COWLES COMMISSION DISCUSSION PAPER: STATISTICS: NO. 343

December 5, 1949

Example of Loss of Efficiency in Structural Estimation

by S. G. Allen

1. Unavailability of statistical data and the high cost in treating large and complex econometric models frequently forces the investigator to sacrifice efficiency in the estimates of structural parameters. In the following, sampling variances are derived for an estimate of a structural parameter under two estimation procedures. The model chosen is trivially simple not only to reduce tedious derivation but to produce results from which some straightforward conclusions are possible.

2. Let

$$(2.1)\qquad y_1(t) + \beta y_2(t) = u_1(t)$$

$$(2.2)\qquad y_2(t) + \gamma_1 z_1(t) + \gamma_2 z_2(t) = u_2(t)$$

be a complete system of linear stochastic equations where

(2.3) $u_1(t)$ and $u_2(t)$ have a nonsingular bivariate normal distribution independent of z_1, z_2, and t with $E\, u_1(t) = E\, u_2(t) = 0$.

Each variable, for simplicity, is measured in deviations from its mean. The z's are exogenous - given functions of time. All expectations in the sequel are conditional for given z.

Variances of the estimate of β will be found assuming:

(2.a) that observations on z_1 and z_2 are used in the estimation of β (call this estimate $\hat{\beta}$).

(2.b) that observations only on z_1 are used (call this estimate β^*).

Under (2.a) the estimate of β using full maximum likelihood and limited information procedures are equivalent. $\hat{\beta}$ in this case is the optimal estimate.

Under (2.b), the likelihood function which generates the observations on the y's is not maximized. In effect the estimation of β proceeds as if (2.2), which is assumed to be the true equation, is replaced by

$$(2.2') \quad y_2(t) + \gamma_1 z_1(t) = u_2^*(t)$$

and as if (2.3) held with respect to $u_1(t)$ and $u_2^*(t)$. Obviously $u_2^*(t) = u_2(t) - \gamma_2 z_2(t)$ is not independent of $z_2(t)$ nor, since the latter is a given function of time by definition, of t.

3. The likelihood function for a sample of size T is

$$(3.1) \quad F\left(\underline{y}(t) \big/ \underline{z}(t)\right) = (2\pi)^{-\frac{T}{2}} \; \Sigma^{-\frac{T}{2}} \exp \operatorname{tr} - \frac{T}{2} \left\{ \alpha \, M \, \alpha' \, \Sigma^{-1} \right\}$$

where

$$\Sigma = (\sigma_{ij}) = E\left\{ u'(t)\, u(t) \right\}, \quad u(t) = \left\{ u_1(t) \; u_2(t) \right\} \qquad i,j = 1, 2$$

$$\alpha = \begin{pmatrix} 1 & \beta & 0 & 0 \\ 0 & 1 & \gamma_1 & \gamma_2 \end{pmatrix}$$

and

$$M = (m_{ij}) = \frac{1}{T} \sum_{t=1}^{T} \begin{pmatrix} y_1^2(t) & y_1(t)\, y_2(t) & y_1(t)\, z_1(t) & y_1(t)\, z_2(t) \\ & y_2^2(t) & y_2(t)\, z_1(t) & y_2(t)\, z_2(t) \\ & & z_1^2(t) & z_1(t)\, z_2(t) \\ & & & z_2^2(t) \end{pmatrix}$$

$$i,j = y_1, y_2, z_1, z_2$$

The six second partial derivatives of $\frac{1}{T}$ log F with respect to the unknown coefficients of (2.1) and (2.2) are

$$(3.2) \quad \frac{\partial^2 \frac{1}{T} \log F}{\partial (\beta, \gamma_1, \gamma_2)^2} = - \begin{pmatrix} \sigma^{11} m_{y_2 y_2} & \sigma^{12} m_{y_2 z_1} & \sigma^{12} m_{y_2 z_2} \\ & \sigma^{22} m_{z_1 z} & \sigma^{22} m_{z_1 z_2} \\ & & \sigma^{22} m_{z_2 z_2} \end{pmatrix}$$

Next (2.1) and (2.2) are employed to transform the elements of (3.2) into moments involving only u's and z's and then the expectations of the latter are found. For more concise notation, let

$$\Gamma = (\gamma_1 \ \gamma_2)$$

$$M_{zz} = \begin{pmatrix} M_{z_1} \\ M_{z_2} \end{pmatrix} = \begin{pmatrix} m_{z_1 z_1} & m_{z_1 z_2} \\ & m_{z_2 z_2} \end{pmatrix}$$

Then

$$(3.3) \quad E \ \frac{\partial^2 \frac{1}{T} \log F}{\partial(\beta, \gamma_1, \gamma_2)^2} = G$$

$$= (g_{ij}) = -\begin{pmatrix} \sigma^{11}(\sigma_{22} + \Gamma M_{zz} \Gamma') & -\sigma^{12} \Gamma M'_{z_1} & -\sigma^{12} \Gamma M'_{z_2} \\ & \sigma^{22} m_{z_1 z_1} & \sigma^{22} m_{z_1 z_2} \\ & & \sigma^{22} m_{z_2 z_2} \end{pmatrix}$$

$$i, j = \beta, \gamma_1, \gamma_2$$

By the theorem of Kendall[1] for sampling variances of maximum likelihood estimators,

$$-T \sigma_{\hat{\beta}\hat{\beta}} = \frac{g_{\beta\beta}}{|G|_{22}}$$

$$= -\frac{\sigma^{22} \ |M_{zz}|}{|M_{zz}| \ (\Gamma M_{zz} \Gamma' \ |\Sigma|^{-1} + \sigma^{11} \sigma^{22} \sigma_{22})}$$

or

$$(3.4) \quad \sigma_{\hat{\beta}\hat{\beta}} = \frac{1}{T} \left(\frac{\sigma_{11}}{\Gamma M_{zz} \Gamma' + \frac{\sigma_{22}}{1-\rho^2}} \right)$$

where ρ is the coefficient of correlation between u_1 and u_2.

4. β^* maximizes the function

$$(4.1) \quad F^*(y(t) \mid z_1(t)) =$$

$$(2\pi)^{-\frac{T}{2}} |\Sigma|^{-\frac{T}{2}} \exp \operatorname{tr} -\frac{T}{2} \begin{pmatrix} 1 & \beta & 0 \\ 0 & 1 & \gamma_1 \end{pmatrix} \begin{pmatrix} m_{y_1 y_1} & m_{y_1 y_2} & m_{y_1 z_1} \\ & m_{y_2 y_2} & m_{y_2 z_1} \\ & & m_{z_1 z_1} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \beta & 1 \\ 0 & \gamma_1 \end{pmatrix}$$

1. Kendall, Advanced Theory of Statistics, Vol. II, p. 36, 37, and 41.

which is not the likelihood function of the sample. Thus Kendall's formula is no longer appropriate to obtain $\sigma_{\beta^* \beta^*}$.

The estimate in this case is

$$\beta^* = -\frac{m_{y_1 z_1}}{m_{y_2 z_1}} .$$

β^*, being the ratio of jointly normally distributed variables, meets the requirements of a theorem of Gurland[2] which can be used to show that β^* is asymptotically normally distributed with mean

$$E\,\beta^* = -\frac{E\, m_{y_1 z_1}}{E\, m_{y_2 z_1}} = \beta$$

and variance

$$(4.2)\quad \sigma_{\beta^* \beta^*} = \left(\frac{\partial \beta^*}{\partial m_{y_1 z_1}} \;\; \frac{\partial \beta^*}{\partial m_{y_2 z_1}} \right) \Lambda \begin{pmatrix} \dfrac{\partial \beta^*}{\partial m_{y_1 z_1}} \\ \dfrac{\partial \beta^*}{\partial m_{y_2 z_1}} \end{pmatrix} \Bigg|_{\substack{m_{y_1 z_1} = E(m_{y_1 z_1}) \\ m_{y_2 z_1} = E(m_{y_2 z_1})}}$$

where Λ is the covariance matrix of $m_{y_1 z_1}$, $m_{y_2 z_1}$. Again (2.1) and (2.2) are employed to find $m_{y_1 z_1}$ and $m_{y_2 z_1}$ in terms of moments of u's and z's in accordance with the true structure. The expectations of the sample statistics are

$$E\, m_{y_1 z_1} = \beta \Gamma M'_{z_1}$$
$$E\, m_{y_2 z_1} = -\Gamma M_{z_1}$$
$$\Lambda = \frac{m_{z_1 z_1}}{T} \begin{pmatrix} (1 \;\; -\beta)\, \Sigma \begin{pmatrix} 1 \\ -\beta \end{pmatrix} & \sigma_{12} - \beta \sigma_{22} \\ & \sigma_{22} \end{pmatrix}$$

---

2. John F. Gurland: Best Asymptotically Normal Estimates - Partial Results, Technical Report 1., ONR No. N 7onr - 2951 XIII, 1948.

The partials required for (4.2) are

$$\left.\frac{\partial \beta^*}{\partial m_{y_1 z_1}}\right|_{\beta^*=\beta} = \frac{1}{\Gamma M'_{z_1}}$$

$$\left.\frac{\partial \beta^*}{\partial m_{y_2 z_1}}\right|_{\beta^*=\beta} = \frac{\beta}{\Gamma M'_{z_1}}, \quad \text{evaluating these at } \beta^* = E\beta^* = \beta.$$

Thus the variance expression is

$$(4.3) \quad \sigma_{\beta^*\beta^*} = \begin{pmatrix} \frac{1}{\Gamma M'_{z_1}} & \frac{\beta}{\Gamma M'_{z_1}} \end{pmatrix} \wedge \begin{pmatrix} \frac{1}{\Gamma M'_{z_1}} \\ \frac{\beta}{\Gamma M'_{z_1}} \end{pmatrix} = \frac{m_{z_1 z_1}}{T} \frac{\sigma_{11}}{(\Gamma M'_{z_1})^2}$$

$$= \frac{\sigma_{11}}{T\,\Gamma\,\bar{M}_{z_1 z_1}\,\Gamma'}, \qquad \bar{M}_{z_1 z_1} = \frac{M'_{z_1} M_{z_1}}{m_{z_1 z_1}}$$

5. $\hat{\beta}$ is readily seen to be the more efficient estimate of β, i.e.,

$$(5.1) \quad \frac{\sigma_{\beta^*\beta^*}}{\sigma_{\hat{\beta}\hat{\beta}}} = \frac{\Gamma M_{zz} \Gamma' + \frac{\sigma_{22}}{1-\rho^2}}{\Gamma \bar{M}_{z_1 z_1} \Gamma'} > 1$$

since

$$(5.2) \quad \Gamma M_{zz} \Gamma' + \frac{\sigma_{22}}{1-\rho^2} - \Gamma \bar{M}_{z_1 z_1} \Gamma'$$

$$= \Gamma (M_{zz} - \bar{M}_{z_1 z_1}) \Gamma' + \frac{\sigma_{22}}{1-\rho^2}$$

$$= \frac{\gamma_2^2 \left| M_{zz} \right|}{m_{z_1 z_1}} + \frac{\sigma_{22}}{1-\rho^2} > 0$$

The expression (5.2), as a measure of the loss of efficiency, shows this loss to be a decreasing function of the degree of correlation of the z's, i.e., $\left| M_{zz} \right|$ approaches zero; in the extreme case $\left| M_{zz} \right| = 0$, (2.2') can be written as

$$(2.2'') \quad y_2(t) + \gamma_1^* z_1(t) = u_2(t)$$

with the assumptions (2.3) holding for the estimation procedure under (2.b). In such cases it appears obvious that the inclusion of observations on both z's adds little explaining power in the estimation over that from using only the one z.

γ_2 in the above expression indicates the cost of ignoring the variable associated with it in the estimation. A priori knowledge of the investigator concerning likely values of γ_1 or γ_2 would aid the investigator in his choice of z's.

It is interesting to observe that $\sigma_{\beta^* \beta^*}$ is independent of σ_{22} and σ_{12} while $\sigma_{\hat{\beta}\hat{\beta}} \rightarrow 0$ as $\rho^2 \rightarrow 1$.

6. The expressions above are highly specific to the simple model assumed. However, similar results might be conjectured for estimation in larger equation systems which can be broken up into two sub-systems having the general appearance of the two single-equation sub-systems (2.1) and (2.2).