ON THE STRONG CONSISTENCY OF ESTIMATORS FOR CERTAIN DISTRIBUTED LAG MODELS WITH AUTOCORRELATED ERRORS

By

Phoebus J. Dhrymes

Discussion Paper Number 13
May/1971

On the Strong Consistency of Estimators for Certain Distributed Lag Models with Autocorrelated Errors*

Phoebus J. Dhrymes

1. Introduction.

In a previous paper [3], the author presented a maximum likelihood procedure for estimating the parameters of the model

$$(1)\quad y_t = \alpha_0 \sum_{i=0}^{\infty} \lambda_0^i x_{t-i} + u_t, \; u_t = \rho_0 u_{t-1} + \varepsilon_t, \qquad t = 1,2,\dots,T$$

where a zero subscript on a parameter symbol will always indicate its true value. In the above, $\{\varepsilon_t : t = 0, \pm 1, \pm 2, \dots\}$ was taken to be a sequence of mutually independent identically distributed (i.i.d.) $N(0,\sigma_0^2)$ variables. It was further assumed that $\lambda \in (0,1)$, $\rho \in (-1,1)$ and that the sequence of explanatory variables $\{x_t : t = 0, \pm 1, \pm 2, \dots\}$ was bounded nonstochastic and such that

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} x_t^2 = c$$

*The research on which this paper is based was in part supported by NSF grant GS 2289 at the University of Pennsylvania and was completed during the author's visit at the University of California, Los Angeles. It is a pleasure for me to express my gratitude to Thomas M. Liggett for invaluable advice at some stages of the preparation of this paper. I would also like to thank E. Malinvaud for critical comments on an earlier draft.

is well defined with $c > 0$.

If we define

$$\text{(2)} \qquad a = \sum_{i=0}^{\infty} \lambda_0^i x_{-i}$$

it was asserted in [3] that a may be consistently estimated by the procedure given there. This claim is false. Lack of consistency is, formally, a consequence of the fact that the "variable" to which it corresponds, viz., λ^t, $t = 1,2,\ldots, T$ has the property

$$\text{(3)} \qquad \sum_{t=1}^{\infty} \lambda^{2t} < \infty .$$

Intuitively, a represents the (nonstochastic) initial conditions of the model in (1). Thus, in view of the stability requirements, it is clear that the farther we are removed from the origin the "less" the position of the system depends on initial conditions. Consequently, additional observations as $T \to \infty$ convey less and less information regarding a.

Consistency for the other parameters, however, is preserved as was asserted in [3]. The proof given there is in many ways deficient. In the following we shall show that, under slightly more restrictive conditions, the estimators obtained by the search procedure given in [3] converge to the true parameters, not only in probability but with probability one[1] as well.

[1]For a definition of this term see [4, Ch. 3]; convergence with probability one is what we mean by strong consistency.

An interesting by-product of the argument used to establish this result is its applicability to problems of estimating, by minimum chi square methods, the parameters of nonlinear models whose error processes are finite order autoregressions. Such problems are occurring with increasing frequency in econometric research.

2. Convergence of Estimators.

Here we shall consider the model in (1) with the stochastic specifications given there but subject to the following additional conditions

(A.1) If we put $\omega = (\alpha, \lambda, \rho)'$, then $\omega \in \Omega$, where Ω is a closed bounded set and in particular $\lambda \in [\delta_1, 1 - \delta_1]$, $\rho \in [-1 + \delta_2, 1 - \delta_2]$, $\delta_1, \delta_2 > 0$ but small.

(A.2) The explanatory sequence $\{x_t : t = 0, \pm 1, \pm 2, \ldots\}$ obeys, for all t, $|x_t| < K$, for some constant K and

$$\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{\infty} x_t x_{t+\tau} = c(\tau)$$

the $c(\tau)$ being well defined constants and $c(0) > 0$.

(A.3) The true parameter, ω_0, is an interior point of Ω, and $1/T$ $\partial^2 L / \partial\omega\partial\omega$ converges to a nonsingular matrix for ω_0, L being a concentrated log likelihood function defined below.

Remark 1: The restrictions imposed by (A.1) are empirically innocuous. In practice we would be searching over an interval, say, [.001,.999] for λ and [-.999,.999] for ρ. The results we shall obtain would, thus, be inapplicable to models in which $|\lambda_0| \geq .999$ or $|\lambda_0| \leq .001$ and $|\rho_0| \geq .999$. It is clear that such restrictions are inconsequential. The condition in (A.3) is needed only in order to establish the asymptotic normality of the resulting estimators. We shall not derive such results here since they have been obtained in [5].

If we partially maximize the (log) likelihood function of the model in (1) with respect to σ^2 we obtain, upon division by T, the concentrated (log) likelihood function

$$(4) \qquad L_T(\omega;y,x) = -\frac{1}{2}[\ln(2\pi)+1] + \frac{1}{2T}\ln(1-\rho^2) - \frac{1}{2}\ln S_T(\omega;y,x)$$

where

$$(5) \qquad S_T(\omega;y,x) = \frac{1}{T}(y-\alpha x^*)' V^{-1}(y-\alpha x^*),\ y = (y_1,y_2,\ldots,y_T)',$$

$$x^* = (x_1^*,x_2^*,\ldots,x_T)',\ x_t^* = \sum_{i=0}^{\infty} \lambda^i x_{t-i} .$$

The function L referred to in (A.3) is simply $TL_T(\omega;y,x)$. We observe that $S_T(\cdot;y,x)$ is bounded away from zero, for $\omega \in \Omega$. The plan of the argument is as follows:

i. First, we show that $S_T(\omega;y,x)$ converges to its limit, say $S(\omega)$, with probability one uniformly in ω. Hence that $L_T(\omega;y,x)$ converges to its limit, say $L(\omega)$, with probability one uniformly in ω.

ii. Second, we show that the sequence of estimators defined by

$$L_T(\hat{\omega}_T : y,x) \geq L_T(\omega;y,x), \forall \omega \in \Omega$$

has at least one limit point, say ω_*, and that $\omega_0 = \omega_*$. Thus, we conclude that $\hat{\omega}_T$ converges to ω_0 with probability one since ω_* is any limit point.

In the course of the argument we shall use a number of results not generally employed in the literature of econometrics, and for that reason we state them as theorems giving appropriate references for their proofs.

THEOREM 1. (Bolzano-Weierstrass). Every bounded infinite set has at least one limit point and there exists a subsequence that converges to it.

Proof. See [1, p. 10] and [10, p. 38].

THEOREM 2. (Arzelà-Ascoli). Let $\{f_n\}$ be a sequence of equicontinuous functions from a compact topological space X to a metric space Y which converge at each point of X to a function f. Then $\{f_n\}$ converges to f uniformly on X.

Proof. For a discussion of a number of variations of this result see [9, pp. 153-155].

THEOREM 3. (Borel-Cantelli). Let Γ be a set of points γ, F be a σ-field, i.e., a collection of subsets of Γ such that if $A_n \in F$ then $A_n^c \in F$, $A_n \cap A_{n'} \in F$, $\bigcup_1^\infty A_n \in F$, A_n^c being the complement of A_n, and $P(\cdot)$ be a probability measure over F. If $A_n \in F$, then

$$\sum_{n=1}^{\infty} P(A_n) < \infty$$

implies

$$P(A_n, \text{ infinitely often}) = 0 .$$

Proof. [2, p. 41].

THEOREM 4. (Birkhoff-Khinchin). Let $\{x_n : n = 0, \pm 1, \pm 2, \ldots\}$ be a strictly stationary process[2] and suppose that

$$E|x_0| < \infty$$

then, with probability one

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=m}^{n+m} x_k = x^*$$

exists and moreover $E(x^*) = E(x)$.

Proof. This is given as Corollary 2 to the Birkhoff-Khinchin theorem in [6, p.129].

In order to show convergence of $L_T(\omega;y,x)$ to $L(\omega)$ with probability one, uniformly in ω, it will suffice to show the same for $S_T(\omega;y,x)$ and $S(\omega)$.

Remark 2: Proceeding as above would also show that minimum chi-square estimators, obtained by the condition

[2]For an explanation of this term, see [4, Ch. 9].

$$(6) \qquad S_T(\hat{\omega}_T : y,x) \leq S_T(\omega;y,x), \ \forall \omega \in \Omega$$

have the same (asymptotic) properties as maximum likelihood estimators in this context.

Remark 3: Note that in (5), ω enters only through α, x^* and V, but not through y, since the latter depends only on ω_0.

We have

$$(7) \quad S_T(\omega;y,x) = \frac{1}{T}(\alpha_0 x_0^* - \alpha x^*)' V^{-1}(\alpha_0 x_0^* - \alpha x^*) + \frac{2}{T}(\alpha_0 x_0^* - \alpha x^*)' V^{-1} u + \frac{1}{T} u' V^{-1} u$$

where x_0^* is a T-element vector whose t^{th}-element is $\Sigma_{i=0}^{\infty} \lambda_0^i x_{t-i}$. It will suffice to show that, as $T \to \infty$, each of the three terms in the right member of (7) converges to its limit uniformly in ω, and when the occasion requires it, with probability one.

Consider the third term first. Thus,

$$(8) \qquad \frac{1}{T} u'V^{-1}u = \frac{1}{T}\sum_{t=1}^{T} u_t^2 - 2\rho \frac{1}{T} \sum_{t=2}^{T} u_t u_{t-1} + \rho^2 \frac{1}{T} \sum_{t=2}^{T-1} u_t^2 .$$

But since $\{u_t : t = 0, \pm 1, \pm 2, \ldots\}$ is a strictly stationary process, by virtue of the i.i.d. assumption regarding $\{\varepsilon_t : t = 0, \pm 1, \pm 2, \ldots\}$, so are $\{u_t^2\}$ and $\{u_t u_{t-1}\}$. We note that

$$(9) \quad E(u_0^2) = \frac{\sigma_0^2}{1-\rho_0^2} < \infty, \quad E|u_0 u_{-1}| < \infty$$

$$\operatorname*{plim}_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} u_t^2 = \frac{\sigma_0^2}{1-\rho_0^2}, \quad \operatorname*{plim}_{T\to\infty} \frac{1}{T} \sum_{t=2}^{T} u_t u_{t-1} = \frac{\sigma_0^2 \rho_0}{1-\rho_0^2} .$$

Consequently, by Theorem 4,

$$\frac{1}{T} \sum_{t=1}^{T} u_t^2, \quad \frac{1}{T} \sum_{t=2}^{T} u_t u_{t-1} ,$$

converge with probability one to some limits, say ξ_0, ξ_1 respectively. But convergence with probability one implies convergence in probability and we conclude

$$(10) \qquad \xi_0 = \frac{\sigma_0^2}{1-\rho_0^2}, \quad \xi_1 = \rho_0 \xi_0 .$$

Therefore, in view of the particularly simple way in which ρ enters (8), we conclude that

$$\frac{1}{T} u'V^{-1}u \to \sigma_0^2 + \frac{\sigma_0^2}{1-\rho_0^2} (\rho - \rho_0)^2$$

with probability one uniformly in ω.

Remark 4: It is worth noting that in the argument above the normality of the $\{u_t\}$ process has not been employed. Indeed, it will never be employed, except in defining the likelihood function. The results follow from the fact that $\{\varepsilon_t : t = 0, \pm 1, \pm 2, \ldots\}$ is a sequence of i.i.d. random variables with zero mean and finite (absolute) moments of certain orders.

We now turn to the second term. We have

$$(11) \qquad \frac{1}{T}(\alpha_0 x_0^* - \alpha x^*)'V^{-1}u = \frac{1}{T}\alpha_0 x_0^{*\prime} V^{-1}u - \frac{1}{T}\alpha x^{*\prime} V^{-1}u .$$

It will suffice to show that the second term in the right member of (11) converges to its limit with probability one, uniformly in ω. The first term depends only on ρ but not on α or λ.

We may write the t^{th} element of x^* as

$$(12) \qquad x_t^* = a\lambda^t + \sum_{i=0}^{t-1} \lambda^i x_{t-i}, \qquad t = 1,2,\ldots,T$$

where a was defined in (2), and note that for all $\omega \in \Omega$

$$|a| < \frac{K}{\delta_1} .$$

Using (12) we can write

$$(13) \qquad \frac{1}{T}\alpha x^{*\prime} V^{-1}u = \frac{1}{T}\alpha a \underline{\lambda}' V^{-1}u + \frac{1}{T}\alpha x^{**\prime} V^{-1}u$$

where x^{**} is a T-element vector the t^{th} element of which is given by $\Sigma_{i=0}^{t-1} \lambda^i x_{t-i}$, $t = 1,2,\ldots,T$ and $\underline{\lambda} = (\lambda,\lambda^2,\ldots,\lambda^T)'$. We observe that

$$(14)\qquad \frac{1}{T}\,\underline{\lambda}'V^{-1}u = \frac{1}{T}\sum_{t=1}^{T}\lambda^t u_t - \rho\,\frac{1}{T}\sum_{t=2}^{T}\lambda^{t-1}u_t - \rho\,\frac{1}{T}\sum_{t=2}^{T}\lambda^t u_{t-1} + \rho^2\,\frac{1}{T}\sum_{t=2}^{T-1}\lambda^t u_t\ .$$

It will suffice to show that $1/T\ \Sigma_{t=1}^{T}\ \lambda^t u_t$ converges to its limit with probability one uniformly in λ. Convergence of the other terms is proved similarly and uniformity of convergence with respect to ρ is obvious from the representation in (14). But

$$u_t = \rho_0^t u_0 + \sum_{i=0}^{t-1}\rho_0^i\varepsilon_{t-i}, \qquad t = 1,2,\ldots$$

and we note that u_0 (the observation on the u-process at "time" zero) is a finite valued random variable. Thus, we may write

$$(15)\qquad \frac{1}{T}\sum_{t=1}^{T}\lambda^t u_t = \frac{u_0}{T}\sum_{t=1}^{T}(\lambda\rho_0)^t + \sum_{t=1}^{T}\lambda^t\left[\frac{1}{T}\sum_{i=0}^{t-1}\rho_0^i\varepsilon_{t-i}\right]$$

and observe that $\Sigma_{t=1}^{T}\ (\lambda\rho_0)^t$ is bounded by some constant independently of λ or T. Since u_0 is a finite valued random variable $\frac{u_0}{T}$ converges to zero with probability one. Hence we need be concerned only with the second term. Define

$$(16)\qquad W_{T,t} = \frac{1}{T}\sum_{i=0}^{t-1}\rho_0^i\varepsilon_{t-i}$$

and observe that

$$(17) \qquad \left| \sum_{t=1}^{T} \lambda^t W_{T,t} \right| \leq \Sigma\lambda^t \sup_t |W_{T,t}| \leq \frac{1}{\delta_1} \sup_t |W_{T,t}| .$$

Then, to show convergence with probability one, uniformly in λ, we must show that $\sup_t |W_{T,t}| \to 0$ with probability one. If $\varphi_1 > 0$ we have, using the Chebyshev inequality for fourth moments, for a random variable, z, having mean zero and finite fourth moment

$$\Pr\{|z| > \varphi_1\} \leq \frac{E(z^4)}{\varphi_1^4} .$$

If the ε-sequence has finite fourth order moments, we have, in view of the definition of $W_{T,t}$ in (16)

$$(18) \qquad \Pr\{|W_{T,t}| > \varphi_1\} \leq \frac{E(W_{T,t})^4}{\varphi_1^4} < \frac{K_1}{\varphi_1^4} \frac{1}{T^4}$$

where K_1 is some constant not depending on t, T or ρ_0. Since

$$(19) \qquad \sum_{T=1}^{\infty} \sum_{t=1}^{T} E(W_{T,t}^4) < K_1 \sum_{T=1}^{\infty} \frac{1}{T^3} < \infty$$

we conclude, by Theorem 3,

$$\sup_{t \leq T} |W_{T,t}| \to 0$$

with probability one. Hence, that $\Sigma_{t=1}^{T} \lambda^t W_{T,t}$ converges to zero with probability one uniformly in λ. We must now deal with

$$(20)\quad \frac{1}{T}\alpha x^{**\prime}V^{-1}u = \frac{1}{T}\alpha\left[\sum_{t=1}^{T}\sum_{i=0}^{t-1}\lambda^i x_{t-i}u_t - \rho\sum_{t=2}^{T}\sum_{i=0}^{t-2}\lambda^i x_{t-1-i}u_t - \rho\sum_{t=2}^{T}\sum_{i=0}^{t-1}\lambda^i x_{t-i}u_{t-1} + \rho^2\sum_{t=2}^{T-1}\sum_{i=0}^{t-2}\lambda^i x_{t-i}u_t\right].$$

It will suffice to show that

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{i=0}^{t-1}\lambda^i x_{t-i}u_t$$

converges to zero with probability one uniformly in λ; uniform convergence with respect to α and ρ is obvious from (20) in view of the fact that the set Ω is bounded (and closed). Using the representation $u_t = \rho_0^t u_0 + \Sigma_{i=0}^{t-1}\rho_0^i\,\varepsilon_{t-i}$, as before, it will suffice to show that

$$(21)\quad \frac{1}{T}\sum_{t=1}^{T}\sum_{j=0}^{t-1}\sum_{i=0}^{t-1}\lambda^i\rho_0^j x_{t-i}\varepsilon_{t-j} = \sum_{i=0}^{T-1}\sum_{j=0}^{T-1}\lambda^i\rho_0^j\frac{1}{T}\sum_{t=\max(i,j)+1}^{T}x_{t-i}\varepsilon_{t-j}$$

converges to zero uniformly in λ. Let

$$(22)\quad W_{T,i,j} = \frac{1}{T}\sum_{t}x_{t-i}\varepsilon_{t-j}\,.$$

We observe that, provided the ε-sequence has finite eighth order moments,

$$(23) \qquad E(W_{T,i,j}^{8}) < K_2 \frac{1}{T^4}$$

where K_2 is some constant not depending on λ, ρ_0, i, j, T or x_t.
Since

$$(24) \qquad \sum_{T=1}^{\infty} \sum_{i=0}^{T-1} \sum_{j=0}^{T-1} E(W_{T,i,j}^{8}) < K_2 \sum_{T=1}^{\infty} \frac{1}{T^2} < \infty$$

we conclude, by Theorem 3,

$$\sup_{i,j} |W_{T,i,j}| \to 0$$

with probability one, and hence that

$$\frac{1}{T} \alpha x^{**'} V^{-1} u \to 0$$

with probability one uniformly in ω.

Remark 5: The construction above proves more than is needed, since it shows convergence to zero with probability one uniformly in ω and ρ_0. Notice also that the bound in (23) uses the fact that the sequence of explanatory variables $\{x_t : t = 0, \pm 1, \pm 2, \ldots\}$ is bounded. This may be relaxed, however, if we assert, instead, that

$$\lim_{T\to\infty} d_T^2 = \infty, \quad \lim_{T\to\infty} \frac{\max\limits_{t\le T}|x_t|}{d_T} = 0, \quad d_T^2 = \sum_{t=1}^{T} x_t^2 ,$$

and in the definition of $S_T(\omega;y,x)$ we divide by d_T instead of T.

Finally, we must show that

$$\frac{1}{T}(\alpha_0 x_0^* - \alpha x^*)' V^{-1}(\alpha_0 x_0^* - \alpha x^*)$$

converges to its limit uniformly in ω. Notice that the quadratic form above does not contain random variables.

Let

$$(25) \qquad g_t(\omega) = \alpha_0 \sum_{i=0}^{\infty} \lambda_0^i x_{t-i} - \alpha \sum_{i=0}^{\infty} \lambda^i x_{t-i}$$

and note that

$$(26) \qquad \frac{1}{T}(\alpha_0 x_0^* - \alpha x^*)' V^{-1}(\alpha_0 x_0^* - \alpha x^*) = \frac{1}{T}\left[\sum_{t=1}^{T} g_t^2(\omega) - 2\rho \sum_{t=2}^{T} g_t(\omega) g_{t-1}(\omega) + \rho^2 \sum_{t=2}^{T-1} g_t^2(\omega)\right].$$

It will suffice to show that $1/T\ \Sigma_{t=1}^{T} g_t^2(\omega)$ converges to its limit uniformly in ω.

We note that

$$(27) \qquad |g_t(\omega)| < \frac{\alpha_0 K}{1-\lambda_0} + \frac{K_3}{\delta_1}$$

for all t, where K_3 is some constant independent of t and ω. Moreover, let $\omega_1, \omega_2 \in \Omega$, then for all t,

$$(28) \quad |g_t(\omega_1) - g_t(\omega_2)| = \Big| (\alpha_2 - \alpha_1) \sum_{i=0}^{\infty} \lambda_2^i x_{t-i}$$

$$+ \alpha_1(\lambda_2 - \lambda_1) \sum_{i=1}^{\infty} \Big(\sum_{j=0}^{i-1} \lambda_2^{i-j-1} \lambda_1^j \Big) x_{t-i} \Big| \leq |\alpha_2 - \alpha_1| \frac{K}{\delta_1} + |\lambda_2 - \lambda_1| \frac{K_4}{\delta_1^2}$$

where K_4 is a constant not depending on ω or t. What (27) and (28) show is that $\{g_t\}$ is a family of continuous functions defined on the compact set Ω which are uniformly bounded and equicontinuous.

Noting that

$$(29) \quad |g_t^2(\omega_2) - g_t^2(\omega_1)| = |g_t(\omega_2) + g_t(\omega_1)| \; |g_t(\omega_2) - g_t(\omega_1)|$$

we see that $\{g_t^2\}$ is also a family of uniformly bounded equicontinuous functions, continuous on the compact set Ω.

Define now

$$(30) \qquad f_{t,T}(\omega) = \frac{1}{T} g_t^2(\omega), \quad f_T(\omega) = \sum_{t=1}^{T} f_{t,T}(\omega)$$

and note that $\{f_T\}$ is a family of continuous functions on Ω, which is uniformly bounded and equicontinuous. In view of assumption (A.2) $f_T(\omega)$ converges pointwise (in ω) to a function $f(\omega)$. We then conclude, in view of Theorem 2 that $\{f_T\}$ converges to its limit uniformly in ω.

We have therefore proved

LEMMA 1. Let $S_T(\omega;y,x)$ be as defined in (5). Then $S_T(\omega;y,x)$ converges to its limit with probability one, uniformly in ω and its limit is given by

$$(31) \qquad S(\omega) = \lim_{T\to\infty} \frac{1}{T} (\alpha_0 x_0^* - \alpha x^*)' V^{-1} (\alpha_0 x_0^* - \alpha x^*) + \sigma_0^2 + \frac{\sigma_0^2}{1-\rho_0^2} (\rho - \rho_0)^2 .$$

Moreover, $L_T(\omega;y,x)$, as defined in (4), converges to its limit with probability one, uniformly in ω, and its limit is given by

$$(32) \qquad L(\omega) = -\frac{1}{2}[\ln(2\pi) + 1] - \frac{1}{2} \ln S(\omega) .$$

Remark 6: It is important to recapitulate what role the various assumptions have played in establishing Lemma 1. The fact that $\{\varepsilon_t : t = 0, \pm 1, \pm 2, \ldots\}$ is a sequence of i.i.d. random variables with finite (absolute) moments of certain order was used in showing that $\{u_t^2\}$, $\{u_t u_{t-1}\}$ were strictly stationary and in invoking certain (ergodic) theorems regarding the convergence with probability one of sample means of processes having finite (absolute) first order moments (Birkhoff-Khinchin theorem). The properties of the ε-sequence were also employed in invoking the Borel-Cantelli lemma (Theorem 3) to show that terms of the form $1/T\ x^{*\prime} V^{-1} u$ converge to zero with probability one uniformly in ω. In this connection we have also employed the boundeness of the sequence of explantory variables. The compactness of the set Ω, of admissible parameters, was utilized in invoking Theorem 2 (Ascoli-Arzelà), as was the boundeness of the explanatory variables. Compactness for α is not essential since we may eliminate it by partial maximization, as was done in [5]. This, however, would lead to more involved, but conceptually identical, arguments.

Remark 7: Introducing into the model additional variables, i.e., considering

$$(33) \qquad y_t = \alpha_0 \sum_{i=0}^{\infty} \lambda_0^i x_{t-i,1} + \sum_{j=2}^{n} \alpha_{0j} x_{tj} + u_t$$

does not change anything substantially. Essentially the same arguments will go through if similar boundedness conditions are placed on the x_{tj}, $j = 2,3,\ldots,n$ and on the additional parameters α_{0j}, $j = 2,3,\ldots,n$.

Remark 8: Increasing the order of autoregression, i.e., if we consider, for finite m,

$$(34) \qquad u_t = \sum_{i=1}^{m} \rho_{0i} u_{t-i} + \varepsilon_t$$

will complicate the argument by forcing us to deal with more involved formulae but will leave the essential features of the proof unaltered.

Let us now show that the estimators defined by

$$L_T(\hat{\omega}_T;y,x) \geq L_T(\omega;y,x), \ \forall \ \omega \in \Omega \tag{35}$$

converges with probability one to ω_0. It will suffice to do so for the minimum chi-square estimators defined by

$$S_T(\hat{\omega}_T;y,x) \leq S_T(\omega;y,x), \ \forall \ \omega \in \Omega \ . \tag{36}$$

We first note that

$$S(\omega_0) = \sigma_0^2, \ S(\omega) \geq S(\omega_0), \ \forall \ \omega \in \Omega \ . \tag{37}$$

Let $\{\hat{\omega}_T\}$ be the sequence of estimators obeying (36). For each sequence $\{u_t\}$ this forms an infinite bounded set and by Theorem 1 it has at least one limit point; let ω_* be such limit point and $\{\hat{\omega}_{T_i}\}$ be a subsequence converging to ω_*.

Because $S_T(\omega;y,x)$ converges to $S(\omega)$ <u>with probability one uniformly in ω</u>, for almost all sequences $\{u_t\}$, we have

$$S(\omega_*) \leq S(\omega_0) \ . \tag{38}$$

But (37) then implies

$$S(\omega_*) = S(\omega_0) \ . \tag{39}$$

For the parameters of the model to be identified it is necessary that (39) imply that $\omega_* = \omega_0$. This is an identifiability condition, else the estimator would not be able to discriminate between ω_* and ω_0. Let us postpone the verification of this condition for the model under consideration and first complete the argument, assuming that (39) implies

(40) $$\omega_* = \omega_0 .$$

Since ω^* is <u>any</u> limit point of $\{\hat{\omega}_T\}$, (40) implies that the sequence, therefore, converges to ω_0. Since (38) and (39) hold with probability one - i.e. for almost all sequences $\{u_t\}$ - we conclude that

(41) $$\Pr\{\lim_{T\to\infty} \hat{\omega}_T = \omega_0\} = 1$$

which states that the estimator, $\hat{\omega}_T$, converges to the true parameter, ω_0, with <u>probability one</u>.

Let us now obtain the conditions under which the model under consideration is identifiable. What we must show is that (39) implies (40). In view of (31), we have

(42) $$\lim_{T\to\infty} \frac{1}{T} (\alpha_0 x_0^* - \alpha_* x^*)' V^{-1} (\alpha_0 x_0^* - \alpha_* x^*) + \sigma_0^2 + \sigma_0^2 \frac{(\rho_* - \rho_0)^2}{1 - \rho_0^2} = \sigma_0^2 .$$

This clearly implies

(43) $$\rho_* = \rho_0, \lim_{T\to\infty} \frac{1}{T} (\alpha_0 x_0^* - \alpha_* x^*)' V_0^{-1} (\alpha_0 x_0^* - \alpha_* x^*) = 0$$

What we must determine is the following: Under what conditions on the sequence $\{x_t\}$ does the second equation in (43) imply $\lambda_* = \lambda_0$, $\alpha_* = \alpha_0$.

It is clear that this is not a vacuous exercise. Consider, for instance, the sequence $x_t = c + \gamma^t$, $t = 1,2,\ldots$ and zero otherwise, for some constant c and $|\gamma| < 1$. It is easy to see that this obeys (A.2) and that all arguments leading to (39) remain perfectly valid. For this sequence the second equation in (43) becomes

$$(1 - \rho_0)^2 \left(\frac{\alpha_0}{1-\lambda_0} - \frac{\alpha_*}{1-\lambda_*} \right)^2 = 0 .$$

This, unfortunately, <u>does not imply</u> $\alpha_* = \alpha_0$, $\lambda_* = \lambda_0$. Thus, for this type of sequence we cannot infer that $\{\hat{\omega}_T\}$ converges to ω_0, nor that (41) is valid.

So, let us see what restrictions we need impose on the x-sequence in order to render the model identifiable. Using the representation in (25) we can write the second equation of (43) as

$$(44) \quad \lim_{T\to\infty} \frac{1}{T} \left[\sum_{t=1}^{T} g_t^2(\omega) - 2\rho_0 \sum_{t=2}^{T} g_t(\omega) g_{t-1}(\omega) + \rho_0^2 \sum_{t=2}^{T} g_{t-1}^2(\omega) \right]$$

$$= \sum_{j=0}^{\infty} \varphi(0,j)c(0,j) + \sum_{i=1}^{\infty} \varphi(i,0)c(i,0) + \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \varphi(i,j)c(i,j)$$

where

(45) $\varphi(0,0) = \alpha_0^2 + \alpha_*^2 - 2\alpha_0\alpha_*$

(46) $\varphi(0,j) = \alpha_0^2\lambda_0^j + \alpha_*^2\lambda_*^j - (\alpha_0\alpha_*)(\lambda_0^j + \lambda_*^j)$

$-\rho_0[\alpha_0^2\lambda_0^{j-1} + \alpha_*^2\lambda_*^{j-1} - \alpha_0\alpha_*(\lambda_*^{j-1} + \lambda_0^{j-1})]$ $\qquad j = 1,2,\ldots$

$\varphi(i,0) = \alpha_0^2\lambda_0^i + \alpha_*^2\lambda_*^i - (\alpha_0\alpha_*)(\lambda_0^i + \lambda_*^i)$

$-\rho_0[\alpha_0^2\lambda_0^{i-1} + \alpha_*^2\lambda_*^{i-1} - (\alpha_0\alpha_*)(\lambda_0^{i-1} + \lambda_*^{i-1})]$ $\qquad i = 1,2,\ldots$

(47) $\varphi(i,j) = [\alpha_0^2\lambda_0^{i+j} + \alpha_*^2\lambda_*^{i+j} - \alpha_0\alpha_*(\lambda_0^i\lambda_*^j + \lambda_*^i\lambda_0^j)]$

$-\rho_0[2\alpha_0^2\lambda_0^{i-1+j} + 2\alpha_*^2\lambda_*^{i-1+j} - (\alpha_0\alpha_*)(\lambda_0^{i-1}\lambda_*^j + \lambda_*^{i-1}\lambda_0^j + \lambda_0^i\lambda_*^{j-1} + \lambda_0^{j-1}\lambda_*^i)]$

$+\rho_0^2[\alpha_0^2\lambda_0^{i+j-2} + \alpha_*^2\lambda_*^{i+j-2} - \alpha_0\alpha_*(\lambda_0^{i-1}\lambda_*^{j-1} + \lambda_*^{i-1}\lambda_0^{j-1})],$

$i,j = 1,2,3,\ldots$

and

(48) $$c(i,j) = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} x_{t-1}x_{t-j} .$$

For the bounded x-sequences we consider here, it is apparent that

(49) $$c(i,j) = c(i - j) = c(j - i) .$$

Denote, then, the "autocovariance" function of the x-sequence by

$$c(\tau) \qquad \tau = 0,1,2,\ldots .$$

Notice that we previously required that $c(0) > 0$.

If $c(\cdot)$ is not a constant function and has the property, say, $c(\tau) = 0$, $\tau \neq 0$, then the identifiability condition holds. To that end the coefficient of $c(0)$ in (44), must vanish. But we observe that for the coefficient of $c(0)$ we must have

$$(50) \qquad \sum_{i=0}^{\infty} \varphi(i,i) = (\alpha_0 - \alpha_*)^2 + \sum_{i=1}^{\infty} \left[(\alpha_0\lambda_0^i - \alpha_*\lambda_*^i) - \rho_0(\alpha_0\lambda_0^{i-1} - \alpha_*\lambda_*^{i-1})\right]^2 = 0$$

which immediately implies

$$(51) \qquad \alpha_0 = \alpha_*, \; \alpha_0\lambda_0^i - \alpha_*\lambda_*^i = \rho_0(\alpha_0\lambda_0^{i-1} - \alpha_*\lambda_*^{i-1}), \qquad i = 1,2,\ldots$$

But the second condition in (51) requires

$$(52) \qquad \lambda_0 = \lambda_* .$$

We have therefore proved

THEOREM 5. Consider the model

$$(53) \qquad y_t = \alpha_0 \sum_{i=0}^{\infty} \lambda_0^i x_{t-i} + u_t, \; u_t = \rho_0 u_{t-1} + \varepsilon_t, \quad t = 1,2,3,\ldots,T.$$

for i.i.d. $N(0,\sigma^2)$ ε_t and subject to the assumptions (A.1), (A.2); then the maximum likelihood and minimum chi-square estimators of the parameters $\omega = (\alpha,\lambda,\rho)'$ defined by (35) and (36), respectively, converge with probability one to ω_0 - the true parameter vector -

provided the identifiability condition that (39) implies (40) is satisfied. To this end it will suffice to place certain restrictions on the "autocovariance function," $c(\cdot)$, of the x-sequence. In particular if $c(\tau) = 0, \tau \neq 0$ identifiability is obtained.

COROLLARY 1. The estimator of σ_0^2 given by $S_T(\hat{\omega}_T;y,x)$ converges to σ_0^2 with probability one.

Proof. Obvious by the uniform convergence of $S_T(\omega;y,x)$ to $S(\omega)$ the ω-continuity of the latter and the fact that $\hat{\omega}_T$ converges to ω_0 with probability one.

Remark 9: If we consider the model of Remark 7 with or without the error specification of Remark 8, nothing of substance will change in the argument leading to Theorem 5, except that the verification of the identifiability condition will be rendered more cumbersome.

Remark 10: The preceding theorem and the argument leading to it are extremely useful in dealing with nonlinear (single equation) estimation problems. Such models may be formulated as

$$(54) \qquad y_t = g(x_t,\theta) + u_t \qquad t = 1,2,\ldots,T$$

where $\{x_t\}$ is the explanatory sequence and θ is a parameter vector constrained to lie in a compact set C. The error process may be specified to be strictly stationary. A typical specification in econometrics may be

(55) $$u_t = \sum_{i=1}^{m} \rho_i u_{t-i} + \varepsilon_t$$

where $\{\varepsilon_t : t = 0, \pm 1, \pm 2, \ldots\}$ is a sequence of i.i.d. random variables with zero mean and finite (absolute) moments of order, say, r.

In the model of (54), since the explanatory sequence is one of fixed numbers, it is notationally convenient to write $g(x_t, \theta)$ as $g_t(\theta)$. In this more suggestive notation we have

(56) $$y_t = g_t(\theta) + u_t \qquad t = 1,2,\ldots,T .$$

The functions $g_t(\cdot)$ may be quite nonlinear (in θ). If Φ is the covariance matrix of $u = (u_1, u_2, \ldots, u_T)'$, the minimum chi-square estimator may be defined as that which (globally) minimizes

(57) $$S_T(\theta; y, x) = \frac{1}{T} [y - g(\theta)]' \Phi^{-1} [y - g(\theta)]$$

where

$$g(\theta) = (g_1(\theta), g_2(\theta), \ldots, g_T(\theta))', \; y = (y_1, y_2, \ldots, y_T)' .$$

Thus the estimator, say $\hat{\theta}_T$, is defined by the condition

(58) $$S_T(\hat{\theta}_T; y, x) \leq S_T(\theta; y, x), \; \forall \, \theta \in C .$$

In order to show that

$$(59) \qquad \Pr\{ \lim_{T\to\infty} \hat{\theta}_T = \theta_0 \} = 1$$

where θ_0 is the true parameter vector one must show that following:

i. $\frac{1}{T}[g(\theta_0) - g(\theta)]'\Phi^{-1}[g(\theta_0) - g(\theta)]$ converges to its limit uniformly in θ, as $T \to \infty$

ii. $\frac{1}{T}[g(\theta_0) - g(\theta)]'\Phi^{-1}u$ converges to zero with probability one uniformly in θ

iii. $\frac{1}{T}u'\Phi^{-1}u$ converges to its (constant) limit with probability one uniformly in θ

iv. The model satisfies an identifiability condition, i.e., if $S(\theta)$ is the limit of $S_T(\theta;y,x)$, then $S(\theta_*) = S(\theta_0)$, $\theta_* \in C$ implies $\theta_* = \theta_0$.

Theorems 1 through 4 should be sufficient to establish the validity of the statements in i., ii. and iii. in the case of bounded explanatory variables. It would be an interesting research problem to determine the minimal conditions on the explanatory sequence that insure convergence with probability one of nonlinear minimum chi-square estimators.

Remark 11: The case of nonlinear least squares, i.e., the special case where, in (57), $\Phi = \sigma_0^2 I$ and the $\{u_t\}$ are i.i.d. random variables with mean zero and variance σ_0^2, has been treated in two important papers by Jennrich [7] and Malinvaud [8]. Thus, Jennrich and Malinvaud have given a solution to the problem of nonlinear least squares for models with i.i.d.

disturbances. Unfortunately, the procedure **employed** does not readily extend to the case of dependent random variables. The approach taken in this paper, however, holds the promise that we may find an equally general solution to the problem posed by the model in (54) and (55), the estimator being defined by (57). Asymptotic normality of the resulting estimates will require the additional assumptions that (a) θ_0 is an interior point of C and (b) $(\partial^2 S_T(\theta;y,x))/\partial\theta\partial\theta$ has a nonsingular limit, for $\theta = \theta_0$. It is for this reason that (A.3) was stated at the beginning at Section 2. Such requirements are quite apparent from the standard mean value theorem applied to $(\partial S_T(\theta;y,x))/\partial\theta$, which is typically employed in establishing the asymptotic distribution of such estimators.

3. Conclusion.

In this paper we have shown that the maximum likelihood and minimum chi-square estimators in the context of the model in (1) converge with probability one to the true parameters, under certain conditions on the explanatory sequence. Moreover, we have indicated how the argument we have employed may be modified in order to show that (nonlinear) minimum chi-square estimators in the context of the model in (54) and (55) converge to the true parameters with probability one, and that their asymptotic distribution is normal.

Finally, it is interesting to note that in the Monte Carlo study reported in [5], using the procedure examined in this paper, for the model

$$y_t = \alpha_1 \sum_{i=0}^{\infty} \lambda_0^i x_{t-i,1} + \alpha_2 x_{t2} + u_t, \; u_t = \rho_0 u_{t-1} + \varepsilon_t$$

one obtains, among other results, the following. For $\lambda_0 = .5$, $\rho_0 = .9$, $\alpha_1 = 2.00$, $\alpha_2 = 5.00$ and 100 replications the mean square error for $\hat{\lambda}_T$ is .002 for sample size 100, .003 for sample size 50. For $\hat{\rho}_T$ the corresponding quantities are .003 and .004. For α_1, they are .009, .010. For α_2, .006, .007.

REFERENCES

1. Buck, R.C: Advanced Calculus, New York: McGraw-Hill, 1956.
2. Breiman, L: Probability, Reading, Massachusetts: Addison-Wesley, 1968.
3. Dhrymes, P.J: "Efficient Estimation of Distributed Lags with Autocorrelated Errors" International Economic Review, vol. 10 (1969) pp. 47-67.
4. ___________: Econometrics: Statistical Foundations and Applications. New York: Harper and Row, 1970.
5. ___________: Distributed Lags: Problems of Formulation and Estimation. San Francisco: Holden-Day, 1971 (in press).
6. Gikhman, I.I. and A.V. Skorokhod: Introduction to the Theory of Random Processes, Philadelphia, W. B. Saunders, 1969.
7. Jennrich, R. J: "Asymptotic Properties of Non-Linear Least Squares," Annals of Mathematical Statistics, vol. 40 (1969) pp. 633-643.
8. Malinvaud, E: "The Consistency of Nonlinear Regressions" Annals of Mathematical Statistics, vol. 41 (1970) pp. 956-969.
9. Royden, H. L: Real Analysis, New York, MacMillan, 1963.
10. Rudin, W: Principles of Mathematical Analysis, New York McGraw-Hill, 1958.

Appendix

In this appendix we give a simple argument, justifying the transition from our equation (19) in the text, to the statement that $\sup_{t \leq T} |W_{T,t}| \to 0$ with probability one. Thus, let (Γ, F, P) be a probability space as in Theorem 3. Define, for any $\varphi > 0$,

$$A^{\varphi}_{T,t} = \{\gamma : |W_{T,t}| > \varphi, \ \gamma \in F\}$$

$$A^{\varphi}_{T} = \cup^{T}_{t=1} A^{\varphi}_{T,t} \ .$$

By the result exhibited in (18) and Theorem 3 we conclude that for any $\varphi > 0$

$$P(A^{\varphi}_{T}, \text{ i.o.}) = 0$$

Let

$$A^{*\varphi}_{T} = \{\gamma : \sup_{t \leq T} |W_{T,t}| > \varphi^*, \ \gamma \in F\}$$

Then $A^{*\varphi}_{T} \subset A^{\varphi}_{T}$ and therefore we conclude, for any $\varphi > 0$,

$$P(A^{*\varphi}_{T}, \text{ i.o.}) = 0 \ .$$

Let

$$w = \lim \sup \{ \sup_{t \leq T} |W_{T,t}| \}$$

and note that w is a nonnegative random variable. Also observe that

$$\Pr\{w > \varphi\} \leq P(A_T^{*\varphi} \text{ i.o}) = 0 .$$

Consequently conclude that

$$\sup_{t \leq T} |W_{T,t}| \to 0$$

with probability one, as claimed.

A similar argument will establish the transition from equation (24) to the statement

$$\sup_{i,j} |W_{T,i,j}| \to 0$$

with probability one.