Annals of Economic and Social Measurement, 6/4, 1977

NOTES

A NOTE ON THE ASYMPTOTIC CRAMER RAO BOUND IN NONLINEAR SIMULTANEOUS EQUATION SYSTEMS*

BY JEAN-JACQUES LAFFONT

1. INTRODUCTION

The study of efficient estimation of nonlinear simultaneous equation systems with additive disturbances requires a good understanding of the complexity of the asymptotic Cramer-Rao bound. The purpose of this note is to provide a clear exposition of the derivation of this bound which corrects an error contained in Jorgenson and Laffont [1974] and to comment on its structure. This discussion leads to a suggestion of research strategy for finding efficient estimators. Finally, an example of use of this bound is given in relation to a limited information maximum likelihood estimator proposed by Amemiya (1975).

2. THE MODEL

We consider the following system of simultaneous equations:

$$f_\alpha(y_{1t}, \ldots, y_{Pt}, x_{1t}, \ldots, x_{K_\alpha t}, \theta_\alpha)^1 = u_{\alpha t} \qquad \begin{aligned} \alpha &= 1, \ldots, P \\ t &= 1, \ldots, T \end{aligned} \tag{1}$$

where

$\{x_{kt}\}, k = 1, \ldots, K_\alpha$ are exogenous variables
$\{y_{pt}\}, p = 1, \ldots, P$ are endogenous variables
$\{\theta_\alpha\}$ is a R_α-vector of parameters.

Let

$$\theta' = (\theta_1, \ldots, \theta_P) = (\theta_1^1, \ldots, \theta_1^{R_1}, \ldots, \theta_P^1, \ldots, \theta_P^{R_P})$$

We consider only the case where there is no constraint across equations, even though the most general case can be dealt with along similar lines.

*The author would like to thank D. Jorgenson for his encouragement.

[1] In the following, the $f_\alpha(\cdot)$ functions are assumed to be differentiable up to the required orders; we need up to third derivatives with respect to some variables.

The random vectors $u_t = [u_{1t}, \ldots, u_{Pt}]'$ are independently identically distributed normal variates such that $\left| \begin{array}{l} Eu_t = 0 \\ Eu_t u_t' = \Omega, \ \forall t \end{array} \right.$ of full rank. Finally, we assume that the Jacobian of the system is never vanishing.

Let $\omega = \text{vec}\ \Omega$ and let $\eta = \begin{bmatrix} \theta \\ \omega \end{bmatrix}$. If L^* is the likelihood function of the system, the asymptotic Cramer-Rao bound R_η is by definition (see Rothenberg (1974)) such that:

$$R_\eta^{-1} = \lim_{T \to \infty} E - \frac{1}{T} \frac{\partial^2 \text{Log}\ L^*}{\partial \eta \partial \eta'}$$

If we are only interested in the parameters θ, we know from Koopmans, Rubin, Leipnik (1950) that:

$$R_\theta^{-1} = \lim_{T \to \infty} E - \frac{1}{T} \frac{\partial^2 \text{Log}\ L^+}{\partial \theta \partial \theta'}$$

where L^+ is the concentrated (in θ) likelihood function.

3. The Asymptotic Cramer Rao Bound With An Unrestricted Ω Matrix

The logarithm of the likelihood function is:

$$\text{(2)} \quad \text{Log}\ L^* = -\frac{PT}{2} \log 2\pi + \frac{T}{2} \log(\det \Omega^{-1}) + \sum_t \log |\det B_t| - \frac{1}{2} \sum_{ipt} f_{it} \Omega^{ip} f_{pt}$$

where

$$f_{it} = f_i(y_{1t}, \ldots, y_{Pt}, x_{1t}, \ldots, x_{K_i t}, \theta_i) \qquad i = 1, \ldots, P$$

Let us denote

$B_{ipt} = (\partial f_i / \partial y_p)_t$ and B_t the matrix of such derivatives.[2]

Then,

$$\frac{\partial L}{\partial B_{ipt}} = B_t^{pi\,3}$$

[2]We adopt the following convention. The differentiation of a numerical function with respect to a column (row) vector is a column (row) vector.

[3]We use the following result. If $A = [a_{ij}]$ is a nonsingular matrix with inverse $A^{-1} = [a^{ij}]$, then $\partial \log |\det A| / \partial a_{ij} = a^{ji}$.

Let us denote

$$H_{i\alpha} = P \lim_{T \to \infty} \frac{1}{T} \sum_t f_{it} \frac{\partial f_{\alpha t}}{\partial \theta_\alpha} \qquad i = 1, \ldots, P$$

(vector) $\qquad \alpha = 1, \ldots, P$

$$H_k = [H_{1k}, \ldots, H_{Pk}]$$

$$\tilde{H}_k = [\tilde{H}_{1k}, \ldots, \tilde{H}_{Pk}] \quad \text{with} \quad \tilde{H}_{jk} = \Omega^{jk} H_{jk}$$

E is a $(P \times P)$ matrix of ones $\qquad j = 1, \ldots, P$

$$F_{ij} = P \lim_{T \to \infty} \frac{1}{T} \sum_t \frac{\partial f_{it}}{\partial \theta_i} \cdot \frac{\partial f_{jt}}{\partial \theta_j'}$$

$$G_{ij} = P \lim_{T \to \infty} \frac{1}{T} \sum_t f_{it} \cdot \frac{\partial^2 f_{jt}}{\partial \theta_i \partial \theta_j'} \qquad i,j = 1, \ldots, P$$

Let

$$J_{\alpha\beta}^{pm} = P \lim_{T \to \infty} \frac{1}{T} \sum_t B_t^{p\beta} B_t^{m\alpha} \frac{\partial^2 f_{\alpha t}}{\partial \theta_\alpha \partial y_p} \cdot \frac{\partial^2 f_{\beta t}}{\partial \theta_\beta' \partial y_m}$$

and

$$J_{\alpha\beta} = \sum_{mp} J_{\alpha\beta}^{pm}$$

Let

$$L_{p\alpha} = P \lim_{T \to \infty} \frac{1}{T} \sum_t B_t^{p\alpha} \cdot \frac{\partial^3 f_{\alpha t}}{\partial \theta_\alpha \partial \theta_\alpha' \partial y_p}$$

After concentrating the logarithm of the likelihood and after a number of manipulations, we finally obtain:

$$(3) \qquad R_\theta^{-1} = C_1 + C_2 + C_2' + C_3 + C_4$$

$$= - \begin{bmatrix} \Omega^{11} H_1 \Omega^{-1} H_1' + \tilde{H}_1 E \tilde{H}_1' & \cdots & \Omega^{1P} H_1 \Omega^{-1} H_P' + \tilde{H}_1 E \tilde{H}_P' \\ & \ddots & \\ & & \Omega^{PP} H_P \Omega^{-1} H_P' + \tilde{H}_P E \tilde{H}_P' \end{bmatrix}$$

$$+ \begin{bmatrix} \Omega^{11} F_{11} & \cdots & \Omega^{1P} F_{1P} \\ & \ddots & \\ \Omega^{P1} F_{P1} & & \Omega^{PP} F_{PP} \end{bmatrix}$$

$$
\left. + \begin{bmatrix} \sum_i \Omega^{i1} G_{i1} & & 0 \\ & \ddots & \\ 0 & & \sum_i \Omega^{iP} G_{iP} \end{bmatrix} + \begin{bmatrix} J_{11} & \cdots & J_{1P} \\ & & \\ J_{P1} & & J_{PP} \end{bmatrix} - \begin{bmatrix} \sum_p L_{1p} & & 0 \\ & \ddots & \\ 0 & & \sum_p L_{Pp} \end{bmatrix} \right\}
$$

4. Some Special Cases

(a). The model (1) is written in a reduced form. This is the case considered by Malinvaud (1970)

$$
f_\alpha(y_{1t}, \ldots, y_{Pt}, x_{1t}, \ldots, x_{K_\alpha t}, \theta_\alpha) = u_{\alpha t} \qquad \begin{matrix} \alpha = 1, \ldots, P \\ t = 1, \ldots, T \end{matrix}
$$

is replaced by:

$$
y_\alpha - g_\alpha(x_{1t}, \ldots, x_{K_\alpha t}, \theta_\alpha) = u_{\alpha t} \qquad \begin{matrix} \alpha = 1, \ldots, P \\ t = 1, \ldots, T \end{matrix}
$$

Then $\partial f_{\alpha t}/\partial \theta_\alpha$ is non random, as well as $\partial^2 f_{\alpha t}/\partial \theta_\alpha \partial \theta'_\alpha$: therefore, if all the derivatives are bounded in the sample space, we have by the law of large numbers:

$$
P \lim \frac{1}{T} \sum_t f_{it} \frac{\partial f_{\alpha t}}{\partial \theta_\alpha} = 0 \qquad \text{so that } C_1 = 0
$$

and

$$
P \lim \frac{1}{T} \sum_t f_{it} \frac{\partial^2 f_{\alpha t}}{\partial \theta_\alpha \partial \theta'_\alpha} = 0 \qquad \text{so that } C'_2 = 0
$$

Moreover,

$$
\frac{\partial^2 f_{\alpha t}}{\partial \theta_\alpha \partial y_p} = 0 \quad \text{and} \quad \frac{\partial^3 f_{\alpha t}}{\partial \theta_\alpha \partial \theta'_\alpha \partial y_p} = 0 \qquad \text{for } \alpha, p = 1, \ldots, P
$$

by definition of a reduced form. Therefore $C_3 = C_4 = 0$. R_θ^{-1} reduces to the matrix C_2.

(b). The Ω matrix is known.

Then we differentiate directly (2) with respect to θ. We botain:

$$
R_\theta^{-1} = C_2 + C'_2 + C_3 + C_4
$$

(c). The model is linear in parameters only.

Then, since

$$\frac{\partial^2 f_{\alpha t}}{\partial \theta_\alpha \partial \theta'_\alpha} = 0, \quad \frac{\partial^3 f_{\alpha t}}{\partial \theta_\alpha \partial \theta'_\alpha \partial y_p} = 0 \qquad \alpha, p = 1, \ldots, P$$

$$C'_2 = C_4 = 0$$

so that

$$R_\theta^{-1} = C_1 + C_2 + C_3$$

Moreover C_2 does not depend on θ.

(d). The model is linear in variables only.

No formal simplification obtains in this case even though C_3 and C_4 depend only on parameters. However, in this case, if there is no constraint on Ω, it is easy to obtain a reduced form and to be back in case (a).

The general form of the inverse of the asymptotic Cramer-Rao bound can be decomposed into five parts

$$R_\theta^{-1} = C_1 + C_2 + C'_2 + C_3 + C_4$$

If there were no endogenous variables on the right of system (1), R_θ^{-1} would reduce to C_2. $C'_2 + C_3 + C_4$ represents the modification due to the existence of endogenous variables on the right when Ω is known; this expression simplifies to C_3 when the model is linear in parameters only. Finally C_1 represents the additional change due to the necessity of estimating Ω. It is not difficult to specialize the results to the linear case considered by Rothenberg and Leenders. When there is no constraint on Ω, it is possible in the linear case to obtain the asymptotic Cramer-Rao bound, from the bound if there were no endogenous variables on the right, by simply replacing the endogenous variables by the systematic part of the reduced form associated with them. In the non linear case, this step is much more complex. A good deal of the difficulty (C_3, C_4) stems from the complexity of the Jacobian of the transformation from u to y.

This result suggests two points. First, even though it is easy to obtain asymptotically efficient estimators when the non linear system is in reduced form (see Malinvaud (1970)), the efficient estimation of non linear "truly" simultaneous systems appears as a substantial additional step. Second, the complexity of the asymptotic Cramer-Rao bound suggests that it will be extremely difficult to prove that we have an asymptotically efficient estimator by showing that its asymptotic matrix of variances and covariances coincides with R_θ^{-1}: the implied research strategy is to prove directly that the proposed estimator solves the same equations as the full information maximum likelihood estimator.

5. An Application

Amemiya (1975) considers the following model that we simplify to a two equation model

$$\begin{cases} y_{1t} = f_t(y_{2t}, \theta) + u_t \\ y_{2t} = \quad X_t'\pi \quad + v_t \end{cases}$$

i.e., model (1) where the only non linear equation is the first equation and where the other part of the system is in reduced form.

He then suggests a non linear limited information maximum likelihood estimator of θ, for which he derives the asymptotic matrix of variances and covariances

$$V_{LI} = P\lim\left[\frac{1}{T}\Omega^{11}G - \frac{1}{T}(\Omega^{11} - \Omega_{11}^{-1})H\right]^{-1}$$

where

$$G = \frac{\partial f}{\partial \theta}[I - v(v'v)^{-1}v']\frac{\partial f}{\partial \theta'}$$

$$H = \frac{\partial f}{\partial \theta}X(X'X)^{-1}X'\frac{\partial f}{\partial \theta}$$

Simple manipulations show that in this case our general formula (3) reduces to:

$$R_{\theta.\pi}^{-1} = \begin{bmatrix} \Omega^{11}H_1\Omega^{-1}H_1' + \tilde{H}_1E\tilde{H}_1' + \Omega^{11}F_{11} + \sum_i \Omega^{ii}G_{ii} & \Omega^{12}F_{12} \\ \Omega^{21}F_{21} & \Omega^{22}F_{22} \end{bmatrix}$$

Therefore,

$R_\theta =$

$$\left[\Omega^{11}H_1\Omega^{-1}H_1' + \tilde{H}_1EH_1' + \sum_i \Omega^{ii}G_{ii} + \Omega^{11}F_{11} - \frac{\Omega^{12}\Omega^{21}}{\Omega^{22}}F_{12}F_{22}^{-1}F_{12}\right]^{-1}$$

In our notations:

$$P\lim\frac{1}{T}\Omega^{11}G = \Omega^{11}P\lim\frac{1}{T}\frac{\partial f}{\partial \theta}\cdot\frac{\partial f}{\partial \theta'}$$

$$- \Omega^{11}P\lim\left(\frac{1}{T}\frac{\partial f}{\partial \theta}v\right)\left(\frac{v'v}{T}\right)^{-1}\left(\frac{1}{T}v'\frac{\partial f}{\partial \theta'}\right)$$

$$= \Omega^{11}F_{11} - \Omega^{11}H_1\Omega^{-1}H_1'$$

Moreover,

$$\frac{\Omega^{12}\Omega^{21}}{\Omega^{22}} F_{12}F_{22}^{-1}F_{21} = (\Omega^{11} - \Omega_{11}^{-1})\, P\lim\left(\frac{1}{T}\,\frac{\partial f}{\partial \theta}\,X\right)\left(\frac{X'X}{T}\right)^{-1}\left(\frac{1}{T}\,X'\,\frac{\partial f}{\partial \theta'}\right)$$

$$= (\Omega^{11} - \Omega_{11}^{-1})\, P\lim \frac{1}{T}\, H$$

Therefore, we have:

$$R_\theta^{-1} = V_{LI}^{-1} + \tilde{H}_1 E \tilde{H}_1' + \sum_i \Omega^{i1} G_{i1}$$

$\tilde{H}_1 E \tilde{H}_1' + \sum_i \Omega^{i1} G_{i1}$ represents the loss of efficiency due to the use of a limited information estimator. We know that nothing is gained from a full information estimator if u and ε are independent. Indeed, it is easy to see that in this case $\partial f/\partial \theta$ is independent of u so that $H_{i1} = G_{i1} = 0$; since, in addition, $\Omega^{21} = 0$ we have

$$\tilde{H}_1 E \tilde{H}_1' = 0 \quad \text{and} \quad \sum_i \Omega^{i1} G_{i1} = 0$$

Then: $V_{LI} = R_\theta$

References

Amemiya, T. (1975): "The Nonlinear Limited Information Maximum Likelihood Estimator and the Modified Nonlinear Two-stage Least Squares Estimator," *Journal of Econometrics*, vol. 3.

Eisenpress, H. and J. Greenstadt (1966): "The Estimation of Nonlinear Econometric Systems," *Econometrica*, vol. 34.

Jorgenson, D. and J. J. Laffont (1974): "Efficient Estimation of Nonlinear Systems of Simultaneous Equations with Additive Disturbances," *Annals of Social and Economic Measurement*, 3, 613-640.

Koopmans, T. Rubin, and I. Leipnick (1950): "Measuring the Equation Systems of Dynamic Economics," chapter II of: KOOPMANS (1950), *"Statistical Inference in Dynamic Economic Models."*

Malinvaud, E. (1970): *"Statistical Methods of Econometrics,"* Amsterdam, North-Holland.

Rothenberg, T. (1974): *"Efficient Estimation with A Priori Information,"* New Haven, Yale University Press.