COWLES COMMISSION DISCUSSION PAPER: STATISTICS No. 342

Asymptotic Properties of Limited Information Estimates Under Generalized Conditions

by H. Chernoff and H. Rubin

1. Introduction.

A method has been developed for obtaining maximum likelihood estimates of the coefficients of a subsystem of a complete linear system of stochastic equations.[1] This method was derived on the assumption that the disturbances of the system have a joint normal distribution with mean zero and are serially uncorrelated. It was also assumed that al' predetermined variables appearing in the complete system are distributed independently of the disturbances. If a scientist wishes to estimate the coefficients of the subsystem and/or the covariances of the estimates he must either be sure on a priori grounds, or he must verify, that the above assumptions are satisfied for his system. This is sometimes difficult or impossible to do since the assumptions are rather stringent. The question then arises as to how relaxed conditions may be obtained so that the statistician may blithely proceed with the computational method already developed and still obtain "good" estimates. It is to be understood that any relaxation in the conditions is an asset to the scientist. On the other hand he may find himself in the position where he has to pay for his asset by verifying conditions which, while considerably weaker, are more complicated to work with. Indeed, the the conditions to be presented in this paper are already a compromise in that while weaker ones do exist they are much more complex.

2. Review of the Original Case.

Suppose that we have a complete system of stochastic equations

---

1. The subsystem can of course be the complete system or even only one equation of a system.

(1) $$\beta_{qy}\, y_t' + \Gamma_{qz}\, z_t' = q_t'$$

At time t the elements of y_t and z_t are presumed observable variables without error, those of y_t being jointly dependent and those of z_t being predetermined.[1] The elements of q_t are unobservable disturbances which have a joint normal distribution with mean 0 and unknown nonsingular covariance matrix Σ_{qq}. The elements of q_t are distributed independently of those of z_t. The disturbances are serially independent, that is, the disturbances at time t are independently distributed of those at time τ, $t \neq \tau$. Since the system is complete q_t is a vector with as many components as y_t and thus β_{qy} is a square matrix which is assumed to be nonsingular. We decompose q into two subvectors $q = \begin{bmatrix} u & r \end{bmatrix}$. Then (1) may be written

(2a) $$\beta_{uy}\, y_t' + \Gamma_{uz}\, z_t' = u_t'$$

(2b) $$\beta_{ry}\, y_t' + \Gamma_{rz}\, z_t' = r_t'$$

Suppose that we are interested in estimating $\alpha_{ux} = \begin{bmatrix} \beta_{uy} & \Gamma_{uz} \end{bmatrix}$. We assume now that the elements of α_{ux} are subject to certain a priori known restrictions which are sufficient for identification of α_{ux} (excluding possibly exceptional values[2] of the coefficients α_{ux} of the equations to be estimated or of the coefficients α_{rx} of the equations not estimated). The maximum likelihood estimate of α_{ux} is that matrix A_{ux} which subject to the above restrictions minimizes

(3) $$V = \frac{\left| A_{ux} M_{xx} A_{ux}' \right|}{\left| A_{ux} W_{xx} A_{ux}' \right|}$$

where $x = \begin{bmatrix} y & z \end{bmatrix}$ and M_{xx} is the moment matrix of the variables x formed according to

~~(4)~~ (4a) $$M_{ab} = \frac{1}{T} \sum_{t=1}^{T} a_t' b_t$$

---

1. For a definition of predetermined variables see Statistics 310.

2. I.e., values confined to a set of measure zero.

(4b) $$W_{xx} = M_{xx} - M_{xz} M_{zz}^{-1} M_{zx} = \begin{bmatrix} W_{yy} & 0 \\ 0 & 0 \end{bmatrix}$$ [1]

(4c) $$W_{yy} = M_{yy} - M_{yz} M_{zz}^{-1} M_{zy}$$

Indeed $V^{T/2}$ is the reciprocal of the likelihood function except for a constant factor. If α_{ux} is subject to restrictions in the form of equalities, then the number of independent parameters involved in α_{ux} is less than the number of elements of α_{ux}. One may select a set of independent parameters so that α_{ux} may be obtained from it. This set may be called the set of unrestricted parameters. The estimate of the sampling covariance of the estimates of the unrestrcited parameters may also be computed. Consider the matrix L of second order partial derivatives of $\frac{1}{2}\log V$ with respect to these parameters. An estimate of the covariance of the estimates of the unrestricted parameters is given by $\frac{1}{T}(-L)^{-1}$. Furthermore, $A_{ux} M_{xx} A'_{ux}$ is an estimate of the covariance matrix Σ_{uu} of u_t.

Under the assumptions at the beginning of section 2 the estimates are consistent if M_{zz} approaches a nonsingular limit in probability[2]. If one were to calculate the estimates by minimising V in (3) when the assumptions are not valid, these estimates will no longer be maximum-likelihood estimates but will be called quasi-maximum-likelihood estimates. We shall investigate various conditions under which quasi-maximum-likelihood estimates are consistent.

3. Loss of Variables.

We assume that all the conditions of section (2) hold, but that the statistician

---

1. W_{yy} has meaning even if M_{zz}^{-1} does not exist. W_{yy} is the observed covariance of the residuals in the least squares regression of the components of y_t on those of z_t.

2. If $X_1, X_2, \ldots X_T, \ldots$ is a sequence of chance variables we say that $X_T \longrightarrow c$ in probability if $\lim_{T\to\infty} \text{Prob}\left\{ | X_t - c | > \epsilon \right\} = 0$ for every $\epsilon > 0$. When we talk of M_{zz} approaching a nonsingular limit we must note (i) that each element of M_{zz} is a chance variable which depends on T and (ii) that each element of M_{zz} must converge to a corresponding element of a nonsingular matrix.

does not have the time series for several of the predetermined variables which do not appear in part (2a) of the equation system (2). In other words the a priori restrictions state that the coefficients of these missing variables in the subsystem $\beta_{uy} y_t' + \Gamma_{uz} z_t' = u_t'$ are all zero. Then the statistician may act as if these variables do not appear at all in the complete system. He may then compute quasi-maximum-likelihood estimates if the system of equations is still identified on the basis of this false assumption. The quasi-maximum-likelihood estimates of the coefficients, the covariances of the estimates of the coefficients, and of Σ_{uu}, so obtained, are all consistent.

In this case there is a loss of efficiency which restrains the statistician from arbitrarily ignoring all the predetermined variables that he can without losing identification. On the other hand in large systems it is a great computational aid to be able to neglect some of these variables.[1] It may be noted here that if the subsystem contains only one equation which has H y's in it, H > 1, then at least H-1 predetermined variables not in the equation must be used for identification purposes. However, if our subsystem contains more than one equation it may frequently occur that one may not need any predetermined variables which do not occur in the subsystem. Thus it is often possible to compute consistent estimates while very little is known about the rest of the system.

If the statistician has a choice of variables to neglect he should neglect those so that W_{yy} will be made as small as possible. As thus stated the above criterion is rather vague for no method has been mentioned of measuring the size of matrices. Suppose that A_{yy} and B_{yy} are two symmetric positive definite matrices. If $y A_{yy} y' \geq y B_{yy} y'$ for all y then we can say that A_{yy} is at least as large as B_{yy}. In general

1. In problems where there are only a few years of observations it has been noted that the use of many predetermined variables may cause a loss of estimability. Then this procedure becomes very important.

such a comparison does not exist since we frequently have $y A_{yy} y' \geq y B_{yy} y'$ for some y and $y A_{yy} y' \leq y B_{yy} y'$ for others.

Every additional z used does decrease W_{yy} in the sense just indicated. If there is a question of which of two z's to use, and one z gives rise to a smaller W than the other we should use that z. Otherwise one will often find that for one z certain parameters will be more efficiently estimated while other parameters will be less efficiently estimated than if the other z were used. It is well to note that W_{yy} represents the unexplained covariance of y's in the least squares regression of the reduced form.

(5) $$y'_t = \Pi_{yz} z'_t + v'_t.$$ [1]

Capital Π;

Here Π_{yz} represent the matrix of coefficients of the regressions of each element of y on the elements of z. To reduce W_{yy} it is desirable to use those z's which help explain most of the variance of the y's in the above regression.

In connection with this let us consider what happens if there is a jointly dependent variable whose coefficients in the subsystem 2a are all zero. Then it can be shown that the expression (3) is not at all affected by forgetting that this variable is involved in the system. This fact helps to remove some of the difficulty in selecting appropriate z's by indicating certain elements of W_{yy} which are simultaneously unnecessary for computations and not indicative of the efficiency of estimation of the parameters involved.

At this point one can add the remark that the nonsingularity of Σ_{qq} can be relaxed to merely the nonsingularity of Σ_{uu}. Thus it is possible to have identities in the equations which are not in the subsystem (2a).

4. Errors in Observations.

The problem of estimation when all variables are subject to errors in observations is quite complex. There are, however, several cases which may be appropriately treated

---

1. See footnote 1 on page 3.

here.

Suppose, for example, that one were estimating the coefficients of

$$y_1 + \beta_{12} y_2 + \gamma_{11} z_1 = u_1 \qquad (6a)$$ [1]

where the rest of the system consisted of the equation

$$y_2 + \gamma_{22} z_2 + \gamma_{23} z_3 = u_2 \qquad (6b)$$

If there are random errors of observation of z_1, which are normal with zero mean and independent of the predetermined variables, then we may assume that what we observe is $y_3 = z_1 + u_3$ instead of z_1. Substituting in (6a) we have the equation

$$y_1 + \beta_{12} y_2 + \gamma_{11} y_3 = u_1 + \gamma_{11} u_3 = u_2^* \qquad (6c)$$

subscript one

which is identifiable in the system (6b) - (6c), and thus γ_{11} and β_{12} can be estimated consistently, by assuming that our inaccurate observations on z_1 are those of a jointly dependent variable, whereas accurate direct observations on z_1 are not available. This argument applies just so long as the new equation is identifiable without knowledge of z_1. It may also be applied to y_1 and y_2.[2]

Another case of importance is that where the predetermined variables not appearing in the subsystem are subject to errors of observation. As a special case suppose z_2 and z_3 were such variables. The observed variables are $z_4 = z_2 + u_4$, $z_5 = z_3 + u_5$, we also let $u_4 = z_6$, $u_5 = z_7$.

Then

$$y_2 + \gamma_{22} z_4 + \gamma_{23} z_5 - \gamma_{22} z_6 - \gamma_{23} z_7 = u_2 \qquad (6d)$$

can be used for the other equation of the system. z_4 and z_5 are observed predetermined variables so long as the errors of observation are independent of the disturbances u_1, u_2 of the equations. Then z_6 and z_7 are also predetermined even though they are not observed. Since z_6 and z_7 are not required for identification they may

---

1. Subscripts t are omitted when there is no ambiguity.

2. A more complete discussion of shock-error models has been given by T. W. Anderson and L. Hurwicz, see <u>Econometrica</u>, Vol. 16, No. 1, January, 1948, p. 36. Abstract of paper given at Washington Meeting, September 6-18, 1947.

be ignored in consideration of the results of section 3.

The case of errors in jointly dependent variables not appearing in the subsystem can be omitted for in section 3 it was stated that such variables have no effect on the likelihood function anyway.

5. Distribution of Disturbances

The conditions on the disturbances q_t are quite strong and not often satisfied. They require (i) that q_t be distributed independently of the predetermined variables, (ii) that the distribution of q_t is the same for all t, (iii) that the distribution of q_t be normal with zero means and (iv) that q_t be distributed independently of q_τ for $t \neq \tau$. In this section a set of eight conditions will be examined with reference to the system (2). These will then be used to relax the original conditions.

Consider

Condition I:

(7a) $$M_{uz} M_{zz}^{-1} M_{zu} \longrightarrow 0 \qquad \text{in probability}$$

(7b) $$E(z_t' u_t) = 0 \qquad \text{for all } t$$

This condition is weaker than the independence mentioned above. It applies only to u_t, stating that u_t and z_t are uncorrelated and that for large samples z_t would "explain" very little of u_t in a least squares regression.

Condition II:

(8) $$M_{uu} \longrightarrow \Sigma_{uu} \text{ (a constant nonsingular matrix) in probability}$$

This, of course, is a much weaker condition than that of having the distribution of u_t fixed for all t. Here all that is required is that the distribution does not fluctuate too wildly.

Condition III:

(9) $$M_{zz} \longrightarrow M \text{ (a constant nonsingular matrix) in probability}$$

Condition IV:

(10) $$\mu_{xz} \longrightarrow \mu \text{ in probability, where}$$

(11) $$\mu_{xz} = \frac{1}{T}\sum_{t=1}^{T} E(x_t' \mid t-1).\ z_t$$ and

$E(x_t' \mid t-1)$ is defined as the conditional expectation of x_t' given all predetermined variables (observed or unobserved) at ~~t~~ time t. If condition III holds, condition IV is equivalent to

insert time

Condition IV A:

(11a) $$\mu_{yz} \rightarrow \mu_1 \qquad \text{in probability}$$

We derive this implication as follows; $\mu_{xz} = \left[\mu_{yz}\ \mu_{zz}\right]$ and $\mu_{zz} = M_{zz}$ because $E(z_t' \mid t-1) = z_t'$. Considering the complete system we have

prime

$$y_t' = -(\beta_{qy})^{-1}\,\Gamma_{qz}\, z_t' + (\beta_{qy})^{-1} q_t'$$

and thus

$$\mu_{yz} = -(\beta_{qy})^{-1}\,\Gamma_{qz}\, M_{zz} + (\beta_{qy})^{-1}\left\{\frac{1}{T}\sum_{t=1}^{T} E(q_t' \mid t-1) z_t\right\}$$

and the condition becomes

Condition IV B:

(12b) $$\frac{1}{T}\sum_{t=1}^{T} E(q_t' \mid t-1)\, z_t \rightarrow \mu_2 \qquad \text{in probability}$$

Condition IV then, would certainly be satisfied if $E(q_t' \mid t-1)$ were zero and condition III held. It would hold in less restrictive cases and use can be made of any of the three forms depending upon convenience.

Condition V:

(13) $$M_{xz} - \mu_{xz} \rightarrow 0 \qquad \text{in probability}$$

This condition also can be considered in alternative forms

Condition V A:

(13a) $$M_{yz} - \mu_{yz} \rightarrow 0 \qquad \text{in probability}$$

Condition V B:

(13b) $$\frac{1}{T}\sum\left\{q_t' - E(q_t' \mid t-1)\right\} z_t \rightarrow 0 \text{ in probability}$$

Condition IV and V represent rather mild restrictions on the behavior of q_t and z_t.

Condition VI:

The equation $A_{ux} \mu_{xz} = 0$ where A_{ux} is subject to the restrictions on α_{yx} defines A_{ux} as a single valued function of μ_{xz} in a neighborhood of $\mu_{xz} = \mu$. Furthermore, this function is continuous at $\mu_{xz} = \mu$.

Under conditions I, II, III, IV, V, VI, the quasi-maximum-likelihood estimates of α_{ux} and Σ_{uu} are consistent.

Condition VII:

The elements of $\sqrt{T}\ M_{uz}$ are asymptotically normally distributed.

In view of the central limit theorem*, this condition constitutes a relatively weak restriction on u and z.

Condition VIII:

(The covariance of $\sqrt{T}\ M_{u_i z_k}$ with $\sqrt{T}\ M_{u_j z_m}$) - $M_{u_i u_j} M_{z_k z_m} \longrightarrow 0$ in probability.

If conditions I to VII are satisfied, then the quasi-maximum-likelihood estimates of α_{ux} are asymptotically normally distributed. If condition VIII is also satisfied then the quasi-maximum-likelihood estimates of the unrestricted parameters of α_{ux} are asymptotically normally distributed with a covariance matrix which is consistently estimated by $(-\frac{1}{T} L^{-1})$.

It is well to note here that a chance variable R_t may be asymptotically normally distributed with zero mean and unit variance while for any given t the variable R_t has no mean and an infinite variance. One may ask what does the asymptotic distribution mean then. It means that if the probability of R_t lying in a certain interval is desired that probability may be approximated if t is large by using the asymptotic distribution instead of that of R_t. Poor approximations may appear essentially only in problems where one tries to deduce the behavior of the tails of the distribution of R_t from the asymptotic distribution.

6. Non-Linear Equations.

We shall now formally construct a system of equations for which quasi-maximum-likelihood estimates are consistent and then indicate in an example how this method applies to the case of non-linear equations.

We assume the existence of a serially independent set of random variables ... q_{t-1}, q_t ; ... and a set of exogenous variables v,[1] such that

(14a) $$y_t = f_t\,(q_t, q_{t-1}, \ldots, v)$$

(14b) $$z_t = g_t\,(q_{t-1}, q_{t-2}, \ldots, v)$$

(14c) $$u_t = h_t\,(q_t, q_{t-1}, \ldots, v)$$

and

$$\alpha_{ux}\, x_t' = \beta_{uy}\, y_t' + \Gamma_{uz}\, z_t' = u_t'$$

Then if conditions I, II, III, IV, V, VI are satisfied, the quasi-maximum-likelihood estimates are consistent estimates of α_{ux} and Σ_{uu}. If condition VII is also satisfied, these estimates of α_{ux} are asymptotically normal and if VIII is satisfied the covariance matrix of the unestimated parameters are consistently estimated by $\frac{1}{T}(-L)^{-1}$. Of course, in order to carry out the computations it is required that y_t and z_t be observable. On the other hand the exact nature of f_t and g_t need not be known.

As an illustrative example consider the following system

The one is part of the subscript.

(15) $$\begin{aligned} &y_{t1} + \alpha v_{t1}\, y_{t2} + \beta y_{t-1,3} + \gamma = q_{t1}\, y_{t-1,1} \\ &y_{t3} + \delta y_{t1}\, y_{t3} + \epsilon v_{t2} + \zeta = q_{t2}\, v_{t1}\, v_{t2} \\ &f\,(y_{t1}, y_{t2}, y_{t3}, y_{t-1,2}, v_{t1}, q_{t3}, q_{t4}, q_{t5}) = 0 \end{aligned}$$

v_{t1}, v_{t2} are exogenous and q_{t1}, q_{t2}, q_{t3}, q_{t4}, q_{t5} are serially independent.

Suppose that enough is known about the nature of f in (15) to know that these three equations can be solved for y_{t1}, y_{t2} and y_{t3} as functions of the other variables and parameters

---

1. For a definition of exogenous variables see Statistics 310. v may consist of a many dimensional chance variable for e.g. $v = (v_{11}, v_{12}, v_{21}, \ldots, v_{t1}, v_{t2}, \ldots)$

$$y_{ti} = k_{ti}\,(y_{t-1,1}, y_{t-1,2}, y_{t-1,3}, q_{t1}, q_{t2} \dots q_{t5}, v_{t1}, v_{t2}) \qquad i = 1, 2, 3$$

Substituting in this expression $y_{t-1,i} = k_{t-1,i}$ (...) etc.

We have

(16) $$y_{ti} = f_{ti}\,(q_{t1}, q_{t2}, \dots, v) \qquad i = 1, 2, 3$$

where $v = (\dots v_{t-1,1}, v_{t-1,2}\ v_{t,1}\ v_{t,2} \dots)$

Now we let

(17a) $$y_t = (y_{t1}, y_{t2}, y_{t3}, v_{t1}\,y_{t2}, y_{t1}\,y_{t3})$$

(17b) $$z_t = (y_{t-1,1}, y_{t-1,2}\ y_{t-1,3}, v_{t1}, v_{t2}, v_{t1}\,v_{t2}, 1)$$

(17c) $$u_t = (q_{t1}\,y_{t-1,1}, q_{t2}\,v_{t1}\,v_{t2})$$

Then the construction is complete according to the formal method described above, giving for the first two equation of (15) the system

$$\alpha_{ux}\,x_t' = u_t'$$

There is a certain amount of arbitrariness involved in selecting y, f and u. First of all, as was stated in section 3, the presence of y_{t3} in y_t does not affect the estimation procedure. It may therefore be omitted. Then again we may have chosen other elements to be inserted in our z's just so long as 1, $y_{t-1,3}$, v_{t2} were included for they are the only terms appearing in the equations with coefficients not known to be zero.

Finally we see that the two equations of our system have served immediately to indicate the construction and the nature of the restrictions on α_{ux}. The applicability of the estimation method indicated then depends only on whether the conditions can be deemed satisfied to a reasonable degree of approximation.