The Annals of Statistics
1980, Vol. 8, No. 4, 778–788

STRONG CONSISTENCY OF LEAST SQUARES ESTIMATORS IN LINEAR REGRESSION MODELS[1]

BY N. CHRISTOPEIT AND K. HELMES

University of Bonn

For the linear regression model $y = X\beta + u$ with stochastic regressor matrix, strong consistency of the least squares estimator of β is proved in the case of martingale difference errors and predetermined regressors and for the case where errors and regressors are orthogonal up to the second order. The results obtained are applied to parameter estimation in autoregressive processes, leading to strong consistency if the errors are quasi-independent up to the fourth order.

1. Introduction. Let x_t and $\varepsilon_t (t = 1, 2, \cdots)$ be sequences of random $K \times 1$ vectors and random variables, respectively, defined on some probability space $(\Omega, \mathbb{F}, P)$. We shall be concerned with the linear regression model

$$\eta_t = x_t'\beta + \varepsilon_t, \qquad t = 1, 2, \cdots \tag{1.1}$$

where β is a nonstochastic $K \times 1$ parameter vector. Introducing the following vectors and matrices:

$$y_T = (\eta_1, \cdots, \eta_T)', \qquad u_T = (\varepsilon_1, \cdots, \varepsilon_T)',$$
$$X_T = (x_1, \cdots, x_T)',$$

(1.1) can be written in the form

$$y_T = X_T\beta + u_T, \qquad T = 1, 2 \cdots. \tag{1.2}$$

The least squares estimator (LSE) of β based on the first T observations is defined as solution of the normal equations

$$X_T'X_T\hat{\beta}_T = X_T'y_T. \tag{1.3}$$

Consider the random time

$$\tilde{T}(\omega) = \min\{T / P_T(\omega) \text{ nonsingular}\},$$

where $P_T = X_T'X_T$. Certainly, $\tilde{T}$ is measurable. Actually, it is a stopping time with respect to the increasing sequence of σ-fields $\mathbb{F}_{t-1} (t = 1, 2, \cdots)$ to be defined below. For fixed T, the LSE is well defined on the ω-set given by $T \geqslant \tilde{T}(\omega)$ through the relation

$$\tilde{\beta}_T = P_T^{-1}X_T'y_T. \tag{1.4}$$

Received May 1977, revised November 1978.

[1]This work was supported by the Sonderforschungsbereiche 21 and 72 at the University of Bonn.

AMS 1970 *subject classifications.* Primary 62J05; secondary 60F15.

Key words and phrases. Least squares estimators, linear regression, strong consistency, autoregressive processes.

Henceforth, we shall assume that the random time $\tilde{T}$ is finite a.e. This means that, with probability one, the LSE is finally given by (1.4) as T increases.

In proving consistency we shall be concerned with the estimation error

$$\hat{\beta}_T - \beta = P_T^{-1}X_T'u_T \tag{1.5}$$

(for $T \geqslant \tilde{T}$). With $d_{Ti}^2 = \Sigma_{t=1}^T x_{ti}^2 = (P_t)_{ii}$, $i = 1, 2, \cdots, K$, this may be written in the form

$$\hat{\beta}_T - \beta = P_T^{-1} \operatorname{diag}(d_{T1}^2, \cdots, d_{TK}^2) z_T, \tag{1.6}$$

where the ith component of the random vector z_T is given by

$$z_{Ti} = d_{Ti}^{-2} \Sigma_{t=1}^T x_{ti}\varepsilon_t. \tag{1.7}$$

More generally, take a nondecreasing function $g : \mathbb{R}_+ \to \mathbb{R}_+$ and put $D_T := \operatorname{diag}(g(d_{T1}^2), \cdots, g(d_{TK}^2))$. Then

$$\hat{\beta}_T - \beta = P_T^{-1} D_T z_T, \tag{1.8}$$

where now

$$z_{Ti} = g^{-1}(d_{Ti}^2) \Sigma_{t=1}^T x_{ti}\varepsilon_t. \tag{1.9}$$

The assumptions to be made below will ensure that these expressions are well defined for T large enough (depending on ω).

Starting from the representation (1.6), (1.7) Drygas [3] has shown that for nonstochastic regressors $x_1, x_2, \cdots$ the LSE is strongly consistent under assumptions that are basically deterministic specializations of the assumptions we are going to introduce. For the one dimensional problem of fitting a straight line a necessary and sufficient condition for strong consistency of the least squares slope estimator has been given by Lai and Robbins [4] in the case of i.i.d. error terms. The case of stochastic regressors is considered in [2]. The assumptions are somewhat stronger than ours; in addition, Anderson and Taylor require that P_T is nonsingular with probability one for some nonrandom time. In our terminology: $\tilde{T}(\omega) = T_0$ a.e. for some T_0. In applications, e.g. in autoregressive models, where the P_T-matrix contains the lagged η-values, this will, however, generally not be the case without further assumptions about the probability distributions involved.

2. The main result. Let us specify the assumptions.

(A). $E(\varepsilon_t/\mathbb{F}_{t-1}) = 0$ for $t = 1, 2, \cdots$, where $\mathbb{F}_{t-1}$ is the σ-field generated by $x_t, \cdots, x_1; \varepsilon_{t-1}, \cdots, \varepsilon_1$.

(B). $\varepsilon_t \in L_2$, and $\sup_t E(\varepsilon_t^2/\mathbb{F}_{t-1}) < \infty$ a.e.

(C). $d_{ti}^2 \to \infty$ a.e. as $t \to \infty$, $i = 1, 2, \cdots, K$.

(D). Let $g : \mathbb{R}_+ \to \mathbb{R}_+$ be a nondecreasing function satisfying

$$\int_c^\infty g^{-2}(t)\, dt = \bar{g} < \infty \tag{2.1}$$

for some $c > 0$ and put

$$D_T = \operatorname{diag}(g(d_{T1}^2), \cdots, g(d_{TK}^2)).$$

Then the matrices $P_T^{-1}D_T (T \geqslant \tilde{T})$ are a.e. bounded uniformly in T.

Examples of such functions are $g(d_{Ti}^2) = d_{Ti}^{1+\delta}$ or

$$g(d_{Ti}^2) = d_{Ti} \log^{(1+\delta)/2} d_{Ti} \qquad \text{for} \quad d_{Ti} \geqslant 1,$$
$$= 0 \qquad \text{else},$$

with some constant $\delta > 0$. Here d_{Ti} denotes the positive square root of d_{Ti}^2. The first example with $\delta = 1$ is the case considered in [2] and [3].

The basic result is contained in

LEMMA 1. *Let z_T be defined by* (1.9). *Under assumptions* (A)–(D)

$$z_{Ti} \to 0 \text{ a.e. as } T \to \infty, \qquad i = 1, \cdots, K.$$

PROOF. Note that by virtue of (C) and (2.1) z_T is well defined for T large enough (depending on ω). For notational simplicity let us suppress the index i in the sequel. Define

$$\tilde{t}(\omega) = \min\{t/d_t^2(\omega) \geqslant c\}.$$

By virtue of (C), $\tilde{t}$ is finite a.e. Moreover, it is a stopping time with respect to the increasing sequence of σ-algebras $\mathbb{F}_{t-1}(t = 1, 2, \cdots)$. In particular, the sets $[\tilde{t} < t]$ are $\mathbb{F}_{t-1}$- measurable for all t. Consider the random variables

$$\zeta_t = \chi_{[\tilde{t}<t]} g^{-1}(d_t^2) x_t \varepsilon_t,$$

where χ_A denotes the indicator of A and the convention $0 \cdot \infty = 0$ has been adopted. ζ_t is $\mathbb{F}_t$-measurable, and, since g^{-1} is decreasing on (c, ∞),

$$x_t^2 g^{-2}(d_t^2) \leqslant \int_{d_{t-1}^2}^{d_t^2} g^{-2}(s)\, ds \leqslant \int_c^\infty g^{-2}(s)\, ds = \bar{g} \text{ a.e.}$$

for $t > \tilde{t}$. Hence

$$\zeta_t^2 \leqslant \bar{g} \varepsilon_t^2 \text{ a.e.},$$

i.e., $\zeta_t \in L_2$. Moreover,

$$E(\zeta_t/\mathbb{F}_{t-1}) = 0.$$

Hence the process $(o_T, \mathbb{F}_T)(T = 1, 2, \cdots)$ defined by

$$w_T = \Sigma_{t=1}^T \zeta_t$$

is a square integrable martingale. The associated increasing process (i.e., the process A that makes $w^2 - A$ a martingale) is given by

$$A_T = \Sigma_{t=1}^T E(\zeta_t^2/\mathbb{F}_{t-1})$$
$$= \Sigma_{t=1}^T \sigma_t^2 x_t^2 g^{-2}(d_t^2) \chi_{[\tilde{t}<t]},$$

where $\sigma_t^2 = E(\varepsilon_t^2/\mathbb{F}_{t-1})$. With $\bar{\sigma}^2 = \sup_t \sigma_t^2 (< \infty$ a.e. by (B)) we obtain the estimate

$$\Sigma_{t=1}^\infty E(\zeta_t^2/\mathbb{F}_{t-1}) \leqslant \bar{\sigma}^2 \Sigma_{t>\tilde{t}} x_t^2 g^{-2}(d_t^2) \leqslant \bar{\sigma}^2 \Sigma_{t>\tilde{t}} \int_{d_{t-1}^2}^{d_t^2} g^{-2}(s)\, ds$$
$$\leqslant \bar{\sigma}^2 \int_c^\infty g^{-2}(s)\, ds < \infty,$$

which shows that $\lim_{T\to\infty} A_T < \infty$ a.e. on Ω; consequently (cf. [5] Proposition

VII-2-3) w_T converges a.e. to a finite limit. By Kronecker's lemma, using (C) and (D) and $\lim_{t\to\infty} g(t) = \infty$,

$$z'_T = g^{-1}(d_T^2)\Sigma_{t=\tilde{t}+1}^T x_t\varepsilon_t \to 0 \qquad \text{a.e. as} \quad T \to \infty.$$

The assertion now follows from the observation that the asymptotic behavior of z_T and z'_T is the same.

Strong consistency of the LSE is now an immediate consequence of Lemma 1 together with formula (1.8) and (D).

THEOREM 1. *Under assumptions* (A)–(D) *the* LSE *converges* a.e. *to the true parameter value.*

Assumption (D) will in general be difficult to verify directly. In the following, we shall give a sufficient condition for (D) to hold. To this end, let $\lambda_{\min}(T)$ and $\lambda_{\max}(T)$ denote the smallest and the largest eigenvalue of P_T, respectively. If $T \geqslant \tilde{T}$, both are positive. Then, for the Euclidean norm

$$\| P_T^{-1} \| \leqslant K^{\frac{1}{2}}\lambda_{\min}^{-1}(T),$$

$$\|D_T\| \leqslant K^{\frac{1}{2}}\max_i g(d_{Ti}^2) \leqslant K^{\frac{1}{2}} g(\lambda_{\max}(T)),$$

hence $\| P_T^{-1}D_T\| \leqslant K \cdot g(\lambda_{\max}(T))/\lambda_{\min}(T)$. Thus, (D) is implied by

(D′). Let $g : \mathbb{R}_+ \to \mathbb{R}_+$ be a nondecreasing function satisfying (2.1). Then

$$\sup_{T\geqslant\tilde{T}} g(\lambda_{\max}(T))/\lambda_{\min}(T) < \infty \text{ a.e.}$$

Choosing $g(t) = t^{(1+\delta)/2}$, we obtain

COROLLARY 1. *Suppose that assumptions* (A)–(C) *hold and that*

$$\sup_{T\geqslant\tilde{T}} \lambda_{\max}^{(1+\delta)/2}(T)/\lambda_{\min}(T) < \infty \text{ a.e.}$$

for some $\delta > 0$. *Then the* LSE *is strongly consistent.*

For $\delta = 1$ this is precisely the condition required in [2].

Assumption (A) implies that $\varepsilon_t (t = 1, 2, \cdots)$ is a martingale difference sequence. As an example of a whole class of processes satisfying (A) and (B) consider disturbances of the form $\varepsilon_t = a_t e_t$, where the random variables a_t and e_t are such that the following conditions are fulfilled:

(1) e_t is independent of $a_1, \cdots, a_t; e_1, \cdots, e_{t-1}$.

(2) $e_t \in L_2$, $E(e_t) = 0$, $\sup_t E(e_t^2) < \infty$.

(3) a_t is a.e. bounded uniformly in t (where the bound may depend on ω).

Then, if in addition the x_t are predetermined, i.e., measurable with respect to the σ-field generated by $\varepsilon_1, \cdots, \varepsilon_{t-1}$, the processes x_t, $\varepsilon_t (t = 1, 2, \cdots)$ satisfy assumptions (A) and (B).

Finally, let us have a short glance at the vectorial case:

$$\eta_t = B'x_t + \varepsilon_t,$$

where η_t, ε_t are now random $G \times 1$-vectors and B is a $K \times G$ parameter matrix.

The formula corresponding to (1.6) is

$$\hat{B}_T - B = P_T^{-1}\text{diag}(d_{T1}^2, \cdots, d_{TK}^2)Z_T,$$

where Z_T is the matrix whose components are given by

$$Z_{Tij} = d_{Ti}^{-2}\Sigma_{t=1}^{T}x_{ti}\varepsilon_{tj},$$
$$i = 1, 2, \cdots, K; j = 1, 2, \cdots, G.$$

Hence the proof of Theorem 1 carries over to the vectorial case.

In applications (cf. Section 3) it can often be shown that

(E) $$\frac{1}{T}P_T \to P \text{ a.e. as } T \to \infty,$$

where P is some nonstochastic positive definite matrix. It is easy to see that (E) implies (C) and (D) (take $g(t) = t$). But in this case we can considerably weaken the assumptions about the stochastic independence of x_t and ε_t.

(A1) $\varepsilon_t \in L_4$, $E(\varepsilon_t) = 0 (t = 1, 2, \cdots)$. $\bar{\sigma}^2 = \sup_t E(\varepsilon_t^2) < \infty$.

(A2) $x_{ti} \in L_4$ $(t = 1, 2, \cdots; i = 1, \cdots, K)$, and

$$\Sigma_{t=1}^{\infty}t^{-2}E(x_{ti}^2)\log^2 t < \infty.$$

(A3) $E(x_{ti}\varepsilon_t) = 0$; $E(x_{ti}^2\varepsilon_t^2) = E(x_{ti}^2)E(\varepsilon_t^2)(t = 1, 2, \cdots)$. $E(x_{ti}x_{si}\varepsilon_t\varepsilon_s) = 0$ for $t \neq s$, $i = 1, \cdots, K$.

THEOREM 2. *Under assumptions* (A1)–(A3) *and* (E) *the* LSE *is strongly consistent.*

PROOF. Starting from formula (1.5) we write

$$\hat{\beta}_T - \beta = \left(\frac{1}{T}P_T\right)^{-1}\frac{1}{T}X'_T u_T$$
$$= \left(\frac{1}{T}P_T\right)^{-1} z_T,$$

where now z_T has the components

$$z_{Ti} = \frac{1}{T}\Sigma_{t=1}^{T}x_{ti}\varepsilon_t$$
$$= \frac{1}{T}\Sigma_{t=1}^{T}\xi_{ti}, \qquad i = 1, \cdots, K.$$

By virtue of (A1)–(A3),

$$E(\xi_{ti}) = 0, \quad E(\xi_{ti}\xi_{si}) = 0 \text{ for } t \neq s,$$

and

$$E(\xi_{ti}^2) = E(x_{ti}^2)E(\varepsilon_t^2) \leqslant \bar{\sigma}^2 E(x_{ti}^2)$$

for all $t = 1, 2, \cdots, i = 1, \cdots$, K. Hence

$$\Sigma_{t=1}^{\infty}t^{-2}E(\xi_{ti}^2)\log^2 t < \infty,$$

and by Satz 3.2.2. in [6], which we cite below for further reference,

$$z_{Ti} \to 0 \text{ a.e. as } T \to \infty, \qquad i = 1, \cdots, K.$$

PROPOSITION 1. *Let $\xi_t (t = 1, 2, \cdots)$ be a sequence of square integrable random variables such that*

$$E(\xi_t) = E(\xi_t \xi_s) = 0 \qquad (s, t = 1, 2, \cdots; t \neq s)$$

and

$$\Sigma_{t=1}^{\infty} t^{-2} E(\xi_t^2) \log^2 t < \infty.$$

Then

$$\frac{1}{T} \Sigma_{t=1}^{T} \xi_t \to 0 \text{ a.e.}$$

REMARKS. In Section 3 we shall work with the following condition which is obviously sufficient for (A2):
(A2′). $x_{ti} \in L_4$, and $E(x_{ti}^2) = 0(t^{\delta})$ for some $0 < \delta < 1$.

3. Autoregressive processes. As an example let us consider the autoregressive process

$$\eta_t = \alpha_1 \eta_{t-1} + \cdots + \alpha_H \eta_{t-H} + \gamma' z_t + \varepsilon_t, \tag{3.1}$$

where $\alpha_h (h = 1, \cdots, H)$ are fixed numbers, γ is a fixed $N \times 1$-vector, $z_t(t = 1, 2, \cdots)$ are nonstochastic exogenous variables and $\varepsilon_t(t = 1, 2, \cdots)$ are random disturbances. We shall first give a proof of strong consistency of the LSE for $(\alpha_1, \cdots, \alpha_H, \gamma')$ which is based on Theorem 2.

The assumptions needed are the following:

(I) The errors $\varepsilon_t \in L_4(t = 1, 2, \cdots)$ and are quasi-independent up to order 4, i.e.,

$$E(\Pi_{i=1}^{4} \varepsilon_{t_i}^{r_i}) = \Pi_{i=1}^{4} E(\varepsilon_{t_i}^{r_i})$$

for every $(t_1, \cdots, t_4)$ and every choice of exponents $0 \leqslant r_i \leqslant 4$ such that $r_1 + \cdots + r_4 \leqslant 4$.

(II) $E(\varepsilon_t) = 0$. $E(\varepsilon_t^2) = \sigma_t^2$; $\sup_t \sigma_t^2 = \bar{\sigma}^2 < \infty$ and $\lim_{T\to\infty}(1/T)\Sigma_{t=1}^{T} \sigma_t^2 = \sigma^2$ exists and is positive. $E(\varepsilon_t^4) = 0(t^{\delta})$ for some $0 < \delta < 1$.

(III) All the roots of the characteristic equation

$$\lambda^H - \alpha_1 \lambda^{H-1} - \cdots - \alpha_H = 0 \tag{3.2}$$

are less than 1 in absolute value.

(IV) The limits

$$\overline{M}_{zz}(\Delta) = \lim_{T\to\infty} \frac{1}{T} \Sigma_{t=1}^{T-\Delta} z_t z_{t+\Delta}'$$

exist and are finite for $\Delta = 0, 1, 2, \cdots$ $\overline{M}_{zz}(0)$ is positive definite. Furthermore,

$$z_{ti}^2 = 0(t^{\delta}), \qquad i = 1, \cdots, N,$$

for some $0 < \delta < 1$.

The concept of quasi-independent disturbances has been used in [7] to show weak consistency of LSE for the coefficients $(\alpha_1, \cdots, \alpha_H, \gamma')$. Note that the exogenous variables z_t need not be bounded.

Introducing the vectors

$$\beta = (\alpha_1, \cdots, \alpha_H, \gamma')', \qquad \eta_-(t) = (\eta_{t-1}, \cdots, \eta_{t-H})',$$
$$x_t = (\eta'_-(t), z'_t)' \tag{3.3}$$

and y_T, u_T, X_T as in Section 1 we obtain the standard form (1.2).

The matrix $(1/T)X'_TX_T = (1/T)\Sigma_{t=1}^T x_t x'_t$ can be decomposed as follows:

$$\frac{1}{T}X'_TX_T = \begin{pmatrix} M_{11} & M_{12} \\ M'_{12} & M_{22} \end{pmatrix}$$

with

$$M_{11} = \frac{1}{T}\Sigma_{t=1}^T \eta_-(t)\eta'_-(t), \qquad M_{12} = \frac{1}{T}\Sigma_{t=1}^T \eta_-(t) z'_t,$$
$$M_{22} = \frac{1}{T}\Sigma_{t=1}^T z_t z'_t.$$

Let us assume for simplicity that the initial values $\eta_-(1) = 0$. Then the solution of (3.1) takes the form

$$\eta_t = \Sigma_{\tau=0}^{t-1} a_\tau \varepsilon_{t-\tau} + \Sigma_{\tau=0}^{t-1} a_\tau \gamma' z_{t-\tau} \tag{3.4}$$
$$= \eta_t^* + w_t.$$

The $a_\tau (\tau = 0, 1, 2, \cdots)$ satisfy the homogeneous difference equations

$$a_\tau = \alpha_1 a_{\tau-1} + \cdots + \alpha_H a_{\tau-H}, \qquad \tau = 1, 2, \cdots,$$

with initial conditions

$$a_0 = 1, a_{-1} = \cdots = a_{1-H} = 0.$$

Assumption (III) implies that $|a_\tau| = 0(\lambda_0^\tau)$, where $|\lambda_{\max}| < \lambda_0 < 1$ and $\lambda_{\max}$ denotes the root of (3.2) with largest absolute value. This means in particular that

$$\Sigma_{\tau=0}^\infty |a_\tau| < \infty \quad \text{and} \quad \Sigma_{\tau=0}^\infty |a_\tau|^2 < \infty. \tag{3.5}$$

In order to abbreviate the notation let us introduce the following symbols for real a_t, ξ_t, ζ_t:

$$[\xi_{-h}\zeta_{-k}]_T = \frac{1}{T}\Sigma_{t=1}^T \xi_{t-h}\zeta_{t-k},$$
$$(a, \xi)_t = \Sigma_{\tau=0}^{t-1} a_\tau \xi_{t-\tau}.$$

The next lemma is basic for the considerations to follow.

LEMMA 2. *Let ξ_t, $\zeta_t (t = 1, 2, \cdots)$ and a_t, $b_t (t = 0, 1, 2, \cdots)$ be sequences of real numbers such that $\Sigma_{\tau=0}^\infty |a_\tau| < \infty$ and $\Sigma_{\theta=0}^\infty |b_\theta| < \infty$. Suppose that the limits*

$\lim_{T\to\infty} [\xi\xi]_T = m_{\xi\xi}(0), \lim_{T\to\infty} [\zeta\zeta]_T = m_{\zeta\zeta}(0)$ *and*

$$\lim_{T\to\infty} \frac{1}{T}\Sigma_{t=1}^{T-\Delta}\xi_t\zeta_{t+\Delta} = m_{\xi\zeta}(\Delta)$$

$$\lim_{T\to\infty} \frac{1}{T}\Sigma_{t=1}^{T-\Delta}\xi_{t+\Delta}\zeta_t = m_{\xi\zeta}(-\Delta)$$

exist and are finite for all $\Delta = 0, 1, 2, \cdots$. *Then for all integers* $h \geqslant 0$, $k \geqslant 0$

$$\lim_{T\to\infty} \left[(a, \xi)_{-h}(b, \zeta)_{-k}\right]_T = \Sigma_{\tau=0}^{\infty}\Sigma_{\theta=0}^{\infty}a_\tau b_\theta m_{\xi\zeta}((h-k)+(\tau-\theta)).$$

PROOF. Define finite signed measures F and G on $\mathbb{Z}^+$ by putting $F(\{\tau\}) = a_\tau$, $G(\{\theta\}) = b_\theta$ and let H be the product measure. Then

$$(3.6)\quad \left[(a, \xi)_{-h}(b, \zeta)_{-k}\right]_T = \frac{1}{T}\Sigma_{t=1}^{T}\int_{\tau\leqslant t-h-1} dF(\tau)\xi_{t-h-\tau}\int_{\theta\leqslant t-k-1}dG(\theta)\zeta_{t-k-\theta}$$

$$= \int dH(\tau,\theta)\frac{1}{T}\Sigma_{t=t^*(\tau,\theta)}^{T}\xi_{t-h-\tau}\zeta_{t-k-\theta}$$

where $t^*(\tau, \theta) = \max(\tau + h + 1, \theta + k + 1)$. The estimate

$$\left|\frac{1}{T}\Sigma_{t=t^*}^{T}\xi_{t-h-\tau}\zeta_{t-k-\theta}\right| \leqslant \frac{1}{T}\left(\Sigma_{t=t^*}^{T}\xi_{t-h-\tau}^2\right)^{\frac{1}{2}}\left(\Sigma_{t=t^*}^{T}\zeta_{t-k-\theta}^2\right)^{\frac{1}{2}}$$

$$\leqslant [\xi\xi]_T^{\frac{1}{2}}[\zeta\zeta]_T^{\frac{1}{2}}$$

$$\leqslant \left(m_{\xi\xi}(0)+c\right)^{\frac{1}{2}}\left(m_{\zeta\zeta}(0)+c\right)^{\frac{1}{2}},$$

where c is a constant, shows that the integrand in (3.4) is bounded by a constant uniformly in T and (τ, θ). Hence by dominated convergence

$$\lim_{T\to\infty} \left[(a, \xi)_{-h}(b, \zeta)_{-k}\right]_T = \int dH(\tau,\theta) \lim_{T\to\infty} \frac{1}{T}\Sigma_{t=t^*(\tau,\theta)}^{T}\xi_{t-h-\tau}\zeta_{t-k-\theta}.$$

For $h + \tau \geqslant k + \theta$, $\Delta = (h + \tau) - (k + \theta)$, we find

$$\lim_{T\to\infty} \frac{1}{T}\Sigma_{t=t^*}^{T}\xi_{t-h-\tau}\zeta_{t-k-\theta} = \lim_{T\to\infty} \frac{1}{T}\Sigma_{t=1}^{T-(h+\tau)}\xi_t\zeta_{t+(h+\tau)-(k+\theta)}$$

$$= \lim_{T\to\infty} \frac{1}{T}\Sigma_{t=1}^{T-\Delta}\xi_t\zeta_{t+\Delta}$$

$$= m_{\xi\zeta}(\Delta).$$

For $k + \theta \geqslant h + \tau$ the limit equals $m_{\xi\zeta}(-\Delta)$. This completes the proof.

THEOREM 3. *Under assumptions* (I)–(IV) *the* LSE *of the parameters* $(\alpha_1, \cdots, \alpha_H, \gamma')$ *is strongly consistent.*

PROOF. We shall verify conditions (A1)–(A3) and (E) of Theorem 2. (A1) is obvious from (II). By (3.3),

$$x_{ti} = \eta_{t-i} \qquad \text{for } i = 1, \cdots, H;$$

$$= z_{t,\,i-H} \qquad \text{for } i = H+1, \cdots, H+N = K.$$

Let us first deal with $1 \leqslant i \leqslant H$. Since $\eta_{t-i} = 0$ for $t \leqslant i$, assume $t \geqslant i + 1$. From (3.4), (I) and (II),

$$E(x_{ti}\varepsilon_t) = 0,$$
$$E(x_{ti}x_{si}\varepsilon_t\varepsilon_s) = 0 \quad \text{for} \quad t \neq s.$$

Further,

$$\begin{aligned} E(x_{ti}^2\varepsilon_t^2) &= \Sigma_{\tau=0}^{t-i-1}a_\tau^2 E(\varepsilon_{t-i-\tau}^2\varepsilon_t^2) + (\Sigma_{\tau=0}^{t-i-1}a_\tau\gamma' z_{t-i-\tau})^2 E(\varepsilon_t^2) \\ &= \left[\Sigma_{\tau=0}^{t-i-1}a_\tau^2 E(\varepsilon_{t-i-\tau}^2) + (\Sigma_{\tau=0}^{t-i-1}a_\tau\gamma' z_{t-i-\tau})^2\right]E(\varepsilon_t^2) \\ &= E(x_{ti}^2)E(\varepsilon_t^2). \end{aligned}$$

This shows (A3). From (II), (IV) and (3.5) we obtain the estimate

$$\begin{aligned} E(x_{ti}^2) &= \Sigma_{\tau=0}^{t-i-1}a_\tau^2\sigma_{t-i-\tau}^2 + (\Sigma_{\tau=0}^{t-i-1}a_\tau\zeta_{t-i-\tau})^2 \\ &\leqslant \bar{\sigma}^2\Sigma_{\tau=0}^{\infty}a_\tau^2 + (\Sigma_{\tau=0}^{\infty}|a_\tau|)(\Sigma_{\tau=0}^{t-i-1}|a_\tau||\zeta_{t-i-\tau}|^2) \\ &\leqslant \text{const} \cdot \left(1 + \left[\Sigma_{\tau=0}^{t-i-1}|a_\tau|t^\delta\right]\right) \\ &\leqslant \text{const} \cdot \left(1 + t^\delta\left[\Sigma_{\tau=0}^{\infty}|a_\tau|\right]\right) \\ &= O(t^\delta), \end{aligned}$$

where $\zeta_\theta = \gamma' z_\theta$, and the const is independent of t. This establishes (A2).

For $H + 1 \leqslant i \leqslant K$, (A3) is an immediate consequence of (I), and (A2) is just the second part of (IV).

It remains to investigate the asymptotic behavior of the matrices $(1/T)P_T$. To begin with, let us calculate the matrix $\overline{M}_{11} = \lim_{T\to\infty} M_{11}$. The elements of M_{11} are of the form $[\eta_{-h}\eta_{-k}]$, $h = 1, \cdots, H$; $k = 1, \cdots, H$. (For notational simplicity, the time index will often be omitted in the sequel.) Inserting (3.4) and multiplying term by term, we arrive at four terms which we shall investigate separately. The term $[\eta^*_{-h}\eta^*_{-k}] = [(a, \varepsilon)_{-h}(a, \varepsilon)_{-k}]$ is of the form considered in Lemma 2 with $\xi_t = \zeta_t = \varepsilon_t(\omega)$ for every ω. Let us show that $m_{\varepsilon(\omega)\varepsilon(\omega)}(\Delta)$ exists and is finite for almost all ω and all $\Delta = 0, 1, 2, \cdots$ For $\Delta > 0$, put $e_t^\Delta = \varepsilon_t\varepsilon_{t+\Delta}$. Then, by virtue of (I),

$$E(e_t^\Delta) = 0, \quad E(e_t^\Delta e_s^\Delta) = 0 \text{ for } t \neq s, \qquad E(e_t^{\Delta 2}) = \sigma_t^2\sigma_{t+\Delta}^2 \leqslant \bar{\sigma}^4;$$

hence, by Proposition 1,

$$m_{\varepsilon\varepsilon}(\Delta) = \lim_{T\to\infty}\frac{1}{T}\Sigma_{t=1}^{T-\Delta}e_t = 0 \text{ a.e.}$$

For $\Delta = 0$, consider the random variables $e_t = \varepsilon_t^2 - \sigma_t^2$. Again, it is easily verified that

$$E(e_t) = 0 \quad \text{and} \quad E(e_te_s) = 0 \text{ for } t \neq s.$$

Moreover, by (II),

$$E(e_t^2) \leqslant E(\varepsilon_t^4) = 0(t^\delta);$$

hence, by Proposition 1,

$$\frac{1}{T}\Sigma_{t=1}^{T}e_t \to 0 \text{ a.e.,}$$

and, again using (II),

$$m_{\varepsilon\varepsilon}(0) = \sigma^2 \text{ a.e.}$$

Now we apply Lemma 2, obtaining

$$\lim_{T\to\infty} \left[\eta^*_{-h}\eta^*_{-k}\right]_T = \sigma^2\Sigma_{\tau=0}^{\infty}a_\tau a_{\tau+|h-k|} =: \bar{F}_{hk} \text{ a.e.}$$

As to the term $[\eta^*_{-h}w_{-k}]$, put $\xi_t = |a_t|^{\frac{1}{2}}\varepsilon_t$, $\zeta_t = \gamma' z_t$. Then again Proposition 1 applies, yielding $m_{\xi\xi}(0) = 0$. $m_{\zeta\zeta}(0)$ exists and is finite by virtue of (IV). For $\Delta \geqslant 0$, consider the random variables $e_t^\Delta = \xi_t\zeta_{t+\Delta}$. Then $E(e_t^\Delta) = 0$, $E(e_t^\Delta e_s^\Delta) = 0$ for $t \neq s$, and

$$\Sigma_{t=1}^{\infty}E\left(e_t^{\Delta 2}\right) = \Sigma_{t=1}^{\infty}|a_t|\sigma_t^2\zeta_{t+\Delta}^2 \leqslant \text{const} \cdot \left(1 + \Sigma_{t=1}^{\infty}|\lambda_0^t|t^\delta\right) < \infty;$$

hence

$$m_{\xi\zeta}(\Delta) = \lim_{T\to\infty} \frac{1}{T}\Sigma_{t=1}^{T-\Delta}e_t^\Delta = 0 \text{ a.e.}$$

Similarly, $m_{\xi\zeta}(-\Delta) = 0$. By Lemma 2 (with a_t replaced by $\text{sign}(a_t)|a_t|^{\frac{1}{2}}$ and $b_t = a_t$)

$$\lim_{T\to\infty}\left[\eta^*_{-h}w_{-k}\right]_T = 0 \text{ a.e.}$$

For $\xi = \zeta = z$, Lemma 2 is applicable by assumption; hence

$$\lim_{T\to\infty}\frac{1}{T}\left[w_{-h}w_{-k}\right]_T = \Sigma_{\tau=0}^{\infty}\Sigma_{\theta=0}^{\infty}a_\tau a_\theta\gamma'\bar{M}_{zz}((h-k)+(\tau-\theta))\gamma = \bar{G}_{hk}.$$

So we arrive at

$$\bar{M}_{11} = \bar{F} + \bar{G}.$$

Next, writing $z_t = (b, z)_t$ with $b_0 = 1, b_\tau = 0$ for $\tau \geqslant 1$, Lemma 2 together with Proposition 1 show that $[\eta^*_{-h}z'] \to 0$ a.e. and $[w_{-h}z']$ converges to some finite limit. Hence M_{12} converges a.e. to some finite nonstochastic limit $\bar{M}_{12}$.

Collecting the results obtained so far we find that $(1/T)P_T$ converges a.e. to the finite nonstochastic matrix

$$\bar{M} = \begin{bmatrix} \bar{F}+\bar{G} & \bar{M}_{12} \\ \bar{M}_{12} & \bar{M}_{zz}(0) \end{bmatrix}.$$

Introducing the vector $w_-(t) = (w_{t-1}, \cdots, w_{t-H})'$ we can write

$$\begin{bmatrix} \bar{G} & \bar{M}_{12} \\ \bar{M}'_{12} & \bar{M}_{zz}(0) \end{bmatrix} = \lim_{T\to\infty}\begin{bmatrix} [w_-w'_-] & [w_-z'_-] \\ [zw'_-] & [zz'] \end{bmatrix}.$$

Hence

$$(u'v')\begin{bmatrix} \bar{G} & \bar{M}_{12} \\ \bar{M}'_{12} & \bar{M}_{zz}(0) \end{bmatrix}\begin{pmatrix} u' \\ v' \end{pmatrix} = \lim_{T\to\infty}\left[(u'v')\begin{pmatrix} w_- \\ z \end{pmatrix}(w'_-z')\begin{pmatrix} u \\ v \end{pmatrix}\right]$$

$$= \lim_{T\to\infty}\left[(u'w_- + v'z)^2\right] \geqslant 0,$$

i.e., the matrix

$$\begin{bmatrix} \bar{G} & \bar{M}_{12} \\ \bar{M}'_{12} & \bar{M}_{zz}(0) \end{bmatrix}$$

is positive semidefinite. $\bar{F}$ can be shown to be positive definite (compare Anderson [1]), and $\bar{M}_{zz}(0)$ is positive definite by assumption (III). Hence $\bar{M}$ is positive definite.

If the ε_t do not possess finite moments of the 4th order, but are independent and identically distributed, then Theorem 1 may be used to derive strong consistency of the LSE. Strictly speaking, replace (I) and (II) by

(I′) The errors $\varepsilon_t(t = 1, 2, \cdots)$ are independent and identically distributed; $\varepsilon_t \in L_2$, $E(\varepsilon_t) = 0$.

THEOREM 4. *Under assumptions* (I′), (III) *and* (IV) *the* LSE *of* $(\alpha_1, \cdots, \alpha_H, \gamma')$ *is strongly consistent.*

For the proof, note that assumptions (A) and (B) of Theorem 1 are satisfied by virtue of (I′) and the representation (3.4). The proof that $(1/T)P_T$ converges a.e. to a nonstochastic nonsingular matrix—from which (C) and (D) will follow—runs along the same lines as in the proof of Theorem 3: every time reference is made to Proposition 1 use Kolmogoroff's law of large numbers for i.i.d. random variables instead.

Acknowledgment. The authors wish to thank the referee for his helpful comments and perspective suggestions.

REFERENCES

[1] ANDERSON, T. W. (1971). *The Statistical Analysis of Time Series.* Wiley, New York.

[2] ANDERSON, T. W. and TAYLOR, J. B. Strong consistency of least squares estimates in dynamic models. *Ann. Statist.* **7** 484-489.

[3] DRYGAS, H. (1976). Weak and strong consistency of least squares estimators in regression models. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **34** 119-127.

[4] LAI, T. L. and ROBBINS, H. (1977). Strong consistency of least squares estimates in regression models. *Proc. Natl. Acad. Sci. U.S.A.* **74** 2667-2669.

[5] NEVEU, J. (1972). *Martingales a Temps Discret.* Masson, Paris.

[6] RÉVÉSZ, P. (1968). *Die Gesetze der großen Zahlen.* Birkhäuser, Basel.

[7] WILLERS, R. (1978). Schwache Konsistenz von Kleinst-Quadrate-Schätzern für Regressions—und Streuungs-parameter in linearen Modellen. Dissertation, Dortmund.

INSTITUTE FÜR ÖKONOMETRIE UND OPERATIONS RESEARCH
UNIVERSITY OF BONN
ADENAUERALLEE 24-42
53 BONN WEST-GERMANY

INSTITUT FÜR ANGEWANDTE MATHEMATIK
UNIVERSITY OF BONN
WEGELERSTR. 6
53 BONN WEST-GERMANY