The Annals of Statistics
1981, Vol. 9, No. 3, 689–693

STRONG CONSISTENCY OF LEAST SQUARES ESTIMATORS IN REGRESSION WITH CORRELATED DISTURBANCES

BY V. SOLO[1]

CSIRO, St. Lucia, Australia

This note considers, under minimal assumptions, the strong consistency of least squares estimates in regression with correlated errors.

Recently Lai et al. (1979) have proven the a.s. convergence of the least squares regression estimator (with nonstochastic regressors and martingale difference sequence disturbances) assuming only the smallest eigenvalue of the $X'X$ matrix tends to ∞. Further, Lai et al. were able to handle some types of autocorrelated noise. The aim of this note is to discuss strong convergence with minimal assumptions on the $X'X$ matrix for some types of autocorrelated noise not handled by Lai et al.

Consider then the regression model

$$y_n = \mathbf{x}_n'\boldsymbol{\beta} + \varepsilon_n$$

where y_n is an observed sequence, $\mathbf{x}_n$ a p-vector of nonstochastic regressors and ε_n a disturbance sequence. We are interested in the least squares estimator

$$\hat{\boldsymbol{\beta}}_n = \mathbf{V}_n^{-1} \sum_1^n \mathbf{x}_s y_s, \qquad \mathbf{V}_n = \sum_1^n \mathbf{x}_s \mathbf{x}_s'.$$

Assume $\mathbf{V}_p$ is positive definite, as then also is $\mathbf{V}_n$, $n \geq p$. Now we can write

$$\hat{\boldsymbol{\beta}}_n = \hat{\boldsymbol{\beta}}_{n-1} + \mathbf{V}_n^{-1}\mathbf{x}_n e_n,$$

$$e_n = y_n - \mathbf{x}_n'\hat{\boldsymbol{\beta}}_{n-1}.$$

Let $\boldsymbol{\alpha}$ be a fixed vector and suppose it can be shown that $\sum_1^n \bar{c}_s e_s$ converges a.s. with $\bar{c}_s = \boldsymbol{\alpha}'\mathbf{V}_s^{-1}\mathbf{x}_s$. Then $\hat{\boldsymbol{\beta}}_n$ converges a.s. and to conclude $\hat{\boldsymbol{\beta}}_n \to \boldsymbol{\beta}$ a.s. it will be enough to show $\hat{\boldsymbol{\beta}}_n \to \boldsymbol{\beta}$ in probability. We now derive two identities that are basic to all that follows. Suppose ε_n is an uncorrelated sequence with unit variance. It is well known, and easily verified directly, that e_n is also an uncorrelated sequence but with variance $1 + \mathbf{x}_n'\mathbf{V}_{n-1}^{-1}\mathbf{x}_n$. Thus is this case, letting c_i be some sequence of constants we have

$$E\left(\sum_1^n c_s e_s\right)^2 = \sum_1^n c_s^2(1 + \mathbf{x}_s'\mathbf{V}_{s-1}^{-1}\mathbf{x}_s). \tag{1}$$

However we can also write

$$\begin{aligned} e_s = y_s - \mathbf{x}_s'\hat{\boldsymbol{\beta}}_{s-1} &= \varepsilon_s - \mathbf{x}_s'(\hat{\boldsymbol{\beta}}_{s-1} - \boldsymbol{\beta}) \\ &= \varepsilon_s - \mathbf{x}_s'\mathbf{V}_{s-1}^{-1} \sum_1^{s-1} \mathbf{x}_t \varepsilon_t \\ &= \sum_1^s a_{st}\varepsilon_t \end{aligned}$$

with $a_{ss} = 1$ and $a_{st} = -\mathbf{x}_s'\mathbf{V}_{s-1}^{-1}\mathbf{x}_t$, $t < s$. Thus

$$\begin{aligned} \sum_1^n c_s e_s = \sum_1^n c_s \sum_1^s a_{st}\varepsilon_t &= \sum_1^n \varepsilon_t \sum_t^n c_s a_{st} \\ &= \sum_1^n \varepsilon_t b_{nt}, \quad \text{say} \end{aligned} \tag{2}$$

Received May, 1979; revised December, 1979.

[1] Now at University of Wisconsin.

AMS 1970 *subject classifications.* Primary G2J05; secondary G0F15.

Key words and phrases. Regression, strong consistency, correlated disturbances, method of subsequences, least squares, asymptotic.

Thus taking variances and equating to (1) yields

$$\textstyle\sum_1^n c_s^2(1 + \mathbf{x}_s'\mathbf{V}_{s-1}^{-1}\mathbf{x}_s) = \sum_1^n b_{nt}^2. \tag{3a}$$

Similarly computing covariances gives, for $n > m$,

$$\textstyle\sum_1^m b_{mt}^2 = \sum_1^m c_s^2(1 + \mathbf{x}_s'\mathbf{V}_{s-1}^{-1}\mathbf{x}_s) = \sum_1^m b_{nt}b_{mt}. \tag{3b}$$

Let us now list three types of disturbance sequence to be considered.

N1. $\{\varepsilon_n\}$ is an uncorrelated sequence with $\sup_i E(\varepsilon_i^2) \le K < \infty$.

N2. $\{\varepsilon_n\}$ is a stationary sequence with autovariance function $\Omega(s)$ and spectrum $f(\omega)$ bounded by K an a.s. constant.

N3. $\{\varepsilon_n\}$ is such that the largest eigenvalue of the covariance matrix of any fixed number of successive values of ε_n is bounded by a constant K. Actually this condition includes N2 and N1, but N2 and N1 are singled out because of their concrete meaning.

Now suppose N1 holds, and let n, m be fixed integers with $n > m$. Then using (3a) and (3b) we have

$$\begin{aligned}
E(\textstyle\sum_{m+1}^n c_s e_s)^2 &= E(\textstyle\sum_1^n c_s e_s - \sum_1^m c_s e_s)^2 = E(\sum_1^n b_{nt}\varepsilon_t - \sum_1^m b_{mt}\varepsilon_t)^2\\
&= E(\textstyle\sum_{m+1}^n b_{nt}\varepsilon_t + \sum_1^m (b_{nt} - b_{mt})\varepsilon_t)^2\\
&\le K\{\textstyle\sum_{m+1}^n b_{nt}^2 + \sum_1^m (b_{nt} - b_{mt})^2\}\\
&= K(\textstyle\sum_{m+1}^n b_{nt}^2 + \sum_1^m b_{nt}^2 - \sum_1^m b_{mt}^2)\\
&= K(\textstyle\sum_1^n b_{nt}^2 - \sum_1^m b_{mt}^2)\\
&= K\textstyle\sum_{m+1}^n c_s^2(1 + \mathbf{x}_s'\mathbf{V}_{s-1}^{-1}\mathbf{x}_s). \qquad (4)
\end{aligned}$$

We next demonstrate that (4) holds also under N2. Writing

$$\textstyle\sum_{m+1}^n b_{nt}\varepsilon_t + \sum_1^m (b_{nt} - b_{mt})\varepsilon_t = \sum_1^n b_{nmt}\varepsilon_t$$

we deduce, since m, n are fixed, that

$$\begin{aligned}
E(\textstyle\sum_{m+1}^n c_s e_s)^2 &= \textstyle\sum_1^n \sum_1^n b_{nmt}b_{nms}\Omega(t-s)\\
&= \int_{-\pi}^{\pi} |\textstyle\sum_1^n b_{nmt}e^{i\omega t}|^2 f(\omega)\, d\omega\\
&\le K(\textstyle\sum_1^n b_{nmt}^2)\\
&= K\textstyle\sum_{m+1}^n c_s^2(1 + \mathbf{x}_s'\mathbf{V}_{s-1}^{-1}\mathbf{x}_s)
\end{aligned}$$

by the arguments just given. Similarly (4) holds under N3.

REMARK 1. If we observe

$$1 + \mathbf{x}_s'\mathbf{V}_{s-1}^{-1}\mathbf{x}_s = \textstyle\sum_1^s a_{st}^2 = v_s^2,$$

say, then it is clear that for any array $\{a_{st}\}$ which is such as to ensure that $\{e_n\}$ is an orthogonal sequence whenever $\{\varepsilon_n\}$ is, we have under N1 or N2 or N3 that

$$E(\textstyle\sum_{m+1}^n c_s e_s)^2 \le K\sum_{m+1}^n c_s^2 v_s^2.$$

Continuing, we observe that (4) is all that is required for Menchoff's inequality (Stout, 1974, page 18) to hold. (In that inequality and its proof replace $(= \cdots)E(x_n^2)$ wherever it occurs by $(\le K \cdots)c_n^2 v_n^2$.) Thus via the method of subsequences (Stout, 1974, page 20) we have the following result.

THEOREM 1. *If* N1 *or* N2 *or* N3 *holds, then*

(5) $$\sum_1^n c_s e_s \quad \textit{converges}\ \text{a.s.}\ \textit{if}\ \ \sum_{p+1}^\infty c_s^2(1 + \mathbf{x}_s'\mathbf{V}_{s-1}^{-1}\mathbf{x}_x)\log^2 s < \infty.$$

REMARK 2. Take $\mathbf{x}_s = \mathbf{0}$ to see that under N1 or N2 or N3

$$\sum_1^\infty c_s\varepsilon_s < \infty \text{ a.s.} \quad \text{if} \quad \sum_1^\infty c_s^2 \log^2 s < \infty.$$

This result was given (implicitly) by Hannan (1978) under N2.

FIRST PROOF. The proof is the same as that of Theorem 2.3.2 of Stout if wherever $(= \cdots)E(x_n^2)$ occurs we replace it by $(\leq K \cdots)c_n^2 v_n^2$ and we realise that orthogonality in that proof is used only to ensure that (4) holds.

SECOND PROOF. The Theorem also follows from Serfling's inequality (Theorem 2.4.2 of Stout, page 25). To see this set

$$g_{a,n} = K\sum_{a+1}^{a+n} c_s^2 v_s^2$$

$$h_{a,n} = K\sum_{a+1}^{a+n} c_s^2 v_s^2 \log^2 s,$$

so by (5), for some constant K',

(6) $$h_{a,n} \leq K' < \infty \quad \forall\ a, n.$$

Also observe that, from (4),

(7) $$E\left(\sum_{a+1}^{a+n} c_s e_s\right)^2 \leq g_{a,n},$$

while

(8) $$g_{a,n} \leq h_{a,n}/\log^2(a+1)$$

and

(9a) $$g_{a,n} + g_{a+n,l} \leq g_{a,n+l},$$

(9b) $$h_{a,n} + h_{a+n,l} \leq h_{a,n+l}.$$

Now Stout requires $g_{a,n}$ to be of the form $g(F_{a,n})$, a functional on the joint distribution $F_{a,n}$ of $x_{a+1} \cdots x_{a+n}$. However a perusal of the proof of Theorem 2.4.2 will show that a set of numbers $g_{a,n}$, $h_{a,n}$ obeying (6)–(9) will do.

To apply Theorem 1 take $c_s = \boldsymbol{\alpha}'\mathbf{V}_s^{-1}\mathbf{x}_s$, call σ_s the smallest eigenvalue of $\mathbf{V}_s$ and observe

$$\mathbf{V}_s^{-1} - \mathbf{V}_{s-1}^{-1} = -\mathbf{V}_{s-1}^{-1}\mathbf{x}_s\mathbf{x}_s'\mathbf{V}_{s-1}^{-1}/(1 + \mathbf{x}_s'\mathbf{V}_{s-1}^{-1}\mathbf{x}_s)$$

so $\mathbf{V}_s^{-1}\mathbf{x}_s = \mathbf{V}_{s-1}^{-1}\mathbf{x}_s/(1 + \mathbf{x}_s'\mathbf{V}_{s-1}^{-1}\mathbf{x}_s)$. Also $\operatorname{tr}(\mathbf{V}_s^{-1}) \leq p/\sigma_s$. Thus

$$\sum_{p+1}^\infty c_s^2(1 + \mathbf{x}_s'\mathbf{V}_{s-1}^{-1}\mathbf{x}_s)\log^2 s \leq \boldsymbol{\alpha}'\boldsymbol{\alpha}\sum_{p+1}^\infty \log^2 s \operatorname{tr}(\mathbf{V}_{s-1}^{-1} - \mathbf{V}_s^{-1})$$

$$\leq p\boldsymbol{\alpha}'\boldsymbol{\alpha}\sum_{p+2}^\infty (\log^2 s - \log^2(s-1))/\sigma_{s-1} + \boldsymbol{\alpha}'\boldsymbol{\alpha}\log^2(p+1)\operatorname{tr}(\mathbf{V}_p^{-1}).$$

Elementary considerations show the first sum is finite iff

$$\sum_{p+2}^\infty \log s/(s\sigma_{s-1}) < \infty.$$

The following Theorem may now be stated.

THEOREM 2. *If* N1 *or* N2 *or* N3 *hold and* $\sigma_n \uparrow \infty$, *then* $\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta} \to \mathbf{0}$ a.s. *if*

$$\sum_{p+2}^\infty (\log^2 s - \log^2(s-1))/\sigma_{s-1} < \infty$$

or equivalently $\sum_{p+2}^\infty \log s/(s\sigma_{s-1}) < \infty$.

PROOF. It is only necessary to show $\hat{\boldsymbol{\beta}}_n \to \boldsymbol{\beta}$ in probability. However let $\boldsymbol{\alpha}$ be a fixed vector and consider that under N2, for example,

$$\begin{aligned}\operatorname{Var}(\boldsymbol{\alpha}'\hat{\boldsymbol{\beta}}_n) &= E(\boldsymbol{\alpha}'\mathbf{V}_p^{-1}\textstyle\sum_1^n \mathbf{x}_s\varepsilon_s)^2 \\ &= \textstyle\sum_1^n\sum_1^n \boldsymbol{\alpha}'\mathbf{V}_n^{-1}\mathbf{x}_s\Omega(s-s')\mathbf{x}_{s'}'\mathbf{V}_n^{-1}\boldsymbol{\alpha} \\ &= \int_{-\pi}^{\pi} |\textstyle\sum_1^n \boldsymbol{\alpha}'\mathbf{V}_n^{-1}\mathbf{x}_s e^{i\omega s}|^2 f(\omega)\, d\omega \\ &\le K\textstyle\sum_1^n (\boldsymbol{\alpha}'\mathbf{V}_n^{-1}\mathbf{x}_s)^2 \\ &= K\,\boldsymbol{\alpha}'\mathbf{V}_n^{-1}\boldsymbol{\alpha} \to 0\end{aligned}$$

if $\sigma_n \uparrow \infty$. Similar calculation follows under N1 and N3.

REMARK 3. For the scalar case $p = 1$, and under N2, this theorem is implicit in Hannan (1978). The proof in the scalar case is straight forward since the fact that $\hat{\beta}_n - \beta_0 \to 0$ a.s. follows via Kronecker's lemma from $\sum_1^\infty x_s\varepsilon_s/V_s < \infty$, which Hannan shows will hold if

$$\textstyle\sum_1^\infty x_s^2 \log^2 s/V_s^2 < \infty;$$

cf. Remark 2 above.

REMARK 4. From Kronecker's lemma it follows that σ_s increases more rapidly than $\log^2 s$. It is instructive to restate the basic theorem of Lai et al. in the present terms.

THEOREM 3. (Modified from Lai et al., 1979). *Let $\bar{c}_i$ be a sequence of constants and let ε_i be such that*

(10) $$\textstyle\sum_1^\infty \bar{c}_i^2 < \infty, \Rightarrow \sum_1^\infty \bar{c}_i\varepsilon_i < \infty \text{ a.s.}$$

Then e_i 'inherits' this property from ε_i in that

$$\textstyle\sum_1^\infty c_i e_i < \infty \text{ a.s.} \quad \textit{if} \quad \sum_1^\infty c_i^2(1 + \mathbf{x}_i'\mathbf{V}_{i-1}^{-1}\mathbf{x}_i) < \infty.$$

REMARK 5. When ε_s are Gaussian with constant variance then $\{e_n\}$ is a martingale difference sequence. Then the claim of Theorem 3 holds from the martingale convergence theorem. This idea, which is effectively that used by Sternby (1977), gives an alternative proof to that of Anderson and Taylor (1976) in the Gaussian case.

THEOREM 4. *Hannan* (1978) *has shown that condition* N4 *below implies* (10).
N4. *$\{\varepsilon_n\}$ is a stationary sequence obeying the following two conditions*

$$E(\varepsilon_n \mid \mathcal{F}_{-\infty}) = 0 \qquad (\varepsilon_n \textit{ is purely nondeterministic})$$

and

$$\textstyle\sum_0^\infty \alpha_j < \infty.$$

In condition N4, *$\mathcal{F}_n$ are the increasing σ-algebras generated by ε_n while the α_j occur as follows. Consider*

$$u_{n,j} = E(\varepsilon_n \mid \mathcal{F}_j) - E(\varepsilon_n \mid \mathcal{F}_{j-1})$$

so that

$$\varepsilon_n = \textstyle\sum_0^\infty u_{n,j} + E(\varepsilon_n \mid \mathcal{F}_{-\infty}).$$

Then stationarity ensures we can write

$$\alpha_j = \sqrt{E}(u_{n,n-j}^2)$$

so that

$$\varepsilon_n = \sum_0^\infty \alpha_j \xi_{n,n-j} + E(\varepsilon_n \mid \mathscr{F}_{-\infty}),$$

where

$$E(\xi_{n,n-j}^2) = 1.$$

REMARK. Hannan shows that ε_n in N4 obeys a Doob-like maximal inequality and this enables (10) to be established.

REFERENCES

ANDERSON, T. W. and TAYLOR, J. B. (1976). Strong consistency of least squares estimates in Normal Linear Regression. *Ann. Stat.* **4** 788–790.

HANNAN, E. J. (1973). Central Limit Theorems for Time Series Regression, *Z. Wahrscheinlichkeitstheorie und verw. Gebiete* **25** 157–170.

HANNAN, E. J. (1978). Rates of convergence for Time Series regression. *Advances Appl. Probability* **10** 740–743.

LAI, T. L., ROBBINS, H. and WEI, C. A. (1979). Strong consistency of Least Squares Estimates in Multiple Regression II. Unpublished manuscript.

STERNBY, J. (1977). On Consistency for the method of least squares using martingale theory. *IEEE, Trans. Autom. Control.* AC-**22** 346–352.

STOUT, W. F. (1974). *Almost Sure Convergence.* Academic.

MATHEMATICS RESEARCH CENTER
UNIVERSITY OF WISCONSIN
610 WALNUT STREET
MADISON, WISCONSIN 53706