COWLES FOUNDATION DISCUSSION PAPER NO. 11

The Efficiency of Estimation in Econometric Models

L. R. Klein

April 20, 1956

The Efficiency of Estimation in Econometric Models

Recent arguments in favor of reversion to conventional application of the method of least squares for single equations in simultaneous systems dealt with in econometrics have stressed the comparative efficiency of these methods. For example, in the discussions at the 1954 meetings of the Econometric Society in Uppsala, H. Theil outlined a theorem showing that the generalized variance of least squares estimates of the parameters in a single equation is at least as small as that of limited information maximum likelihood estimates.* It has long been suggested that it may be rewarding to gain some efficiency at the expense of bias by using the least squares method, but the formal proof of the superior efficiency has only just been set out by Theil. Of course, we must know more about the "utility function" of the users of estimated models before we can judge about the relative importance of bias and efficiency.

* "Report of the Uppsala Meeting, August 2-4, 1954" Econometrica, Vol. 23, April, 1955, pp. 204-205.

In this paper, I should like to suggest that Theil's results and, in fact, much of the discussion on the relative merits of different methods is misplaced. In systems of equations, with several parameters in each equation, it is misleading to look at individual parameters, one by one, or even restricted groups of them in reaching overall judgments about the importance of bias or efficiency. Users of econometric models are often not really interested in particular structural parameters by themselves. They are interested in the solution to the system, under alternative sets of conditions. In other, more technical, words they are interested in the reduced forms of the estimated system.** It is the difference between partial and general analysis that is involved. It is conceivable that partial analysis is an end, in itself, for some problems--possibly those of a purely pedagogical nature--but most problems call for a more complete analysis of the system.

** Theil points out to me that for treatment of structural change and appraisal of the a priori "reasonableness" of parameter estimates one would be primarily interested in the individual structural estimates.

The transformation of a structural system to its reduced form can be associated with the process of forecasting. We shall use this term in a general sense in this paper, that is in the sense of making estimates of endogenous economic magnitudes outside the realm of past experience. In this sense, a wide variety of problems of empirical economic analysis are forecasting problems.

Propositions valid for partial systems may not carry over when complete systems are studied. Or even propositions valid for structural parameters may not be valid for reduced form parameters. A beautiful property of the maximum likelihood method of estimation is that its characteristics are preserved under single-valued transformation of variables. Let θ be an unknown parameter, $\hat{\theta}$ its maximum likelihood estimate, and $f(\theta)$ a single-valued transformation function. Then it follows that the maximum likelihood estimate of $f(\theta)$ is given by $f(\hat{\theta})$. As a result of this proposition, the desirable features of maximum likelihood estimates of structural parameters remain as desirable features when the parameters and estimates are transformed into reduced form coefficients. An analogous property does not hold for the method of least squares in general.

Bias--Structural Equations and Reduced Forms

First, let us consider the question of bias. By comparing one-by-one the coefficients of a system estimated by the method of least squares with corresponding coefficients estimated by some unbiased method, investigators sometimes conclude superficially that the amounts of bias are unimportant. From studies of numerical methods of solving linear equation systems with parameters subject to error, we learn, however, that small errors in coefficients may lead to sizeable errors in the final solution.

Suppose that we have a linear equation system

$$By_t + \Gamma z_t = u_t \qquad \text{, y jointly dependent, z predetermined,} \tag{1}$$

with reduced form

$$Y_t = B^{-1}\Gamma z_t + B^{-1} u_t. \tag{2}$$

We shall denote a set of unbiased or consistent estimates as $\hat{B}$ and $\hat{\Gamma}$. Least squares estimates will then be written as

$$\hat{B} + D(B) \quad , \quad \hat{\Gamma} + D(\Gamma).$$

$D(B)$ and $D(\Gamma)$ are discrepancy matrices showing how the least squares estimates differ from the set $\hat{B}$, $\hat{\Gamma}$. For a given z_t-vector we shall then be interested in the effect of the discrepancies on the solution vector.

A first order approximation to the solution of

$$[\hat{B} + D(B)]\,[\hat{y}_t + D(y_t)] \;+\; [\hat{\Gamma} + D(\Gamma)]z_t = 0$$

is given by

$$(3) \qquad D(y_t) = -\hat{B}^{-1}\,[D(\Gamma)\; z_t + D(B)\; \hat{y}_t] \;.$$

Except for the possibility that errors in the Γ-matrix, $D(\Gamma)$, compensate errors in the B-matrix, $D(B)$, these original errors will be reflected in errors in the vector, $D(y_t)$ after multiplication by $\hat{B}^{-1}$. In some problems $\hat{B}^{-1}$ can be a large factor, as will be illustrated below in a simple example.

Consider, for illustrative purposes, the simple multiplier model

$$(4) \qquad C_t = \alpha\, Y_t + u_t \;,$$

$$(5) \qquad Y_t = C_t + I_t \;.$$

C_t = consumption (endogenous)

Y_t = income (endogenous)

I_t = investment (exogenous)

u_t = random disturbance

The unbiased estimate of the marginal propensity to consume is a, and the biased least squares estimate is a + e.* The bias, when looked at from a partial point of view is simply e, but when considered from the more general point of view of the whole system it is $\frac{e}{1 - a - e}$, as can be seen from

* Haavelmo derives an explicit expression for the bias in this model and finds it positive under fairly general conditions. T. Haavelmo. "Methods of Measuring the Marginal Propensity to Consume." Journal of the American Statistical Association, Vol. 42, 1947, pp. 105-122.

$$C_t = (a + e)\,(C_t + I_t)$$

(6) $$C_t = \frac{a + e}{1 - a - e}\; I_t = \left[\frac{a}{1 - a} + \frac{e}{(1 - a)\,(1 - a - e)}\right] I_t$$

$$= \left[a + \frac{e}{1 - a - e}\right] \hat{Y}_t$$

where $$\hat{Y}_t = \frac{1}{1 - a}\; I_t \; .$$

In the reduced form, or multiplier, equation the bias is magnified by the factor $\frac{1}{1 - a - e}$, which will be of the order of magnitude of the multiplier.** Another way of looking at the matter is to observe that the percentage bias is smaller in absolute value in the structural equation than in the reduced form equation.

** Haavelmo, op. cit., shows numerically how a difference of .06 in the marginal propensity to consume becomes a difference of .68 in the multiplier.

$$\left|\frac{e}{a}\right| < \left|\frac{e}{a(1 - a - e)}\right| \; , \quad \text{as long as} \quad |\,1 - a - e\,| < 1.$$

In this simple case, it is clearly seen that a discrepancy in the structural equation, e, is magnified by the multiplier, $\frac{1}{1 - a - e}$. The multiplier plays the role of $\hat{B}^{-1}$ in the more general formulation (3), showing the possibility of comparatively small discrepancies becoming comparatively large in the final result.

Least Squares Efficiency--Structural Equation and Reduced Form

This simple model of the multiplier process is useful in providing an example to show that Theil's proposition about the efficiency of least-squares methods cannot be extended to reduced form parameters. The efficiency properties of least-squares estimation of the parameter in the structural equation, is not preserved under transformation to the reduced form, just as we know that the point values of least squares estimates vary with the direction of minimization of squared residuals.

The variance of the direct least-squares estimate of the marginal propensity to consume is given by

$$\text{var } (a + e) = \text{var } \frac{\Sigma C_t Y_t}{\Sigma Y_t^2} = \text{var } \frac{\Sigma \, (\alpha Y_t + u_t) Y_t}{\Sigma \, Y_t^2}$$

$$= \text{var } (a + \frac{\Sigma u_t Y_t}{\Sigma Y_t^2}) = \text{var } \frac{\Sigma u_t Y_t}{\Sigma Y_t^2} \quad .$$

Applying a first order approximation formula of a function of random variables to the right hand expression we get the classical type result.*

* This is, of course the same result that Theil derives by a different method in his comparison of the efficiency of least squares and limited information estimates. He deals with the estimated variance determined from the sample observations.

(7) $$\text{var } (a + e) = \frac{\text{var } (u)}{\Sigma Y_t^2} \quad ,$$

in the limit as the sample size $\to \infty$. In deriving this result the standard independence assumptions are used about u_t and Y_t for unequal subscript values. Thus for large samples the classical formula is used even though u_t and Y_t are not independent (for equal subscript values).

In forecasting Y_t from estimates of the reduced form equation

$$Y_t = \frac{1}{1-\alpha} I_t + \frac{1}{1-\alpha} u_t ,$$

we are interested in the variance of the estimated multiplier

$$\text{est.} \frac{1}{1-\alpha} .$$

The least-squares estimate of the structural equation leads to an estimate of the multiplier as

$$\ell.\ .s.\ \text{est.} \frac{1}{1-\alpha} = \frac{1}{1-a-e} ,$$

while the corresponding unbiased estimate will be written as

$$\text{m.}\ell.\text{est.} \frac{1}{1-\alpha} = \frac{1}{1-\alpha} \qquad (\text{m.}\ell. = \text{maximum likelihood}).$$

We approximate the variance of the least squares estimate as

$$(8) \qquad \text{var} \frac{1}{1-a-e} = \frac{\text{var}(a+e)}{(1-a-e)^4} = \frac{\left(\frac{1}{1-a-e}\right)^2 \text{var}(u)}{(1-a-e)^2 \Sigma Y_t^2}$$

$$= \frac{\left(\frac{1}{1-a-e}\right)^2 \text{var}(u)\ (\Sigma Y_t^2)^2}{(\Sigma Y_t^2 - \Sigma C_t Y_t)^2 \Sigma Y_t^2} = \frac{\text{var}(u)}{(1-a-e)^2} \ \frac{\Sigma Y_t^2}{(\Sigma I_t Y_t)^2} .$$

The maximum likelihood estimator is simply that multiplier value calculated from the least squares regression of Y_t on I_t ,

$$(9) \qquad \mathrm{var}\left(\frac{1}{1-a}\right) = \frac{\left(\frac{1}{1-\alpha}\right)^2 \mathrm{var}\ (u)}{\Sigma\, I_t^2} .$$

With a positive bias, $e > o$, we have the inequality

$$\frac{\mathrm{var}\ u}{(1-\alpha)^2} \leq \mathrm{plim} \frac{\mathrm{var}\ (u)}{(1-a-e)^2}$$

since $\mathrm{plim}\ a = \alpha$
and we assume $|1 - a - e| < 1$.

$$\text{Also } \frac{1}{\Sigma\, I_t^2} \leq \frac{\Sigma\, Y_t^2}{(\Sigma\, I_t Y_t)^2} ;$$

$$\text{hence} \quad \mathrm{var}\left(\frac{1}{1-a}\right) \leq \mathrm{plim}\ \mathrm{var}\left(\frac{1}{1-a-e}\right) .$$

Thus our example shows clearly that the efficiency properties of least-squares structural estimates do not carry over to the reduced form equations. This example is, of course, special as well as simple. The maximum-likelihood estimate of the reduced form equation is simultaneously a least squares estimate of the reduced form, a full information maximum likelihood estimate of the whole system, and a limited information maximum likelihood estimate of the consumption function. All three of these identical estimates are therefore superior to least squares structural estimates, in a sense, for this model. In the next section, we shall compare efficiency of least squares and full and limited information maximum likelihood estimates of the reduced forms in more general cases, including overidentification.

The Efficiency of A Priori Restrictions

It has frequently been remarked that by imposing *a priori* restrictions on an economic model we gain efficiency of estimation because more information is brought to bear on the problem than is the case in the absence of such restrictions.* It is the purpose of this section to give a formal proof of this point

* See e.g. T. C. Koopmans and Wm. C. Hood, "The Estimation of Simultaneous Linear Economic Relationships." Studies in Econometric Method, ed. by Wm. C. Hood and T. C. Koopmans, (N. Y.: John Wiley and Sons) 1953 esp. p. 176.

of view and to interpret it.

Liu has posed the seemingly paradoxical proposition: Least squares estimates of the reduced form parameters, ignoring all *a priori* restrictions, lead to a higher point on the likelihood function, assuming normally distributed disturbances, than do restricted maximum likelihood estimates; therefore the unrestricted estimates are to be preferred.** Put in another way, he observes

** T. C. Liu, "A Simple Forecasting Model for the U. S. Economy", International Monetary Fund Staff Papers, Vol. IV, August, 1955, pp. 464-65, Liu's argument is not purely in terms of the point reached on the likelihood function with and without restriction. He also argues that structural systems are in general underidentified, leading to the result that we can do no better than to make unrestricted least squares estimates of the reduced forms.

that squared discrepancies between predicted and actual values of endogenous variables in the model will be smaller over the sample period if calculated from (unrestricted) least squares estimates of the reduced form equations than if calculated from any set of estimates of structural parameters using *a priori* restrictions.*** If the least squares values of the reduced forms give better

*** In the case of exact identification, least squares estimates of the reduced form parameters will coincide with fully restricted maximum likelihood estimates.

"predictions" over the sample period, should we not expect them to give better predictions outside the sample?

While the least squares estimates of the reduced forms are *consistent* and while they lead to minimal squared residuals; they do not lead to *efficient* estimates of the reduced form parameters. We shall now turn to the proof of the proposition that the more one uses valid information in the form of *a priori* restrictions imposed on the system, the more efficient are the estimates.

Let us write a linear model as

(1) $$B\, y_t + \Gamma z_t = u_t \quad ,$$

with reduced form

(2) $$y_t = -B^{-1}\Gamma\, z_t + B^{-1}u_t \ .$$

y_t, z_t and u_t are column vectors with n,m, and n components, respectively. B is a square, nonsingular, matrix of order nxn, while Γ is a rectangular of order nxm.

We can also write the reduced form equation as

(10) $$y_t = \pi z_t + v_t \ ,$$

obscuring the relation between its coefficients and those of the structural equations. On multiplying both sides of this equation by B, however, we find

(11) $$B\pi = -\Gamma$$

by equating coefficients of like variables in the reduced form and structural equations. Some elements of Γ are zero; therefore $B\pi$ has the same zero elements. These are the restrictions. Let us write, symbolically,

(12) $$(B\pi)_r = 0$$

to show that r elements of Γ are zero, and that corresponding elements of $B\pi$ are zero.

The logarithm of the likelihood function of the entire system can be written as

(13) $$L = \text{const.} \ - \frac{T}{2} \log \mid \Sigma_v \mid \ - \frac{1}{2} \sum_{t=1}^{T} v_t' \, \Sigma_v^{-1} v_t + \lambda' \ (B\pi)_r \ .$$

Σ_v = matrix of variances and covariance of elements of v_t .
λ is a vector of Lagrange multipliers with r non-zero elements; $t = 1,2,\ldots T$. If there were no restrictions on the system, variances and covariances of est. π sould be given by

$$(14) \qquad \left\| -\frac{\partial^2 L}{\partial\pi_{ij}\partial\pi_{k\ell}} \right\|^{-1} = \left\| \sigma^{ij} M_{zz} \right\|^{-1} = \left\| \sigma_{ij} \; M_{zz}^{-1} \right\| .$$

σ^{ij} = typical element of Σ_v^{-1} .

σ_{ij} = typical element of Σ_v .

M_{zz} = moment matrix of predetermined variables, z_t. The elements of this inverse matrix are the variances and covariances of unrestricted least squares estimates of π. For any particular equation in this complete set of reduced forms, the appropriate variance-covariance matrix is $\sigma_{ii}M^{-1}_{zz}$. To determine the variance of forecast for any endogenous variable, outside the sample values, we have

$$(15) \qquad S_F^2 = \sigma_{ii} \; (1 + z_F' M^{-1}_{zz} z_F).$$

Suppose now that we have r restrictions on the maximization of the likelihood function, then the variances and covariances of est. π are given by the N.W. principal minor, $\hat{A}$, on the right hand side of the following expression

$$(16) \qquad \left\| \begin{array}{c|c} -\dfrac{\partial^2 L}{\partial\pi_{ij}\partial\pi_{k\ell}} & \Delta \\ \hline \Delta' & 0 \end{array} \right\|^{-1} = \left\| \begin{array}{c|c} \hat{A} & \hat{\Delta} \\ \hline \hat{\Delta}' & \Omega \end{array} \right\|$$

The bordering matrix Λ consists of elements of λ obtained as $-\dfrac{\partial^2 L}{\partial\pi_{ij}\partial b_{k\ell}}$

<u>Theorem</u>: Consider an mxm principal minor of $\hat{A}$ corresponding to the i-th reduced form equation. Denote this principal minor by $\hat{A}_{ip}$.

It follows that

$$z_F' \hat{A}_{ip} z_F \leq z_F' M_{zz}^{-1} z_F \quad .$$

<u>Proof</u>: Denote a vector of mn elements by w.

$$\text{(17} \qquad \text{Let} \quad \left\|\begin{array}{c|c} -\dfrac{\partial^2 L}{\partial\pi_{ij}\partial\pi_{k\ell}} & \lambda \\ \hline \lambda' & o \end{array}\right\| = \Delta \ ,$$

where λ is a vector.

Form the difference

$$\delta = \frac{\sum\limits_{i,j=1}^{mn} \Delta oo.ij \ w_i w_j}{\Delta oo} - \frac{\sum\limits_{i,j=1}^{mn} \Delta_{ij} \ w_i w_j}{\Delta}$$

Δoo is formed from Δ by deleting the last row and column.

$\Delta oo.ij$ is formed from Δoo by deleting the i-th row and j-th column.

$$\delta = \frac{\Delta \sum\limits_{i,j=1}^{mn} \Delta oo.ij \ w_i w_j - \Delta oo \sum\limits_{i,j=1}^{mn} \Delta_{ij} \ w_i w_j}{\Delta\Delta oo} \quad .$$

By Jacobi's theorem on determinants* we have

---

* See e.g. A. C. Aitken, <u>Determinants and Matrices</u>. (London: Oliver and Boyd) 1942, pp. 98-99.

$$\Delta \, \Delta_{oo.ij} = \Delta_{oo} \, \Delta_{ij} - \Delta_{oj} \, \Delta_{oi} \qquad ;$$

therefore

$$(18) \qquad \delta = - \frac{\sum_{i,j=1}^{mn} \Delta_{oj} \Delta_{oi} w_i w_j}{\Delta \Delta oo} = \frac{\left(\sum_{i=1}^{mn} \Delta_{oi} w_i\right)^2}{\Delta \Delta oo} \geq 0 \quad .$$

The difference, δ, is positive since Δ is a positive definite quadratic form. Determinants of positive definite quadratic forms obtained by successive rows and columns of bordering alternate in sign; hence

$$- \Delta \Delta_{oo} > 0 \quad .$$

Thus by bordering

$$(19) \qquad \left\| - \frac{\partial^2 L}{\partial \pi_{ij} \partial \pi_{k\ell}} \right\| = \Delta_{oo}$$

with one row and column, we find

$$(20) \qquad w' (\Delta^{-1})_{oo} w \leq w' (\Delta oo)^{-1} w \quad .$$

By bordering Δ with a row and column to form $({}_{11}\Delta)$, we similarly find

$$(21) \qquad w' [({}_{11}\Delta)^{-1}]_{oo.oo} w \leq w' (\Delta^{-1})_{oo} w \quad ,$$

and so on for successive borderings. By letting all elements of w vanish except the m elements in z_F corresponding to the mxm principal minor of $\hat{A}$ associated with i-th reduced form equation, we have

$$(22) \qquad z_F' \hat{A}_{ip} z_F \leq z_F' M_{zz}^{-1} z_F \quad ,$$

as was to be proved.

The total variance of forecast from a system of structural equations is composed of two factors, the variance of disturbances σ_{ii} and a quadratic form (plus unity) in the assumed values of the predetermined variables. The matrix of this quadratic form when multiplied by σ_{ii} is the variance-covariance matrix of reduced form parameter estimates. By adding information to the system in the form of restrictions on the parameters, we decrease the magnitude of this quadratic form. The estimated value of forecast error will reflect differences in estimation procedure, the calculated variance of residuals being lowest when no restrictions are used. Nevertheless, the underlying or inherent error using the true value of σ_{ii} in the formula for prediction error will be smaller, the more one employs a priori restrictions in calculating the structural characteristics.

The succession of inequalities derived shows that full information maximum likelihood estimates will be more efficient than limited information maximum likelihood estimates since more restrictions are used with the former than with the latter set. Both of these methods, in turn, will be more efficient than unrestricted least squares estimates of the reduced forms.

In connection with limited information estimates a word of explanation is in order. When any single equation in a system is being estimated, this method will produce a set of reduced form estimates, separately involving each of the endogenous variables in that equation. In so far as a single endogenous variable appears in several different structural equations, there will be several possible reduced form equations that could serve as alternative forecasting equations. The best among these will be that set which yields the smallest quadratic form

$$z_F' \hat{A}_{ip} z_F$$

This will be inferior to the full information maximum likelihood estimates. Another way of using limited information estimates in practical forecasting, is by solving algebraically for reduced form equations in a structural system, each equation of which has been estimated by the method of limited information.*

---

* L. R. Klein and A. S. Goldberger, An Econometric Model of the United States, 1929-1952. (Amsterdam: North-Holland Publishing Co.) 1955.

The foregoing results on the efficiency of the use of a priori information follow closely a development by Samuelson called "the generalized Le Chatelier principle."* Samuelson shows that the sensitivity of an economic variable to

---

* P. A. Samuelson, Foundations of Economic Analysis (Cambridge: Harvard University Press) 1947, pp. 36-39.

---

a parameter change (price elasticity of demand e.g.) decreases as the number of restraints imposed upon the system increases. He proves this proposition by bordering a matrix of second derivatives of a function being maximized. In a sense, this paper extends his result from inequalities on diagonal elements of inversed bordered matrices to quadratic forms of principal minors.**

** Samuelson's demonstration appears to contain some minor compensating errors, but the final results are quite correct.

If we were interested solely in variances of individual coefficients in the reduced forms, his result would be directly applicable. Since the forecast error involves an entire quadratic form, his proposition must be extended. It may also be remarked that Samuelson's proposition is not a perfect analogue for ours. His bordered matrices are those familiar in establishing conditions for extremes, while ours would have to be bordered further to take that form.