigma_s^{-1} \circ \frac{\partial S_s^{\cdot}}{\partial a_i}\ ;\ \frac{\partial S_s^{\cdot}}{\partial a_i} = \left\{\frac{\partial F_r}{\partial a_i}(ZZ'F_r' - ZY')\right\} + \left\{''\right\}' \tag{49}$$

(Cf. (42) above).

$$\frac{\partial^2 MD}{\partial a_i \partial a_j} = s\ \Sigma_s^{-1} \circ \frac{\partial^2 S_s^{\cdot}}{\partial a_i \partial a_j} = s\ \Sigma_s^{-1} \circ \left[\left(\frac{\partial^2 F_r}{\partial a_i \partial a_j} E + \frac{\partial F_r}{\partial a_i} ZZ' \frac{\partial F_r'}{\partial a_j}\right) + ('')'\right] \tag{50}$$

(Cf. (45) above). Note that in the actual estimation process, Σ_s is estimated from S_s of the previous iteration.

$\underline{MD_{ay}}$

Starting from (49)

$$\frac{\partial^2 MD}{\partial a_i \partial Y_{jt}} = s\ \Sigma_s^{-1} \circ \left[\left(-\frac{\partial F_r}{\partial a_i} Z \frac{\partial Y'}{\partial Y_{jt}}\right) + ('')'\right] \tag{51}$$

Analysis of (12) for MDSLS

As in the case of SLS above, MD_{aa} is dominated by $ZZ' = T\ O(CS\ of\ Z)$. The curvature of the MD function increases with T and with covariance and size of Z, with the converse effects on S(a) via MD_{aa}^{-1}. From (51) we conclude that $MD_{ay}\ S_s \otimes I_T\ MD_{ya}$ will be O(T). The effect of the relative "noise" level in the model on S(a) is provided directly by S_s. The complete formula for (12) is $O(\frac{1}{T})$.

6. CONCLUSIONS

The general formula (12) for the sampling deviation covariance matrix S(a) is asymptotic, since it could only be derived using probability limits. It is an approximation, since it uses only second-order terms of the Taylor expansion of the function being maximized or minimized, and equivalently, only first-order terms of the Taylor expansion of the first-order partials of this function. The success of this approximation hinges partly on the sampling deviations of a′ and Y being small. The approximation seems to be tolerable, since in our Monte Carlo research so far we get reasonable results using (12), with both large and small "noise" models.

Formula (12) for S(a) provides more detail related to sample data and their variability than is found in the fully asymptotic formula for each method studied. Hence it may be closer to the true $\Sigma(a)$ for small samples. Our research to this point indicates that this is so for the full systems methods studied. But for 2SLS the conventional asymptotic formula appears to do slightly better.

We began by analyzing the three main causes behind sampling error in estimation: small T, small CS of Z, large Σ. It was reasoned that a formula for S(a), with some relevance to the small sample case, should include these three causes. Our formula (12) does so for each of the methods studied.

Should our Monte Carlo research continue to confirm that (12) is a useful formula--and this research will be submitted for publication when completed--the formula will then be tested on nonlinear models. Because of its ease of computation it may prove to be a useful tool along with Monte Carlo testing for appraising the relative quality of different estimators for both linear and nonlinear models.

The University of Western Ontario

REFERENCES

BROWN, T. MERRITT, 1959, "Simplified Full Maximum Likelihood and Comparative Structural Estimates," _Econometrica_ 27, 638-653.

________________, 1960, "Simultaneous Least Squares: A Distribution Free Method of Equation System Structure Estimation," _International Economic Review_ 1, 173-191.

________________, 1973, "Conversion of Simultaneous Least Squares to a Malinvaud Minimum Distance Estimator: MDSLS," Research Report 7305, Department of Economics, University of Western Ontario, 1-15.

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

DHRYMES, PHOEBUS, J., 1972, "Asymptotic Properties of Simultaneous Least Squares Estimators," _International Economic Review_ 13, 201-211.

KOOPMANS, TJALLING C., ed., 1950, _Statistical Inference in Dynamic Economic Models_ (John Wiley and Sons, New York).

MALINVAUD, E., 1970, _Statistical Methods of Econometrics_, Second revised ed. (North-Holland Publishing Company, Amsterdam).