COWLES COMMISSION DISCUSSION PAPER: STATISTICS NO. 330

April 4, 1949

EXTENT OF LEAST SQUARES BIAS IN ESTIMATING A SINGLE STOCHASTIC EQUATION IN A COMPLETE SYSTEM

By

Jean Bronfenbrenner

Notation

Let

(1) $$\beta y^{*\prime} + \gamma z^{*\prime} = u_1$$

be a single equation (say the first) in a complete system of G stochastic difference equations

(2) $$\beta y' + \Gamma z' = u'.$$

As usual y $[= (y^* \; y^{**})]$ is the vector of endogenous variables, z $[=(z^* z^{**})]$ is the vector of exogenous variables (which we assume to be random variables depending only on time), and u is the vector of serially uncorrelated residuals with expected value zero. Let $\Sigma \; [= (\sigma_{ij})]$ be the covariance matrix of the residuals and let it be independent of time.

The right subscript i, attached to a vector, will stand for the i^{th} element of the vector. The left subscript i will indicate the ommission of the i^{th} element of the vector. The right subscript $a_1 \; a_2 \; \ldots \; a_n$ attached to a matrix will stand for rows $a_1, a_2, \ldots, a_n$ of the matrix.

If H is the number of elements in $y^{*\prime}$, let the first H rows of the reduced form of (2) be given by

$$y^{*\prime} = (\Pi^* \; \Pi^{**}) \begin{pmatrix} z^{*\prime} \\ z^{**\prime} \end{pmatrix} + v^{*\prime}$$

$$= \Pi z' + v^{*\prime} = -(\beta^{-1})_{12 \ldots H} \Gamma z' + (\beta^{-1})_{12 \ldots H} u'$$

Let the covariance matrix of ${}_1v^*$ be $\Omega = (\beta^{-1})_{2 \ldots H} \; \Sigma \; [(\beta^{-1})_{2 \ldots H}]'$.

Let $M_{pq} \left[= \frac{1}{T} \sum_t p'(t)\, q(t) \right]$ be the moment matrix (in the sample) of the variable vectors p and q, and let $\mathcal{M}_{pq}$ be the probability limit of this matrix.

Let the vector z^o be the residual of z^{**} from the regression of z^{**} on z^*

$$z^{o\prime} = z^{**\prime} - \mathcal{M}_{z^{**} z^*} \mathcal{M}^{-1}_{z^* z^*} z^{*\prime}.$$

Derivation of the Expression for Least Squares Bias

If the sample least squares regression of y_1 on ${}_1y^*$ and z^* is

$$y_1 + k_{y^*}\, {}_1y^{*\prime} + k_{z^*}\, z^{*\prime} = 0$$

then k_{y^*} and k_{z^*} are defined by

$$\begin{pmatrix} k_{y^*} \\ k_{z^*}' \end{pmatrix} = - \begin{pmatrix} M_{{}_1y^*\,{}_1y^*} & M_{{}_1y^* z^*} \\ M_{z^*\,{}_1y^*} & M_{z^* z^*} \end{pmatrix}^{-1} \begin{pmatrix} M_{{}_1y^* y_1} \\ M_{z^* y_1} \end{pmatrix}$$

From (1), under the normalization $\beta_1 = 1$, we obtain the following expression for the value of y_1 at time t

$$y_1(t) = -\, {}_1y^*(t)\, {}_1\beta' - z^*(t)\, \gamma' + u_1.$$

Thus for a sample of size T

$$\begin{pmatrix} M_{{}_1y^* y_1} \\ M_{z^* y_1} \end{pmatrix} = \frac{1}{T} \sum_{t=1}^{T} \begin{pmatrix} {}_1y^{*\prime}(t) \\ z^{*\prime}(t) \end{pmatrix} \left[-\, {}_1y^*(t)\, {}_1\beta' - z^*(t)\, \gamma' + u_1(t) \right]$$

$$= - \begin{pmatrix} M_{{}_1y^*\,{}_1y^*} \\ M_{z^*\,{}_1y^*} \end{pmatrix} {}_1\beta' - \begin{pmatrix} M_{{}_1y^* z^*} \\ M_{z^* z^*} \end{pmatrix} \gamma' + \begin{pmatrix} M_{{}_1y^* u_1} \\ M_{z^* u_1} \end{pmatrix}$$

$$= -\begin{pmatrix} M_{{}_1y^*\,{}_1y^*} & M_{{}_1y^*z^*} \\ M_{z^*\,{}_1y^*} & M_{z^*z^*} \end{pmatrix} \begin{pmatrix} ({}_1\beta' \\ (\gamma' \end{pmatrix} + \begin{pmatrix} M_{{}_1y^*\,u_1} \\ M_{z^*\,u_1} \end{pmatrix}$$

Substituting in (3) we have

$$(4) \quad \begin{pmatrix} k_{y^*}' \\ k_{z^*}' \end{pmatrix} = \begin{pmatrix} {}_1\beta' \\ \gamma' \end{pmatrix} - \begin{pmatrix} M_{{}_1y^*\,{}_1y^*} & M_{{}_1y^*\,z^*} \\ M_{z^*\,{}_1y^*} & M_{z^*z^*} \end{pmatrix}^{-1} \begin{pmatrix} M_{{}_1y^*\,u_1} \\ M_{z^*\,u_1} \end{pmatrix}$$

To obtain the probability limit of the least squares bias we now wish to find the probability limits χ_{y^*} and χ_{z^*} of k_{y^*} and k_{z^*}. This requires the assumption that M_{zz} has a finite probability limit μ_{zz}.

The value of ${}_1y^{*\prime}$ at time t is given by

$${}_1y^{*\prime}(t) = \Pi_{2 \,\ldots\, H}\, z'(t) + {}_1v^{*\prime}.$$

Since z is independent of u and hence of v and since the probability limit of a rational function of random variables is the same rational function of the probability limits of the variables, we have $\mu_{zu} = \mu_{zv} = 0$, and we may write the probability limits of the moment matrices in (4) as follows, in terms of submatrices of μ_{zz}

$$\mu_{{}_1y^*\,{}_1y^*} = \Pi_{2 \,\ldots\, H}\, \mu_{zz}\, (\Pi_{2 \,\ldots\, H})' + \Omega$$

$$\mu_{{}_1y^*\,z^*} = \Pi_{2 \,\ldots\, H}\, \mu_{zz^*}$$

$$\mu_{{}_1y^*\,u_1} = (\beta^{-1})_{2 \,\ldots\, H}\, (\Sigma_1)'$$

$$\mu_{z^*\,u_1} = 0$$

Making use of a well known method for the inversion of a partitioned matrix we obtain

$$(5) \quad \chi'_{y^*} = {}_1\beta' - \psi^{-1}(\beta^{-1})_{2\ldots H}(\Sigma_1)'$$

$$(6) \quad \chi'_{z^*} = \gamma' + \mu^{-1}_{z^*z^*}\mu_{z^*z}(\pi_{2\ldots H})'\psi^{-1}(\beta^{-1})_{2\ldots H}(\Sigma_1)'$$

where

$$\psi = \pi_{2\ldots H}\mu_{zz}(\pi_{2\ldots H})' + \Omega - \pi_{2\ldots H}\mu_{zz^*}\mu^{-1}_{z^*z^*}\mu_{z^*z}(\pi_{2\ldots H})'$$

$$= \pi_{2\ldots H}(\mu_{zz} - \mu_{zz^*}\mu^{-1}_{z^*z^*}\mu_{z^*z})(\pi_{2\ldots H})' + \Omega$$

$$= \pi^{**}_{2\ldots H}\mu_{z^oz^o}(\pi^{**}_{2\ldots H})' + \Omega$$

Implications and Consequences

From (5) and (6) we observe that the size of the bias depends on the entire β matrix as well as on the Σ matrix and that part of the variance of z^{**} which is not explained by z^*. This makes it impossible to use these expressions to estimate in advance of calculation how large the least squares bias is likely to be. However, it is still possible to say something about the conditions under which the bias will be small.

If σ_1, the standard deviation of u_1, is zero, (i.e. $u_1 \equiv 0$ and the equation (1) is exact), then it follows immediately that the bias is zero since Σ_1 will necessarily vanish. This result is, of course, intuitively obvious, since if all observations lay on a given hyperplane it would be impossible to obtain a different hyperplane by regression procedures, whichever the direction in which the sum square residual were minimized.

It is interesting to note that the bias does not depend directly on the moments of the z's (or of the y's) but simply on certain functions of these moments namely the moments of z^o. This suggests that given a sample from which we wish to estimate (1) we may learn less about the size of least squares bias by observing such ordinary criteria as the multiple correlation coefficient or the standard errors of the regression coefficients than by looking directly at the variance of the residuals of the z^{**}'s from their regression on the z^{*}'s.

It is also interesting that the matrix ψ enters the expression as an inverse. ψ is that part of the variance and covariance of ${}_1y^*$ which is not explained by a linear regression on z^*. The appearance of ψ^{-1} may perhaps be rationalized as follows. The more variation in ${}_1y^*$ depends on z^*, which also strongly effects y_1, the more y_1 and ${}_1y^*$ are constrained to move together. On the other hand the more variation in ${}_1y^*$ depends on factors orthogonal to z^*, the more capable ${}_1y^*$ becomes of the "free variation", in Haavelmo's words, which is required under least squares assumptions. Thus it is conducive to small bias when the variance in ${}_1y^*$ unexplained by z^* is large.