COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station

New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 698

A ZERO-ONE RESULT FOR THE LEAST SQUARES ESTIMATOR

Donald W.K. Andrews

March 1984

A ZERO-ONE RESULT FOR THE LEAST SQUARES ESTIMATOR

Donald W. K. Andrews

Cowles Foundation for Research in Economics
Yale University

March, 1984

ABSTRACT

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 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 (1978, 1979), and Drygas (1976)). However, for more general error processes and random regressors the results are more piecemeal (see Anderson and Taylor (1979), Christopeit and Helmes (1979), Eicker (1979), Nelson (1980), and Robinson (1978)). In this note we prove a simple result for the LS estimator which yields a synthesis of known and unknown results for the strong consistency or 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. Thus, different stochastic environments can be categorized dichotomously according to whether the LS estimator is strongly consistent or whether it converges to the true parameter vector with probability zero. 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 modified 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 (1973) for the specification of a regression model 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 subscript 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 (1933) or, for example, Billingsley (1979)).

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 which 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 (1974), but cf. Andrews (1984)) which are often utilized in statistical models. Moreover, strong mixing does not imply stationarity.

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

$$\alpha(s) \equiv \sup_{\ell \geq 1} \sup_{A \in \mathcal{B}_{1\ell}, B \in \mathcal{B}_{\ell+s,\infty}} |P(A \cap B) - P(A)P(B)| . \tag{1}$$

Note, if $\langle Z_i \rangle$ are independent, then $\alpha(s) = 0$, $\forall s \geq 1$, and if $\langle Z_i \rangle$ are m-dependent, then $\alpha(s) = 0$, $\forall s > m$.

[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.

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 (1969)). However, 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 (1971), cf. Kolmogorov and Rozonov (1960)). 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 (1969) prove the 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, $\mathcal{T}$, generated by the random vectors $\langle Z_i \rangle$ as

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

An event in $\mathcal{T}$ is called a <u>tail event</u>. Tail events are determined by random vectors arbitrarily far out in the sequence $\langle Z_i \rangle$. Examples for scalar Z_i include:

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

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

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

(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 $A \in \mathcal{T}$, where $\mathcal{T}$ is the tail σ-field, then $P(A)$ equals 0 or 1.

Proof of Theorem 1. Let $\mathcal{B}$ be the σ-field generated by the whole sequence of random vectors $\langle Z_i \rangle$. We will show

$$P(A \cap B) - P(A)P(B) = 0 \ , \quad \forall A \in \mathcal{T} \text{ and } \forall B \in \mathcal{B} \ . \tag{3}$$

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

To show (3), let $B \in \mathcal{B}_{1\ell}$ and $A \in \mathcal{T}$. Then the strong mixing of $\langle Z_i \rangle$ and the observation that $A \in \mathcal{B}_{\ell+s,\infty}$ imply

$$|P(A \cap B) - P(A)P(B)| \le \alpha(s) \ . \tag{4}$$

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

$$|P(A \cap B) - P(A)P(B)| = 0 \ , \quad A \in \mathcal{T} \text{ and } B \in \mathcal{B}_{1\ell} \ , \text{ for all } \ell \ . \tag{5}$$

Let $\mathcal{M} = \{C \in \mathcal{B}: P(A \cap C) - P(A)P(C) = 0, \ \forall A \in \mathcal{T}\}$. By (5)

$$\mathcal{M} \supset \bigcup_{\ell=1}^{\infty} \mathcal{B}_{1\ell} \ .$$

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

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

where $\sigma(\bigcup_{\ell=1}^{\infty} \mathcal{B}_{1\ell})$ is the σ-field generated by $\bigcup_{\ell=1}^{\infty} \mathcal{B}_{1\ell}$. But, $\sigma(\bigcup_{\ell=1}^{\infty} \mathcal{B}_{1\ell}) = \mathcal{B}$, so $\mathcal{M} = \mathcal{B}$ and (3) follows. □

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

$$y_{1n} = X_{1n}\beta_n^0 + u_{1n} , \quad n = 1, 2, \ldots , \tag{6}$$

where y_{1n} is the n-vector, with i^{th} element y_i , of the first n values of the dependent random variable; 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; β_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 x_i be the vector with elements x_{ij}^n , $j = 1, \ldots, k_n$, $n = 1, 2, \ldots$, and let $\langle y_i, x_i \rangle \equiv \{(y_i, x_i) : i = 1, 2, \ldots\}$ denote the infinite sequence of dependent and regressor variables corresponding to observations i , for $i = 1, 2, \ldots$. We assume

A1) $\langle y_i, x_i \rangle$ 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 its largest element.

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

A2) $k_n^2 \cdot \| (X'_{1n} X_{1n})^{-} \| \xrightarrow{n\to\infty} 0$ almost surely (a.s.),

A3) $X'_{mn} X_{mn}$ is non-singular for n sufficiently large, a.s., $\forall m = 1, 2, \ldots$, and

A4) $\sup\limits_{n \geq 1} \frac{1}{k_n} \sum\limits_{j=1}^{k_n} |x_{ij}^n| < \infty$ a.s., $\forall i = 1, 2, \ldots$,

where k_n is the number of regressors when the sample size is n , $(\cdot)^-$ denotes a generalized inverse, 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 for n sufficiently large a.s., and converges to the 0 matrix a.s. (see Anderson and Taylor (1979) and Lai, Robbins, and Wei (1978, 1979)).

In general, a LS estimator $\hat{\beta}_n$ can be expressed as

$$\hat{\beta}_n \equiv (X'_{1n} X_{1n})^{-} X'_{1n} y_{1n} .$$

By A3, $\hat{\beta}_n$ is a.s. unique for n sufficiently large. 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 β_n of most interest are the true regression parameter vectors β_n^0 . However, for the following result to hold, parameter vectors β_n^0 which are "true" in some sense need not even exist.[2] The vectors β_n are assumed to be sufficiently well-behaved as $n \to \infty$ that the corresponding "regression function" for the i^{th} observation, viz., $x_i^{n'}\beta_n$ (where $x_i^n \equiv (x_{i1}^n, \ldots, x_{ik_n}^n)'$), does not blow-up as additional regressors are added:

A5) $\sup_{n \geq 1} |x_i^{n'}\beta_n| < \infty$ 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. Under assumptions A1-A5,

$$||\hat{\beta}_n - \beta_n|| \xrightarrow{n\to\infty} 0 \tag{7}$$

with probability zero or probability one.

The use of this result is discussed in the Introduction, and is not repeated here. However, we mention that it may be of particular interest in models with unidentified parameters. In such models (with a fixed number of regressors) it might be thought that the LS estimator converges to different points in the set of observationally equivalent, true, parameter values with non-trivial probabilities. The theorem implies this is incorrect--the probability of convergence to any such parameter value is either 0 or 1.

Before proving the theorem, we state a Lemma. Let $v_i^n = y_i - x_i^{n'}\beta_n$, and $v_{rs}^n = (v_r^n, \ldots, v_s^n)'$, for $1 \leq r \leq s \leq n$. Define G to be the set of sample paths ω for which the conditions defined in A2-A5 hold.

---

[2] In such a case equation (6) still holds (tautologically) by taking $u_{1n} \equiv y_{1n} - X_{1n}\beta_n^0$, for any given β_n^0 .

Note, $P(G) = 1$. Let $S_1 = \{\omega : \|(X'_{1n}X_{1n})^- X'_{1n}v^n_{1n}\| \xrightarrow{n\to\infty} 0\}$, and $S_{2m} = \{\omega : \|(X'_{mn}X_{mn})^- X'_{mn}v^n_{mn}\| \xrightarrow{n\to\infty} 0\}$.

Lemma 1. For all positive integers m , $G \cap S_1 = G \cap S_{2m}$.

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 $S_{2m} \in \mathcal{B}_{m,\infty}$, the σ-field generated by (y_i, x_i) , $i = m, m+1, \ldots$. Further, $G \cap H = G \cap \overline{\lim_{m\to\infty}} S_{2m}$, where $\overline{\lim_{m\to\infty}} S_{2m} \equiv \lim_{m\to\infty} \bigcup_{\ell=m}^{\infty} S_{2\ell} \in \mathcal{T} \equiv \bigcap_{\ell=1}^{\infty} \mathcal{B}_{\ell,\infty}$. 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 $S_{3m} = \{\omega : \|(X'_{1n}X_{1n})^- X'_{mn}v^n_{mn}\| \xrightarrow{n\to\infty} 0\}$. 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 ,

$$\|(X'_{1n}X_{1n})^- X^{n\prime}_{1m}v^n_{1(m-1)}\| \xrightarrow{n\to\infty} 0 \quad \text{for all} \quad \omega \in G , \tag{8}$$

where X^n_{1m} 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^- (which is a.s. unique for n large), and η_r be the r^{th} element of the k_n-vector $(X'_{1n}X_{1n})^- X^{n\prime}_{1m}v^n_{1(m-1)}$. Then,

$$\max_{r\le k_n} |\eta_r| = \Big|\sum_{\ell=1}^{k_n}\sum_{i=1}^{m-1} c^{r\ell}x^n_{i\ell}(y_i - x^{n\prime}_i\beta_n)\Big|$$

$$\le k_n \cdot \|C^-\| \cdot \sum_{i=1}^{m-1}\left(\frac{1}{k_n}\sum_{\ell=1}^{k_n}|x^n_{i\ell}|\right)\cdot|y_i - x^{n\prime}_i\beta_n| \xrightarrow{n\to\infty} 0 , \tag{9}$$

for all $\omega \in G$, using A2, A4, and A5. This gives (8) and (i) is proved.

To show (ii) we introduce the following notation: $A = X_{mn}'X_{mn}$, $B = X_{1m}^{n}$, $d = X_{mn}'v_{mn}^{n}$, $g = C^{-1}d$, 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:

$$\|g\| \xrightarrow{n\to\infty} 0 \quad \text{implies} \quad \|A^{-1}d\| \xrightarrow{n\to\infty} 0 \,, \quad \text{for all} \quad \omega \in G \,. \tag{10}$$

Assume $\|g\| \xrightarrow{n\to\infty} 0$, for all $\omega \in G$. By A3 we can take n sufficiently large that C^{-1} and A^{-1} exist. Then, a well-known (and easily verified) equality for matrix inverses yields

$$A^{-1} \equiv (C - B'B)^{-1} = C^{-1} + C^{-1}B'(I_{m-1} - BC^{-1}B')^{-1}BC^{-1} \,, \tag{11}$$

and so,

$$\|A^{-1}d\| \le \|C^{-1}d\| + \|C^{-1}B'QBC^{-1}d\| \,. \tag{12}$$

Also,

$$\|BC^{-1}B'\| = \max_{r,s<m} \Big| \sum_{\ell=1}^{k_n} \sum_{j=1}^{k_n} x_{r\ell} c^{\ell j} x_{sj} \Big|$$

$$\le k_n^2 \cdot \|C^{-1}\| \cdot \max_{r,s<m} \Big(\frac{1}{k_n} \sum_{\ell=1}^{k_n} |x_{r\ell}|\Big)\Big(\frac{1}{k_n} \sum_{j=1}^{k_n} |x_{sj}|\Big) \xrightarrow{n\to\infty} 0 \,, \tag{13}$$

for all $\omega \in G$, by assumptions A2 and A4, where $x_{r\ell}$ is the $(r,\ell)^{th}$ element of B and $c^{\ell j}$ is the $(\ell,j)^{th}$ 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

$$||C^{-1}B'QBg|| \leq \max_{r\leq k_n} \sum_{j=1}^{k_n} \sum_{i=1}^{m-1} \sum_{\ell=1}^{m-1} \sum_{s=1}^{k_n} |c^{rj}x_{ij}q_{i\ell}x_{\ell s}g_s|$$

$$\leq k_n^2||C^{-1}||\cdot||Q||\cdot||g||\cdot\left(\sum_{i=1}^{m-1} \frac{1}{k_n} \sum_{j=1}^{k_n} |x_{ij}|\right)^2$$

$$\xrightarrow{n\to\infty} 0 \, , \tag{14}$$

for all $\omega \in G$, where $q_{i\ell}$ is the $(i,\ell)^{th}$ element of Q, g_s is the s^{th} element of g, and the convergence to zero follows using A2, A4, the result $||Q|| \xrightarrow{n\to\infty} 1$, and the assumption that $||g|| \xrightarrow{n\to\infty} 0$. Equations (12) and (14) yield the desired result (10).

The converse of (10) follows by the same argument as used to prove (10), noting that

$$C^{-1} \equiv (A+B'B)^{-1} = A^{-1} - A^{-1}B'(I_{m-1}+BA^{-1}B')^{-1}BA^{-1} \, , \tag{15}$$

and provided $k_n^2||A^{-1}|| \xrightarrow{n\to\infty} 0$, for all $\omega \in G$. The latter follows using (11), 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 (14). □

4. Acknowledgments

I would like to thank Peter C. B. Phillips for helpful comments.

REFERENCES

Anderson, T. W. and J. B. Taylor (1979). "Strong consistency of least squares estimates in dynamic models," Ann. Statist. 7 484-489.

Andrews, D. W. K. (1984). "Non-strong mixing autoregressive processes," forthcoming, J. Appl. Prob.

Bártfai, P. and P. Révész (1967). "On a zero-one law," Z. Wahrscheinlichkeitstheorie verw. Gebiete 7 43-47.

Billingsley, P. (1979). Probability and Measure (J. Wiley and Sons, New York).

Chanda, K. C. (1974). "Strong mixing properties of linear stochastic processes," J. Appl. Prob. 11 401-408.

Christopeit, N. and K. Helmes (1979). "A convergence theorem for random linear combinations of independent normal random variables," Ann. Statist. 7 795-800.

Drygas, H. (1976). "Weak and strong consistency of the least squares estimators in regression models," Z. Wahrscheinlichkeitstheorie verw. Gebeite 34 119-127.

Eicker, F. (1979). "Consistent parameter estimation in mixed autoregressions and in general linear stochastic regressions," manuscript, Dortmund University.

Huber, P. J. (1973). "Robust regression: Asymptotics, conjectures, and Monte Carlo," Ann. Statist. 1 799-821.

Ibragimov, I. A. and Yu. V. Linnik (1971). Independent and Stationary Sequences of Random Variables (Wolters-Noordhoff, Groningen, The Netherlands).

Iosifescu, M. and R. Theodorescu (1969). Random Processes and Learning (Springer-Verlag, New York).

Kolmogorov, A. N. (1933). Grundbegriffe der Wahrscheinlichkeitsrechnung (Springer, Berlin).

Kolmogorov, A. N. and Y. A. Rozonov (1960). "On strong mixing conditions for stationary Gaussian processes," Theor. Probability Appl. 5 204-208.

Nelson, P. I. (1980). "A note on strong consistency of least squares estimators in regression models with Martingale difference errors," Ann. Statist. 8 1057-1064.

Lai, T. L., H. Robbins, and C. Z. Wei (1978). "Strong consistency of least squares estimates in multiple regression," Proc. Natl. Acad. Sci. USA 75 3034-3036.

Lai, T. L., H. Robbins, and C. Z. Wei (1979). "Strong consistency of least squares estimates in multiple regression II," manuscript, Department of Mathematical Statistics, Columbia University.

Robinson, P. M. (1978). "On consistency in time series analysis," Ann. Statist. 6 215-223.