*The Annals of Statistics*
1985, Vol. 13, No. 4, 1498–1508

# ASYMPTOTIC PROPERTIES OF LEAST-SQUARES ESTIMATES IN STOCHASTIC REGRESSION MODELS

BY C. Z. WEI[1]

*University of Maryland*

Strong consistency of least-squares estimates in stochastic regression models is established under the assumption that the underlying model can be reparametrized so that the new design vectors are weakly correlated. An application to fixed-width interval estimation in stochastic approximation schemes is also discussed.

**1. Introduction and summary.** Consider the multiple regression model

$$y_n = \beta_1 x_{n1} + \cdots + \beta_p x_{np} + \varepsilon_n, \qquad n = 1, 2\ldots, \tag{1.1}$$

where the $\varepsilon_n$ are unobservable random errors, $\beta_1, \ldots, \beta_p$ are unknown parameters, and $y_n$ is the observed response corresponding to the design vector $x_n = (x_{n1}, \ldots, x_{np})'$. Then

$$\mathbf{b}_n = (b_{n1}, \ldots, b_{np})' = \left(\sum_1^n \mathbf{x}_i \mathbf{x}_i'\right)^{-1} \sum_1^n \mathbf{x}_i y_i$$

denotes the least-squares estimate of $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ based on the observations $\mathbf{x}_1, y_1, \ldots, \mathbf{x}_n, y_n$, assuming that $\sum_1^n \mathbf{x}_i \mathbf{x}_i'$ is nonsingular. Throughout the sequel, we shall assume that $\{\varepsilon_n\}$ is a martingale difference sequence with respect to an increasing sequence of $\sigma$ fields $\{\mathscr{F}_n\}$, i.e., $\varepsilon_n$ is $\mathscr{F}_n$-measurable and $E(\varepsilon_n | \mathscr{F}_{n-1}) = 0$ for every $n$. We shall also assume that the design vector at stage $n$ depends on the previous observations $\mathbf{x}_1, y_1, \ldots, \mathbf{x}_{n-1}, y_{n-1}$; i.e., $\mathbf{x}_n$ is $\mathscr{F}_{n-1}$-measurable. The asymptotic properties of the least-squares estimates were recently studied by Lai and Wei (1982) who in particular established the following theorem concerning the strong consistency of $b_n$.

THEOREM A. *Assume in the regression model* (1.1) *that*

$$\sup_n E(|\varepsilon_n|^\gamma | \mathscr{F}_{n-1}) < \infty \quad a.s. \text{ for some } \gamma > 2, \tag{1.2}$$

*and*

$$\lambda_{\min}(n) \to \infty \quad a.s. \quad \text{and} \quad \log \lambda_{\max}(n) = o(\lambda_{\min}(n)) \quad a.s., \tag{1.3}$$

*where* $\lambda_{\min}(n)$ *and* $\lambda_{\max}(n)$ *denote the minimum and maximum eigenvalues of* $\sum_1^n \mathbf{x}_i \mathbf{x}_i'$, *respectively. Then*

$$\mathbf{b}_n \to \boldsymbol{\beta} \quad a.s. \tag{1.4}$$

Received August 1983; revised February 1985.
[1]Research supported by the National Science Foundation under grant NSF-MCS-8103448A01.
*AMS* 1980 *subject classifications*. Primary 62J05; secondary 62L20.
*Key words and phrases*. Stochastic regressors, least squares, stochastic approximation, strong consistency, martingales.

Lai and Wei (1982) also showed that, without further assumptions on $\mathbf{x}_n$, (1.3) is the best possible condition for the strong consistency of $\mathbf{b}_n$. They construct a counterexample in which $P\{\mathbf{b}_n \to \boldsymbol{\beta}\} = 0$ and in which (1.3) is only violated marginally in the sense that

$$\lambda_{\min}(n) \to \infty \quad \text{a.s.} \tag{1.5}$$

and $\log \lambda_{\max}(n)/\lambda_{\min}(n)$ converges to a positive random variable.

There are many interesting applications, such as stochastic approximation and optimal control of a linear dynamic system, in which (1.5) holds, rather than (1.3), as shown in Section 3. [See also Lai and Wei (1985).] This motivates a study of the asymptotic behavior of least-squares estimates when (1.3) is relaxed. However, we impose the additional condition that there exists a linear transformation $\mathbf{A}$ such that the new design vectors $\mathbf{z}_n = \mathbf{A}\mathbf{x}_n$ have the property:

$$\liminf_{n\to\infty} \lambda_{\min}\left(\mathbf{D}_n^{-1}\left(\sum_1^n \mathbf{z}_i\mathbf{z}_1'\right)\mathbf{D}_n^{-1}\right) > 0 \quad \text{where } \mathbf{D}_n = \left\{\operatorname{diag}\left(\sum_1^n \mathbf{z}_i\mathbf{z}_i'\right)\right\}^{1/2} \tag{1.6}$$

Thus, the linear transformation $\mathbf{A}$ induces a reparametrization of the model (1.1) so that the new design vectors $\mathbf{z}_n$ form a "weakly correlated" design in the sense of (1.6). As far as strong consistency is concerned, this reparametrization reduces the original problem to a univariate case. That is why the improvement is possible. More precisely, when such a reparametrization is available, the following theorem shows that $\mathbf{b}_n$ is still strongly consistent when (1.5) holds, or more generally, when (1.3) is weakened to condition (1.7) below.

THEOREM 1. *Suppose that in the regression model* (1.1), (1.2) *holds and*

$$\lambda_{\min}(n) \to \infty \quad a.s. \quad \text{and} \quad \{\log \lambda_{\max}(n)\}^{2\delta} = o(\lambda_{\min}(n)) \quad a.s. \tag{1.7}$$

*for some* $\delta > 1/\min(\gamma, 4)$.

*Suppose that there exists a nonsingular matrix* $\mathbf{A}$ *such that the random vectors* $\mathbf{z}_n = \mathbf{A}\mathbf{x}_n$ *satisfy* (1.6). *Then* $\mathbf{b}_n \to \boldsymbol{\beta}$ a.s. *If the vectors* $\mathbf{z}_n = (z_{n1}, \dots, z_{np})'$ *satisfy, in addition to* (1.6), *the condition*:

$$z_{ni}^2 = o\left(\left(\sum_{k=1}^n z_{ki}^2\right)^c\right) \quad \text{for some } 0 < c < 1 \quad \text{and} \quad i = 1, \dots, p, \tag{1.8}$$

*then we can weaken* (1.7) *to the assumption*

$$\lambda_{\min}(n) \to \infty \quad \text{a.s.} \quad \text{and} \quad \log\log \lambda_{\max}(n) = o(\lambda_{\min}(n)) \quad \text{a.s.}, \tag{1.9}$$

*and still obtain the strong consistency of* $\mathbf{b}_n$.

The proof of Theorem 1 is given in Section 2. An application of Theorem 1 to fixed width interval estimation in stochastic approximation schemes is discussed in Section 3.

**2. Proof of Theorem 1.** The proof of Lemma 1 below is straightforward.

LEMMA 1. *Let* **B** *and* **C** *be* $p \times p$ *matrices. If* **C** *is symmetric and nonnegative definite, then*

$$\lambda_{\max}(\mathbf{C})\lambda_{\max}(\mathbf{B}'\mathbf{B}) \geq \lambda_{\max}(\mathbf{B}'\mathbf{C}\mathbf{B}) \geq \lambda_{\min}(\mathbf{B}'\mathbf{C}\mathbf{B}) \geq \lambda_{\min}(\mathbf{C})\lambda_{\min}(\mathbf{B}'\mathbf{B}). \tag{2.1}$$

LEMMA 2. *Let* $\{\varepsilon_n, \mathscr{F}_n\}$ *be a martingale difference sequence satisfying* (1.2). *Let* $\{u_n\}$ *be a sequence of random variables such that* $u_n$ *is* $\mathscr{F}_{n-1}$*-measurable. Define* $s_n^2 = \sum_1^n u_i^2$. *Then*

$$\sum_1^n u_i \varepsilon_i = O\left(s_n (\log s_n)^{\delta}\right) \quad a.s., \tag{2.2}$$

$\forall\, \delta > 1/\min(\gamma, 4)$. *Furthermore, if*

$$u_n^2 = o\left(s_n^{2c}\right) \quad \textit{for some } 0 < c < 1, \tag{2.3}$$

*then*

$$\sum_1^n u_i \varepsilon_i = O\left(s_n (\log\log s_n)^{1/2}\right) \quad a.s. \tag{2.4}$$

PROOF. Without loss of generality, we can assume each $u_k$ is a bounded random variable. Otherwise, choose $a_k$ so that $\sum_1^\infty P[|u_k| > a_k] < \infty$ and replace $u_k$ by $u_k I_{[|u_k| \leq a_k]}$. By Strassen's imbedding theorem (Strassen, 1966 or Hall and Heyde, 1980, page 269), there is a Brownian motion $\{W(t),\, t \geq 0\}$ together with a sequence of nonnegative random variables $\{T_i,\, i \geq 1\}$ such that

$$\sum_1^n Y_i = W\left(\sum_1^n T_i\right) \quad \text{a.s.}, \tag{2.5}$$

$$E(T_n | \mathscr{F}_{n-1}) = E\left(Y_n^2 | \mathscr{F}_{n-1}\right) \quad \text{a.s.}, \tag{2.6}$$

and

$$E(T_n^r | \mathscr{F}_{n-1}) \leq c_r E\left(|Y_n|^{2r} | \mathscr{F}_{n-1}\right) \quad \text{a.s.} \quad \text{for all } r > 1, \tag{2.7}$$

where $Y_i = u_i \varepsilon_i$ and $c_r$ only depends on $r$.

Fix $r = \min(\gamma/2, 2)$. Then $2 \geq r > 1$ and $2\delta r > 1$. Let $Z_k = T_k - E(T_k | \mathscr{F}_{k-1})$, $t_k = s_k^2 (\log s_k)^{2\delta} / (\log\log s_k)$ and $V_n = \sum_1^n Z_k / t_k$. Since $s_k$ is $\mathscr{F}_{k-1}$-measurable,

$\{V_n,\ n \geq 1\}$ is a martingale. By (2.7),

$$\begin{aligned}\sum_1^{\infty} E(|Z_k|^r|\mathscr{F}_{k-1})/t_k^r &\leq 2^r \sum_1^{\infty} E(|T_k|^r|\mathscr{F}_{k-1})/t_k^r \\ &\leq 2^r c_r \sum_1^{\infty} E\big(|Y_k|^{2r}|\mathscr{F}_{k-1}\big)/t_k^r \\ &= 2^r c_r \sum_1^{\infty} E\big(|\varepsilon_k|^{2r}|\mathscr{F}_{k-1}\big)|u_k|^{2r}/t_k^r \\ &\leq 2^r c_r \sum_1^{\infty} \big(E(|\varepsilon_k|^{\gamma}|\mathscr{F}_{k-1})\big)^{(2r)/\gamma} u_k^2 (\log\log s_k)^r / s_k^2 (\log s_k)^{2\delta r} \\ &< \infty \quad \text{a.s.}\end{aligned}$$

on the event $\{s_n^2 \to \infty\}$ by (1.2). Thus by the local martingale convergence theorem (Chow, 1965) and Kronecker's lemma,

$$\sum_1^n Z_k = o(t_n) \quad \text{a.s.} \quad \text{on } \{s_n^2 \to \infty\}. \tag{2.8}$$

Hence by (2.6) and (2.8),

$$\begin{aligned}\sum_1^n T_k &= \sum_1^n E(T_k|\mathscr{F}_{k-1}) + \sum_1^n Z_k \\ &= \sum_1^n E\big(Y_k^2|\mathscr{F}_{k-1}\big) + \sum_1^n Z_k \\ &= O\big(s_n^2\big) + o(t_n) \quad \text{a.s.} \quad \text{on}\{s_n^2 \to \infty\} \\ &= o(t_n) \quad \text{a.s.} \quad \text{on } \{s_n^2 \to \infty\}.\end{aligned}$$

By the law of the iterated logarithm for Brownian motion,

$$\begin{aligned}\sum_1^n u_i\varepsilon_i &= W\left(\sum_1^n T_i\right) \\ &= O\left(\left(\sum_1^n T_i\right)\log\log\left(\sum_1^n T_i\right)\right)^{1/2} \quad \text{a.s.} \\ &= o(t_n \log\log t_n)^{1/2} \quad \text{a.s.} \quad \text{on } \{s_n^2 \to \infty\} \\ &= o\big(s_n(\log s_n)^{\delta}\big) \quad \text{a.s.} \quad \text{on } \{s_n^2 \to \infty\}.\end{aligned} \tag{2.9}$$

On the event $\{\sum_1^{\infty} u_i^2 < \infty\}$, by the local martingale convergence theorem

$$\sum_1^n u_i\varepsilon_i = O(1) \quad \text{a.s.} \tag{2.10}$$

From (2.9) and (2.10), (2.2) follows. Moreover, if (2.3) holds, a standard Chebyshev's inequality argument shows that

$$\sum_{k=1}^{\infty} E\Big\{(u_k\varepsilon_k)^2 I_{[(u_k\varepsilon_k)^2 > s_k^2/(\log s_k)(\log\log s_k)^4]}|\mathscr{F}_{k-1}\Big\}\log\log s_k/s_k^2$$

$$\le S\sup_k E(|\varepsilon_k|^r|\mathscr{F}_{k-1}) \sum_{k=1}^{\infty} |u_k|^r(\log s_k)^{r/2-1}(\log\log s_k)^{2r-3}/s_k^r$$

$$\le \sup_k E(|\varepsilon_k|^r|\mathscr{F}_{k-1}) \sum_{k=1}^{\infty} u_k^2(\log s_k)^{r/2-1}(\log\log s_k)^{2r-3}/s_k^{2+(r-2)(1-c)}$$

$$< \infty \quad \text{a.s.} \quad \text{on the event } \left\{\sum_1^{\infty} u_k^2 = \infty\right\}.$$

Hence by (2.10) and Theorem 3.2 of Jain et al. (1975), (2.4) follows. □

PROOF OF THEOREM 1. First note that

$$\begin{aligned}
\mathbf{b}_n - \boldsymbol{\beta} &= \left(\sum_1^n \mathbf{x}_i\mathbf{x}_i'\right)^{-1}\sum_1^n \mathbf{x}_i\varepsilon_i \\
&= \left\{\mathbf{A}^{-1}\left(\sum_1^n \mathbf{z}_i\mathbf{z}_i'\right)(\mathbf{A}')^{-1}\right\}^{-1}\mathbf{A}^{-1}\sum_1^n \mathbf{z}_i\varepsilon_i \\
&= \mathbf{A}'\left(\sum_1^n z_i z_i'\right)^{-1}\sum_1^n \mathbf{z}_i\varepsilon_i \\
&= \mathbf{A}'\mathbf{D}_n^{-1}\left\{\mathbf{D}_n^{-1}\left(\sum_1^n \mathbf{z}_i\mathbf{z}_i'\right)\mathbf{D}_n^{-1}\right\}^{-1}\mathbf{D}_n^{-1}\sum_1^n \mathbf{z}_i\varepsilon_i.
\end{aligned}$$

Hence by (1.6),

$$\|\mathbf{b}_n - \boldsymbol{\beta}\| = O\left(\|\mathbf{D}_n^{-1}\|\left\|\mathbf{D}_n^{-1}\sum_1^n \mathbf{z}_i\varepsilon_i\right\|\right) \quad \text{a.s.} \tag{2.11}$$

Since

$$\sum_1^n \mathbf{x}_i\mathbf{x}_i' = \mathbf{A}^{-1}\left(\sum_1^n \mathbf{z}_i\mathbf{z}_i'\right)(\mathbf{A}')^{-1}$$

and

$$\sum_1^n \mathbf{z}_i\mathbf{z}_i' = \mathbf{D}_n^{-1}\left\{\mathbf{D}_n^{-1}\left(\sum_1^n \mathbf{z}_i\mathbf{z}_i'\right)\mathbf{D}_n^{-1}\right\}^{-1}\mathbf{D}_n^{-1},$$

Lemma 1, (1.6), and (1.7) imply that

$$\lambda_{\min}(\mathbf{D}_n) \to \infty \quad \text{a.s.} \tag{2.12}$$

and

$$\{\log\lambda_{\max}(\mathbf{D}_n)\}^{\delta} = o(\lambda_{\min}^2(\mathbf{D}_n)) \quad \text{a.s.}$$

By Lemma 2, (2.11), and (2.12),

$$\|\mathbf{b}_n - \boldsymbol{\beta}\| = O\big(\{\log \lambda_{\max}(\mathbf{D}_n)\}^{\delta}/\lambda^2_{\min}(\mathbf{D}_n)\big)$$
$$= o(1) \quad \text{a.s.}$$

Moreover, if (1.9) holds then similar argument shows that (1.4) also holds. □

**3. Applications to fixed-width interval estimation in stochastic approximation schemes.** Consider the general regression model

$$y_n = y(x_n) = M(x_n) + \varepsilon_n, \qquad n = 1,2,\dots, \tag{3.1}$$

where the $x_n$ are real, the errors $\varepsilon_1, \varepsilon_2, \dots$ are i.i.d. random variables with mean 0, variance $\sigma^2$, and $E|\varepsilon_1|^3 < \infty$. Assume that the regression function $M(x)$ is a Borel function satisfying the following three conditions:

$$M(\theta) = 0 \quad \text{for a unique } \theta \text{ and } M'(\theta) = \beta \text{ exists and is positive;} \tag{3.2}$$

$$\inf_{\delta \le |x-\theta| \le \delta^{-1}} \{M(x)(x-\theta)\} > 0 \quad \text{for all } 0 < \delta < 1; \tag{3.3}$$

$$|M(x)| \le C|x| + D \quad \text{for some } C, D > 0 \text{ and all } x. \tag{3.4}$$

A class of stochastic approximation procedures, originally proposed by Robbins and Monro (1951), for finding the root $\theta$ and choosing the levels $x_n$ is represented by recursions of the form

$$x_{n+1} = x_n - by_n/n \tag{3.5}$$

where $x_1$ and $b > 0$ are constants. Blum (1954) showed that $x_n \to \theta$ a.s. and Sacks (1958) showed that if $2b\beta > 1$ then $n^{1/2}(x_n - \theta)$ is asymptotically normal with mean 0 and variance $b^2\sigma^2/(2b\beta - 1)$. In practice we may want to terminate the procedure when $x_n$ is sufficiently close to $\theta$ with high probability. We propose here a stopping rule giving a fixed-width confidence interval for $\theta$. Let $\mathbf{x}_n = (1, x_n)'$. Define

$$(\hat\alpha_n, \hat\beta_n)' = \left(\sum_1^n \mathbf{x}_i\mathbf{x}_i'\right)^{-1} \sum_1^n \mathbf{x}_i y_i,$$
$$\hat\sigma_n^2 = n^{-1}\sum_1^n (y_i - \hat\alpha_n - \hat\beta_n x_i)^2. \tag{3.6}$$

Given $0 < \eta < \frac{1}{2}$, let $K_\eta$ be the upper $100(1-\eta)$-percentile of the standard normal distribution. For any $d > 0$, define the stopping rule

$$N(d) = \inf\{n | v_n \le d^2 n\}, \tag{3.7}$$

where

$$v_n = K_\eta^2 b^2 \hat\sigma_n^2/(2b\hat\beta_n - 1).$$

As an application of Theorem 1, we obtain:

THEOREM 2. *Suppose that* (3.1)–(3.4) *are satisfied and that* $M$ *is continuously differentiable in some neighborhood of* $\theta$. *Consider the stochastic approxi-*

*mation scheme* (3.5) *with* $2b > 1$. *Define* $\hat{\alpha}_n, \hat{\beta}_n, \hat{\sigma}_n$ *by* (3.6) *and* $N(d)$ *by* (3.7). *Then*

$$(3.8)\quad \begin{aligned} &(1+\theta^2)\lambda_{\min}\left(\sum_1^n \mathbf{x}_i\mathbf{x}_i'\right) \sim \sum_1^n (x_i - \bar{x}_n)^2 \sim (\log n)\sigma^2 b^2/(2b\beta - 1) \quad a.s., \\ &\lambda_{\max}\left(\sum_1^n \mathbf{x}_i\mathbf{x}_i'\right) \sim (1+\theta^2)n \quad a.s.; \end{aligned}$$

$$(3.9)\qquad \hat{\alpha}_n \to -\beta\theta \quad a.s. \quad and \quad \hat{\beta}_n \to \beta \quad a.s.;$$

$$(3.10)\qquad \hat{\sigma}_n^2 \to \sigma^2 \quad a.s.;$$

$$(3.11)\qquad \lim_{d\to\infty} d^2N(d) = K_\eta^2 b^2\sigma^2/(2b\beta - 1) \quad a.s.;$$

$$(3.12)\qquad \lim_{d\to 0} P\{|x_{N(d)} - \theta| \le d\} = 1 - 2\eta.$$

REMARKS. (a) Sielken (1973) proposed a stopping rule for the stochastic approximation scheme (3.5). He used strongly consistent estimates for $\beta$ and $\sigma^2$ which were proposed by Burkholder (1956). The estimates require that at the $n$th step, an observation be taken on $y(x_n)$ and $y(x_n + j_n)$, where $\{j_n\}$ is a sequence of positive constants such that $j_n n^\lambda$ converges to a positive limit for some $0 < \lambda < \frac{1}{2}$. Specifically, the estimates $\tilde{\beta}_n$ and $\tilde{\sigma}_n^2$ of $\beta$ and $\sigma^2$ are defined by

$$\tilde{\beta}_n = n^{-1}\sum_1^n \{y(x_n + j_n) - y(x_n)\}/j_n$$

and

$$\tilde{\sigma}_n^2 = \tfrac{1}{2}\left\{n^{-1}\sum_1^n \left[y^2(x_n) + y^2(x_n + j_n)\right]\right\}.$$

(b) Stroup and Braun (1982) argued that Burkholder's variance estimate is distorted by $y(x_n + j_n)$ and proposed a modified least-squares estimate for $\beta$ and an unbiased estimate for $\sigma^2$. The estimates require that at the $n$th step, two independent observations $y_{n,1}$ and $y_{n,2}$ be taken at the same level $x_n$. The estimates $\beta_n$ and $\sigma_n^2$ of $\beta$ and $\sigma^2$ are defined by

$$\beta_n = \sum_{i=1}^n \bar{y}_i(x_i - \bar{x}_n) \Big/ \sum_{i=1}^n (x_i - \bar{x}_n)^2$$

and

$$\sigma_n^2 = n^{-1}\sum_{i=1}^n \left[(y_{i,1} - \bar{y}_i)^2 + (y_{i,2} - \bar{y}_i)^2\right],$$

where $\bar{y}_i = (y_{i,1} + y_{i,2})/2$ and $\bar{x}_n = n^{-1}\Sigma_1^n x_i$.

(c) Both estimates in (a) and (b) require two observations at each step, which is not required in our procedure. Although the variance estimate of Stroup and

Braun is unbiased, the squared error of their estimate is larger than ours asymptotically. More specifically, assume $\tau = E|\varepsilon_1^2 - \sigma^2|^2 < \infty$. Then

$$\lim|\hat{\sigma}_n^2 - \sigma^2|^2 = \tau/2 \quad \text{a.s.},$$

and

$$\lim|\sigma_n^2 - \sigma^2|^2 = \tau/2 + 4\sigma^4 \quad \text{a.s.}$$

(d) The crucial step for establishing (3.11) and (3.12) is to show that the estimates of $\beta$ and $\sigma$ are strongly consistent. Applying a theorem due to Lai and Robbins, which is a particular case of Theorem A, Stroup and Braun (1982) claim that $\beta_n$ is strongly consistent because

$$\sum_1^n (x_i - \bar{x}_n)^2/\log n \to \infty \quad \text{a.s.} \tag{3.13}$$

However, (3.13) contradicts (3.8). The usefulness of Theorem 1 for such settings is clear. [See also Stroup and Braun (1984).]

(e) Instead of using stopping rules similar to (3.7), Stroup and Braun (1982) proposed a new stopping rule defined by

$$N_k = \inf\{n | u_n(k) < c\}$$

where $u_n(k) = \sum_{i=1}^k \bar{y}_{n-k+i}^2 / k\sigma_n^2$ and $c > 0$ a constant. Their results rely on their Theorem 2 which requires the restrictive assumption that $\varepsilon_i$ are normal random variables. [Cf., Stroup and Braun (1984).]

Before proving Theorem 2, we need two lemmas. The first one is easy and its proof is omitted.

LEMMA 3. *If* $\lim_{n\to\infty} \sum_1^n a_i^2/\log n = c > 0$, *then*

$$\sum_1^n a_i = o\left((n \log n)^{1/2}\right).$$

LEMMA 4. *Suppose that in the regression model* (1.1) *and* (1.2) *hold. Suppose that there is a nonsingular* $\mathbf{A}$ *such that* $\mathbf{z}_n = \mathbf{A}\mathbf{x}_n$ *satisfy* (1.10). *Let* $\hat{\sigma}_n^2 = n^{-1}\sum_{i=1}^n (y_i - \mathbf{b}_n'\mathbf{x}_i)^2$. *If* $\lim_{n\to\infty} E(\varepsilon_n^2 | \mathscr{F}_{n-1}) = \sigma^2$ *a.s. for some* $\sigma > 0$ *and*

$$\lambda_{\min}(n) \to \infty \quad \text{and} \quad \{\log \lambda_{\max}(n)\}^{2\delta} = o(n) \quad a.s. \tag{3.14}$$

*for some* $\delta > 1/min(\gamma, 4)$, *then*

$$\hat{\sigma}_n^2 \to \sigma^2 \quad a.s. \tag{3.15}$$

*Moreover, if* $\mathbf{z}_n$ *further satisfies* (1.8) *and*

$$\lambda_{\min}(n) \to \infty \quad \text{and} \quad \log\log \lambda_{\max}(n) = o(n) \quad a.s., \tag{3.16}$$

*then* (3.15) *holds.*

PROOF. Apply identity

$$\hat{\sigma}_n^2 = n^{-1}\left(\sum_1^n \varepsilon_i^2\right) - n^{-1}(\mathbf{b}_n - \boldsymbol{\beta})'\left(\sum_1^n \mathbf{x}_i\mathbf{x}_i'\right)(\mathbf{b}_n - \boldsymbol{\beta})$$

and a similar argument as that in the proof of Lemma 3 of Lai and Wei (1982). □

PROOF OF THEOREM 2. We only have to show (3.8)–(3.10), since (3.11) and (3.12) follow easily from these and Sielken's (1973) arguments. Let $\xi^2 = \sigma^2 b^2/(2b\beta - 1)$. By Theorem 4 of Lai and Robbins (1979), it follows that

$$\lim_{n\to\infty} \sum_1^n (x_i - \theta)^2/\log n = \xi^2 \quad \text{a.s.,} \tag{3.17}$$

and

$$\limsup_{n\to\infty} |x_n - \theta|(n/2\log\log n)^{1/2} = |\xi| \quad \text{a.s.} \tag{3.18}$$

Hence

$$\begin{aligned} \sum_1^n (x_i - \bar{x}_n)^2 &= \sum_1^n (x_i - \theta)^2 - n(\bar{x}_n - \theta)^2 \\ &= \sum_1^n (x_i - \theta)^2 + O(\log\log n) \\ &\sim (\log n)\xi^2 \quad \text{a.s.} \end{aligned} \tag{3.19}$$

Now

$$\begin{aligned} \lambda_{\max}(n) + \lambda_{\min}(n) &= n + \sum_1^n x_i^2 = n(1 + \bar{x}_n^2) + \sum_1^n (x_i - \bar{x}_n)^2, \\ \lambda_{\max}(n) \cdot \lambda_{\min}(n) &= n\sum_1^n x_i^2 - \left(\sum_1^n x_i\right)^2 = n\sum_1^n (x_i - \bar{x}_n)^2. \end{aligned} \tag{3.20}$$

Since $\bar{x}_n \to \theta$ a.s., $\lambda_{\min}(n) \sim (1 + \theta^2)^{-1}\sum_1^n (x_i - \bar{x}_n)^2$ a.s. and $\lambda_{\max}(n) \sim n(1 + \theta^2)$ a.s. This completes the proof of (3.8). For the proof of (3.9), we claim that (3.9) is equivalent to the strong consistency of the least-squares estimate $(\tilde{\alpha}_n, \tilde{\beta}_n)$ in the model

$$y_i' = \alpha + \beta x_i + \varepsilon_i \tag{3.21}$$

where $\alpha = -\beta\theta$. We show this by proving that

$$|\hat{\alpha}_n - \tilde{\alpha}_n| + |\hat{\beta}_n - \tilde{\beta}_n| \to 0 \quad \text{a.s.} \tag{3.22}$$

Let $g(x) = M(x) - (\alpha + \beta x)$. Then

$$\hat{\beta}_n - \tilde{\beta}_n = \sum_1^n (x_i - \bar{x}_n) g(x_i) \Big/ \sum_1^n (x_i - \bar{x}_n)^2,$$

and

$$\hat{\alpha}_n - \tilde{\alpha}_n = \overline{g(x_n)} + (\hat{\beta}_n - \hat{\beta}_n)\bar{x}_n. \tag{3.23}$$

Lai and Robbins (1981, page 338) have already shown that

$$\sum_1^n (x_i - \bar{x}_n) g(x_i) = o\left(\sum_1^n (x_i - \bar{x}_n)^2\right) \quad \text{a.s.}$$

Hence $\hat{\beta}_n - \tilde{\beta}_n = o(1)$ a.s. and consequently by (3.23) and the fact that $g(x_n) \to 0$ a.s., $\hat{\alpha}_n - \tilde{\alpha}_n = o(1)$ a.s. This completes the proof of (3.22).

Now let

$$\mathbf{A} = \begin{pmatrix} 1 & 0 \\ -\theta & 1 \end{pmatrix}, \qquad \mathbf{z}_n = \mathbf{A}\mathbf{x}_n = \begin{pmatrix} 1 \\ x_n - \theta \end{pmatrix}$$

and

$$\mathbf{D}_n^2 = \begin{pmatrix} n & 0 \\ 0 & \sum_1^n (x_i - \theta)^2 \end{pmatrix}.$$

By (3.17) and Lemma 3,

$$\mathbf{R}_n = \mathbf{D}_n^{-1}\left(\sum_1^n \mathbf{z}_i \mathbf{z}_i'\right)\mathbf{D}_n^{-1} \to \mathbf{I}_2 \quad \text{a.s.}$$

Thus (1.6) is satisfied. By (3.8), (1.7) is satisfied with $\delta = 5/12$ and $\gamma = 3$. By Theorem 1 and (3.22), (3.9) is proved. Using a similar reduction, we can prove (3.10) by applying Lemma 4. □

**Acknowledgment.** The author would like to thank his colleague P. Smith for some helpful discussions.

## REFERENCES

BLUM, J. (1954). Approximation methods which converge with probability one. *Ann. Math. Statist.* **25** 382–386.

BURKHOLDER, D. (1956). On a class of stochastic approximation processes. *Ann. Math. Statist.* **27** 1044–1059.

CHOW, Y. S. (1965). Local convergence of martingales and the law of large numbers. *Ann. Math. Statist.* **36** 552–558.

HALL, P. and HEYDE, C. C. (1980). *Martingale Limit Theorem and Its Application.* Academic Press, New York.

JAIN, N. C., JOGDEO, K. and STOUT, W. (1975). Upper and lower functions for martingales and mixing processes. *Ann. Probab.* **3** 119–145.

LAI, T. L. and ROBBINS, H. (1979). Adaptive design and stochastic approximation. *Ann. Statist.* **7** 1196–1221.

LAI, T. L. and ROBBINS, H. (1981). Consistency and asymptotic efficiency of slope estimates in stochastic approximation schemes. *Z. Wahrsch. Verw. Gebiete.* **56** 329–360.

LAI, T. L. and WEI, C. Z. (1982). Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. *Ann. Statist.* **10** 154–166.

LAI, T. L. and WEI, C. Z. (1985). Asymptotically efficient self-tuning regulators. Technical Report, Department of Statistics, Columbia University.

ROBBINS, H. and MONRO, S. (1951). A stochastic approximation method. *Ann. Math. Statist.* **22** 400–407.

SACKS, J. (1958). Asymptotic distribution of stochastic approximation procedures. *Ann. Math. Statist.* **29** 373–405.

SIELKEN, R. L. (1973). Stopping times for stochastic approximation procedures. *Z. Wahrsch. Verw. Gebiete.* **27** 79–86.

STRASSEN, V. (1966). Almost sure behaviour of sums of independent random variables and martingales. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* **2** 315–343.

STROUP, D. F. and BRAUN, H. I. (1982). On a new stopping rule for stochastic approximation. *Z. Wahrsch. Verw. Gebiete.* **60** 535–556.

STROUP, D. F. and BRAUN, H. I. (1984). Correction to "On a new stopping rule for stochastic approximation." *Z. Wahrsch. Verw. Gebiete.* **67** 237.

STATISTICS PROGRAM
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF MARYLAND
COLLEGE PARK, MARYLAND 20742