Cowles Foundation Paper #621

Econometric Theory, **1**, 85–96 Printed in the United States of America

A ZERO-ONE RESULT FOR THE LEAST SQUARES ESTIMATOR

DONALD W. K. ANDREWS

Yale University

The least squares estimator for the linear regression model is shown to converge to the true parameter vector either with probability one or with probability zero. In the latter case, it either converges to a point not equal to the true parameter with probability one, or it diverges with probability one. These results are shown to hold under weak conditions on the dependent random variable and regressor variables. No additional conditions are placed on the errors. The dependent and regressor variables are assumed to be weakly dependent—in particular, to be strong mixing. The regressors may be fixed or random and must exhibit a certain degree of independent variability. No further assumptions are needed. The model considered allows the number of regressors to increase without bound as the sample size increases. The proof proceeds by extending Kolmogorov's 0-1 law for independent random variables to strong mixing random variables.

1. INTRODUCTION

The linear regression model is the most widely used tool of econometrics. The least squares (LS) estimator of this model is optimal under certain model assumptions, and in consequence, is utilized extensively. The simplifying assumptions used for optimality may not hold in economic applications, however, so the statistical properties of the LS estimator under more general model assumptions are of great importance. In response, there has been considerable interest in extending the results for strong consistency of the LS estimator. These results are more or less complete for the case of independent identically distributed L^2 errors and fixed regressors. (See Lai, Robbins, and Wei [17, 18], and Drygas [9].) For more general error processes and random regressors, however, the results are more piecemeal (see Anderson and Taylor [1]; Chen, Lai, and Wei [7]; Christopeit and Helmes [8]; Eicker [10]; Nelson [19]; and Robinson [21]). In this note we prove a simple result for the LS estimator that yields a *synthesis* of results concerning strong consistency and inconsistency of the LS estimator.

Simply stated, we find that the LS estimator converges to the true parameter vector either with probability one or with probability zero. Further, in the case of inconsistency, the LS estimator either converges to a parameter value different from the true value with probability one, or it diverges with probability one. Thus, different stochastic environments can be categorized trichotomously according to whether the LS estimator is strongly consistent, whether it converges almost surely to a parameter vector that is not true, or whether it diverges almost surely. This shows that known results for strong consistency and inconsistency are incomplete; a more complete categorization is possible. The 0-1 result is also of independent theoretical interest, since it may be found useful in proofs or in suggesting results to be proven. With regard to the latter, it delimits the alternative possibilities in situations where strong consistency is at issue.

The regression model considered here is quite general. The regressors may be fixed or random, and the number of regressors may increase without bound as the sample size increases. Allowing this flexibility in the model seems appropriate for economic applications. In such applications, some regressors are necessarily fixed, e.g., dummy variables, while other regressors are random and may be treated as such, or may be conditioned on and treated as fixed. Further, it is often the case that the number of regressors chosen to be included in an economic regression model is limited by the statistical problem of degrees of freedom, rather than by a belief motivated by economic theory that only a fixed number of regressors belong in the model. In such cases, the number of variables included in the regression model is usually related to the sample size. The possibility of such a relationship is incorporated in the model considered below. (Also see Huber [11, 12] and Yohai and Maronna [22] for the specification of regression models where the number of regressors is related to the sample size.)

For the 0-1 result, the dependent variable and regressors must be weakly dependent (more explicitly, strong mixing is assumed). That is, the "dependence" between variables is assumed to die out as the difference in time subscripts of the variables become infinitely large. Further, the regressors must exhibit a certain degree of independent variability. No assumptions of independence, identical distribution, or normality of the errors are made. In fact, no assumptions at all are made on the errors except that of weak dependence (which follows from the assumption of weak dependence of the variables in the model). Exogeneity of the regressors is not imposed. Thus, the true parameter vector may or may not be identified.

The LS estimator 0-1 result is obtained by first proving a 0-1 law for sequences of strong mixing random vectors. This law is an extension of Kolmogorov's classical 0-1 law for sequences of independent random variables (see Kolmogorov [15] or, for example, Billingsley [5]).

2. A ZERO-ONE LAW

First we define the concept of strong mixing. Strong mixing is an assumed property of the regression model variables. It implies that realizations of variables at a particular date in the past have an impact on current variables that dies out as time passes. This is a realistic assumption for many economic situations.[1] It is considerably weaker than other assumptions, such as independence, m-dependence, or auto-regressive moving average structure (see Chanda [6], but cf. Andrews [2, 3]) that are often utilized in statistical models. Moreover, strong mixing does not imply stationarity.

Let $\langle Z_i \rangle$ denote a sequence of random vectors $\mathbf{Z}_i$, $i = 1, 2, \ldots$, of arbitrary (possibly infinite) dimensions. Let $\mathscr{B}_{ij}$ denote the σ-field generated by the random vectors $\mathbf{Z}_i, \mathbf{Z}_{i+1}, \ldots, \mathbf{Z}_j$. That is, $\mathscr{B}_{ij}$ is the collection of all events determined by $\mathbf{Z}_i, \mathbf{Z}_{i+1}, \ldots, \mathbf{Z}_j$. $\langle \mathbf{Z}_i \rangle$ is called *strong mixing* if $\alpha(s) \downarrow 0$ as $s \to \infty$, where

$$\alpha(s) \equiv \sup_{l \geqslant 1} \sup_{A \in \mathscr{B}_{1l}, B \in \mathscr{B}_{l+s,\infty}} \left| P(A \cap B) - P(A)P(B) \right|. \tag{1}$$

Note, if $\langle \mathbf{Z}_i \rangle$ are independent, then $\alpha(s) = 0$, $\forall\, s \geqslant 1$, and if $\langle \mathbf{Z}_i \rangle$ are m-dependent, then $\alpha(s) = 0$, $\forall\, s > m$.

The following 0-1 law for strong mixing random vectors is an extension of Kolmogorov's 0-1 law for independent random variables. It is known that Kolmogorov's 0-1 law applies to ϕ-mixing random vectors (see Iosifescu and Theodorescu [14]). But ϕ-mixing is a much stronger assumption than strong mixing. For example, Gaussian random variables are ϕ-mixing only if they are m-dependent (see Ibragimov and Linnik [13], cf. Kolmogorov and Rozonov [16]). A simple first-order autoregressive Gaussian sequence is not ϕ-mixing. The extension of the 0-1 law to strong mixing random vectors has not been noted in the literature. Bártfai and Révész [4] prove that 0-1 law for sequences which they call δ-mixing and δ-mixing in mean. It can be shown that strong mixing sequences are δ-mixing in mean. Hence, the 0-1 law holds for strong mixing random vectors. Below we give an alternative proof of this result. The proof avoids Bártfai and Révész's use of the powerful machinery of the Martingale convergence theorem. Thus the proof is more direct, and hopefully, more clear.

Define the tail σ-field, $\mathscr{T}$, generated by the random vectors $\langle \mathbf{Z}_i \rangle$ as

$$\mathscr{T} = \bigcap_{l=1}^{\infty} \mathscr{B}_{l,\infty}. \tag{2}$$

An event in $\mathscr{T}$ is called a *tail event*. Tail events are determined by random vectors arbitrarily far out in the sequence $\langle \mathbf{Z}_i \rangle$. Examples for scalar Z_i include:

(i) $\{Z_i > b$ for infinitely many $i\}$, for some constant b,

(ii) $\left\{\overline{\lim_{i\to\infty}}\, Z_i \in [a, b]\right\}$, for some constants a and b,

(iii) $\left\{\sum_{i=1}^{n} Z_i \text{ converges as } n \to \infty\right\}$,

(iv) $\{Z_i, i = 1, 2, \ldots$ converges as $i \to \infty\}$.

The 0-1 law concerns the probabilities of different tail events.

THEOREM 1. If the sequence of random vectors $\langle Z_i \rangle$ *is strong mixing, and* $A \in \mathscr{T}$, *where* $\mathscr{T}$ *is the tail* σ*-field, then* P(A) *equals* 0 *or* 1.

The proof of Theorem 1 is in Section 4.

3. RESULTS FOR THE LEAST SQUARES ESTIMATOR

In this section we use the strong mixing 0-1 law to prove that the LS estimator in a regression model converges to the true parameter vector with probability 0 or 1. The regressors may be fixed or random and their number may increase with the sample size. The model is written as

$$\mathbf{y}_{1n} = \mathbf{X}_{1n}\boldsymbol{\beta}_n^0 + \mathbf{u}_{1n}, \qquad n = 1, 2, \ldots, \tag{3}$$

where $\mathbf{y}_{1n}$ is the n-vector, with ith element y_i, of the first n values of the dependent random variable; $\mathbf{X}_{1n}$ is the $n \times k_n$ matrix, with (i, j)th element x_{ij}^n, of the first n values of the k_n regressors; $\boldsymbol{\beta}_n^0$ is the unobserved R^{k_n}-valued true parameter vector (for the model with k_n regressors); and u_{1n} is the n-vector of the first n (unobserved) random errors. Let $\mathbf{x}_i$ be the vector with elements x_{ij}^n, $j = 1, \ldots, k_n$, $n = 1, 2, \ldots,$, and let $\langle y_i, \mathbf{x}_i \rangle \equiv \{(y_i, \mathbf{x}_i): i = 1, 2, \ldots\}$ denote the infinite sequence of dependent and regressor variables corresponding to observations i, for $i = 1, 2, \ldots$.

Since regression models with increasing numbers of regressors are somewhat novel, we motivate their consideration by providing an example that generates the model of (3) as the true model. The results of general equilibrium theory imply that many economic variables depend, more or less, on "everything else" in the economy. Thus, for a regression model with a dependent variable which is economic in nature, the number of relevant explanatory variables may be infinite. Of course, certain variables will have large explanatory power while most will have only minute power. In this situation, the true model can be written as

$$y_{1n} = \mathbf{X}_{1n}^{\infty\prime}\beta^0 + \tilde{u}_{1n}, \tag{4}$$

where y_{1n} is as in equation (3), $\mathbf{X}_{1n}^{\infty}$ is the $(n \times \infty)$ matrix of regressor variables corresponding to the first n observations on the infinite number of regressors, β^0 is the $(\infty \times 1)$ true parameter vector, and $\tilde{u}_{1n}$ is the $(n \times 1)$ vector of mean zero (say), strong mixing errors. It is impossible to estimate the entire infinite-dimensional vector β^0 with a sample of finite size. Hence, depending upon the available sample size, a greater or lesser number of variables is included in the regression in practice. The remainder are lumped in with the error term. The model actually estimated, then, is given by equation (3), which includes k_n regressors. $\mathbf{X}_{1n}$ is given by the first k_n columns of $\mathbf{X}_{1n}^{\infty}$, β_n^0 is given by the first k_n elements of β^0, and u_{1n} equals the sum of $\tilde{u}_{1n}$ and that part of the regression function that is ignored when the sample size is n. Within this framework, the parameter β_n^0 is "true" in a clear and meaningful sense. Notice that the LS estimator of β_n^0 is biased in this model if the included and excluded regressors are collinear, as is likely. It still may be strongly consistent, however, since the small sample bias may be reduced as more and more regressors are included in the model.

For most economic applications, asymptotics based on an increasing number of regressors mimic reality closer than do conventional asymptotics. For consistency results, this makes increasing regressor asymptotics preferable. Such asymptotics also allow one to explore the effects of different ratios of the number of regressors to the number of observations, and the effects of lumping part of the regression function in with the error (see Huber's discussion, pp. 164–70 of [12], especially with respect to bias). On the other hand, it is not necessarily the case that increasing regressor asymptotics yield better distributional approximations for statistics than do conventional asymptotics, and such approximations constitute the most important use of asymptotic theory. Thus, both increasing regressor asymptotics and conventional asymptotics have a role to play in econometric theory.

We now return to the analysis of the model given by equation (3). This model may be generated by a model of the type given in equation (4), or it may arise in some other manner. We assume

(A1) $\langle y_t, x_t \rangle$ is a strong mixing sequence.

The distance between an estimator and its estimand is measured by the supremum norm, denoted $\|\cdot\|$, of their difference. The supremum norm of a vector or matrix is simply the greatest modulus of any of its elements.

The regressor variables are assumed to satisfy conditions that ensure a certain degree of independent variability:

(A2) For any given $m \geqslant 1$, $X'_{mn}X_{mn}$ is nonsingular for n sufficiently large, almost surely (a.s.),

(A3) $k_n^2 \cdot \|(X'_{1n}X_{1n})^{-1}\| \xrightarrow{n \to \infty} 0$ a.s., (A4) $\sup_{n\geq 1}\frac{1}{k_n}\sum_{j=1}^{k_n}|x_{ij}^n| < \infty$ a.s., $\forall\, i = 1, 2, \ldots,$

where k_n is the number of regressors when the sample size is n, and X_{mn} is the $(n - m + 1) \times k_n$ matrix of regressors for observations $m, m + 1, \ldots, n$ when the sample size is n. In the case of a fixed number of regressors, (A4) is redundant, and (A2) and (A3) reduce to a condition commonly used in consistency proofs, viz., that $(X'_{1n}X_{1n})^{-1}$ exists and converges to the 0 matrix a.s. (See Anderson & Taylor [1] and Lai, Robbins, & Wei [17, 18].) Note that this assumption eliminates the possibility of lack of identification due to deficient rank of the regressor matrix.

The LS estimator $\hat{\beta}_n$ is defined as

$$\hat{\beta}_n \equiv (X'_{1n}X_{1n})^{-1}X'_{1n}y_{1n}, \quad \text{for } n = 1, 2, \ldots.$$

We consider the convergence to zero of the difference between $\hat{\beta}_n$ and vectors β_n in R^{k_n}, for $n = 1, 2, \ldots$. Of course, the vectors $\{\beta_n\}$ of most interest are the true regression parameter vectors $\{\beta_n^0\}$.[2] We consider arbitrary vectors $\{\beta_n\}$ (which *could* be taken to be the true parameter vectors $\{\beta_n^0\}$), however, because in the case where the LS estimator is not strongly consistent, we are still interested in its behavior. Does it converge to the true parameter vector with probability between zero and one? Does it converge to some incorrect parameter vector with positive probability? Or, does it diverge with positive probability? To answer these questions, we need to know the probability that $\|\hat{\beta}_n - \beta_n\| \xrightarrow{n\to\infty} 0$, both for the true parameter vectors $\{\beta_n^0\}$, and for arbitrary vectors $\{\beta_n\}$.

The vectors $\{\beta_n\}$, which we consider, are assumed to be sufficiently well behaved as $n \to \infty$ that the corresponding "regression function" for the ith observation, viz., $x_i^{n\prime}\beta_n$ (where $x_i^n \equiv (x_{i1}^n, \ldots, x_{ik_n}^n)'$), does not blow-up as additional regressors are added:

(A5) $\sup_{n\geq 1}|x_i^{n\prime}\beta_n| < \infty$ a.s., $\forall\, i = 1, 2, \ldots.$

If the number of regressors is fixed, then A5 is redundant.

We now prove the main result for the LS estimator $\hat{\beta}_n$.

THEOREM 2 Let $\{\beta_n\}$ *be any sequence of vectors (in* R^{k_n} *for all* n) *that satisfies* A5. *Then, under assumptions* A1–A4,

$$\|\hat{\beta}_n - \beta_n\| \xrightarrow{n\to\infty} 0 \qquad \textbf{(5)}$$

with probability zero or probability one. In consequence, if $\{\beta_n^0\}$ *is the sequence of true parameter vectors, then* $\hat{\beta}_n$ *is strongly consistent for* $\{\beta_n^0\}$ *(i.e.,*

$\|\hat{\beta}_n - \beta_n^0\| \xrightarrow{n\to\infty} 0$ *a.s.), or* $\hat{\beta}_n$ *converges to some other sequence of parameter vectors* $\{\beta_n^1\}$ *with probability one (i.e.,* $\|\hat{\beta}_n - \beta_n^1\| \xrightarrow{n\to\infty} 0$ *a.s.), or* $\hat{\beta}_n$ *diverges with probability one (i.e., for all sequences* $\{\beta_n\}$, $\|\hat{\beta}_n - \beta_n\| \not\xrightarrow{n\to\infty} 0$ *with probability one).*

The proof of Theorem 2 is in Section 4.

Theorem 2 may be viewed as a possibility theorem. The result of the theorem states that it is possible to categorize the convergence properties of the LS estimator into just three categories. The underlying stochastic environment determines which category obtains. The first category is that of strong consistency. Numerous papers are devoted to establishing conditions which ensure that the LS estimator falls in this category (see the Introduction for references). The second and third categories describe the only possible inconsistent behavior of the LS estimator, viz., either almost sure convergence to some sequence of parameter vectors $\{\beta_n^1\}$ which is not true, or almost sure divergence. This categorization provides immediate information about the behavior of the LS estimator under misspecification. It also may be useful in extending known results for strong consistency, since the demonstration that $\hat{\beta}_n$ converges to the true parameter sequence with some positive probability, combined with Theorem 2, yields strong consistency.

The three categories of Theorem 2 can be exemplified quite easily if we consider a fixed number of regressors. Suppose the true model is given by (3) with $k_n = k$ and $\beta_n^0 = \beta^0$, for all n. Assume the regressors are independent of the errors, both are independent identically distributed (iid), the errors have finite mean, and A3 holds. In this case, it is well known that the LS estimator is strongly consistent. Alternatively, suppose the true model is as above, but the underlying regressors X_{1n} are measured with error. That is, the observed regressors are $\tilde{X}_{1n} = X_{1n} + V_{1n}$, where X_{1n} and V_{1n} are $(n \times k)$ matrices of iid, finite mean random variables which are independent of μ_{1n}. In this case, the LS estimator $\hat{\beta}_n = (\tilde{X}'_{1n}\tilde{X}_{1n})^{-1}\tilde{X}'_{1n}y_{1n}$ converges with probability one to $\beta^0 - (E(1/n)\tilde{X}'_{1n}\tilde{X}_{1n})^{-1} \cdot E((1/n)\tilde{X}'_{1n}V_{1n})\beta^0 \neq \beta^0$. This illustrates the second category of Theorem 2. The third category arises, for example, if the model is as above (with X_{1n} observed) but the errors have infinite mean. Then,

$$\|\hat{\beta}_n - \beta\| = \left\|\left(\frac{1}{n} X'_{1n}X_{1n}\right)^{-1} \frac{1}{n} X'_{1n}u_{1n} + (\beta^0 - \beta)\right\|,$$

$$\left(\frac{1}{n} X'_{1n}X_{1n}\right)^{-1} \xrightarrow{n\to\infty} (EX'_{11}X_{11})^{-1}$$

a.s. by the strong law of large numbers (SLLN), and $\|(1/n)X'_{1n}u_{1n}\| \xrightarrow{n\to\infty} \infty$ a.s. by the converse to the SLLN. Hence $\hat{\beta}_n$ diverges almost surely.

We mention that the results of Theorem 2 may be of particular interest in models with unidentified parameters, perhaps due to endogeneity of some regressors. In such models it might be thought that the LS estimator converges to different points in the set of observationally equivalent, true, parameter vectors with nontrivial probabilities. The theorem implies that this is incorrect—the probability of convergence to any such parameter value is either 0 or 1. This phenomena is nicely illustrated by a result of Phillips [20]. Consider a simultaneous equation that is unidentified due to the lack of association between the included endogenous variables and the excluded exogenous variables. (Note, this situation is quite similar to our second example.) The LS estimator of the endogenous variables coefficients falls in category two, that is, almost sure convergence to the wrong parameter vector, unless there is no simultaneity present (i.e., unless the true endogenous variables parameter vector equals zero). In the latter case, the LS estimator converges to the true parameter vector almost surely. Interestingly, as Phillips [20] shows, LIML and 2SLS do not exhibit the same performance in the latter case—their distributions are invariant with respect to the sample size (under the assumption of normal errors).

4. PROOFS

PROOF OF THEOREM 1. Let $\mathscr{B}$ be the σ-field generated by the whole sequence of random vectors $\langle Z_t \rangle$. We will show

$$P(A \cap B) - P(A)P(B) = 0, \qquad \forall\, A \in \mathscr{T} \text{ and } \forall\, B \in \mathscr{B}. \tag{6}$$

If true, take $B = A$ ($\in \mathscr{T} \subset \mathscr{B}$) to get $P(A) = P(A)^2$, and $P(A)$ equals 0 or 1, as desired.

To show (6), let $B \in \mathscr{B}_{1l}$ and $A \in \mathscr{T}$. Then the strong mixing of $\langle Z_t \rangle$ and the observation that $A \in B_{l+s,\infty}$ imply

$$|P(A \cap B) - P(A)P(B)| \leqslant \alpha(s). \tag{7}$$

(7) holds for all s, and $\alpha(s) \downarrow 0$ as $s \to \infty$, hence

$$|P(A \cap B) - P(A)P(B)| = 0, \qquad \text{for } A \in \mathscr{T} \text{ and } B \in \beta_{1l}, \text{ for all } l. \tag{8}$$

Let $\mathscr{M} = \{C \in \mathscr{B} \colon P(A \cap C) - P(A)P(C) = 0, \forall\, A \in \mathscr{T}\}$. By (8)

$$\mathscr{M} \supset \bigcup_{l=1}^{\infty} \mathscr{B}_{1l}.$$

Now, the monotone class theorem applies because (i) $\bigcup_{l=1}^{\infty} \mathscr{B}_{1l}$, being an increasing union of fields, is a field, and (ii) the continuity of P implies that $\mathscr{M}$ is closed under increasing unions (and by simple algebra $\mathscr{M}$ is closed under complements as well). Thus

$$\mathscr{M} \supset \sigma\left(\bigcup_{l=1}^{\infty} \mathscr{B}_{1l}\right),$$

where $\sigma(\bigcup_{l=1}^{\infty} \mathscr{B}_{1l})$ is the σ-field generated by $\bigcup_{l=1}^{\infty} \mathscr{B}_{1l}$. But, $\sigma(\bigcup_{l=1}^{\infty} \mathscr{B}_{1l}) = \mathscr{B}$, so $\mathscr{M} = \mathscr{B}$ and (6) follows. ∎

Before proving Theorem 2, we state a lemma. Let $v_\iota^n = y_\iota - x_\iota^{n\prime}\beta_n$, and $v_{rs}^n = (v_r^n, \ldots, v_s^n)'$, for $1 \leqslant r \leqslant s \leqslant n$. Define G to be the set of sample paths ω for which the conditions defined in A2–A5 hold. Note, $P(G) = 1$. Let $S_1 = \{\omega: \|(X_{1n}'X_{1n})^{-1}X_{1n}'v_{1n}^n\| \xrightarrow{n\to\infty} 0\}$, and $S_{2m} = \{\omega: \|(X_{mn}'X_{mn})^{-1}X_{mn}'v_{mn}^n\| \xrightarrow{n\to\infty} 0\}$.

LEMMA 1. *For all positive integers* m, $\mathrm{G} \cap \mathrm{S}_1 = \mathrm{G} \cap \mathrm{S}_{2m}$.

PROOF OF THEOREM 2. Let $H = \{\omega: \|\hat{\beta}_n - \beta_n\| \xrightarrow{n\to\infty} 0\}$. It is easy to see that $G \cap H = G \cap S_1$. Lemma 1 gives $G \cap H = G \cap S_{2m}$ for $m = 1, 2, \ldots$, where $S_{2m} \in \mathscr{B}_{m,\infty}$, the σ-field generated by (y_ι, x_ι), $i = m, m+1, \ldots$. Further, $G \cap H = G \cap \overline{\lim}_{m\to\infty} S_{2m}$, where $\overline{\lim}_{m\to\infty} S_{2m} \equiv \lim_{m\to\infty} \bigcup_{l=m}^{\infty} S_{2l} \in \mathscr{T} \equiv \bigcap_{l=1}^{\infty} \mathscr{B}_{l,\infty}$. Since $P(G) = 1$, we get $P(H) = P(\overline{\lim}_{m\to\infty} S_{2m})$, and the latter is 0 or 1 by A1 and Theorem 1, since $\overline{\lim}_{m\to\infty} S_{2m}$ is a tail event. ∎

PROOF OF LEMMA 1. Define $S_{3m} = \{\omega: \|(X_{1n}'X_{1n})^{-1}X_{mn}'v_{mn}^n\| \xrightarrow{n\to\infty} 0\}$. We show (i) $G \cap S_1 = G \cap S_{3m}$, and (ii) $G \cap S_{3m} = G \cap S_{2m}$. To show (i) it suffices to show, for any fixed m,

$$\|(X_{1n}'X_{1n})^{-1}X_{1m}^{n\prime}v_{1(m-1)}^n\| \xrightarrow{n\to\infty} 0 \qquad \text{for all } \omega \in G, \tag{9}$$

where $\mathbf{X}_{1m}^n$ is the $(m-1) \times k_n$ matrix of regressors for observations $1, \ldots, m-1$ when the sample size is n. Let $C = X_{1n}'X_{1n}$, c^{ij} be the (i,j)th element of C^{-1}, and η_r be the rth element of the k_n-vector $(X_{1n}'X_{1n})^{-1}X_{1m}^{n\prime}v_{1(m-1)}^n$. Then,

$$\begin{aligned}\max_{r\leqslant k_n} |\eta_r| &= \left|\sum_{l=1}^{k_n}\sum_{\iota=1}^{m-1} c^{rl}x_{\iota l}^n(y_\iota - x_\iota^{n\prime}\beta_n)\right| \\ &\leqslant k_n \cdot \|C^{-1}\| \cdot \sum_{\iota=1}^{m-1}\left(\frac{1}{k_n}\sum_{l=1}^{k_n}|x_{\iota l}^n|\right) \cdot |y_\iota - x_\iota^{n\prime}\,\beta_n| \xrightarrow{n\to\infty} 0,\end{aligned} \tag{10}$$

for all $\omega \in G$, using (A3)–(A5). This gives (9), and (i) is proved.

To show (ii), we introduce the following notation: $A = X'_{mn}X_{mn}$, $B = X^n_{1m}$, $d = X'_{mn}v^n_{mn}$, $g = C^{-1}d$, and $Q = (I_{m-1} - BC^{-1}B')^{-1}$, where I_{m-1} is the $(m-1)$-dimensional identity matrix, and the dependence on n of each of the quantities is suppressed for notational simplicity. First we show:

$$\|g\| \xrightarrow{n\to\infty} 0 \quad \text{implies} \quad \|A^{-1}d\| \xrightarrow{n\to\infty} 0, \qquad \forall\, \omega \in G. \tag{11}$$

Assume $\|g\| \xrightarrow{n\to\infty} 0$, for all $\omega \in G$. By (A2) we can take n sufficiently large that A^{-1} exists. Then, a well-known (and easily verified) equality for matrix inverses yields

$$A^{-1} \equiv (C - B'B)^{-1} = C^{-1} + C^{-1}B'(I_{m-1} - BC^{-1}B')^{-1}BC^{-1}, \tag{12}$$

and so,

$$\|A^{-1}d\| \leqslant \|C^{-1}d\| + \|C^{-1}B'QBC^{-1}d\|. \tag{13}$$

Also,

$$\begin{aligned}\|BC^{-1}B'\| &= \max_{r,s<m}\left|\sum_{l=1}^{k_n}\sum_{j=1}^{k_n} x_{rl}c^{lj}x_{sj}\right| \\ &\leqslant k_n^2 \cdot \|C^{-1}\| \cdot \max_{r,s<m}\left(\frac{1}{k_n}\sum_{l=1}^{k_n}|x_{rl}|\right)\left(\frac{1}{k_n}\sum_{j=1}^{k_n}|x_{sj}|\right) \xrightarrow{n\to\infty} 0, \end{aligned} \tag{14}$$

for all $\omega \in G$, by assumptions (A3) and (A4) where x_{rl} is the (r, l)th element of B and c^{lj} is the (l,j)th element of C^{-1}. This, plus the fact that Q has a fixed number of elements for all n, implies $\|Q\| \xrightarrow{n\to\infty} 1$, for all $\omega \in G$. Thus, we have

$$\begin{aligned}\|C^{-1}B'QBg\| &\leqslant \max_{r\leqslant k_n}\sum_{j=1}^{k_n}\sum_{i=1}^{m-1}\sum_{l=1}^{m-1}\sum_{s=1}^{k_n}|c^{rj}x_{ij}q_{il}x_{ls}g_s| \\ &\leqslant k_n^2\|C^{-1}\| \cdot \|Q\| \cdot \|g\| \cdot \left(\sum_{i=1}^{m-1}\frac{1}{k_n}\sum_{j=1}^{k_n}|x_{ij}|\right)^2 \xrightarrow{n\to\infty} 0, \end{aligned} \tag{15}$$

for all $\omega \in G$, where q_{il} is the (i, l)th element of Q, g_s is the sth element of g, and the convergence to zero follows using (A3), (A4), the result $\|Q\| \xrightarrow{n\to\infty} 1$, and the assumption that $\|g\| \xrightarrow{n\to\infty} 0$. Equations (13) and (15) yield the desired result (11).

The converse of (11) follows by the same argument as used to prove (11), noting that

$$C^{-1} \equiv (A + B'B)^{-1} = A^{-1} - A^{-1}B'(I_{m-1} + BA^{-1}B')^{-1}BA^{-1}, \tag{16}$$

and provided $k_n^2\|A^{-1}\| \xrightarrow{n\to\infty} 0$, for all $\omega \in G$. The latter follows using (12), the triangle inequality, and the result $\|C^{-1}B'QBC^{-1}\| \xrightarrow{n\to\infty} 0$ as shown by an argument analogous to that of (15). ■

ACKNOWLEDGMENTS

I would like to thank Peter C. B. Phillips, an associate editor, and a referee for helpful comments

NOTES

1 Note, if one views the economy as evolving via some infinite sequence of events each of which alters, irrevocably, the future course of the economy, then the strong mixing assumption is not appropriate

2. For the true parameter vectors $\{\beta_n^0\}$ of equation (3) to be meaningful, either the model of (3) must be interpretable in terms of a more complete model, such as that of (4), or some assumptions need to be placed on the errors u_{1n}—for example, assumptions of mean zero, median zero, or identical distribution If the model of (3) or the model which generates (3) is misspecified, then $\{\beta_n^0\}$ may lose its meaning, but we are still interested in the behavior of the LS estimator (since the behavior of estimators under misspecification is an important property in general) For increased generality, then, no assumptions are placed on the errors in (3), and the results below hold whether or not the model is correctly specified

REFERENCES

1. Anderson, T W., & Taylor, J B (1979) Strong consistency of least squares estimates in dynamic models. *Annals of Statistics*, *7*, 484–489.
2. Andrews, D.W K (1984). Non-strong mixing autoregressive processes. *Journal of Applied Probability*, *21*, 930–934
3. Andrews, D W.K (1985) A nearly independent, but non-strong mixing, triangular array. *Journal of Applied Probability* (forthcoming)
4 Bártfai, P , & Révész, P. (1967). On a zero-one law. *Zeitshrift fur Wahrscheinlichkeitstheorie und verwande Gebiete*, *7*, 43–47.
5 Billingsley, P (1979) *Probability and Measure*. New York J. Wiley & Sons
6 Chanda, K C (1974) Strong mixing properties of linear stochastic processes. *Journal of Applied Probability*, *11*, 401–408
7 Chen, G. J., Lai, T L , & Wei, C Z (1981) Convergence systems and strong consistency of least squares estimates in regression models *Journal of Multivariate Analysis*, *11*, 319–333.
8 Christopeit, N , & Helmes, K. (1979). A convergence theorem for random linear combinations of independent normal random variables. *Annals of Statistics*, *7*, 795–800
9 Drygas, H (1976). Weak and strong consistency of the least squares estimators in regression models. *Zeitshrift fur Wahrscheinlichkeitstheorie und verwande Gebiete*, *34*, 119–127.
10 Eicker, F. (1979) Consistent parameter estimation in mixed autoregressions and in general linear stochastic regressions Dortmund University (manuscript).
11. Huber, P. J (1973). Robust regression. Asymptotics, conjectures, and Monte Carlo. *Annals of Statistics*, *1*, 799–821.
12 Huber, P J (1981). *Robust Statistics* New York. J. Wiley & Sons
13 Ibragimov, I A., & Linnik, Yu. V. (1971). *Independent and Stationary Sequences of Random Variables* Groningen, The Netherlands: Wolters-Noordhoff.
14. Iosifescu, M , & Theodorescu, R (1969). *Random Processes and Learning* New York Springer-Verlag.
15 Kolmogorov, A N (1933). *Grundbegriffe de Wahrscheinlichkeitsrechnung*. Berlin· Springer

16 Kolmogorov, A. N., & Rozonov, Y A. (1960). On strong mixing conditions for stationary Gaussian processes *Theory of Probability and its Applications*, *5*, 204–208.
17 Lai, T L., Robbins, H, & Wei, C Z, (1978) "Strong consistency of least squares estimates in multiple regression" *Proceedings of the National Academy of Sciences USA* **75**, 3034–3036.
18 Lai, T L, Robbins, H., & Wei, C Z (1979) Strong consistency of least squares estimates in multiple regression II *Journal of Multivariate Analysis*, *9*, 343–361.
19 Nelson, P I (1980). A note on strong consistency of least squares estimators in regression models with Martingale difference errors. *Annals of Statistics*, *8*, 1057–1064
20 Phillips, P.C.B. (1984) The exact distribution of LIML: I. *International Economic Review*, *25*, 249–261
21. Robinson, P. M (1978) On consistency in time series analysis. *Annals of Statistics*, *6*, 215–223.
22 Yohai, V J, & Maronna, R. A (1979) Asymptotic behavior of *M*-estimators for the linear model *Annals of Statistics*, *7*, 258–268.