Stat.

Instituut voor Actuariaat & Econometrie

Report AE9/76

On the estimation of a single non-linear structural equation of an incomplete simultaneous equation system

by H.J. Bierens

Preliminary report

July, 1976

Universiteit van Amsterdam

University. Institute of actuarial sciences and econometrics.

UNIVERSITY OF AMSTERDAM
INSTITUTE OF ACTUARIAL SCIENCES AND ECONOMETRICS
Jodenbreestraat 23
Amsterdam

Report AE9/76

On the estimation of a single non-linear structural equation of an incomplete simultaneous equation system[1)]

by H.J. Bierens

AMS(MOS) subject classification scheme (1970): 62P20, 62F10, 62E20, 60F05, 60F15.

Keywords: estimation, non-linear, structural equation, incomplete system.

Abstract: This paper considers the estimation of a non-linear structural equation without using explicite information about the other equations of the system. The non-linearity may appear in the disturbance variable as well. It is assumed that this disturbance variable is symmetrically distributed. The estimator involved is consistent and asymptotically normally distributed and its asymptotic variance matrix can be estimated consistently.

Preliminary report July, 1976

---

1) This research is supported by the Foundation for Economic Research of the University of Amsterdam.

1 INTRODUCTION

In this paper we consider a consistent and asymptotically normally distributed estimator that is appropriate for estimating the parameters of a single (non-linear) structural equation of a simultaneous equation system without using explicite information about the other equations of the system. In contrast with the model considered by Amemiya (1974), the structural equation involved may be non-linear in the disturbance variable as well. Besides, Amemiya's non-linear two-stage least-squares estimator needs some explicite information about the other equations in the system in order to select an appropriate set of instrumental variables [see Jorgenson and Laffont (1974) and Amemiya (1975)], and is therefore less useful for models with errors in variables. (Note that the "errors in variables" model can be considered as a structural equation of an incomplete simultaneous equation system).

One of our mean assumptions is that the disturbance variable is symmetrically distributed, so that, roughly speaking, the true values of the parameters can be identified using the well known property that the imaginary part of the characteristic function of this disturbance variable is zero everywhere.

2 THE MODEL AND ITS ESTIMATOR

The model to be considered has the structure

$$y_j=f(z_j,u_j,\theta_o) \quad , \quad j=1,2,\ldots \tag{1}$$

where the y_j's are observable scalar random variables, the u_j's are non-observable scalar random disturbance variables, the z_j's are observable p-component vectors consisting partly of endogenous variables correlated with the disturbance variable u_j, and partly of exogenous variables, and θ_o is an unknown q-component parameter vector. The exogenous components of z_j may be considered as random variables taking the value of the exogenous variable involved with probability one. If equation (1) can be solved for u_j and if y_j is one of the components of z_j, this model can be written as

$$g(z_j,\theta_o)=u_j \quad , \quad j=1,2,\ldots \tag{2}$$

We now assume:

(a) *The function* $g(z,\theta)$ *is continuous on the product space* $R^p\times\Theta$ *and for every* $z\varepsilon R^p$ *twice continuous differentiable on* Θ , *where* Θ *is a compact subset of the* q*-dimensional real space* R^q *and* $\theta_o\varepsilon\Theta$. *The* z_j*'s are independent* p*-variate distributed random vectors with unknown d.f.* $F_j(z)$, *respectively. The* u_j*'s are independent identically distributed random variables and their common d.f. is symmetric.*

Let

$$S_n(t,\theta)=\frac{1}{n}\sum_{j=1}^{n}\sin(tg(z_j,\theta)) \quad . \tag{3}$$

Then

$$ES_n(t,\theta_o)=\frac{1}{n}\sum_{j=1}^{n}E\sin(tu_j)=0 \quad \text{for every} \quad t \tag{4}$$

since $E\sin(tu_j)$ is the imaginary part of the characteristic function of the symmetric error distribution. On the other hand, if

(b) *for every* $\theta\varepsilon\Theta\setminus\{\theta_o\}$ *there exists at least one positive integer* m

such that $\lim \frac{1}{n}\sum_{j=1}^{n}Eg(z_j,\theta)^{2m-1}$ *exists and is non-zero,* then for all sufficient large n and every $\beta>0$,

$$ES_n(t,\theta)\neq 0 \quad \text{for some} \quad t\varepsilon(0,\beta] \quad \text{if} \quad \theta\varepsilon\Theta\backslash\{\theta_o\} \tag{5}$$

since then

$$(\partial/\partial t)^{2m-1}ES_n(t,\theta)=(-1)^{m+1}\frac{1}{n}\sum_{j=1}^{n}\int g(z,\theta)^{2m-1}\cos(tg(z,\theta))dF_j(z) \tag{6}$$

is non-zero at $t=0$. Thus under the assumptions (a) and (b) ,

$$\int_o^\beta \{ES_n(t,\theta)\}^2dt>0 \quad \text{if} \quad \theta\varepsilon\Theta\backslash\{\theta_o\} \quad \text{and} \quad n \quad \text{is large} , \tag{7}$$

while

$$\int_o^\beta \{ES_n(t,\theta_o)\}^2dt=0 \quad \text{for all} \quad n . \tag{8}$$

This suggests the following estimator:

$$\hat{\theta}_n : \int_o^\beta S_n(t,\hat{\theta}_n)^2dt=\inf_{\theta\varepsilon\Theta}\int_o^\beta S_n(t,\theta)^2dt \tag{9}$$

At least one of the solutions of (9) is measurable [see Jennrich (1969, lemma 2)] because of the continuity of $\int_o^\beta S_n(t,\theta)^2dt$ on Θ and the compactness of Θ .

Assumption (b) plays an important rôle in our analysis. Its plausibility needs therefore attention. As an example, let us consider the linear model with errors in variables:

$$g(z_j,\theta_o)=z_j'\theta_o=u_j \tag{10}$$

Assuming that the first component of z_j is the dependent variable, we have to put the first component of θ_o equal to one. The same applies for other θ's to be considered. Therefore we assume that $\Theta=\{1\}\times\Theta^*$, where Θ^* is a compact subset of R^{p-1} . Furthermore, let z_j be p-variate normal distributed with mean μ_j and variance Σ :

$$z_j \sim N_p(\mu_j,\Sigma) , \tag{11}$$

where the matrix Σ may contain zero columns and rows corresponding with "non-random" components of z_j . Then $\mu_j'\theta_o$ must be zero because of the symmetry of the distribution of the u_j's.

Thus

$$\mu_j'\theta_o=0 \quad ; \quad z_j'\theta \sim N(\mu_j'\theta,\theta'\Sigma\theta) \tag{12}$$

and consequently

$$ES_n(t,\theta)=\{\frac{1}{n}\sum_{j=1}^{n} \sin (t\mu_j'\theta)\}\exp(-\tfrac{1}{2}t^2\theta'\Sigma\theta) \tag{13}$$

Let $\theta\varepsilon\Theta\backslash\{\theta_o\}$ be given and put $v_j=z_j-\mu_j$. Then

$$z_j'\theta=\mu_j'\theta+v_j'\theta \quad ; \quad v_j'\theta \sim N(0,\theta'\Sigma\theta) \tag{14}$$

and thus

$$\begin{aligned} E(z_j'\theta)^{2m-1} &= \sum_{k=0}^{m-1}\begin{bmatrix}2m-1\\2k\end{bmatrix}(\mu_j'\theta)^{2k} E(v_j'\theta)^{2(m-k)-1} \\ &+ \sum_{k=1}^{m}\begin{bmatrix}2m-1\\2k-1\end{bmatrix}(\mu_j'\theta)^{2k-1} E(v_j'\theta)^{2(m-k)} \\ &= \sum_{k=1}^{m}\begin{bmatrix}2m-1\\2k-1\end{bmatrix}(\mu_j'\theta)^{2k-1} E(v_1'\theta)^{2(m-k)} \end{aligned} \tag{15}$$

since $E(v_j'\theta)^{2(m-k)-1}=0$ and $E(v_j'\theta)^{2(m-k)}$ is independent of j . Hence, assumption (b) implies in this case that there exists at least one positive integer k such that $\lim \frac{1}{n}\sum_{j=1}^{n}(\mu_j'\theta)^{2k-1} \neq 0$, which in its turn implies that for all sufficient large n and every $\beta>0$,

$$\frac{1}{n}\sum_{j=1}^{n} \sin(t\mu_j'\theta)\neq 0 \quad \text{for some} \quad t\varepsilon(0,\beta] \quad \text{if} \quad \theta\varepsilon\Theta\backslash\{\theta_o\} \tag{16}$$

Since $\frac{1}{n}\sum_{j=1}^{n}\sin(t\mu_j'\theta)$ is the imaginary part of the characteristic function of the empirical distribution of $\mu_1'\theta$,..., $\mu_n'\theta$, (16) and consequently assumption (b) imply in this case that this empirical distribution is non-symmetric if $\theta\varepsilon\Theta\backslash\{\theta_o\}$ and n is large. Moreover, if this empirical distribution is non-symmetric if $\theta\varepsilon\Theta\backslash\{\theta_o\}$ and n is large, it is easily checked that there exists a positive integer m such that $\frac{1}{n}\sum_{j=1}^{n}E(z_j'\theta)^{2m-1}\neq 0$. This result suggests that assumption (b) is not very restrictive.

3 CONSISTENCY

First we shall give conditions such that (7) also holds in the limit. Of course this limit must exist:

(c) $s(t,\theta)=\lim \frac{1}{n}\sum_{j=1}^{n} E\sin(tg(z_j,\theta))$ *exists for every* $\theta\varepsilon\Theta$ *and every* t *in a compact interval containing zero.*

Let $\theta\varepsilon\Theta\backslash\{\theta_o\}$ and let m be the smallest positive integer for which assumption (b) is satisfied. Then for every positive integer $k<m$,

$$\{(\partial/\partial t)^{2k-1} ES_n(t,\theta)\}_{t=0}=(-1)^{k+1}\frac{1}{n}\sum_{j=1}^{n} Eg(z_j,\theta)^{2k-1}\rightarrow 0 \tag{17}$$

while for $k=0,1,2,\ldots, m$

$$\{(\partial/\partial t)^{2k} ES_n(t,\theta)\}_{t=0}=0 \tag{18}$$

Therefore the Taylor expansion of $ES_n(t,\theta)$ yields in the limit

$$\lim ES_n(t,\theta)=(-1)^{m+1}(\lim \frac{1}{n}\sum_{j=1}^{n} Eg(z_j,\theta)^{2m-1})\frac{t^{2m-1}}{(2m-1)!}+$$

$$(\lim R_n(t,\theta))\frac{t^{2m}}{(2m)!} \tag{19}$$

where

$$|R_n(t,\theta)|\leq\frac{1}{n}\sum_{j=1}^{n} Eg(z_j,\theta)^{2m} \quad . \tag{20}$$

Now put $a=\lim \frac{1}{n}\sum_{j=1}^{n} Eg(z_j,\theta)^{2m-1}$ and $b=\limsup \frac{1}{n}\sum_{j=1}^{n} Eg(z_j,\theta)^{2m}$

Then it follows from (19) and (20) that

$$(-1)^{m+1}\frac{t^{2m-1}}{(2m-1)!}(a-\frac{bt}{2m})\leq s(t,\theta)\leq(-1)^{m+1}\frac{t^{2m-1}}{(2m-1)!}(a+\frac{bt}{2m}) \quad ,$$

so that $|s(t,\theta)|>0$ if $t\varepsilon(0,\frac{2m|a|}{b})$ and $b<\infty$. Hence,

$$\int_0^\beta s(t,\theta)^2 dt>0 \quad \text{if} \quad \theta\varepsilon\Theta\backslash\{\theta_o\} \quad , \tag{21}$$

provided that

(d) *for every* $\theta\varepsilon\Theta\backslash\{\theta_o\}$, $\limsup \frac{1}{n}\sum_{j=1}^{n} Eg(z_j,\theta)^{2m}<\infty$, *where* m *is the smallest positive integer for which assumption* (b) *is satisfied.*

Next we shall give an additional condition such that

$$\text{plim} \sup_{\theta\varepsilon\Theta,t\varepsilon[0,\beta]} |S_n(t,\theta)-ES_n(t,\theta)|=0 \tag{22}$$

and

$$\lim \sup_{\theta\varepsilon\Theta,t\varepsilon[0,\beta]} |ES_n(t,\theta)-s(t,\theta)|=0 \tag{23}$$

Let N be a neighbourhood of $\theta^*\varepsilon\Theta$ and let $t^*\varepsilon[0,\beta]$.
Then for every $\theta\varepsilon N$ and every $t\varepsilon[0,\beta]$,

$$\begin{aligned}
|S_n(t,\theta)-S_n(t^*,\theta^*)| &\leq \frac{1}{n}\sum_{j=1}^{n} |tg(z_j,\theta)-t^*g(z_j,\theta^*)| \\
&\leq |t-t^*|\frac{1}{n}\sum_{j=1}^{n} |g(z_j,\theta^*)|+t\frac{1}{n}\sum_{j=1}^{n} |g(z_j,\theta)-g(z_j,\theta^*)| \\
&\leq |t-t^*|\frac{1}{n}\sum_{j=1}^{n} |g(z_j,\theta^*)|+\beta|\theta-\theta^*|\frac{1}{n}\sum_{j=1}^{n} \sup_{\theta\varepsilon N}|(\partial/\partial\theta)g(z_j,\theta)| \\
&\leq |t-t^*|+\beta|\theta-\theta^*|+|t-t^*|\frac{1}{n}\sum_{j=1}^{n} g(z_j,\theta^*)^2 + \\
&\quad \beta|\theta-\theta^*|\frac{1}{n}\sum_{j=1}^{n} \sup_{\theta\varepsilon\Theta}|(\partial/\partial\theta)g(z_j,\theta)|^2 \quad ,
\end{aligned} \tag{24}$$

where the first and the thirth inequality in (24) follow from the mean value theorem. Moreover, it is obvious from the strong law of large numbers that

$$S_n(t,\theta)-ES_n(t,\theta)\to 0 \qquad \text{a.e.} \tag{25}$$

and consequently

$$\text{plim } \{S_n(t,\theta)-ES_n(t,\theta)\}=0 \quad , \tag{26}$$

since $S_n(t,\theta)$ is the mean of n independent uniformly bounded random variables. Comparing (23) and (24) with part (ii) of theorem A in the appendix, we see that (22) holds if

(e) *for every* $\theta^* \varepsilon \Theta$ *there exists a neighbourhood* $N=\{\theta\varepsilon\Theta:|\theta-\theta^*|<\delta\}$,

$\delta>0$, *such that* $\lim \frac{1}{n}\sum_{j=1}^{n} Eg(z_j,\theta^*)^2$ *and*

$\lim \frac{1}{n}\sum_{j=1}^{n} E \sup_{\theta\varepsilon N}|(\partial/\partial\theta)g(z_j,\theta)|^2$ *exist, while* $\limsup \frac{1}{n}\sum_{j=1}^{n} Eg(z_j,\theta^*)^4$

and $\limsup \frac{1}{n}\sum_{j=1}^{n} E \sup_{\theta\varepsilon N}|(\partial/\partial\theta)g(z_j,\theta)|^4$ *are finite,*

since then by the weak law of large numbers

$$\operatorname{plim}\frac{1}{n}\sum_{j=1}^{n} g(z_j,\theta^*)^2=\lim\frac{1}{n}\sum_{j=1}^{n} Eg(z_j,\theta^*)^2 \tag{27}$$

and

$$\operatorname{plim}\frac{1}{n}\sum_{j=1}^{n} \sup_{\theta\varepsilon N}|(\partial/\partial\theta)g(z_j,\theta)|^2=\lim\frac{1}{n}\sum_{j=1}^{n} E\sup_{\theta\varepsilon N}|(\partial/\partial\theta)g(z_j,\theta)|^2 \tag{28}$$

Furthermore, taking expectations of the random variables in (24) and using part (iii) of theorem A it follows by the same kind of argument that (23) also holds. Hence,

$$\operatorname{plim} \sup_{\theta\varepsilon\Theta,t\varepsilon[0,\beta]} |S_n(t,\theta)-s(t,\theta)|=0 \quad , \tag{29}$$

Moreover, (23) implies that $s(t,\theta)$ is uniformly continuous on $[0,\beta]\times\Theta$, since $ES_n(t,\theta)$ is continuous and $[0,\beta]\times\Theta$ is compact.

These results imply that

$$\operatorname{plim} \sup_{\theta\varepsilon\Theta}\left|\int_0^\beta S_n(t,\theta)^2 dt-\int_0^\beta s(t,\theta)^2 dt\right|=0 \tag{30}$$

and $\int_0^\beta s(t,\theta)^2 dt$ is uniformly continuous on Θ .

From (30) it follows now

$$\text{plim}\,\{\int_0^\beta S_n(t,\hat\theta_n)^2dt-\int_0^\beta s(t,\hat\theta_n)^2dt\}=0 \quad , \tag{31}$$

and together with (9) ,

$$0\leq\int_0^\beta S_n(t,\hat\theta_n)^2dt\leq\int_0^\beta S_n(t,\theta_o)^2dt\to 0 \quad \text{in prob.}, \tag{32}$$

since $\int_0^\beta s(t,\theta_o)^2dt=0$. Hence

$$\text{plim}\int_0^\beta s(t,\hat\theta_n)^2dt=0 \tag{33}$$

which, together with (21) and the continuity of $\int_0^\beta s(t,\theta)^2dt$, implies

$$\text{plim}\;\hat\theta_n=\theta_o \quad . \tag{34}$$

We have proved by now:

THEOREM 1. *Under the assumptions* (a) *through* (e), $\hat\theta_n$ *is weakly consistent.*

A stronger result can be obtained if we assume that

$$\frac{1}{n}\sum_{j=1}^{n} F_j(z)\to F(z) \quad \text{vaguely}^{1)} \quad , \tag{35}$$

where F is a distribution function and the F_j's are distribution functions defined in assumption (a), because then by theorem B in the appendix

$$s(t,\theta)=\int\sin(tg(z,\theta))dF(z) \tag{36}$$

and

$$\sup_{\theta\varepsilon\Theta,t\varepsilon[0,\beta]}|S_n(t,\theta)-s(t,\theta)|\to 0 \quad \text{a.e.} \tag{37}$$

so that

$$\sup_{\theta\varepsilon\Theta}|\int_0^\beta S_n(t,\theta)^2dt-\int_0^\beta s(t,\theta)^2dt|\to 0 \quad \text{a.e.} \tag{38}$$

---

1) See Chung (1974, p. 80 and 85) for a definition of vague convergence.

Thus if the assumptions (c) and (e) are replaced by (35), then

$\hat{\theta}_n \to \theta_o$ a.e. (39)

THEOREM 2. *Let the assumptions* (a), (b) *and* (d) *be satisfied.*

If $\frac{1}{n}\sum_{j=1}^{n} F_j(z)$ *converges vaguely to a distribution function, then* $\hat{\theta}_n$ *is strongly consistent.*

4 ASYMPTOTIC NORMALITY

Deriving the limiting distribution of our estimator $\hat{\theta}_n$, we need the following Taylor expansion of $(\partial/\partial\theta)\int_0^\beta S_n(t,\hat{\theta}_n)^2 dt$[1] :

$$(\partial/\partial\theta)\int_0^\beta S_n(t,\hat{\theta}_n)^2 dt=(\partial/\partial\theta)\int_0^\beta S_n(t,\theta_o)^2 dt+(\hat{\theta}_n-\theta_o)'(\partial/\partial\theta)(\partial/\partial\theta)\int_0^\beta S_n(t,\theta_n^*)dt \quad (40)$$

If

(f) Θ *is convex and* θ_o *is an interior point of* Θ

then it follows from Jennrich (1969, lemma 3) that at least one of the θ_n^*'s for which (40) holds is measurable and that

$$|\theta_n^*-\theta_o|\leq|\hat{\theta}_n-\theta_o| \quad \text{a.e.} \quad (41)$$

Moreover, by the same kind of argument as in the proof of Jennrich (1969, theorem 7) it follows that under the conditions of theorem 1 ,

$$\sqrt{n}(\partial/\partial\theta')\int_0^\beta S_n(t,\hat{\theta}_n)^2 dt\rightarrow 0 \quad \text{in prob.} \quad , \quad (42)$$

where 0 is a zero vector:
(Note that under the conditions of theorem 2 , (42) holds a.e.)

Hence, if $\sqrt{n}(\partial/\partial\theta')\int_0^\beta S_n(t,\theta_o)^2 dt$ converges in distribution to a q-variate normal distribution with zero mean vector and variance matrix Σ , and if $\text{plim}\,(\partial/\partial\theta)(\partial/\partial\theta)\int_0^\beta S_n(t,\theta_n^*)^2 dt=\Gamma$, where Γ is a non-singular $q\times q$ matrix, then $\sqrt{n}(\hat{\theta}_n-\theta_o)$ converges in distribution to a q-variate normal distribution with zero mean vector and variance matrix $\Gamma^{-1}\Sigma(\Gamma')^{-1}$.

---

1) The notation $(\partial/\partial\theta)f(\theta^*)$ denotes a row vector of partial derivatives of the function $f(\theta)$ in the point θ^* , while $(\partial/\partial\theta')f(\theta^*)$ denotes its transpose. The notation $(\partial/\partial\theta)(\partial/\partial\theta)f(\theta^*)$ denotes the matrix of second partial derivatives of $f(\theta)$ in the point θ^* .

First we consider the asymptotic normality of $\sqrt{n}(\partial/\partial\theta')\int_0^\beta S_n(t,\theta_o)^2 dt$. For $m=1,2,\ldots,q$ we have

$$|\sqrt{n}(\partial/\partial\theta_m)\int_0^\beta S_n(t,\theta_o)^2 dt - 2\int_0^\beta \sqrt{n}\, S_n(t,\theta_o)E(\partial/\partial\theta_m)S_n(t,\theta_o)dt|$$

$$\leq 2\sup_{t\varepsilon[0,\beta]} |(\partial/\partial\theta_m)S_n(t,\theta_o)-E(\partial/\partial\theta_m)S_n(t,\theta_o)|\int_0^\beta |\sqrt{n}\, S_n(t,\theta_o)|dt \ . \quad (43)$$

Furthermore, if $t\varepsilon[0,\beta]$ and $t^*\varepsilon[0,\beta]$,

$$|(\partial/\partial\theta_m)S_n(t,\theta_o)-(\partial/\partial\theta_m)S_n(t^*,\theta_o)|\leq$$

$$\leq \frac{1}{n}\sum_{j=1}^n |t\cos(tu_j)-t^*\cos(t^*u_j)||(\partial/\partial\theta_m)g(z_j,\theta_o)|$$

$$\leq (1+\beta)|t-t^*|\frac{1}{n}\sum_{j=1}^n |(\partial/\partial\theta_m)g(z_j,\theta_o)|$$

$$\leq (1+\beta)|t-t^*|\{1+\frac{1}{n}\sum_{j=1}^n \sup_{\theta\varepsilon N}|(\partial/\partial\theta)g(z_j,\theta)|^2\} \ , \quad (44)$$

where N is a neighbourhood of θ_o , and

$$\mathrm{var}\{(\partial/\partial\theta_m)S_n(t,\theta_o)\}\leq\beta^2 \frac{1}{n^2}\sum_{j=1}^n E\{(\partial/\partial\theta_m)g(z_j,\theta_o)\}^2 \leq \beta^2 \frac{1}{n^2}\sum_{j=1}^n E\sup_{\theta\varepsilon N}|(\partial/\partial\theta)g(z_j,\theta)|^2. \quad (45)$$

From (44), (45), assumption (e) and part (ii) of theorem A in the appendix it follows now that

$$\mathrm{plim} \sup_{t\varepsilon[0,\beta]} |(\partial/\partial\theta_m)S_n(t,\theta_o)-E(\partial/\partial\theta_m)S_n(t,\theta_o)|=0 \quad (46)$$

Since $|\sqrt{n}\, S_n(t,\theta)|\leq\frac{1}{2}+\frac{1}{2}nS_n(t,\theta)^2$, we have

$$E\int_0^\beta |\sqrt{n}\, S_n(t,\theta_o)|dt\leq\frac{1}{2}\beta+\frac{1}{2}\frac{1}{n}\sum_{j=1}^n\int_0^\beta E\{\sin(tu_j)\}^2 dt\leq\beta \quad (47)$$

and thus by Chebishev's inequality

$$P\{\int_0^\beta |\sqrt{n}\, S_n(t,\theta_o)|dt>\frac{\beta}{\varepsilon}\}<\varepsilon \quad (48)$$

for every $\varepsilon>0$. Hence we obtain from (43), (46) and (48)

$$\mathrm{plim}\{\sqrt{n}(\partial/\partial\theta_m)\int_0^\beta S_n(t,\theta_o)^2 dt-2\int_0^\beta \sqrt{n}\, S_n(t,\theta_o)E(\partial/\partial\theta_m)S_n(t,\theta_o)dt\}=0 \quad (49)$$

using the same kind of argument as in the proof of theorem 4.4.6.b in Chung (1974) .

Now let

$$c_{m,n}(t)=E(\partial/\partial\theta_m)S_n(t,\theta_o) \quad , \quad c_n(t)'=(c_{1,n}(t) \quad ,\ldots, \quad c_{q,n}(t)) \quad ;$$

$$x_n'=(\int_o^\beta S_n(t,\theta_o)c_{1,n}(t)dt,\ldots, \quad \int_o^\beta S_n(t,\theta_o)c_{q,n}(t)dt) \tag{50}$$

and choose $\zeta\varepsilon R^q$ arbitrarily . Then

$$x_n'\zeta=\int_o^\beta S_n(t,\theta_o)c_n(t)'\zeta dt=\frac{1}{n}\sum_{j=1}^n\int_o^\beta \sin(tu_j)c_n(t)'\zeta dt \tag{51}$$

The u_j's are symmetrically distributed. Therefore

$$E\int_o^\beta \sin(tu_j)c_n(t)'\zeta dt=0 \quad . \tag{52}$$

Moreover,

$$\mathrm{var}(x_n'\zeta)=\frac{1}{n^2}\sum_{j=1}^n E\{\int_o^\beta \sin(tu_j)c_n(t)'\zeta dt\}^2=\frac{1}{n}\int_o^\beta\int_o^\beta \psi(t_1,t_2)\zeta'c_n(t_1)c_n(t_2)'\zeta dt_1dt_2$$

$$=\frac{1}{n}\zeta'B_n\zeta \quad \text{, say} \tag{53}$$

where

$$\psi(t_1,t_2)=E \sin(t_1u_j)\sin(t_2u_j) \tag{54}$$

and

$$B_n=(b_n^{(i,j)})=(\int_o^\beta\int_o^\beta \psi(t_1,t_2)c_{i,n}(t_1)c_{j,n}(t_2)dt_1dt_2) \tag{55}$$

We assume now

(g) $c(t)=\lim \frac{1}{n}\sum_{j=1}^n tE\cos(tu_j)(\partial/\partial\theta')g(z_j,\theta_o)$ *exists for every* t *in a compact interval.* $c(t)'=(c^{(1)}(t),\ldots, c^{(q)}(t))$

Then $c(t)=\lim c_n(t)$ uniformly on $[0,\beta]$, which can be proved using (44), part (iii) of theorem A and assumption (e). Hence

$$\lim B_n=B=(b_{i,j})=(\int_o^\beta\int_o^\beta \psi(t_1,t_2)c^{(i)}(t_1)c^{(j)}(t_2)dt_1dt_2) \tag{56}$$

and

$$\sum_{j=1}^n E\left|\frac{\frac{1}{n}\int_o^\beta \sin(tu_j)c_n(t)'\zeta dt}{\sqrt{\mathrm{var}(\zeta'x_n)}}\right|^3 \leq n^{-\frac{1}{2}}(\zeta'B_n\zeta)^{-3/2}\{\int_o^\beta |c_n(t)'\zeta|dt\}^3$$

$$=O(n^{-\frac{1}{2}}) \quad \text{if} \quad \zeta'B\zeta>0 \tag{57}$$

From (51) through (57) it follows now by Liapounov's central limit theorem that for every $\zeta \varepsilon R^q$ such that $\zeta' B\zeta > 0$,

$$\sqrt{n}\, x_n'\zeta \to N(0, \zeta' B\zeta) \text{ in distr.} \tag{58}$$

while $\operatorname{plim} \sqrt{n}\, x_n'\zeta = 0$ if $\zeta' B\zeta = 0$. Hence

$$\sqrt{n}\, x_n \to N_q(0, B) \text{ in distr.,} \tag{59}$$

where B may be singular .

Comparing this result with (50) and (49), we see now:

THEOREM 3. *Under the assumptions* (a), (e) *and* (g),

$\sqrt{n}(\partial/\partial\theta')\int_0^\beta S_n(t,\theta_0)^2 dt \to N_q(0, 4B)$ *in distr., where* B *is defined in* (56).

5 THE LIMITING DISTRIBUTION OF THE ESTIMATOR

Next we consider the probability limit of $(\partial/\partial\theta)(\partial/\partial\theta)\int_0^\beta S_n(t,\theta_n^*)^2 dt$. Let N be a neighbourhood of θ_o and put

$$a_n=\frac{1}{n}\sum_{j=1}^{n}\sup_{\theta\varepsilon N}|(\partial/\partial\theta)g(z_j,\theta)|^2 \quad ; \tag{60}$$

$$b_n=\sum_{k=1}^{q}\sum_{m=1}^{q}\sup_{\theta\varepsilon N}|(\partial/\partial\theta_k)(\partial/\partial\theta_m)g(z_j,\theta)| \tag{61}$$

By the mean value theorem and the trivial inequalities $|x_1y_1-x_2y_2|\leq|x_1||y_1-y_2|+|y_2||x_1-x_2|$ and $|xy|\leq\frac{1}{2}(x^2+y^2)$ we then have for every $\theta\varepsilon N$ and every $t\varepsilon[0,\beta]$

$$|(\partial/\partial\theta_m)S_n(t,\theta)-(\partial/\partial\theta_m)S_n(t,\theta_o)|\leq\beta^2|\theta-\theta_o|a_n+\beta|\theta-\theta_o|b_n \quad ; \tag{62}$$

$$|(\partial/\partial\theta_m)S_n(t,\theta)|\leq\tfrac{1}{2}\beta+\tfrac{1}{2}\beta a_n \quad ; \tag{63}$$

and

$$|(\partial/\partial\theta_k)(\partial/\partial\theta_m)S_n(t,\theta)|\leq\beta^2 a_n+\beta b_n \tag{64}$$

so that

$$|(\partial/\partial\theta_k)(\partial/\partial\theta_m)S_n(t,\theta)^2-2\{(\partial/\partial\theta_k)S_n(t,\theta_o)\}\{(\partial/\partial\theta_m)S_n(t,\theta_o)\}|$$

$$\leq 2|\{(\partial/\partial\theta_k)S_n(t,\theta)\}\{(\partial/\partial\theta_m)S_n(t,\theta)\}-\{(\partial/\partial\theta_k)S_n(t,\theta_o)\}\{(\partial/\partial\theta_m)S_n(t,\theta_o)\}|$$

$$+2|S_n(t,\theta)||(\partial/\partial\theta_k)(\partial/\partial\theta_m)S_n(t,\theta)|$$

$$\leq 2(\beta+\beta a_n)(\beta^2 a_n+\beta b_n)|\theta-\theta_o|+2(\beta^2 a_n+\beta b_n)\sup_{t\varepsilon[0,\beta]}|S_n(t,\theta)| \tag{65}$$

Assumption (e) implies that $\text{plim}\, a_n$ exists and is finite. If we assume

(h) *There exists a neighbourhood* N *of* θ_o *such that for* $k,m=1,2,\ldots,q$, $\lim\frac{1}{n}\sum_{j=1}^{n}E\sup_{\theta\varepsilon N}|(\partial/\partial\theta_k)(\partial/\partial\theta_m)g(z_j,\theta)|$ *converges and*

$\limsup\frac{1}{n}\sum_{j=1}^{n}E\sup_{\theta\varepsilon N}\{(\partial/\partial\theta_k)(\partial/\partial\theta_m)g(z_j,\theta)\}^2$ *is finite* , then by the weak law of large numbers, plim b_n exists and is finite. Since by (41) plim $\theta_n^* = \theta_o$ and thus by (29) ,

$$\text{plim} \sup_{t\varepsilon[0,\beta]} |S_n(t,\theta_n^*)| = \text{plim} \sup_{t\varepsilon[0,\beta]} |s(t,\theta_n^*)| = \sup_{t\varepsilon[0,\beta]} |s(t,\theta_o)| = 0 \quad , \tag{66}$$

it follows from (65) that

$$\text{plim}\left[(\partial/\partial\theta_k)(\partial/\partial\theta_m)\int_0^\beta S_n(t,\theta_n^*)^2 dt - 2\int_0^\beta \{(\partial/\partial\theta_k)S_n(t,\theta_o)\}\{(\partial/\partial\theta_m)S_n(t,\theta_o)\}dt\right] = 0. \tag{67}$$

Furthermore,

$$\text{plim}\left[\int_0^\beta \{(\partial/\partial\theta_k)S_n(t,\theta_o)\}\{(\partial/\partial\theta_m)S_n(t,\theta_o)\}dt - \int_0^\beta c_{k,n}(t)c_{m,n}(t)dt\right] = 0 \tag{68}$$

by (46), (50) and (63), while assumption (g) implies

$$\lim\int_0^\beta c_{k,n}(t)c_{m,n}(t)dt = \int_0^\beta c^{(k)}(t)c^{(m)}(t)dt \tag{69}$$

Hence

$$\text{plim}(\partial/\partial\theta_k)(\partial/\partial\theta_m)\int_0^\beta S_n(t,\theta_n^*)^2 dt = 2\int_0^\beta c^{(k)}(t)c^{(m)}(t)dt \tag{70}$$

Let

$$A = (a_{k,m}) = (\int_0^\beta c^{(k)}(t)c^{(m)}(t)dt \tag{71}$$

Then it follows from (40), (42), (70), (71) and theorem 3

$$2\sqrt{n}(\hat{\theta}_n - \theta_o)A \rightarrow N_q(0,4B) \text{ in distr.} \tag{72}$$

so that

$$\sqrt{n}(\hat{\theta}_n - \theta_o) \rightarrow N_q(0,A^{-1}BA) \text{ in distr.} \tag{73}$$

provided that

(i) *the matrix* A *, defined by* (71) *, is non-singular.*

This assumption is not very restrictive, as will be shown below. Since E $S_n(t,\theta_o)=0$ for every t , it follows from (50) that

$$(\partial/\partial\theta)(\partial/\partial\theta)\int_0^\beta \{E\ S_n(t,\theta_o)\}^2 dt = 2(\int_0^\beta c_{n,k}(t)c_{n,m}(t)dt) = 2A_n \quad , \text{ say } . \tag{74}$$

This matrix is the Hessian of the function $\int_0^\beta \{E\,S_n(t,\theta)\}^2 dt$ in the minimum point θ_o (see (7) and (8)). We may therefore expect that A_n is positive definite if n is large and thus that $A=\lim A_n$ is positive definite, and hence non-singular.

We have proved :

THEOREM 4. *Under the conditions of theorem* 1 *and the assumptions* (c) *through* (i) ,

$\sqrt{n}(\hat{\theta}_n-\theta_o)\to N_q(0,A^{-1}BA^{-1})$ *in distr.* ,

where A *is defined by* (71) *and* B *by* (56).

6 A WEAKLY CONSISTENT ESTIMATOR OF THE ASYMPTOTIC VARIANCE MATRIX

In this section we shall derive a weakly consistent estimator of the variance matrix $A^{-1}BA^{-1}$.

From the second inequality in (65) it follows that

$$\operatorname{plim}\left|\int_0^\beta\{(\partial/\partial\theta_k)S_n(t,\hat{\theta}_n)\}\{(\partial/\partial\theta_m)S_n(t,\hat{\theta}_n)\}dt-\int_0^\beta\{(\partial/\partial\theta_k)S_n(t,\theta_o)\}\{(\partial/\partial\theta_m)S_n(t,\theta_o)\}dt\right|$$
$$\leq\beta(\beta+\beta.\operatorname{plim} a_n)(\beta^2\operatorname{plim} a_n+\beta\operatorname{plim} b_n)\operatorname{plim}|\hat{\theta}_n-\theta_o|=0 \quad (75)$$

Thus if we put

$$\hat{A}_n=\left(\int_0^\beta\{(\partial/\partial\theta_k)S_n(t,\hat{\theta}_n)\}\{(\partial/\partial\theta_m)S_n(t,\hat{\theta}_n)\}dt\right) \quad (76)$$

then it follows from (68), (69), (71) and (75)

$$\operatorname{plim}\hat{A}_n=A\ . \quad (77)$$

Let

$$\hat{\psi}_n(t_1,t_2)=\frac{1}{n}\sum_{j=1}^n\sin(t_1g(z_j,\hat{\theta}_n))\sin(t_2g(z_j,\hat{\theta}_n)) \quad (78)$$

and

$$\psi_n^0(t_1,t_2)=\frac{1}{n}\sum_{j=1}^n\sin(t_1g(z_j,\theta_o))\sin(t_2g(z_j,\theta_o))=\frac{1}{n}\sum_{j=1}^n\sin(t_1u_j)\sin(t_2u_j) \quad (79)$$

Then for every $t_1\varepsilon[0,\beta]$ and every $t_2\varepsilon[0,\beta]$,

$$|\hat{\psi}_n(t_1,t_2)-\psi_n^0(t_1,t_2)|\leq 2\beta\frac{1}{n}\sum_{j=1}^n|g(z_j,\hat{\theta}_n)-g(z_j,\theta_o)|$$
$$\leq 2\beta|\hat{\theta}_n-\theta_o|\frac{1}{n}\sum_{j=1}^n\sup_{\theta\varepsilon N}|(\partial/\partial\theta)g(z_j,\theta)|$$
$$\leq\beta|\hat{\theta}_n-\theta_o|\{1+\frac{1}{n}\sum_{j=1}^n\sup_{\theta\varepsilon N}|(\partial/\partial\theta)g(z_j,\theta)|^2\} \quad (80)$$

if $\hat{\theta}_n\varepsilon N$, where N is a neighbourhood of θ_o . Since $\operatorname{plim}\hat{\theta}_n=\theta_o$, it follows from assumption (e) and from (80) that

$$\operatorname{plim}\sup_{t_1\varepsilon[0,\beta],\ t_2\varepsilon[0,\beta]}|\hat{\psi}_n(t_1,t_2)-\psi_n^0(t_1,t_2)|=0 \quad (81)$$

Moreover, by the weak law of large numbers ;

$$\text{plim } \psi_n^o(t_1,t_2)=\psi(t_1,t_2) \quad . \tag{82}$$

Since $\psi(t_1,t_2)$, defined by (54), is uniformly continuous on $[0,\beta]\times[0,\beta]$ and since

$$|\psi_n^o(t_1^*,t_2^*)-\psi_n^o(t_1,t_2)|\leq|t_1-t_1^*|\frac{1}{n}\sum_{j=1}^{n}|u_j|+|t_2-t_2^*|\frac{1}{n}\sum_{j=1}^{n}|u_j| \tag{83}$$

it follows from assumption (e) and part (ii) of theorem A in the appendix that (82) holds uniformly on $[0,\beta]\times[0,\beta]$:

$$\text{plim } \sup_{t_1\varepsilon[0,\beta],\ t_2\varepsilon[0,\beta]} |\psi_n^o(t_1,t_2)-\psi(t_1,t_2)|=0 \tag{84}$$

and thus by (81)

$$\text{plim } \sup_{t_1\varepsilon[0,\beta],\ t_2\varepsilon[0,\beta]} |\hat{\psi}_n(t_1,t_2)-\psi(t_1,t_2)|=0 \tag{85}$$

Let

$$\hat{B}_n=\left(\int_0^\beta\int_0^\beta\hat{\psi}_n(t_1,t_2)\{(\partial/\partial\theta_k)S_n(t_1,\hat{\theta}_n)\}\{(\partial/\partial\theta_m)S_n(t_2,\hat{\theta}_n)\}dt_1dt_2\right) \tag{86}$$

Then, analogous to (77), it is easily proved that

$$\text{plim } \hat{B}_n=B \ . \tag{87}$$

THEOREM 5. *Let the matrices* $\hat{A}_n$ *and* $\hat{B}_n$ *be defined by* (76) *and* (86), *respectively. Under the conditions of theorem* 4,

$$\text{plim } \hat{A}_n^{-1}\hat{B}_n\hat{A}_n^{-1}=A^{-1}BA^{-1}.$$

APPENDIX

The following two theorems are generalisations of the theorems 4 and 1, respectively in Jennrich (1969).

THEOREM A.

Let $(\phi_n(\theta))$ be a sequence of (random) functions on a compact subset Θ of a Euclidean space. Let $(\zeta_n(\theta))$ be a sequence of (random) functions on Θ such that for every $\theta_o \varepsilon \Theta$ and every $\delta > 0$ in a bounded interval,

$$\sup_{\theta \varepsilon U_\delta(\theta_o)} |\phi_n(\theta) - \phi_n(\theta_o)| \leq \delta \zeta_n(\theta_o) \quad \text{(a.e.)} \quad ,$$

where $U_\delta(\theta_o) = \{\theta \varepsilon \Theta : |\theta - \theta_o| < \delta\}$.

If for every $\theta_o \varepsilon \Theta$,

(i) $\phi_n(\theta_o) \to 0$ a.e. and $\zeta_n(\theta_o) \to \zeta(\theta_o)$ a.e.

or

(ii) $\text{plim}\ \phi_n(\theta_o) = 0$ and $\text{plim}\ \zeta_n(\theta_o) = \zeta(\theta_o)$

or

(iii) $\lim\ \phi_n(\theta_o) = 0$ and $\lim\ \zeta_n(\theta_o) = \zeta(\theta_o)$,

respectively, where $\zeta(\theta_o)$ is positive, finite and non-random, then

(i) $\sup_{\theta \varepsilon \Theta} |\phi_n(\theta)| \to 0$ a.e.

or

(ii) $\text{plim} \sup_{\theta \varepsilon \Theta} |\phi_n(\theta)| = 0$

or

(iii) $\lim \sup_{\theta \varepsilon \Theta} |\phi_n(\theta)| = 0$,

respectively

PROOF. Choose an $\varepsilon > 0$ and let $\delta(\theta_o)$ be such that $0 \leq \delta(\theta_o) . \zeta(\theta_o) < \varepsilon$ and $\delta(\theta_o) < \varepsilon$

(i) Let $(\Omega, \mathcal{F}, \mathcal{P})$ be the probability space. The random functions $\phi_n(\theta)$ and $\zeta_n(\theta)$ may now be written as $\phi_n(\omega, \theta)$ and $\zeta_n(\omega, \theta)$ respectively, where $\omega \varepsilon \Omega$. Since $\phi_n(\theta_o) \to 0$ a.e. and $\zeta_n(\theta_o) \to \zeta(\theta_o)$ a.e. for every $\theta_o \varepsilon \Theta$, there exists a function $n_o(\omega, \varepsilon, \theta_o)$ and a null set $N(\theta_o)$ such that $|\zeta_n(\omega,\theta_o)-\zeta(\theta_o)|\leq 1$ and $|\phi_n(\omega,\theta_o)|\leq\varepsilon$ if $n\geq n_o(\omega,\varepsilon,\theta_o)$ and $\omega\varepsilon\Omega\backslash N(\theta_o)$. Hence,

$$|\phi_n(\omega,\theta)|\leq|\phi_n(\omega,\theta)-\phi_n(\omega,\theta_o)|+|\phi_n(\omega,\theta_o)|\leq\delta(\theta_o)|\zeta_n(\theta_o)-\zeta(\theta_o)|+\delta(\theta_o)\zeta(\theta_o)+\varepsilon\leq 3\varepsilon$$

if $n\geq n_o(\omega,\varepsilon,\theta_o)\omega\varepsilon\Omega\backslash N(\theta_o)$ and $\theta\varepsilon U_\delta(\theta_o)$.

The compactness of Θ implies that there exists a finite sequence $\theta_1,\ldots,\theta_{m(\varepsilon)}$ of points in Θ such that

$$\Theta=\bigcup_{i=1}^{m(\varepsilon)} U_{\delta(\theta_i)}(\theta_i)$$

Now put $n_*(\omega,\varepsilon)=\max_{i=1,2,\ldots,m(\varepsilon)} n_o(\omega,\varepsilon,\theta_i)$ and $N_\varepsilon=\bigcup_{i=1}^{m(\varepsilon)} N(\theta_i)$.

Then for every $\theta\varepsilon\Theta$,

$|\phi_n(\omega,\theta)|\leq 3\varepsilon$ if $n\geq n_*(\omega,\varepsilon)$ and $\omega\varepsilon\Omega\backslash N_\varepsilon$.

Hence, $\sup_{\theta\varepsilon\Theta}|\phi_n(\theta)|\to 0$ a.e. since ε is arbitrary .

(ii) First we note that a sequence of random variables converges to zero in probability if and only if every subsequence contains a further subsequence that converges to zero a.e. . Let (n_k) be a subsequence. Then for every $\theta_i\varepsilon\{\theta_1,\ldots,\theta_{m(\varepsilon)}\}$ there exists a further subsequence (n_{k_j}) such that

$$\phi_{n_{k_j}}(\theta_i)\to 0 \quad \text{a.e. and} \quad \zeta_{n_{k_j}}(\theta_i)\to\zeta(\theta_i) \quad \text{a.e.}$$

Since we consider a finite sequence of points θ_i , this further subsequence can be chosen equally for all the θ_i's . However, (n_{k_j}) may depend on ε .

Hence, by the argument in the proof of (i) , there exists a function $j_*(\omega,\varepsilon)$ and a null set N_ε such that

$\sup_{\theta\varepsilon\Theta}|\phi_{n_{k_j}}(\omega,\theta)|\leq 3\varepsilon$ if $j\geq j_*(\omega,\varepsilon)$ and $\omega\varepsilon\Omega\backslash N_\varepsilon$,

which excludes the possibility that there exist positive numbers ε and δ and a subsequence (u_k) such that

$$P\{\sup_{\theta\in\Theta}|\phi_{n_k}(\theta)|\leq\varepsilon\}\leq 1-\delta \quad \text{for} \quad k=1,2,\ldots,$$

Hence, $\lim P\{\sup_{\theta\in\Theta}|\phi_n(\theta)|\leq\varepsilon\}=1$, which was to be proved.

(iii) This part of the theorem follows directly from part (i).

THEOREM B.

Let $\phi(x,\theta)$ be a continuous and uniformly bounded function on $X\times\Theta$, where X is a Euclidean space and Θ is a compact subset of a Euclidean space Y . Let $x_1,\ldots,x_n,\ldots,$ be a sequence of independent random variables in X with distribution functions $F_1(x),\ldots,F_n(x),\ldots,$ respectively.

If $\frac{1}{n}\sum_{j=1}^{n} F_j(x)$ converges vaguely to a distribution function $F(x)$, then

$$\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{j=1}^{n}\phi(x_j,\theta)-\int\phi(x,\theta)dF(x)\right|\to 0 \quad \text{a.e.}$$

This theorem can be proved in the same way as theorem 1 in Jennrich (1969). The proof is therefore left to the reader.

REFERENCES

Amemiya, T., 1974, The nonlinear two-stage least-squares estimator, Journal of Econometrics, 2, 105-110

Amemiya, T., 1975, The nonlinear limited-information maximum-likelihood estimator and the modified nonlinear two-stage least-squares estimator, Journal of Econometrics, 3, 375-386

Chung, K.L., 1974, A Course in Probability Theory, (Academic Press, New York-London)

Jennrich, R.I., 1969, Asymptotic properties of non-linear least squares estimators, The Annals of Mathematical Statistics, 40, 633-643

Jorgenson, D.W. and Laffont, J., 1974, Efficient estimation of nonlinear simultaneous equations with additive disturbances, Annals of Social and Economic Measurement, 3, 615-640