Volume Title: Annals of Economic and Social Measurement, Volume 3, number 4

Volume Author/Editor: Sanford V. Berg, editor

Volume Publisher: NBER

Volume URL: http://www.nber.org/books/aesm74-4

Publication Date: October 1974

Chapter Title: Efficient Estimation of Nonlinear Simultaneous Equations with Additive Disturbances

Chapter Author: Dale W. Jorgenson, Jean-Jacques Laffont

Chapter URL: http://www.nber.org/chapters/c10204

Chapter pages in book: (p. 65 - 90)

Annals of Economic and Social Measurement, 3/4, 1974

EFFICIENT ESTIMATION OF NONLINEAR SIMULTANEOUS EQUATIONS WITH ADDITIVE DISTURBANCES

BY DALE W. JORGENSON AND JEAN-JACQUES LAFFONT

This paper develops a theory of CUAN estimation for systems of nonlinear simultaneous equations with additive disturbances. We first derive the Cramer-Rao lower bound for the variance of a CUAN estimator. The method of maximum likelihood can be used to generate an estimator that attains this bound. We show that minimum distance and instrumental variables estimators cannot generally attain the Cramer-Rao bound.

1. INTRODUCTION

The statistical theory of estimation for systems of linear simultaneous equations is based on the construction of consistent, uniformly asymptotically normal (CUAN) estimators.[1] Within this class it is natural to select estimators that are, in addition, efficient; we refer to such estimators as best consistent uniformly asymptotically normal (Best CUAN) estimators.[2] The purpose of this paper is to develop a theory of CUAN estimation for systems of nonlinear simultaneous equations with additive disturbances.[3]

The theory of CUAN estimation for systems of linear simultaneous equations can be summarized as follows: estimators can be constructed that attain the Cramer-Rao lower bound for the variance of a CUAN estimator.[4] The ordinary least squares estimator for the reduced form is CUAN, but not generally Best CUAN. Best CUAN estimators can be constructed by the method of maximum likelihood, the minimum distance method, and the method of efficient instrumental variables.[5]

Malinvaud has developed a theory of CUAN estimation for systems of nonlinear simultaneous equations with an explicit reduced form having additive disturbances.[6] For this class of nonlinear systems the ordinary least squares estimator for the reduced form is CUAN, but not generally Best CUAN; Malinvaud shows that Best CUAN estimators can be constructed by the method of maximum likelihood and the minimum distance method. Hausman has shown that a Best CUAN estimator can be constructed for a closely related class of models by the method of efficient instrumental variables.[7]

Our first step in developing a theory of CUAN estimation for systems of nonlinear simultaneous equations with additive disturbances is to derive the

[1] The statistical theory of CUAN estimation is discussed by Rao (1973), pp. 344–351.

[2] Best CUAN estimators are discussed by Rao (1973), pp. 350–351.

[3] This specification for simultaneous equations models is considered by Eisenpress and Greenstadt (1966).

[4] A complete review of the theory of CUAN estimation for systems of linear simultaneous equations models is presented by Malinvaud (1970), pp. 348–366, and Rothenberg (1974).

[5] See Malinvaud (1970), pp. 675–678, for a discussion of maximum likelihood and minimum distance estimators, and Brundy and Jorgenson (1971) for a discussion of efficient instrumental variables estimators.

[6] See Malinvaud (1970), pp. 348–366.

[7] See Hausman (1974).

Cramer-Rao lower bound to the variance of a CUAN estimator. The method of maximum likelihood can be used to generate a Best CUAN estimator. As for linear systems, the burden associated with the conventional approach to computation of the maximum likelihood estimator, based on the Newton-Raphson method or the method of scoring, is very substantial.[8]

We can distinguish two alternative lines of attack on the problem of reducing the computational burden for Best CUAN estimation of systems of nonlinear simultaneous equations. First, the computation of the maximum likelihood estimator can be simplified. Rothenberg and Leenders have shown that the first step of the Newton-Raphson method is Best CUAN, provided that the initial parameter value is a consistent estimator.[9] Although Rothenberg and Leenders apply this result only to systems of linear simultaneous equations, the proposition holds for nonlinear systems as well.

In this paper we concentrate on a second line of attack, namely, construction of estimators by methods, such as minimum distance or instrumental variables, that are easier to compute. Amemiya has proposed a minimum distance estimator for a single equation in a system of nonlinear simultaneous equations.[10] We extend his method of estimation to systems of nonlinear simultaneous equations and his proof that the resulting estimator is CUAN. However, we show by means of an example that the minimum distance estimator is not generally Best CUAN.

We also develop an instrumental variables estimator for a system of nonlinear simultaneous equations, extending the efficient instrumental variables estimator for linear systems developed by Brundy and Jorgenson.[11] We show that the resulting estimator is CUAN and, in fact, asymptotically equivalent to our minimum distance estimator. Again, the efficient instrumental variables estimator is not Best CUAN.

We conclude that minimum distance and instrumental variables estimators can be constructed that are CUAN, but that these estimators are not generally Best CUAN. Further research on Best CUAN estimation for systems of nonlinear simultaneous equations should be focused on simplifying the computation of the maximum likelihood estimator.[12]

2. The Model

We consider the following system of simultaneous equations:

$$(1) \qquad \begin{aligned} y_{1t} &= f_1(z_{1t}, \beta) + u_{1t} \\ y_{Pt} &= f_P(z_{Pt}, \beta) + u_{Pt} \qquad t = 1 \ldots T \end{aligned}$$

or in vector form:

$$y_t = f(z_t, \beta) + u_t \qquad t = 1 \ldots T$$

[8] The Newton-Raphson method is described by Eisenpress and Greenstadt (1966).

[9] See Rothenberg and Leenders (1964).

[10] See Amemiya (1974). Minimum distance estimators for a single equation in a system of simultaneous equations are also discussed by Edgerton (1972), Kelejian (1972), and Zellner, Huang and Chau (1965).

[11] See Brundy and Jorgenson (1971, 1973).

[12] Important progress along these lines is reported by Berndt, Hall, Hall, and Hausman elsewhere in this issue.

with

$$y_t = [y_{1t}, \ldots, y_{Pt}]'; \qquad z_t = [z'_{1t}, \ldots, z'_{Pt}]'$$

$$u_t = [u_{1t}, \ldots, u_{Pt}]'; \qquad f(\,.\,) = [f_1(\,.\,), \ldots, f_P(\,.\,)]'$$

$y_{1t}, \ldots, y_{Pt}$ are the endogenous variables: for each $i = 1, \ldots, P$, z_{1t} is a Q_i-vector of endogenous and (nonrandom) exogenous variables: β is an R-vector of unknown parameters: f_i is a nonlinear function with continuous second derivatives (R_i is the number of elements of β in f_i). $u_t, t = 1, \ldots, T$ are random vectors such that $Eu_{it} = 0, i = 1, \ldots, P$; $t = 1, \ldots, T$, $Eu_t u'_t = \Omega$ of full rank and $Eu_t u'_{t'} = 0$ if $t \neq t'$.

This form of the model can be obtained from a model with different parameters β_i in each equation. If there are constraints on these parameters, they are solved to obtain a minimal set of parameters β. We assume that the constraints can be solved uniquely at least in a neighborhood of the true value β^0. We assume further that the parameter β is identifiable.[13]

The model is now rewritten differently to use the simplifying Kronecker notation. Let $Y_i = [y_{i1}, \ldots, y_{iT}]'$, $i = 1, \ldots, P$ and $Y = [Y_1, \ldots, Y_P]'$. Let Z be the stacked vector of variables appearing on the right of system (1). Let $F(Z, \beta) = [f_1(z_{11}, \beta), \ldots, f_1(z_{1t}, \beta), \ldots, f_P(z_{P1}, \beta), \ldots, f_P(z_{PT}, \beta_P)]'$

$$U = [u_{11}, \ldots, u_{1T}, \ldots, u_{P1}, \ldots, u_{PT}].$$

Then (1) can be rewritten

$$Y = F(Z, \beta) + U \text{ with } EUU' = \Omega \otimes I_T. \tag{2}$$

We choose a specific notation for the set of exogenous variables (independent of U) which are K^* in number.

$$X^* = \begin{bmatrix} x_{11} & x_{21} & \ldots & x_{K^*1} \\ \vdots & & & \\ x_{1T} & x_{2T} & \ldots & x_{K^*T} \end{bmatrix}$$

X will be in this Section a matrix of K variables constructed from X^* with $\max_i R_i \leq K \leq K^*$. We refer to assumptions specified in Section 2 as A_0.

3. Cramer-Rao Bound

3.1. *Introduction*

The comparison of different full information methods[14] to estimate nonlinear econometric systems with additive disturbances requires an explicit form of the Cramer-Rao bound. In this Section, we derive the Cramer-Rao bound when the matrix of variances and covariances of errors is unknown, completely known, or known to be diagonal.

[13] Identifiability for systems of nonlinear simultaneous equations is discussed by Fisher (1966), pp. 127–167.

[14] We restrict ourselves to cases without constraints across equations to allow comparisons with the Rothenberg–Leenders (1964) results in the linear case, but there is no substantive difficulty extending these derivations to more general cases. We use the notation of Eisenpress–Greenstadt (1966).

For this Section we rewrite the model more symmetrically:

$$f_\alpha(y_1, \ldots, y_P, x_1, \ldots, x_{K^*}, \theta_\alpha) = u_\alpha \qquad \alpha = 1, \ldots, P \tag{3}$$

where:

$\{x_k\} k = 1, \ldots, K^*$ are predetermined variables

$\{y_p\} p = 1, \ldots, P$ are endogenous variables

$\{\theta_\alpha\}$ is a R_α-vector of parameters $\alpha = 1, \ldots, P$.

We assume that the Jacobian of the system is nevervanishing (it is clearly a strong assumption) and we assume the multinormal distribution for the errors so that we can derive the logarithm of the likelihood function.

3.2. *Unrestricted Ω Matrix*

The logarithm of the likelihood function is:

$$L^* = -\frac{PT}{2}\log 2\pi + \frac{T}{2}\log(\det \Omega^{-1}) + \sum_t \log|\det B_t| - \frac{1}{2}\sum_{ipt} f_{it}\Omega^{ip} f_{pt} \tag{4}$$

where

$$f_{it} = f_i[y_{1t}, \ldots, y_{Pt}, x_{1t}, \ldots, x_{K^*t}, \theta_\alpha]$$

$$B_{ipt} \equiv \left(\frac{\partial f_i}{\partial y_p}\right)_t$$

and B_t is the matrix of such derivatives. We concentrate the likelihood:

$$\frac{\partial L^*}{\partial \Omega^{ip}} = \frac{T}{2}\hat{\Omega}_{pi} - \frac{1}{2}\sum_t f_{it} \cdot f_{pt} = 0 \tag{5}$$

$$\hat{\Omega}_{pi} = \hat{\Omega}_{ip} = \frac{1}{T}\sum_t f_{it} \cdot f_{pt}. \tag{6}$$

Substituting (6) in (4) we obtain the nonconstant part of the concentrated likelihood.

$$L = -\frac{T}{2}\log(\det \hat{\Omega}) + \sum_t \log|\det B_t|. \tag{7}$$

We want to obtain:

$$\lim_{T\to\infty} E - \frac{1}{T}\frac{\partial^2 \log L}{\partial\theta\partial\theta'} \qquad (\theta' = [\theta'_1, \ldots, \theta'_\alpha, \ldots, \theta'_P])$$

which will be the inverse of the Cramer-Rao bound for the parameters θ.[15]

(8)[16]
$$\frac{\partial L}{\partial \theta_\alpha} = \sum_{ip} \frac{\partial L}{\partial \hat{\Omega}_{ip}} \cdot \frac{\partial \hat{\Omega}_{ip}}{\partial \theta_\alpha} + \sum_{ipt} \frac{\partial L}{\partial B_{ipt}} \cdot \frac{\partial B_{ipt}}{\partial \theta_\alpha}$$

(9)[17]
$$\frac{\partial L}{\partial \hat{\Omega}_{ip}} = -\frac{T}{2} \hat{\Omega}^{ip}$$

(10)
$$\frac{\partial L}{\partial B_{ipt}} = B_t^{pi}$$

(11)
$$\frac{\partial L}{\partial \theta_\alpha} = -\frac{T}{2} \sum_{ip} \hat{\Omega}^{ip} \cdot \frac{\partial \hat{\Omega}_{ip}}{\partial \theta_\alpha} + \sum_{ipt} B_t^{pi} \cdot \frac{\partial B_{ipt}}{\partial \theta_\alpha}$$

From (6)

(12)
$$\frac{\partial \hat{\Omega}_{ip}}{\partial \theta_\alpha} = \frac{1}{T} \sum_t \left(\frac{\partial f_{it}}{\partial \theta_\alpha} \cdot f_{pt} + f_{it} \cdot \frac{\partial f_{pt}}{\partial \theta_\alpha} \right).$$

with the following simplifications (since f_{it} depends only on θ_i and not on θ_j with $j \neq i$):

(13)
$$\frac{\partial \hat{\Omega}_{\alpha\alpha}}{\partial \theta_\alpha} = \frac{2}{T} \sum_t f_{\alpha t} \cdot \frac{\partial f_{\alpha t}}{\partial \theta_\alpha}$$

(14)
$$\frac{\partial \hat{\Omega}_{i\alpha}}{\partial \theta_\alpha} = \frac{\partial \hat{\Omega}_{\alpha i}}{\partial \theta_\alpha} = \frac{1}{T} \sum_t f_{it} \cdot \frac{\partial f_{\alpha t}}{\partial \theta_\alpha} \qquad \text{if } i \neq \alpha$$

(15)
$$\frac{\partial \hat{\Omega}_{ij}}{\partial \theta_\alpha} = 0 \qquad \text{if } i \neq \alpha \text{ and } j \neq \alpha$$

(16)
$$\frac{\partial B_{ipt}}{\partial \theta_\alpha} = \frac{\partial^2 f_{it}}{\partial \theta_\alpha \partial y_p}$$

with the following simplifications:

(17)
$$\frac{\partial B_{\alpha pt}}{\partial \theta_\alpha} = \frac{\partial^2 f_{\alpha t}}{\partial \theta_\alpha \partial y_p}$$

(18)
$$\frac{\partial B_{ipt}}{\partial \theta_\alpha} = 0 \qquad \text{if } i \neq \alpha$$

[15] See Koopmans and Hood (1953).

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

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

Then:

$$\frac{\partial L}{\partial \theta_\alpha} = -\sum_i \hat{\Omega}^{i\alpha}\left(\sum_t f_{it}\cdot\frac{\partial f_{\alpha t}}{\partial \theta_\alpha}\right) + \sum_{pt} B_t^{p\alpha}\frac{\partial^2 f_{\alpha t}}{\partial \theta_\alpha \partial y_p} \tag{19}$$

We can now obtain the second derivatives:[18]

$$\begin{aligned}\frac{\partial^2 L}{\partial \theta_\alpha \partial \theta'_\beta} = &-\sum_i\left(\sum_t f_{it}\cdot\frac{\partial f_{\alpha t}}{\partial \theta_\alpha}\right)\frac{\partial \hat{\Omega}^{i\alpha}}{\partial \theta'_\beta} - \sum_i \hat{\Omega}^{i\alpha}\left(\sum_t \frac{\partial f_{\alpha t}}{\partial \theta_\alpha}\cdot\frac{\partial f_{it}}{\partial \theta'_\beta} + f_{it}\frac{\partial^2 f_{\alpha t}}{\partial \theta_\alpha \partial \theta'_\beta}\right)\\ &+\sum_{pt}\frac{\partial^2 f_{\alpha t}}{\partial \theta_\alpha \partial y_p}\frac{\partial B_t^{p\alpha}}{\partial \theta'_\beta} + \sum_{pt} B_t^{p\alpha}\frac{\partial^3 f_{\alpha t}}{\partial \theta_\alpha \partial \theta'_\beta \partial y_p}\end{aligned} \tag{20}$$

We use the general formula:

$$\frac{\partial M_{ip}^{-1}}{\partial \theta_\alpha} = -\sum_{hm} M_{ih}^{-1}\cdot\frac{\partial M_{hm}}{\partial \theta_\alpha}\cdot M_{mp}^{-1} \tag{21}$$

using (13), (14), (15)

$$\begin{aligned}\frac{\partial \hat{\Omega}^{i\alpha}}{\partial \theta'_\beta} = \frac{\partial (\hat{\Omega}^{-1})_{i\alpha}}{\partial \theta'_\beta} &= -\sum_{hm}\hat{\Omega}^{ih}\frac{\partial \hat{\Omega}_{hm}}{\partial \theta'_\beta}\hat{\Omega}^{m\alpha}\\ &= -\sum_m \hat{\Omega}^{i\beta}\frac{\partial \hat{\Omega}_{\beta m}}{\partial \theta'_\beta}\hat{\Omega}^{m\alpha} - \sum_h \hat{\Omega}^{ih}\frac{\partial \hat{\Omega}_{h\beta}}{\partial \theta'_\beta}\hat{\Omega}^{\beta\alpha}\\ &= -\sum_m \hat{\Omega}^{i\beta}\hat{\Omega}^{m\alpha}\left(\frac{1}{T}\sum_t f_{mt}\frac{\partial f_{\beta t}}{\partial \theta'_\beta}\right) - \sum_h \hat{\Omega}^{ih}\hat{\Omega}^{\beta\alpha}\left(\frac{1}{T}\sum_t f_{ht}\cdot\frac{\partial f_{\beta t}}{\partial \theta'_\beta}\right)\end{aligned} \tag{22}$$

The first element of the right hand side of (20) is then:

$$\frac{1}{T}\sum_{im}\hat{\Omega}^{i\beta}\hat{\Omega}^{m\alpha}\left(\sum_t f_{it}\frac{\partial f_{\alpha t}}{\partial \theta_\alpha}\right)\left(\sum_t f_{mt}\frac{\partial f_{\beta t}}{\partial \theta'_\beta}\right) + \frac{1}{T}\sum_{ih}\hat{\Omega}^{ih}\hat{\Omega}^{\beta\alpha}\left(\sum_t f_{it}\frac{\partial f_{\alpha t}}{\partial \theta_\alpha}\right)\left(\sum_t f_{ht}\frac{\partial f_{\beta t}}{\partial \theta'_\beta}\right) \tag{23}$$

$$\begin{aligned}\frac{\partial B_t^{p\alpha}}{\partial \theta'_\beta} &= -\sum_{hm} B_t^{ph}\frac{\partial B_{hmt}}{\partial \theta'_\beta}\cdot B_t^{m\alpha}\\ &= -\sum_m B_t^{p\beta}\frac{\partial^2 f_{\beta t}}{\partial \theta'_\beta \partial y_m}\cdot B_t^{m\alpha}\end{aligned} \tag{24}$$

The third element of the right hand side of (20) becomes

$$-\sum_{pmt} B_t^{p\beta}B_t^{m\alpha}\frac{\partial^2 f_{\alpha t}}{\partial \theta_\alpha \partial y_p}\cdot\frac{\partial^2 f_{\beta t}}{\partial \theta'_\beta \partial y_m} \tag{25}$$

The second term of the right hand side of (20) becomes:

$$-\hat{\Omega}^{\beta\alpha}\sum_t\frac{\partial f_{\alpha t}}{\partial \theta_\alpha}\frac{\partial f_{\beta t}}{\partial \theta'_\beta} \qquad \text{if } \beta \neq \alpha \tag{26}$$

[18] We do not use "prime indices", so that the sign ' must always be interpreted as a transposition.

or

$$(27)\qquad -\hat{\Omega}^{\alpha\alpha}\sum_t \frac{\partial f_{\alpha t}}{\partial \theta_\alpha}\cdot\frac{\partial f_{\alpha t}}{\partial \theta'_\alpha} - \sum_i \hat{\Omega}^{i\alpha}\sum_t f_{it}\cdot\frac{\partial^2 f_{\alpha t}}{\partial\theta_\alpha\partial\theta'_\alpha} \qquad \text{if } \beta=\alpha$$

We compute now

$$\lim \frac{1}{T}\frac{\partial^2 L}{\partial\theta_\alpha\partial\theta'_\beta}.$$

The first term of the right hand side of $1/T\,(20)$ is:

$$(28)\quad \sum_{im}\hat{\Omega}^{i\beta}\hat{\Omega}^{m\alpha}\left(\frac{1}{T}\sum_t f_{it}\frac{\partial f_{\alpha t}}{\partial\theta_\alpha}\right)\left(\frac{1}{T}\sum_t f_{mt}\frac{\partial f_{\beta t}}{\partial\theta'_\beta}\right)+\sum_{ih}\hat{\Omega}^{ih}\hat{\Omega}^{\beta\alpha}\left(\frac{1}{T}\sum_t f_{it}\frac{\partial f_{\alpha t}}{\partial\theta_\alpha}\right)\left(\frac{1}{T}\sum_t f_{ht}\frac{\partial f_{\beta t}}{\partial\theta'_\beta}\right)$$

We assume that:

$$(29)\qquad \frac{1}{T}\sum_t f_{it}\frac{\partial f_{\alpha t}}{\partial\theta_\alpha}\xrightarrow{P} H_{i\alpha}\qquad \text{(vector)}$$

From (28) we obtain:

$$(30)\qquad \sum_{im}\Omega^{i\beta}\Omega^{m\alpha}H_{i\alpha}H'_{m\beta}+\sum\Omega^{ih}\Omega^{\beta\alpha}H_{i\alpha}H'_{h\beta}$$

since the estimations $\hat{\Omega}^{ij}$ of Ω^{ij} are consistent. In matrix notation:

$$(31)\qquad C_1=\begin{bmatrix}\sum_{im}\Omega^{i1}\Omega^{m1}[H_{i1}H'_{m1}] & \cdots & \sum_{im}\Omega^{iP}\Omega^{m1}H_{i1}H'_{mP}\\ \quad+\sum\Omega^{ih}\Omega^{11}[H_{i1}H'_{h1}] & & +\sum_{ih}\Omega^{ih}\Omega^{P1}H_{i1}H'_{hP}\\ & \ddots & \\ & & \sum_{im}\Omega^{iP}\Omega^{mP}H_{iP}H'_{mP}\\ & & +\sum_{ih}\Omega^{ih}\Omega^{PP}H_{iP}H'_{hP}\end{bmatrix}$$

Let

$$(32)\qquad G_{i\alpha}=\operatorname{plim}\frac{1}{T}\sum_t f_{it}\frac{\partial^2 f_{\alpha t}}{\partial\theta_\alpha\partial\theta'_\alpha}\qquad F_{ij}=\operatorname{plim}\frac{1}{T}\sum_t\frac{\partial f_{it}}{\partial\theta_i}\cdot\frac{\partial f_{jt}}{\partial\theta'_j}$$

$$i=1,\ldots,P;\qquad j=1,\ldots,P;\qquad \alpha=1,\ldots,P.$$

The second term of the right hand side of (20):

$$(33)\quad \operatorname{plim}\frac{1}{T}\begin{bmatrix}\frac{\partial f_1}{\partial\theta_1} & & 0\\ & \ddots & \\ 0 & & \frac{\partial f_P}{\partial\theta_P}\end{bmatrix}[\hat{\Omega}^{-1}\otimes I]\begin{bmatrix}\frac{\partial f_1}{\partial\theta'_1} & & 0\\ & \ddots & \\ 0 & & \frac{\partial f_P}{\partial\theta'_P}\end{bmatrix}+\begin{bmatrix}\sum_i\Omega^{i1}G_{i1} & & 0\\ & \ddots & \\ 0 & & \sum_i\Omega^{iP}G_{iP}\end{bmatrix}$$

$$=\begin{bmatrix}\Omega^{11}F_{11} & \cdots & \Omega^{1P}F_{1P}\\ & \ddots & \\ & & \Omega^{PP}F_{PP}\end{bmatrix}+\begin{bmatrix}\sum_i\Omega^{i1}G_{i1} & & 0\\ & \ddots & \\ 0 & & \sum_i\Omega^{iP}G_{iP}\end{bmatrix}=C_2+C'_2$$

The third term of the right hand side of (20):

$$(34)\quad C_3 = \operatorname{plim} \frac{1}{T} \begin{bmatrix} \sum_t \left(\sum_p B_t^{p1} \frac{\partial^2 f_{1t}}{\partial \theta_1 \partial y_p} \right) \left(\sum_p B_t^{p1} \frac{\partial^2 f_{1t}}{\partial \theta_1' \partial y_p} \right) & \cdots & \sum_t \left(\sum_p B_t^{pP} \frac{\partial^2 f_{1t}}{\partial \theta_1 \partial y_p} \right) \cdot \left(\sum_m B_t^{m1} \frac{\partial^2 f_{Pt}}{\partial \theta_P' \partial y_m} \right) \\ & \ddots & \\ & & \sum_t \left(\sum_p B_t^{pP} \frac{\partial^2 f_{Pt}}{\partial \theta_P' \partial y_p} \right) \left(\sum_p B_t^{pP} \frac{\partial^2 f_{Pt}}{\partial \theta_P' \partial y_p} \right) \end{bmatrix}$$

If

$$J_{jp}^{\alpha} = \operatorname{plim} \frac{1}{T} \sum_t B_t^{p\alpha} \frac{\partial^2 f_{jt}}{\partial \theta_j \partial y_p} \qquad \text{for } j, p, \alpha = 1, \ldots, P$$

$$(35)\quad C_3 = \begin{bmatrix} \sum_p J_{1p}^1 J_{1p}^{1'} & & \sum_p J_{1p}^P J_{Pp}^{1'} \\ & \ddots & \\ \sum_p J_{Pp}^P J_{Pp}^{P'} & & \end{bmatrix}$$

The fourth term of the right hand side of (21):

$$(36)\quad \operatorname{plim} \frac{1}{T} \begin{bmatrix} \sum_t \sum_p B_t^{p1} \frac{\partial^3 f_{1t}}{\partial \theta_1 \partial \theta_1' \partial y_p} & & 0 \\ & \ddots & \\ 0 & & \sum_t \sum_p B_t^{pP} \frac{\partial^3 f_{Pt}}{\partial \theta_P \partial \theta_P' \partial y_p} \end{bmatrix}$$

$$(37)\quad C_4 = \begin{bmatrix} \sum_p L_{1p} & & 0 \\ & \ddots & \\ 0 & & \sum_p L_{Pp} \end{bmatrix}$$

where

$$(38)\quad L_{jp} = \operatorname{plim} \frac{1}{T} \sum_t B_t^{pj} \frac{\partial^3 f_{jt}}{\partial \theta_j \partial \theta_j' \partial y_p} \qquad \text{for } j = 1, \ldots, P; \qquad p = 1, \ldots, P.$$

$$(39)\quad \operatorname{plim} -\frac{1}{T} \frac{\partial^2 L}{\partial \theta \partial \theta'}$$

$$= -\begin{bmatrix} \sum_{im} \Omega^{i1} \Omega^{m1} (H_{i1} H_{m1}') + \sum_{ih} \Omega^{ih} \Omega^{11} (H_{i1} H_{h1}') & \cdots & \sum_{im} \Omega^{iP} \Omega^{m1} (H_{i1} H_{mP}') + \sum_{ih} \Omega^{ih} \Omega^{P1} (H_{i1} H_{hP}') \\ & \ddots & \\ & & \sum_{im} \Omega^{iP} \Omega^{mP} (H_{iP} H_{mP}') + \sum_{ih} \Omega^{ih} \Omega^{PP} (H_{iP} H_{hP}') \end{bmatrix}$$

$$+\begin{bmatrix} \Omega^{11}F_{11} & & \Omega^{1P}F_{1P} \\ & \ddots & \\ & & \Omega^{PP}F_{PP} \end{bmatrix} + \begin{bmatrix} \sum_i \Omega^{i1}G_{i1} & & 0 \\ & \ddots & \\ 0 & & \sum_i \Omega^{iP}G_{iP} \end{bmatrix}$$

$$-\begin{bmatrix} \sum_p J^1_{1p}J^{1'}_{1p} & & \sum_p J^P_{1p}J^{1'}_{Pp} \\ & \ddots & \\ & & \sum_p J^P_{Pp}J^{P'}_{Pp} \end{bmatrix} - \begin{bmatrix} \sum_p L_{1p} & & 0 \\ & \ddots & \\ 0 & & \sum_p L_{Pp} \end{bmatrix}$$

$$= -C_1 + C_2 + C_2' - C_3 - C_4$$

For the special case of the reduced form considered by Malinvaud (1970)

$$\frac{\partial f_{\alpha t}}{\partial \theta_\alpha}$$

is nonrandom as well as

$$\frac{\partial^2 f_{\alpha t}}{\partial \theta_\alpha \partial \theta'_\alpha}.$$

If all the derivatives are bounded in the sample space for all $i, \alpha = 1, \ldots, P$

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

$$\frac{1}{T}\sum_t f_{it}\frac{\partial^2 f_{\alpha t}}{\partial \theta_\alpha \partial \theta'_\alpha} \to 0 \qquad \text{so that } C_2' \xrightarrow{P} 0.$$

Since there is no endogenous variable on the right hand side

$$\frac{\partial^2 f_{\alpha t}}{\partial \theta_\beta \partial y_p} = 0 \quad \alpha = 1, \ldots, P; \quad \beta = 1, \ldots, P; \quad p = 1, \ldots, P \quad \text{so that } C_3 = 0$$

$$\frac{\partial^3 f_{\alpha t}}{\partial \theta_\alpha \partial \theta'_\alpha \partial y_p} = 0 \quad \alpha = 1, \ldots, P; \quad p = 1, \ldots, P \qquad \text{so that } C_4 = 0.$$

Finally:

$$\text{(40)} \quad \text{plim} -\frac{1}{T}\frac{\partial^2 L}{\partial \theta \partial \theta'} = C_2 = \text{plim}\frac{1}{T}\begin{bmatrix} \frac{\partial f_1}{\partial \theta'_1} & & 0 \\ & \ddots & \\ 0 & & \frac{\partial f_P}{\partial \theta'_P} \end{bmatrix} [\hat{\Omega}^{-1} \otimes I] \begin{bmatrix} \frac{\partial f_1}{\partial \theta_1} & & 0 \\ & \ddots & \\ 0 & & \frac{\partial f_P}{\partial \theta_P} \end{bmatrix}.$$

Malinvaud (1970) shows that the maximum likelihood estimator reaches this Cramer-Rao bound. Similarly, the bound is attained by a minimum distance estimator weighted by a matrix S which converge to Ω. When the model is linear in parameters

$$C_2' = 0 \qquad C_4 = 0$$

C_3 is only a function of variables.

$$\text{(41)} \qquad \operatorname{plim} -\frac{1}{T}\frac{\partial^2 L}{\partial\theta\partial\theta'} = -C_1 + C_2 + C_3$$

3.3. *Restricted Ω Matrix*

The logarithm of the likelihood is:

$$\text{(42)} \qquad L = \frac{-NT}{2}\log(2\pi) + \frac{T}{2}\log(\det\Omega^{-1}) + \sum\log|\det B_t| - \frac{1}{2}\sum_{ipt} f_{it}\Omega^{ip}f_{pt}$$

We first consider the case where Ω is known so that we have only to differentiate with respect to θ.

$$\text{(43)} \qquad \frac{\partial}{\partial\theta_\alpha}\sum_{ipt} f_{it}\Omega^{ip}f_{pt} = \sum_t\left[\sum_p \Omega^{\alpha p}f_{pt} + \sum_i \Omega^{i\alpha}f_{it}\right]\frac{\partial f_{\alpha t}}{\partial\theta_\alpha} \qquad \alpha = 1,\ldots,P$$

$$\text{(44)} \qquad \frac{\partial}{\partial\theta_\alpha}\sum_t \log|\det B_t| = \sum_t\sum_p B_t^{p\alpha}\frac{\partial^2 f_{\alpha t}}{\partial\theta_\alpha\partial y_p}$$

If $\alpha \neq \beta$:

$$\text{(45)} \qquad \frac{\partial^2 L}{\partial\theta_\alpha\partial\theta'_\beta} = -\frac{1}{2}\sum_t(\Omega^{\alpha\beta} + \Omega^{\beta\alpha})\frac{\partial f_{\alpha t}}{\partial\theta_\alpha}\cdot\frac{\partial f_{\beta t}}{\partial\theta'_\beta} + \sum_t\sum_p\frac{\partial^2 f_{\alpha t}}{\partial\theta_\alpha\partial y_p}\frac{\partial B_t^{p\beta}}{\partial\theta'_\beta}$$

$$= -\sum_t \Omega^{\alpha\beta}\frac{\partial f_{\alpha t}}{\partial\theta_\alpha}\cdot\frac{\partial f_{\beta t}}{\partial\theta'_\beta} + \sum_t\sum_p\frac{\partial^2 f_{\alpha t}}{\partial\theta_\alpha\partial y_p}\cdot\frac{\partial B_t^{p\beta}}{\partial\theta'_\beta}$$

If $\alpha = \beta$:

$$\text{(46)} \qquad \frac{\partial^2 L}{\partial\theta_\alpha\partial\theta'_\alpha} = -\sum_t \Omega^{\alpha\alpha}\frac{\partial f_{\alpha t}}{\partial\theta_\alpha}\cdot\frac{\partial f_{\alpha t}}{\partial\theta'_\alpha} + \sum_t\sum_p\frac{\partial^2 f_{\alpha t}}{\partial\theta_\alpha\partial y_p}\cdot\frac{\partial B_t^{p\alpha}}{\partial\theta_\alpha} + \sum_t\sum_p B_t^{k\alpha}\cdot\frac{\partial^3 f_{\alpha t}}{\partial\theta_\alpha\partial\theta'_\alpha\partial y_p}$$

$$-\sum_i\Omega^{i\alpha}\sum_t f_{it}\cdot\frac{\partial^2 f_{\alpha t}}{\partial\theta_\alpha\partial\theta'_\alpha}$$

Using (24) and (26) we obtain the matrix form:

$$\lim -\frac{1}{T}\frac{\partial^2 L}{\partial\theta_\alpha\partial\theta'_\alpha} = C_2 + C'_2 - C_3 - C_4$$

using our previous notation. Asymptotically, the gain represented by the knowledge of Ω corresponds to $-C_1$.

Next, we consider the case where Ω is known to be diagonal. The nonconstant part of the concentrated likelihood function is then:

$$L = \sum_t \log |\det B_t| - \frac{T}{2} \sum_{\alpha=1}^{P} \log \hat{\Omega}_{\alpha\alpha}$$

$$\hat{\Omega}_{\alpha\alpha} = \frac{1}{T} \sum_t f_{\alpha t} f_{\alpha t}$$

C_3 and C_4 will remain as in the general case. We have to compute

$$\frac{\partial}{\partial \theta \partial \theta'} \sum_{\alpha=1}^{P} \log \hat{\Omega}_{\alpha\alpha}$$

$$\frac{\partial}{\partial \theta_\alpha} \log \hat{\Omega}_{\alpha\alpha} = \hat{\Omega}^{\alpha\alpha} \cdot \frac{\partial \hat{\Omega}_{\alpha\alpha}}{\partial \theta_\alpha} = \hat{\Omega}^{\alpha\alpha} \cdot \frac{2}{T} \sum_t f_{\alpha t} \cdot \frac{\partial f_{\alpha t}}{\partial \theta_\alpha}$$

$$\frac{\partial}{\partial \theta_\beta} \log \hat{\Omega}_{\alpha\alpha} = 0 \qquad \text{if } \beta \neq \alpha$$

$$\frac{\partial^2}{\partial \theta_\alpha \partial \theta'_\alpha} \log \hat{\Omega}_{\alpha\alpha} = \hat{\Omega}^{\alpha\alpha} \frac{2}{T} \left(\sum_t \frac{\partial f_{\alpha t}}{\partial \theta_\alpha} \frac{\partial f_{\alpha t}}{\partial \theta'_\alpha} + \sum_t f_{\alpha t} \frac{\partial^2 f_{\alpha t}}{\partial \theta_\alpha \partial \theta'_\alpha} \right)$$

$$- \frac{2}{T} \left(\sum_t f_{\alpha t} \frac{\partial f_{\alpha t}}{\partial \theta_\alpha} \right) \left(2 \sum_m \hat{\Omega}^{\alpha\alpha} \frac{\partial \hat{\Omega}_{\alpha m}}{\partial \theta'_\alpha} \hat{\Omega}^{m\alpha} \right)$$

$$= \frac{2\hat{\Omega}^{\alpha\alpha}}{T} \left(\sum_t \frac{\partial f_{\alpha t}}{\partial \theta_\alpha} \cdot \frac{\partial f_{\alpha t}}{\partial \theta'_\alpha} + \sum_t f_{\alpha t} \frac{\partial^2 f_{\alpha t}}{\partial \theta_\alpha \partial \theta'_\alpha} \right)$$

$$- 4\hat{\Omega}^{\alpha\alpha} \sum_m \hat{\Omega}^{m\alpha} \left(\frac{1}{T} \sum_t f_{\alpha t} \frac{\partial^2 f_{\alpha t}}{\partial \theta_\alpha} \right) \left(\frac{1}{T} \sum_t f_{mt} \cdot \frac{\partial f_{\alpha t}}{\partial \theta_\alpha} \right)$$

$$C_1^* = \operatorname{plim} \begin{bmatrix} 2\hat{\Omega}^{11} \frac{1}{T} \sum_t f_{1t} \frac{\partial f_{1t}}{\partial \theta_1} \sum_m \hat{\Omega}^{m1} \frac{1}{T} \sum_t f_{mt} \frac{\partial f_{1t}}{\partial \theta'_1} & & 0 \\ & \ddots & \\ 0 & & 2\hat{\Omega}^{PP} \frac{1}{T} f_{Pt} \frac{\partial f_{Pt}}{\partial \theta_P} \sum_m \hat{\Omega}^{mP} \cdot \frac{1}{T} f_{mt} \frac{\partial f_{Pt}}{\partial \theta'_1} \end{bmatrix}$$

$$C_2^* = \operatorname{plim} \frac{1}{T} \begin{bmatrix} \frac{\partial f_1}{\partial \theta_1} & & 0 \\ & \ddots & \\ 0 & & \frac{\partial f_P}{\partial \theta_P} \end{bmatrix} \begin{bmatrix} \hat{\Omega}^{11} & & 0 \\ & \ddots & \\ 0 & & \hat{\Omega}^{PP} \end{bmatrix} \otimes I \begin{bmatrix} \frac{\partial f_1}{\partial \theta'_1} & & 0 \\ & \ddots & \\ 0 & & \frac{\partial f_P}{\partial \theta'_P} \end{bmatrix}$$

$$C_2^{*\prime} = \operatorname{plim} \frac{1}{T} \begin{bmatrix} \hat{\Omega}^{11} \sum_t f_{1t} \frac{\partial f_{1t}}{\partial \theta_1 \partial \theta'_1} & & 0 \\ & \ddots & \\ 0 & & \hat{\Omega}^{PP} \sum_t f_{Pt} \frac{\partial f_{Pt}}{\partial \theta_P \partial \theta'_P} \end{bmatrix}$$

So that using our notation we obtain:

$$(46^*)\quad \operatorname{plim} -\frac{1}{T}\frac{\partial^2 L}{\partial\theta\partial\theta'} = -\begin{bmatrix} 2\Omega^{11}H_{11}\sum_m \Omega^{m1}H^P_{m1} & & 0 \\ & \ddots & \\ 0 & & 2\Omega^{PP}H_{PP}\sum_m \Omega^{mP}H'_{mP} \end{bmatrix}$$

$$+\begin{bmatrix} \Omega^{11}F_{11} & & 0 \\ & \ddots & \\ 0 & & \Omega^{PP}F_{PP} \end{bmatrix} + \begin{bmatrix} \Omega^{11}G_{11} & & 0 \\ & \ddots & \\ 0 & & \Omega^{PP}G_{PP} \end{bmatrix}$$

$$-\begin{bmatrix} \sum_p J^1_{1p}J^{1'}_{1p} & \cdots & \sum_p J^P_{1p}J^{1'}_{Pp} \\ & \ddots & \\ & & \sum_p J^P_{Pp}J^{P'}_{Pp} \end{bmatrix}$$

$$-\begin{bmatrix} \sum_p L_{1p} & & 0 \\ & \ddots & \\ 0 & & \sum_p L_{Pp} \end{bmatrix}$$

3.4. *Conclusion*

The general form of the Cramer-Rao bound can be decomposed in five parts:

$$-C_1 + C_2 + C'_2 - C_3 - C_4$$

where C_2 would be the Cramer-Rao bound if there were no endogenous variables on the right, C_3, C'_2, C_4 represent the modification due to the existence of endogenous variables on the right when Ω is known, and $-C_1$ 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 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 nonlinear case, the derivation is much more complicated.

4. Minimum Distance

We next consider a family of minimum distance estimators of the parameter β in the system of nonlinear equations (2). We obtain the minimum distance estimator by minimizing

$$J(\beta) = [Y - F(Z,\beta)]'S[Y - F(Z,\beta)]$$

where

$$S = (I \otimes X)[\hat{\Omega} \otimes X'X]^{-1}(I \otimes X')$$

with X defined above, $\hat{\Omega}$ a consistent estimator of order $0(T^{-1/2})$ of Ω.

A1. The parameter space is compact and the matrix $X'X$ is of full rank with probability one.

Proposition 1. Under A_0, A_1, a minimum distance estimator exists.

A2. $u_t, t = 1, \ldots, T$ are identical independently distributed random vectors.

A3. $\lim_{T\to\infty} \frac{1}{T} X'X$ exists and is equal to the nonsingular matrix M.

A4. $\operatorname{plim} \frac{1}{T} X' \frac{\partial f_i}{\partial \beta'} = H_i$ uniformly in β. Then $\operatorname{plim} \frac{1}{T} X' \frac{\partial f}{\partial \beta'} = \begin{bmatrix} H_1 \\ \vdots \\ H_P \end{bmatrix} = H$ of rank R uniformly in β, with the notation:

$$\left(\frac{\partial f_i}{\partial \beta'}\right)' = \frac{\partial f_i'}{\partial \beta} = \begin{bmatrix} \frac{\partial f_i}{\partial \beta^1}(z_{i1}, \beta), & \ldots, & \frac{\partial f_i}{\partial \beta^1}(z_{iT}, \beta) \\ \vdots & & \\ \frac{\partial f_i}{\partial \beta^R}(z_{i1}, \beta), & \ldots, & \frac{\partial f_i}{\partial \beta^R}(z_{iT}, \beta) \end{bmatrix}.$$

Proposition 2. Under A0, A2 to A4 a minimum distance estimator is consistent.

Proof.

$$F(Z, \hat{\beta}) - F(Z, \beta^0) = \begin{bmatrix} \frac{\partial f_1}{\partial \beta'} \\ \vdots \\ \frac{\partial f_P}{\partial \beta'} \end{bmatrix}_{\tilde{\beta}} (\hat{\beta} - \beta^0) \qquad \text{with } \tilde{\beta} \text{ between } \hat{\beta} \text{ and } \beta^0. \tag{47}$$

Note that $F(Z, \beta) = Y - U$. Multiplying each member of (47) by

$$\left(\frac{\hat{\Omega} \otimes X'X}{T}\right)^{-1/2} \left(\frac{I \otimes X'}{T}\right)$$

gives:

$$\left(\frac{\Omega \otimes X'X}{T}\right)^{-1/2} \cdot \left(\frac{I \otimes X'}{T}\right) U - \left(\frac{\hat{\Omega} \otimes X'X}{T}\right)^{-1/2} \left(\frac{I \otimes X'}{T}\right) (F(Z, \hat{\beta}) - Y) \tag{48}$$

$$= \left(\frac{\hat{\Omega} \otimes X'X}{T}\right)^{-1/2} \left(\frac{I \otimes X'}{T}\right) \begin{bmatrix} \frac{\partial f_1}{\partial \beta'} \\ \vdots \\ \frac{\partial f_P}{\partial \beta'} \end{bmatrix} (\hat{\beta} - \beta^0)$$

From A3, $\operatorname{plim}_{T\to\infty} \frac{X'X}{T} = M$ and $\left(\frac{I \otimes X'}{T}\right)U = \begin{bmatrix} \frac{1}{T}X'U_1 \\ \vdots \\ \frac{1}{T}X'U_P \end{bmatrix}$ which goes to zero

in probability when $T \to \infty$, by Chebychev's theorem (A2–A3). Let

$$\alpha_1 = \left(\frac{\hat{\Omega} \otimes X'X}{T}\right)^{-1/2}\left(\frac{I \otimes X'}{T}\right)U; \qquad \operatorname{plim}_{T\to\infty} \alpha_1 = 0$$

Let

$$\alpha_2 = \left(\frac{\hat{\Omega} \otimes X'X}{T}\right)^{-1/2}\left(\frac{I \otimes X'}{T}\right)(F(Z, \beta^0) - Y).$$

By definition of $\hat{\beta}$, $J(\hat{\beta}) \leq J(\beta^0)$. Then, $0 \leq \alpha_2'\alpha_2 \leq \alpha_1'\alpha_1$. Since $\alpha_1 \xrightarrow{p} 0$ $\alpha_1'\alpha_1 \xrightarrow{p} 0$, therefore $\alpha_2'\alpha_2 \xrightarrow{p} 0$ and $\alpha_2 \xrightarrow{p} 0$.

The left hand side of (48) converges to 0 in probability when $T \to \infty$. Also

$$\frac{1}{T}[I \otimes X'] \begin{bmatrix} \frac{\partial f_1}{\partial \beta'} \\ \vdots \\ \frac{\partial f_P}{\partial \beta'} \end{bmatrix} = \begin{bmatrix} \frac{1}{T}X'\frac{\partial f_1}{\partial \beta'} \\ \vdots \\ \frac{1}{T}X'\frac{\partial f_P}{\partial \beta'} \end{bmatrix}$$

which by **A4** converges to the full rank matrix H when $T \to \infty$. Consequently, we see on (48) that $\hat{\beta} \xrightarrow{p} \beta$.· Q.E.D.

A5 $\operatorname{plim}_T \frac{1}{T}X'\frac{\partial^2 f_i}{\partial \beta^j \partial \beta'} = G_j^i$ uniformly in β, $i = 1, \ldots, P$, $j = 1, \ldots, R$

A6
$$\begin{bmatrix} \frac{X'U_1}{\sqrt{T}} \\ \vdots \\ \frac{X'U_P}{\sqrt{T}} \end{bmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}[0, \Omega \otimes M]$$

Proposition 3. Under A0 to A6. $\sqrt{T}(\hat{\beta} - \beta^0) \xrightarrow{\mathcal{D}} \mathcal{N}[0, (H'(\Omega \otimes M)^{-1}H)^{-1}]$

Proof.

$$\frac{\partial J}{\partial \beta}\bigg|\beta^0 = \frac{\partial J}{\partial \beta}\bigg|\hat{\beta} + \frac{\partial^2 J}{\partial \beta \partial \beta'}\bigg|\tilde{\beta}(\hat{\beta} - \beta^0),$$

with $\tilde{\beta}$ between β^0 and $\hat{\beta}$. Then:

(49)
$$\sqrt{T}(\beta_0 - \hat{\beta}) = -\left[\frac{1}{T}\frac{\partial^2 J}{\partial \beta \partial \beta'}\bigg|\tilde{\beta}\right]^{-1}\frac{1}{\sqrt{T}}\frac{\partial J}{\partial \beta}\bigg|\beta^0$$

$$J(\beta) = [Y - F(Z, \beta)]'[I \otimes X][\hat{\Omega}^{-1} \otimes (X'X)^{-1}][I \otimes X'][Y - F(Z, \beta)]$$

$$(50)\quad \frac{\partial J(\beta)}{\partial \beta} = -2\begin{bmatrix} \frac{\partial f_1'}{\partial \beta} \\ \vdots \\ \frac{\partial f_P'}{\partial \beta} \end{bmatrix}[I \otimes X][\hat{\Omega}^{-1} \otimes (X'X)^{-1}][I \otimes X'][Y - F(Z, \beta)]$$

$$\frac{\partial^2 J(\beta)}{\partial \beta \partial \beta'} = 2\begin{bmatrix} \frac{\partial f_1'}{\partial \beta} \\ \vdots \\ \frac{\partial f_P'}{\partial \beta} \end{bmatrix}[I \otimes X][\hat{\Omega}^{-1} \otimes (X'X)^{-1}][I \otimes X']\begin{bmatrix} \frac{\partial f_1}{\partial \beta'} \\ \vdots \\ \frac{\partial f_P}{\partial \beta'} \end{bmatrix} - 2\begin{bmatrix} a_1 \\ \vdots \\ a_P \end{bmatrix}$$

with

$$a_j = [Y - F(Z, \beta)]'[I \otimes X][\hat{\Omega}^{-1} \otimes (X'X)^{-1}][I \otimes X']\begin{bmatrix} \frac{\partial^2 f_1}{\partial \beta^j \partial \beta'} \\ \vdots \\ \frac{\partial^2 f_P}{\partial \beta^j \partial \beta'} \end{bmatrix}$$

To obtain the asymptotic distribution of $\hat{\beta}$ we will first derive the asymptotic distribution of the pseudoestimator $\bar{\beta}$ obtained by replacing $\hat{\Omega}$ by Ω. $[J \to \bar{J}]$. By assumption $\hat{\Omega} = \Omega + \Delta_1$ with $\Delta_1 \sim 0(T^{-1/2})$ so

$$\hat{\Omega}^{-1} = \Omega^{-1} - \Omega^{-1}\Delta_1\Omega^{-1} + \ldots = \Omega^{-1} - \Delta_2 \quad \text{with } \Delta_2 \sim 0(T^{-1/2})$$

or

$$\hat{\Omega}^{ij} = \Omega^{ij} - \delta_{ij} \quad \text{with } \delta_{ij} \sim 0(T^{-1/2}) \quad i = 1, \ldots, P, \quad j = 1, \ldots, P.$$

For $\bar{\beta}$, a_j becomes:

$$(51)\qquad \frac{U'[I \otimes X]}{T}\left[\Omega^{-1} \otimes \left(\frac{X'X}{T}\right)^{-1}\right]\begin{bmatrix} \frac{1}{T}X'\frac{\partial^2 f_1}{\partial \beta^j \partial \beta'} \\ \frac{1}{T}X'\frac{\partial^2 f_1}{\partial \beta^j \partial \beta'} \end{bmatrix}.T$$

As already shown, the first member of (51) converges in probability to zero as $T \to \infty$, the second member to the matrix $\Omega^{-1} \otimes M^{-1}$ by A3, and the third member to the matrix

$$\begin{bmatrix} G_j^1 \\ \vdots \\ G_j^P \end{bmatrix} \text{ so that (49)} \xrightarrow{P} 0 \text{ as } T \to \infty.$$

Now,

$$\frac{1}{2}\operatorname{plim}\frac{1}{T}\frac{\partial^2 J}{\partial\beta\partial\beta'}\bigg|\beta^0 = \operatorname{plim}\begin{bmatrix}\frac{1}{T}\cdot\frac{\partial f_1'}{\partial\beta}X\\ \vdots\\ \frac{1}{T}\cdot\frac{\partial f_P'}{\partial\beta}X\end{bmatrix}\left[\Omega\otimes\frac{X'X}{T}\right]^{-1}\begin{bmatrix}\frac{1}{T}X'\frac{\partial f_1}{\partial\beta'}\\ \vdots\\ \frac{1}{T}X'\frac{\partial f_P}{\partial\beta'}\end{bmatrix}$$

$\tilde{\beta}$ which lies between $\hat{\beta}$ and β^0 converges to β^0 when $T\to\infty$ because of the consistency of $\hat{\beta}$. Then:[19]

$$\operatorname*{plim}_{T\to\infty}\frac{1}{T}\frac{\partial^2 J}{\partial\beta\partial\beta'}\bigg|\tilde{\beta} = \operatorname*{plim}_{T\to\infty}\frac{1}{T}\frac{\partial^2 J}{\partial\beta\partial\beta'}\bigg|_{\beta_0}$$

But:

$$\frac{1}{2}\operatorname*{plim}_{T\to\infty}\frac{1}{T}\frac{\partial^2 J}{\partial\beta\partial\beta'}\bigg|_{\beta_0} = H'(\Omega\otimes M)^{-1}H.$$

From (50)

$$-\frac{1}{2}\frac{1}{\sqrt{T}}\frac{\partial\bar{J}}{\partial\beta} = \begin{bmatrix}\frac{1}{T}\frac{\partial f_1'}{\partial\beta}X\\ \vdots\\ \frac{1}{T}\frac{\partial f_P'}{\partial\beta}X\end{bmatrix}\left[\Omega\otimes\frac{X'X}{T}\right]^{-1}\begin{bmatrix}\frac{1}{\sqrt{T}}X'U_1\\ \vdots\\ \frac{1}{\sqrt{T}}X'U_P\end{bmatrix}$$

Using A2, A3, A6, we can derive:

$$\operatorname{Var}\operatorname*{plim}_{T\to\infty}\left[-\frac{1}{2}\frac{1}{\sqrt{T}}\frac{\partial\bar{J}}{\partial\beta}\bigg|\beta^0\right] = H'(\Omega\otimes M)^{-1}.$$

Finally from (49) $\operatorname{Var}\operatorname{plim}\sqrt{T}(\tilde{\beta}-\beta^0) = (H'(\Omega\otimes M)^{-1}H)^{-1}$ so that from A6

$$\sqrt{T}(\tilde{\beta}-\beta_0)\xrightarrow{\mathcal{D}}\mathcal{N}[0,(H'(\Omega\times M)^{-1}H)^{-1}].$$

Let us now consider $\hat{\beta}$:

$$\frac{\partial J}{\partial\beta}\bigg|\beta^0 = -2\begin{bmatrix}\hat{\Omega}^{11}\cdot\frac{\partial f_1'}{\partial\beta}X(X'X)^{-1}X'U_1 & \cdots & \hat{\Omega}^{1P}\frac{\partial f_P'}{\partial\beta}X(X'X)^{-1}X'U_P\\ & \ddots & \\ & & \hat{\Omega}^{PP}\frac{\partial f_P'}{\partial\beta}X(X'X)^{-1}X'U_P\end{bmatrix}$$

$$= -2\begin{bmatrix}\Omega^{11}\cdot\frac{\partial f_1'}{\partial\beta}X(X'X)^{-1}X'U_1-\delta_{11}\frac{\partial f_1'}{\partial\beta}X(X'X)^{-1}X'U_1+\cdots & \cdots & \\ & \ddots & \\ & & \Omega^{PP}\cdot\frac{\partial f_P'}{\partial\beta}X(X'X)^{-1}X'U_P-\delta_{PP}\cdot\frac{\partial f_P'}{\partial\beta}X(X'X)^{-1}X'U_P\end{bmatrix}$$

[19] Amemiya (1972), p. 10, Lemma 4: let $f_T(\omega,\theta)$ be a measurable function on a measurable space Ω and for each ω in Ω a continuous function for θ in a compact set H. If $f_T(\omega,\theta)$ converges to $f(\theta)$ a.e. uniformly for all θ in H and if $\theta_T(\omega)$ converges to θ_0 a.e., then $f_T(\omega,\theta_T(\omega))$ converges to $f(\theta_0)$ a.e.

Let us consider an element a_j.

$$a_j = U'[\Omega^{-1} \otimes X(X'X)^{-1}X'] \begin{bmatrix} \dfrac{\partial^2 f_1}{\partial\beta^j \partial\beta'} \\ \vdots \\ \dfrac{\partial^2 f_P}{\partial\beta^j \partial\beta'} \end{bmatrix}$$

$$= \left[\sum_{k=1}^{P} \hat{\Omega}^{k1} U_k' X(X'X)^{-1}X' \begin{bmatrix} \dfrac{\partial^2 f_1}{\partial\beta^j \partial\beta^1} \\ \\ \dfrac{\partial^2 f_P}{\partial\beta^j \partial\beta^1} \end{bmatrix}, \ldots\ldots, \sum_{k=1}^{P} \hat{\Omega}^{kP} U_k' X'(X'X)^{-1}X' \begin{bmatrix} \dfrac{\partial^2 f_1}{\partial\beta^j \partial\beta^R} \\ \vdots \\ \dfrac{\partial^2 f_P}{\partial\beta^j \partial\beta^R} \end{bmatrix} \right]$$

The matrix $X(X'X)^{-1}X'\,\partial^2 f_h/\partial\beta^j\partial\beta^i$ is the projection of $\partial^2 f_h/\partial\beta^j\partial\beta^i$ on X, so that it is independent of u. Then $U_k'X(X'X)^{-1}X'(\partial^2 f_h/\partial\beta^j\partial\beta^i) \sim 0(T^{1/2})$. A current element of $\partial^2 J/\partial\beta\partial\beta'$ is then of the form

$$(\Omega^{ij} - \delta^{ij})\frac{\partial f_i'}{\partial\beta} X(X'X)^{-1}X' \frac{\partial f_j}{\partial\beta'} - \Omega^{ij} \cdot 0(T^{1/2}) - \delta^{ij} \cdot 0(T^{1/2})$$

or:

$$\Omega^{ij} \cdot \frac{\partial f_i'}{\partial\beta} X(X'X)^{-1}X' \frac{\partial f_j}{\partial\beta'}[1 + 0(T^{-1/2}) + 0(T^{-1})].$$

The matrix $\partial^2 J/\partial\beta\partial\beta'$ can then be rewritten symbolically:

$$A(I + 0(T^{-1/2})) \text{ with } A = \left(\Omega^{ij} \cdot \frac{\partial f_i'}{\partial\beta} X(X'X)^{-1}X' \cdot \frac{\partial f_j}{\partial\beta'}\right)$$

Then:

$$[A(I + 0(T^{-1/2}))]^{-1} = A^{-1} - 0(T^{-1/2})A^{-1} = A^{-1} - 0(T^{-3/2}).$$

$$\frac{\partial J}{\partial\beta} = B + 0(T^0) \text{ with } B = -2 \begin{bmatrix} \dfrac{\partial f_1'}{\partial\beta} \\ \vdots \\ \dfrac{\partial f_R'}{\partial\beta} \end{bmatrix} \hat{\Omega} \otimes X(X'X)^{-1}X' \cdot U; \quad B \sim 0(T^{1/2})$$

so that

$$\beta_0 - \hat{\beta} = [A^{-1} - 0(T^{-3/2})][B - 0(T^0)]$$
$$= A^{-1}B - 0(T^{-1}) - 0(T^{-1}) - 0(T^{-3/2}).$$

By Cramer's theorem (1971, p. 254) $\hat{\beta} \xrightarrow{a} \bar{\beta}$.

$$\hat{\beta} - \beta_0 = \bar{\beta} - \beta_0 + 0(T^{-1}).$$

Since $\sqrt{T}(\hat{\beta} - \beta_0)$ is normal, the asymptotic distribution of $T^{1/2}(\hat{\beta} - \beta_0)$ is the same as that of $\sqrt{T}(\bar{\beta} - \beta_0)$ because the difference of these two quantities $\sqrt{T}0(T^{-1})$ has zero probability limit. Q.E.D.

5. Instrumental Variables

To generate a family of instrumental variables estimators of the parameter β in the system of nonlinear equations (2), we linearize the system around the true value, say β^0:

$$(52)\qquad \begin{aligned} y_{1t} - f_1(z_{1t}, \beta_1^0) &= \sum_{j=1}^{R_1} f_{1jt} \cdot (\beta_{1j} - \beta_{1j}^0) + u_{1t} \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ y_{Pt} - f_P(z_{Pt}, \beta_P^0) &= \sum_{j=1}^{R_P} f_{Pjt} \cdot (\beta_{Pj} - \beta_{Pj}^0) + u_{Pt} \qquad t = 1, \ldots, T \end{aligned}$$

with $f_{ijt} = \partial f_i / \partial \beta_{ij}(z_{it}, \beta_i^0)$.

In general f_{ijt} which depends on endogenous variables is correlated with errors. Consider the estimation of (52) by the instrumental variables method.[20]

We denote the set of instrumental variables

$$\overline{W} = \begin{bmatrix} W_{11} & W_{12} & \cdots & W_{1P} \\ \hline W_{P1} & W_{P2} & \cdots & W_{PP} \end{bmatrix}$$

where each submatrix W_{ij} has R_j columns. (52) can be rewritten $Y = F(\beta - \beta^0)$ with

$$F = \begin{bmatrix} \begin{matrix} f_{111} & \cdots & f_{1R_11} \\ f_{11T} & \cdots & f_{1R_1T} \end{matrix} & & 0 \\ & \ddots & \\ 0 & & \begin{matrix} f_{P11} & \cdots & f_{PR_P1} \\ f_{P1T} & \cdots & f_{PR_PT} \end{matrix} \end{bmatrix} = \begin{bmatrix} \left.\dfrac{\partial f_1}{\partial \beta_1'}\right|\beta^0 & & 0 \\ & \ddots & \\ 0 & & \left.\dfrac{\partial f_P}{\partial \beta_P'}\right|\beta^0 \end{bmatrix}$$

The instrumental variables estimator is then

$$\hat{\beta} - \beta^0 = (\overline{W}'F)^{-1}\overline{W}'(Y - Y_0) = (\overline{W}'F)^{-1}\overline{W}'U$$

so that:

$$\operatorname{Var}\sqrt{T}(\hat{\beta} - \beta_0) = \operatorname{plim}\left(\frac{1}{T}\overline{W}'F\right)^{-1}\frac{\overline{W}'}{T}(\Omega \otimes I)\overline{W}\left(\frac{1}{T}F'\overline{W}\right)^{-1}.$$

If we choose:

$$\overline{W} = \overline{X}(\Omega \otimes X'X)^{-1}\overline{X}'F \text{ with } \overline{X} = I \otimes X$$

[20] It is only a pseudomodel (since β^0 is not known), for which we construct a pseudoestimator.

we find

$$\operatorname{plim} \frac{1}{T}(F'\bar{X}'(\Omega \otimes X'X)^{-1}\bar{X}'F)^{-1}\frac{1}{T}F'\bar{X}'(\Omega \otimes X'X)^{-1}\bar{X}'(\Omega \otimes I)\bar{X}(\Omega \otimes X'X)^{-1}\bar{X}'F$$

$$\left(\frac{1}{T}F'\bar{X}(\Omega \otimes X'X)^{-1}\bar{X}'F\right)^{-1}$$

since

$$\bar{X}'(\Omega \otimes I)\bar{X} = (I \otimes X')(\Omega \otimes I)(I \otimes X) = \Omega \otimes X'X$$

$$\operatorname{plim} \frac{1}{T}(F'(I \otimes X)(\Omega \otimes X'X)^{-1}(I \otimes X')F)^{-1}.$$

It then appears that the minimum distance estimator is asymptotically equivalent to the pseudoestimator with a specific choice of instrumental variables.

If now we can find the best set of instrumental variables $\bar{W}^*$, the best choice of X will be X^* such that:

$$\bar{W}^* = \bar{X}^*(\Omega \otimes X^{*\prime}X^*)^{-1}\bar{X}^{*\prime}F \tag{53}$$

at least asymptotically so that F and Ω can be replaced by consistent estimators. The search for a best set of instrumental variables will reveal the nature of the difficulty.

In the linear case an efficient set of instrumental variables is:

$$W_{ij} = \hat{\Omega}^{ij}W_j$$

where $\hat{\Omega}^{ij}$ is a typical element of $\hat{\Omega}^{-1}$ (with $\hat{\Omega}$ being a consistent estimator of Ω) and W_j is a consistent estimator of the systematic part of variables in the j-th equation independent of the errors U. So, by analogy, we can attempt to construct consistent estimators of the systematic part of derivatives $\partial f_i/\partial \beta_{ij}(z_{it}, \beta_i^0)$ which are not correlated with the U. It is then clear that we want X to be variables independent of U but nevertheless as closed as possible to $\partial f_i/\partial \beta_i'|_{\beta_0}$, $i = 1, \ldots, P$. Since there is no constraint on the number of X, as many powers of X as possible seem ideal. However, after some n the powers become probably useless, the n depending on the degree of nonlinearity of the derivatives. Moreover, this method leads to a huge matrix $X'X$ which we have to invert, and leads to X which are collinear. Our suggestion is then the following one:

1. Find a consistent estimator of β using a NL2SLS estimator for example with a minimum set of X ($\operatorname{Max}_i P_{ij}$), taken as a subset (eventually) of exogenous variables.
2. Simulate the model to obtain values of endogenous variables.
3. Use the results of 1 and 2 to approximate the derivatives of f_i, $i = 1, \ldots, P$.

Malinvaud (1970b) restricts himself to the case where a reduced form is available and shows that a minimum distance estimator with $S = \Omega^{-1}$ or a consistent estimator of Ω gives the best minimum distance estimator. Moreover,

if normality is assumed, it is asymptotically efficient. In his case the

$$\frac{\partial f_i}{\partial \beta_i'}(z_{it}, \beta_i^0)$$

are not correlated with the U, so that if a consistent estimator β_i of β_i^0 is available, it is not necessary to project them to eliminate the dependence on U.

They are the best possible auxiliary variables since they obviously maximize asymptotically

$$\operatorname{tr} \frac{\partial f_i'}{\partial \beta_i}\bigg|_{\beta_0} \cdot X(X'X)^{-1}X' \cdot \frac{\partial f_i}{\partial \beta_i'}\bigg|_{\beta_0}$$

But:

$$\operatorname*{plim}_{T\to\infty} \operatorname{tr} \frac{\partial f_i'}{\partial \beta_i}\bigg|_{\beta_0} X(X'X)^{-1}X' \frac{\partial f_i}{\partial \beta_i'}\bigg|_{\beta_0} = \left\| \frac{\partial f_i'}{\partial \beta_i}\bigg|_{\beta_0} \right\|$$

Malinvaud tells us that asymptotically it is not worth the trouble since $X = I$ does as well.

6. Efficiency

6.1. *Introduction*

We have developed an explicit form for the Cramer-Rao lower bound for the variance of a CUAN estimator of the parameter β in the system of nonlinear simultaneous equations (2). This bound is attained by the full information maximum likelihood estimator. We have shown that the minimum distance and efficient instrumental variables estimators are asymptotically equivalent. We next consider the relative efficiency of the minimum distance estimator for a system of nonlinear simultaneous equations and for a single equation in such a system.

6.2. *Minimum Distance Versus Maximum Likelihood*

We show with an example that the minimum distance estimator of Section 4 does not generally attain the Cramer-Rao bound. It is sufficient to prove that one element in the inverse of the matrix of variances and covariances of the minimum distance estimator is different from the corresponding element in the inverse of the corresponding matrix for the Cramer-Rao bound. We consider the system of nonlinear simultaneous equations:

$$\begin{aligned} y_1 + \alpha_1 y_2^2 + \beta_1 x_1 &= u_1 \\ y_2 + \beta_2 x_2 &= u_2 \end{aligned} \qquad u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

$$Eu'u = \begin{bmatrix} \Omega_{11} & 0 \\ 0 & \Omega_{22} \end{bmatrix} \quad \text{with } \Omega_{11}\Omega_{22} = 1$$
(Therefore $\Omega^{11} = \Omega_{22}$ and $\Omega^{22} = \Omega_{11}$)

Element (1, 1) in the inverse of the Cramer-Rao bound:

a) From (46*) we have: $2\Omega^{11}\cdot\Omega^{11}H_{11}H_{11}' + 2\Omega^{11}\Omega^{21}H_{11}H_{21}'$

$$H_{11} = \lim\frac{1}{T}\sum_t f_{1t}\frac{\partial f_{1t}}{\partial\theta_1} = \lim\frac{1}{T}\sum_t u_{1t}\begin{bmatrix} y_{2t}^2 \\ x_{1t}\end{bmatrix} = \begin{bmatrix}0\\0\end{bmatrix}$$

Then:

$$C_1(1,1) = 0$$

b) From (46*) we have: $\Omega^{11}F_{11}$

$$F_{11} = \lim\frac{1}{T}\sum_t\frac{\partial f_{1t}}{\partial\theta_1}\cdot\frac{\partial f_{1t}}{\partial\theta_1'} = \lim\frac{1}{T}\sum_t\begin{bmatrix} y_{2t}^4 & / \\ / & /\end{bmatrix};\ \lim\frac{1}{T}\sum_t y_{2t}^4 = \lim\frac{1}{T}\sum_t(u_2-\beta_2x_2)^4$$

$$= \lim\frac{1}{T}\left(\sum_t u_{2t}^4 - 4\sum_t u_{2t}^3\beta_2x_{2t} + 6\sum_t u_{2t}^2\beta_2^2x_{2t}^2 - 4\sum_t u_{2t}\beta_2^3x_{2t}^3 + \sum_t\beta_2^4x_2^4\right)$$

$$= \sigma_2^4 + 6\beta_2^2\Sigma_{22}^{22} + \beta_2^4\Sigma_{22}^4$$

with

$$\sigma_2^4 = \lim\frac{1}{T}\sum_t u_{2t}^4 \qquad \Sigma_{22}^4 = \lim\frac{1}{T}\sum_t x_{2t}^4,\ \Sigma_{22}^{22} = \lim\frac{1}{T}\sum_t x_{2t}^2u_{2t}^2$$

$$C_2(1,1) = \Omega^{11}[\sigma_2^4 + 6\beta_2^2\Sigma_{22}^{22} + \beta_2^4\Sigma_{22}^4]$$

Clearly $C_2' = 0$ and $C_4 = 0$.

c) From (46*) we have: $J_{11}^1\cdot J_{11}^{1'} + J_{12}^1\cdot J_{12}^{1'}$:

Here:

$$\begin{bmatrix}B_{11t} & B_{12t}\\ B_{21t} & B_{22t}\end{bmatrix} = \begin{bmatrix}1 & 2\alpha_1y_{2t}\\ 0 & 1\end{bmatrix};\begin{bmatrix}B_t^{11} & B_t^{12}\\ B_t^{21} & B_t^{22}\end{bmatrix} = \begin{bmatrix}1 & -2\alpha_1y_{2t}\\ 0 & 1\end{bmatrix}$$

$$J_{11}^1 = \lim\frac{1}{T}\sum_t B_t^{11}\frac{\partial^2 f_{1t}}{\partial\theta_1\partial y_1} = \lim\frac{1}{T}\sum_t 1\begin{bmatrix}0\\0\end{bmatrix} = 0$$

$$J_{12}^2 = \lim\frac{1}{T}\sum_t B_t^{21}\frac{\partial^2 f_{1t}}{\partial\theta_1\partial y_2} = \lim\frac{1}{T}\sum_t 0\cdot\begin{bmatrix}2y_{2t}\\0\end{bmatrix} = 0$$

Then: $C_3(1,1) = 0$.

Finally, we obtain the element (1, 1) of the inverse of the Cramer-Rao matrix:

$$\Omega^{11}(\sigma_2^4 + 6\beta_2^2\Sigma_{22}^{22} + \beta_2^4\Sigma_{22}^4)$$

Element (1.1) of the inverse of the asymptotic matrix of variances and covariances of the minimum distance estimator is:

$$\lim\frac{1}{T}\Omega^{11}\begin{bmatrix}y_2^{2'}\\x_1'\end{bmatrix}[x_1\quad x_2]\begin{bmatrix}x_1'x_1 & x_1'x_2\\ x_2'x_1 & x_2'x_2\end{bmatrix}\begin{bmatrix}x_1'\\x_2'\end{bmatrix}[y_2^2\quad x_1]$$

Let

$$\Sigma_{221} = \lim\frac{1}{T}\sum_t x_{2t}^2x_{1t} \qquad \Theta_{21} = \lim\frac{1}{T}\sum_t u_{2t}^2x_{1t}$$

$$\Sigma_{222} = \lim \frac{1}{T} \sum_t x_{2t}^2 x_{2t} \qquad \Theta_{22} = \lim \frac{1}{T} \sum_t u_{2t}^2 x_{2t}$$

The element (1. 1) is then:

$$\frac{\Omega^{11}}{\Sigma_{22}\Sigma_{11} - \Sigma_{12}^2} \cdot [\beta_2^4(\Sigma_{221}^2\Sigma_{22} + \Sigma_{222}^2\Sigma_{11} - 2\Sigma_{222}\Sigma_{221}\Sigma_{12}) + 2\beta_2^2(\Theta_{21}\Sigma_{22}\Sigma_{221} + \Theta_{22}\Sigma_{11}\Sigma_{222} - \Theta_{21}\Sigma_{12}\Sigma_{222} - \Theta_{22}\Sigma_{12}\Sigma_{221}) + \Theta_{21}^2\Sigma_{22} + \Theta_{22}^2\Sigma_{11} - 2\Theta_{21}\Theta_{22}\Sigma_{12}]$$

which differs from the corresponding element of the Cramer-Rao bound.

6.3. *Limited Information Versus Full Information*

We can also consider the relative asymptotic efficiency of the minimum distance estimator for a system of nonlinear simultaneous equations proposed in Section 4 with the corresponding estimator for a single equation developed by Amemiya (1974). We consider only the case without restrictions across equations. A minimum distance estimator is obtained for each equation $i = 1, \ldots, P$ by minimizing

$$[y_i - f_i(z_i, \beta_i)]'X(X'X)^{-1}X'[y_i - f_i(z_i, \beta_i)]$$

for a given choice of X. To each choice of X corresponds also a minimum distance estimator as defined in Section 4. We will show that the corresponding estimator for a system of nonlinear simultaneous equations is always asymptotically better (or as good) as the estimator for a single equation.

The asymptotic matrix of variances and covariances of the single equation estimator, given by Amemiya (1974), is:

$$\begin{bmatrix} \Omega_{11}^{-1}H_1'M^{-1}H_1 & & 0 \\ & \ddots & \\ 0 & & \Omega_{PP}^{-1}H_P'M^{-1}H_P \end{bmatrix}^{-1}$$

$$= \operatorname*{plim}_{T\to\infty} T \begin{bmatrix} \Omega_{11}^{-1}\dfrac{\partial f_1'}{\partial \beta_1}\Big|_\beta X(X'X)^{-1}X'\dfrac{\partial f_1}{\partial \beta_1'}\Big|_{\beta_0} & & 0 \\ & \ddots & \\ 0 & & \Omega_{PP}^{-1}\dfrac{\partial f_P'}{\partial \beta_P}\Big|_{\beta_0} X(X'X)^{-1}X'\dfrac{\partial f_P}{\partial \beta_P'} \end{bmatrix}^{-1}$$

The asymptotic matrix of variances and covariances of the estimator for a system of equations, given above, is:

$$(H'(\Omega \otimes M)^{-1}H)^{-1} = \begin{bmatrix} \Omega^{11}H_1'M^{-1}H_1 & \Omega^{12}H_1'M^{-1}H_2 & \\ \Omega^{21}H_2'M^{-1}H_1 & \Omega^{22}H_2'M^{-1}H_2 & \\ & & \ddots \\ & & \Omega^{PP}H_P'M^{-1}H_P \end{bmatrix}^{-1}$$

$$= \operatorname*{plim}_T T \begin{bmatrix} \Omega^{11} \dfrac{\partial f_1'}{\partial \beta_1}\Big|_{\beta_0} X(X'X)^{-1}X' \dfrac{\partial f_1}{\partial \beta_1'}\Big|_{\beta_0} & \cdots & \Omega^{1P} \dfrac{\partial f_1'}{\partial \beta_1}\Big|_{\beta_0} X(X'X)^{-1}X' \dfrac{\partial f_P}{\partial \beta_P'}\Big|_{\beta_0} \\ \Omega^{P1} \dfrac{\partial f_P'}{\partial \beta_P}\Big|_{\beta_0} X(X'X)^{-1}X' \dfrac{\partial f_1}{\partial \beta_1'}\Big|_{\beta_0} & \cdots & \Omega^{PP} \dfrac{\partial f_P'}{\partial \beta_P}\Big|_{\beta_0} X(X'X)^{-1}X' \dfrac{\partial f_P}{\partial \beta_P'}\Big|_{\beta_0} \end{bmatrix}^{-1}$$

It is then clear that if we replace $\partial f_j/\partial \beta_j'|_{\beta_0}$ by the familiar Z_j (the set of variables on the right hand side of the j-th structural equation), the *formal* analogy with the classical comparison of 3SLS and 2SLS is complete. Note that here X is not necessarily the set of all exogenous variables. Therefore, the usual proof for linear systems implies our result.

We can also deduce that the two estimators coincide when Ω is diagonal and when

$$\operatorname*{plim}_{T\to\infty} \frac{1}{T} \frac{\partial f_i'}{\partial \beta_i}\Big|_{\beta_0} X$$

is invertible for $i = 1, \ldots, P$. The invertibility condition means in particular that the matrix is square, i.e., that in each equation there are as many independent variables X as unknowns. We have seen that A4 requires

$$\operatorname{Rank} \operatorname*{plim} \frac{1}{T} \frac{\partial f_i'}{\partial \beta_i}\Big|_{\beta_0} X \geq R_i.$$

If the matrix X is restricted to exogenous variables, the condition we find is similar to the just identification in the linear case. However, in the nonlinear case we know that it is not a necessary condition of identification (Fisher, 1966).[21] Since we do not have to restrict ourselves to exogenous variables and can use powers of the X or fitted values, the condition

$$\operatorname*{plim} \frac{1}{T} \frac{\partial f_i'}{\partial \beta_i}\Big|_{\beta_0} X \geq R_i$$

is not really a constraint so long as the model is truly nonlinear. The condition

$$\operatorname*{plim} \frac{1}{T} \frac{\partial f_i'}{\partial \beta_i}\Big|_{\beta_0} X = R_i$$

imposes a limit on the number of elements of X to use so that the two estimators are equivalent.

We derive directly the result for the two-equations case

$$y_1 = f_1(z_1, \beta_1) + u_1$$

$$y_2 = f_2(z_2, \beta_2) + u_2.$$

The single-equation estimator is obtained by minimizing

$$(y_i - f_i(z_i, \beta_i))' X(X'X)^{-1}X'(y_i - f_i(z_i, \beta_i))$$

[21] This condition has been obtained also in Edgerton (1972), as a necessary condition for the workability of 2SLS methods suggested by Goldfeld and Quandt (1968). Note that it is not really a constraint for the method since as many powers of the X as necessary can be introduced to satisfy it.

and the asymptotic matrix of variances and covariances is

$$\Omega_{ii}\left[\operatorname{plim}\frac{1}{T}\frac{\partial f_i'}{\partial \beta_i}\bigg|_{\beta_0} X(X'X)^{-1}X'\frac{\partial f_i}{\partial \beta_i'}\bigg|_{\beta_0}\right]^{-1}$$

We are led to compare

$$K = \Omega_{11}\left[\operatorname{plim}\frac{1}{T}\frac{\partial f_1'}{\partial \beta_1}\bigg|_{\beta_0} X(X'X)^{-1}X'\frac{\partial f_1}{\partial \beta_1'}\bigg|_{\beta_0}\right]^{-1}$$

and the upper left corner of the inverse of

$$\operatorname*{plim}_{T\to\infty}\begin{bmatrix}\frac{1}{T}\frac{\partial f_1'}{\partial \beta_1}\Big|_{\beta_0} X & 0\\ 0 & \frac{1}{T}\frac{\partial f_2'}{\partial \beta_2}\Big|_{\beta_0} X\end{bmatrix}\left\{\begin{bmatrix}\Omega_{11} & \Omega_{12}\\ \Omega_{21} & \Omega_{22}\end{bmatrix}\otimes\frac{X'X}{T}\right\}^{-1}$$

$$\begin{bmatrix}\frac{1}{T}X'\frac{\partial f_1}{\partial \beta_1'}\Big|_{\beta_0} & 0\\ 0 & \frac{1}{T}X'\frac{\partial f_2}{\partial \beta_2'}\Big|_{\beta_0}\end{bmatrix}$$

or

$$\operatorname*{plim}_{T\to\infty}\begin{bmatrix}\Omega^{11}\frac{\partial f_1'}{\partial \beta_1}\Big|_{\beta_0} X(X'X)^{-1}X'\frac{\partial f_1}{\partial \beta_1'}\Big|_{\beta_0} & \Omega^{12}\frac{\partial f_1'}{\partial \beta_1}\Big|_{\beta_0} X(X'X)^{-1}X'\frac{\partial f_2}{\partial \beta_2'}\Big|_{\beta_0}\\ \Omega^{21}\frac{\partial f_2'}{\partial \beta_2}\Big|_{\beta_0} X(X'X)^{-1}X'\frac{\partial f_1}{\partial \beta_1'}\Big|_{\beta_0} & \Omega^{22}\frac{\partial f_2'}{\partial \beta_2}\Big|_{\beta_0} X(X'X)^{-1}X'\frac{\partial f_2}{\partial \beta_2'}\Big|_{\beta_0}\end{bmatrix}^{-1}$$

The upper left corner we look for is then:

$$\begin{aligned}K_1 = \Bigg\{\operatorname{plim}\frac{1}{T}\Bigg\{&\Omega^{11}\frac{\partial f_1'}{\partial \beta_1}\Big|_{\beta_0} X(X'X)^{-1}X'\frac{\partial f_1}{\partial \beta_1'}\Big|_{\beta_0} - \Omega^{12}\frac{\partial f_1'}{\partial \beta_1}\Big|_{\beta_0} X(X'X)^{-1}X'\frac{\partial f_2}{\partial \beta_2'}\Big|_{\beta_0}\\ &\times\left[\Omega^{22}\frac{\partial f_2'}{\partial \beta_2}\Big|_{\beta_0} X(X'X)^{-1}X'\frac{\partial f_2}{\partial \beta_2'}\Big|_{\beta_0}\right]^{-1}\\ &\times\Omega^{21}\left[\frac{\partial f_2'}{\partial \beta_2}\Big|_{\beta_0} X(X'X)^{-1}X'\frac{\partial f_1}{\partial \beta_1'}\Big|_{\beta_0}\right]\Bigg\}\Bigg\}^{-1}.\end{aligned}$$

$$\begin{aligned}K_1^{-1} = \operatorname{plim}\frac{1}{T}\Bigg\{&\frac{1}{\Omega^{11}}\cdot\frac{\partial f_1'}{\partial \beta_1}\Big|_{\beta_0} X(X'X)^{-1}X\frac{\partial f_1}{\partial \beta_1'}\Big|_{\beta_0} + 0\left[\frac{\partial f_1'}{\partial \beta_1}\Big|_{\beta_0} X(X'X)^{-1}X'\frac{\partial f_1}{\partial \beta_1'}\Big|_{\beta_0}\right.\\ &-\frac{\partial f_1'}{\partial \beta_1}\Big|_{\beta_0} X(X'X)^{-1}X'\frac{\partial f_2}{\partial \beta_2'}\Big|_{\beta_0}\left(\frac{\partial f_2'}{\partial \beta_2}\Big|_{\beta_0} X(X'X)^{-1}X'\frac{\partial f_2}{\partial \beta_2}\Big|_{\beta_0}\right)^{-1}\\ &\left.\times\frac{\partial f_2'}{\partial \beta_2}\Big|_{\beta_0} X(X'X)^{-1}X'\frac{\partial f_1}{\partial \beta_1'}\Big|_{\beta_0}\right]\Bigg\}\end{aligned}$$

because:

$$\frac{\Omega^{12}\Omega^{21}}{\Omega^{22}} = \frac{\Omega_{12}\Omega_{21}}{\Omega_{11}(\Omega_{11}\Omega_{22} - \Omega_{12}\Omega_{21})} = \frac{\Omega_{12}^2}{\Omega_{11}^2\Omega_{22} - \Omega_{11}\Omega_{12}^2} = 0$$

and:

$$\Omega^{11} - \frac{1}{\Omega^{11}} = \frac{\Omega_{22}}{\Omega_{11}\Omega_{22} - \Omega_{12}^2} - \frac{1}{\Omega_{11}} = \frac{\Omega_{12}^2}{\Omega_{11}^2\Omega_{22} - \Omega_{11}\Omega_{12}^2} = 0$$

We inverted a matrix

$$\begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix}^{-1} = \begin{bmatrix} K_1 & — \\ — & — \end{bmatrix}$$

where $\begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix}$ is a positive definite matrix: so is K_1; so is K_1^{-1}; and

$$K_1 = (B + \theta C)^{-1} \qquad \theta \geq 0,$$

B is definite positive and C is semidefinite positive, $B^{-1} - (B + \theta C)^{-1}$ is semidefinite positive. If C is definite positive and $\theta > 0$, the minimum distance estimator is strictly better than the corresponding single-equation estimator.

If Ω is diagonal the minimum distance estimator coincides with the single-equation estimator. It is also true if $C = 0$.

$$C = \frac{\partial f_1'}{\partial \beta_1}\bigg|_{\beta_0} X(X'X)^{-1}X' \frac{\partial f_1}{\partial \beta_1'}\bigg|_{\beta_0} - \frac{\partial f_1'}{\partial \beta_1'}\bigg|_{\beta_0} X(X'X)^{-1}X' \frac{\partial f_2}{\partial \beta_2'}\bigg|_{\beta_0}$$

$$\times \left(\frac{\partial f_2'}{\partial \beta_2}\bigg|_{\beta_0} X(X'X)^{-1}X' \frac{\partial f_2}{\partial \beta_2'}\bigg|_{\beta_0}\right)^{-1} \cdot$$

$$\times \frac{\partial f_2'}{\partial \beta_2}\bigg|_{\beta_0} X(X'X)^{-1}X' \frac{\partial f_1}{\partial \beta_1'}\bigg|_{\beta_0}.$$

If $X'(\partial f_2/\partial \beta_2')|_\beta$ is invertible—i.e., square and nonsingular—then $C = 0$. There are as many exogenous variables as unknowns in the second equation; this yields just identification in the linear case.

6.4. *Conclusion*

We conclude that except for the case of linearity in the variables, the minimum distance and efficient instrumental variables estimators are CUAN but not Best CUAN. On the other hand these estimators appear to be an interesting step in the estimation of nonlinear systems with constraints across equations, since they provide consistent estimators that incorporate all of the constraints. A consistent estimator can be used to initialize a one-step linearized maximum likelihood estimator. This estimator is asymptotically equivalent to the maximum

likelihood estimator, just as in the case of systems of linear simultaneous equations considered by Rothenberg and Leenders.

Harvard University
and University of Montreal

REFERENCES

Amemiya, T. (1972). "Regression Analysis When the Dependent Variable is Truncated Normal," Technical Report No. 59. Institute for Mathematical Studies in the Social Sciences, Stanford University.

Amemiya, T. (1974). "The Nonlinear Two Stage Least Squares Estimator," *Journal of Econometrics*, forthcoming.

Brundy, J. M. and D. W. Jorgenson (1971). "Efficient Estimation of Simultaneous Equations by Instrumental Variables," *Review of Economics and Statistics*. Vol. 53.

——— and ——— (1973). "Consistent and Efficient Estimation of Systems of Simultaneous Equations" in P. Zarembka (ed.). *Frontiers in Econometrics*. New York. Academic Press. pp. 215–244.

Cramer, H. (1971). *Mathematical Methods of Statistics*. Princeton University Press.

Dhrymes, P. (1970). *Econometrics - Statistical Foundations and Applications*. New York. Harper & Row.

Edgerton, D. L. (1972). "Some Properties of Two Stage Least Squares as Applied to Non-Linear Models," *International Economic Review*. Vol. 13.

Eisenpress, H. and J. Greenstadt (1966). "The Estimation of Non Linear Econometric Systems," *Econometrica*. Vol. 34.

Fisher, F. M. (1966). *The Identification Problem in Econometrics*. New York. McGraw-Hill.

Goldfeld, S. M. and R. E. Quandt (1968). "Nonlinear Simultaneous Equations: Estimation and Prediction," *International Economic Review*. Vol. 10.

Hausman, J. (1974). "An Instrumental Variable Approach to Full-Information Estimators for Linear and Certain Non-linear Econometric Models," *Econometrica*. forthcoming.

Kalejian, J. J. (1972). "Two-Stage Least Squares and Econometric Systems Linear in Parameters but Non-Linear in Endogenous Variables," *Journal of the American Statistical Association*, Vol. 66.

Koopmans, T. C., and W. Hood (1953), "The Estimation of Simultaneous Linear Economic Relationships," in T. C. Koopmans and W. Hood (eds.). *Studies in Econometric Method*. New York. Wiley.

Malinvaud, E. (1970a), "The Consistency of Non-Linear Regression," *The Annals of Mathematical Statistics*. Vol. 41.

———(1970b). *Statistical Methods of Econometrics*. Amsterdam. North-Holland Publishing Company.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*. New York. Wiley.

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

Rothenberg, T. and C. T. Leenders (1964). "Efficient Estimation of Simultaneous Equation Systems," *Econometrica*. Vol. 32.

Zellner, A., D. Huang, and L. Chau (1965). "Further Analysis of the Short Run Consumption Function with Emphasis on the Role of Liquid Assets," *Econometrica*. Vol. 33.