The Annals of Statistics
1982, Vol. 10, No. 1, 154–166

LEAST SQUARES ESTIMATES IN STOCHASTIC REGRESSION MODELS WITH APPLICATIONS TO IDENTIFICATION AND CONTROL OF DYNAMIC SYSTEMS

BY TZE LEUNG LAI[1] AND CHING ZONG WEI[2]

Columbia University and University of Maryland

Strong consistency and asymptotic normality of least squares estimates in stochastic regression models are established under certain weak assumptions on the stochastic regressors and errors. We discuss applications of these results to interval estimation of the regression parameters and to recursive on-line identification and control schemes for linear dynamic systems.

1. Introduction. Consider the multiple regression model

$$y_n = \beta_1 x_{n1} + \cdots + \beta_{p1} x_{np} + \varepsilon_n, \quad n = 1, 2, \cdots \tag{1.1}$$

where the ε_n are unobservable random errors, $\beta_1, \cdots, \beta_p$ are unknown parameters, and y_n is the observed response corresponding to the design levels $x_{n1}, \cdots, x_{np}$. Let $\mathbf{x}_n = (x_{n1}, \cdots, x_{np})'$ and let $\mathbf{X}_n = (x_{ij})_{1\le i\le n, 1\le j\le p}$, $\mathbf{Y}_n = (y_1, \cdots, y_n)'$. Then

$$\mathbf{b}_n = (b_{n1}, \cdots, b_{np})' = (\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{X}_n'\mathbf{Y}_n \tag{1.2}$$

denotes the least squares estimate of $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)'$ based on the observations $\mathbf{x}_1, y_1, \cdots, \mathbf{x}_n, y_n$, assuming that $\mathbf{X}_n\mathbf{X}_n'$ is nonsingular. Throughout the sequel we shall assume that $\{\varepsilon_n\}$ is a martingale difference sequence with respect to an increasing sequence of σ-fields $\{\mathscr{F}_n\}$; i.e., ε_n is $\mathscr{F}_n$-measurable and $E(\varepsilon_n \mid \mathscr{F}_{n-1}) = 0$ for every n. An important example is the case where the ε_n are independent random variables with zero means.

While the statistical properties of the least squares estimate $\mathbf{b}_n$ are relatively well understood in the case where the design levels x_{ij} are non-random constants, there is a much less definitive theory for the case where the $\mathbf{x}_n$ are sequentially determined random vectors. Examples of stochastic regressors $\mathbf{x}_n$ arise in time series models, dynamic input-output systems, adaptive stochastic approximation schemes, stochastic control and other applications. In Section 3 below, we consider some of these applications.

An important feature in these applications is that the design vector $\mathbf{x}_n$ at stage n depends on the previous observations $\mathbf{x}_1, y_1, \cdots, \mathbf{x}_{n-1}, y_{n-1}$; i.e., $\mathbf{x}_n$ is $\mathscr{F}_{n-1}$-measurable. This in turn implies that $\{\sum_1^n \mathbf{x}_i\varepsilon_i, \mathscr{F}_n, n \ge 1\}$ is a martingale transform since $\{\varepsilon_n\}$ is a martingale difference sequence with respect to $\{\mathscr{F}_n\}$. Noting that $\mathbf{X}_n'\mathbf{X}_n = \sum_1^n \mathbf{x}_i\mathbf{x}_i'$ and that

$$\mathbf{b}_n = \boldsymbol{\beta} + (\textstyle\sum_1^n \mathbf{x}_i\mathbf{x}_i')^{-1} \sum_1^n \mathbf{x}_i\varepsilon_i, \tag{1.3}$$

the statistical properties of the least squares estimate $\mathbf{b}_n$ are related to the martingale transform $\sum_1^n \mathbf{x}_i\varepsilon_i$ and the random matrix $\sum_1^n \mathbf{x}_i\mathbf{x}_i'$. When $E(\varepsilon_n^2 \mid \mathscr{F}_{n-1}) = \sigma^2$ and $E(\mathbf{x}_n'\mathbf{x}_n) < \infty$ for all n, the random matrix $\sigma^2 \sum_1^n \mathbf{x}_i\mathbf{x}_i'$ is simply the conditional covariance of the martingale transform $\sum_1^n \mathbf{x}_i\varepsilon_i$, i.e.,

$$\textstyle\sum_1^n E\{(\mathbf{x}_i\varepsilon_i)(\mathbf{x}_i\varepsilon_i)' \mid \mathscr{F}_{i-1}\} = \sigma^2 \sum_1^n \mathbf{x}_i\mathbf{x}_i'.$$

Received June 1981; revised August 1981.

[1] Research supported by the National Science Foundation and the National Institute of Health.

[2] Research supported by the National Science Foundation.

AMS 1970 *subject classifications.* Primary 62J05, 62M10; secondary 60F15, 60G45, 93B30, 93C40.

Key words and phrases. Stochastic regressors, least squares, system identification, adaptive control, dynamic models, strong consistency, asymptotic normality, martingales.

In recent years there has been considerable interest in the question of strong consistency of the least squares estimate $\mathbf{b}_n$ in stochastic regression models, both in the statistical and in the engineering literature (cf. Anderson and Taylor, 1979; Christopeit and Helmes, 1980; Drygas, 1976; Lai and Robbins, 1979, 1981; Ljung, 1976, 1977; and Moore, 1978). In engineering applications, the least squares estimate $\mathbf{b}_n$ is often used in its recursive form of the Kalman filter type, viz.,

$$\mathbf{b}_{n+1} = \mathbf{b}_n + \{(y_{n+1} - \mathbf{x}'_{n+1}\mathbf{b}_n)/(1 + \mathbf{x}'_{n+1}\mathbf{V}_n\mathbf{x}_{n+1})\}\mathbf{V}_n\mathbf{x}_{n+1}, \tag{1.4a}$$

$$\mathbf{V}_{n+1} = \mathbf{V}_n - \mathbf{V}_n\mathbf{x}_{n+1}\mathbf{x}'_{n+1}\mathbf{V}_n/(1 + \mathbf{x}'_{n+1}\mathbf{V}_n\mathbf{x}_{n+1}), \tag{1.4b}$$

for the recursive on-line identification of dynamic systems; cf. Goodwin and Payne (1977), and Ljung (1977). The matrix $\mathbf{V}_n$ in the above recursion is equal to $(\mathbf{X}'_n\mathbf{X}_n)^{-1}$. This recursive representation of $\mathbf{b}_n$ provides a simple algorithm for successively updating the least squares estimate, and the problem of convergence of the recursive scheme to the true parameter $\boldsymbol{\beta}$ is equivalent to the strong consistency problem for $\mathbf{b}_n$.

In adaptive control systems, the sequentially updated least squares estimate $\mathbf{b}_n$ is often used to decide on the input at the next stage. The underlying idea here is that in many stochastic control problems, the optimal controller has a simple recursive form when the parameter $\boldsymbol{\beta}$ of the system is known; cf. Åström (1970), Box and Jenkins (1970), and Goodwin and Payne (1977). Replacing $\boldsymbol{\beta}$ by the least squares estimate $\mathbf{b}_n$ in the optimal controller at each stage n, one may hope that the performance of such an adaptive controller approaches that of the optimal controller assuming known $\boldsymbol{\beta}$ if $\mathbf{b}_n$ should converge to $\boldsymbol{\beta}$ with probability 1. Thus, the strong consistency of $\mathbf{b}_n$ is of basic interest in these applications, as will be illustrated in Section 3 below.

For the case where the design levels x_{ij} are nonrandom constants, the strong consistency of $\mathbf{b}_n$ was recently established by Lai, Robbins and Wei (1978, 1979) under the minimal assumption

$$(\mathbf{X}'_n\mathbf{X}_n)^{-1} \to \mathbf{0} \tag{1.5}$$

on the design constants x_{ij} when the errors ε_n form a martingale difference sequence such that $\sup_n E(\varepsilon_n^2 \mid \mathscr{F}_{n-1}) < \infty$ a.s. In particular, if $\varepsilon_1, \varepsilon_2, \cdots$ are i.i.d. with zero mean and variance $\sigma^2 > 0$, then (1.5) is both necessary and sufficient for the strong consistency of $\mathbf{b}_n$ in the fixed design case (i.e., where the x_{ij} are non-random constants).

Throughout the sequel we shall let $\lambda_{\max}(n)$ denote the maximum eigenvalue of $\mathbf{X}'_n\mathbf{X}_n$ and $\lambda_{\min}(n)$ denote the minimum eigenvalue of $\mathbf{X}'_n\mathbf{X}_n$. The condition (1.5) is equivalent to $\lambda_{\min}(n) \to \infty$. While this condition will be shown in Section 2 to be not sufficient for the strong consistency of $\mathbf{b}_n$ when $\mathbf{x}_n$ are sequentially determined random vectors, the following theorem establishes the strong consistency of $\mathbf{b}_n$ in stochastic regression models under the slightly stronger assumption that $\lambda_{\min}(n)$ tends to infinity faster than $\log \lambda_{\max}(n)$.

THEOREM 1. *Suppose that in the regression model* (1.1), $\{\varepsilon_n\}$ *is a martingale difference sequence with respect to an increasing sequence of σ-fields* $\{\mathscr{F}_n\}$ *such that*

$$\sup_n E(|\varepsilon_n|^\alpha \mid \mathscr{F}_{n-1}) < \infty \quad \text{a.s.} \quad \textit{for some } \alpha > 2. \tag{1.6}$$

Moreover, assume that the design levels $x_{n1}, \cdots, x_{np}$ *at stage n are* $\mathscr{F}_{n-1}$*-measurable random variables such that*

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

Then the least squares estimate $\mathbf{b}_n$ *converges* a.s. *to* $\boldsymbol{\beta}$; *in fact,*

$$\max_j |b_{nj} - \beta_j| = O(\{(\log \lambda_{\max}(n))/\lambda_{\min}(n)\}^{1/2}) \quad \text{a.s.} \tag{1.8}$$

The condition (1.7) on the stochastic regressors is in some sense the weakest possible. In Section 2 we give an example in which the design levels $x_{n1} \cdots, x_{np}$ are $\mathscr{F}_{n-1}$-measurable such that $\log \lambda_{\max}(n)/\lambda_{\min}(n)$ converges a.s. to a finite nonzero limit but $\mathbf{b}_n$ fails to be strongly consistent. Recently, assuming that $\lambda_{\min}(n) \to \infty$ a.s., Anderson and Taylor (1979) established the strong consistency of $\mathbf{b}_n$ under the condition $\lambda_{\max}(n) = O(\lambda_{\min}((n))$ a.s., while Christopeit and Helmes (1980) weakened the Anderson-Taylor condition to $(\lambda_{\max}(n))^r = O(\lambda_{\min}(n))$ a.s. for some $r > \frac{1}{2}$. Theorem 1 therefore provides a substantial improvement of their results, and its proof, which will be given in Section 2, involves ideas very different from theirs.

In Section 3 we apply Theorem 1 to some problems in the literature on the identification and control of linear dynamic systems. Another related application of the strong consistency of least squares estimates in stochastic regression models can be found in the recent work of Lai and Robbins (1979, 1981) on adaptive stochastic approximation schemes.

Questions of asymptotic normality of least squares estimates in stochastic regression models are discussed in Section 4, where we also consider related problems concerning confidence regions for $\boldsymbol{\beta}$.

2. A quadratic form and the proof of Theorem 1. Throughout the sequel, let $\|\mathbf{x}\|$ denote the Euclidean norm of a k-dimensional vector $\mathbf{x} = (x_1, \cdots, x_k)'$, i.e., $\|\mathbf{x}\|^2 = \sum_i^k x_i^2$. Viewing a $k \times k$ matrix $\mathbf{A}$ as a linear operator, we define $\|\mathbf{A}\| = \sup_{\|\mathbf{x}\|=1} \|\mathbf{Ax}\|$. Thus, $\|\mathbf{A}\|^2$ is equal to the maximum eigenvalue of $\mathbf{A}'\mathbf{A}$; cf. Rao (1973, page 62). As will be shown below, Theorem 1 follows easily from

LEMMA 1. *Let $\{\varepsilon_n\}$ be a martingale difference sequence with respect to an increasing sequence of σ-fields $\{\mathscr{F}_n\}$ such that*

$$\sup_n E(\varepsilon_n^2 \mid \mathscr{F}_{n-1}) < \infty \quad \text{a.s.} \tag{2.1}$$

Let $x_{n1}, \cdots, x_{np}$ be $\mathscr{F}_{n-1}$-measurable random variables for every n. Let $\mathbf{e}_n = (\varepsilon_1, \cdots, \varepsilon_n)'$, $\mathbf{X}_n = (x_{ij})_{1\le i\le n,\, 1\le j\le p}$, and define $N = \inf\{n : \mathbf{X}_n'\mathbf{X}_n$ is nonsingular$\}$; $\inf \phi = \infty$. Assume that $N < \infty$ a.s., and for $n \ge N$, define

$$Q_n = \mathbf{e}_n'\mathbf{X}_n(\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{X}_n'\mathbf{e}_n. \tag{2.2}$$

Let $\lambda_{\max}(n)$ denote the maximum eigenvalue of $\mathbf{X}_n'\mathbf{X}_n$. Then $\lambda_{\max}(n)$ is nondecreasing in n.

(i) *On $\{\lim_{n\to\infty} \lambda_{\max}(n) < \infty\}$ $Q_n = O(1)$ a.s.*

(ii) *On $\{\lim_{n\to\infty} \lambda_{\max}(n) = \infty\}$ we have for every $\delta > 0$*

$$Q_n = o((\log \lambda_{\max}(n))^{1+\delta}) \quad \text{a.s.} \tag{2.3}$$

(iii) *If the assumption* (2.1) *is replaced by the stronger assumption* (1.6), *then on $\{\lim_{n\to\infty} \lambda_{\max}(n) = \infty\}$ the conclusion* (2.3) *can be strengthened to*

$$Q_n = O(\log \lambda_{\max}(n)) \quad \text{a.s.} \tag{2.4}$$

REMARK. The quadratic form Q_n in $\varepsilon_1, \cdots, \varepsilon_n$ with coefficients given by the positive definite symmetric matrix $\mathbf{X}_n(\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{X}_n'$ as defined in (2.2) has the following geometric interpretation. Let $\hat{\mathbf{e}}_n$ denote the projection of $\mathbf{e}_n$ into the linear space spanned by the column vectors of $\mathbf{X}_n$. Then

$$Q_n = \|\hat{\mathbf{e}}_n\|^2. \tag{2.5}$$

The proof of Lemma 1 depends on a recursive representation of Q_n (see (2.16) later) and on the following lemma on martingale transforms and quadratic forms.

LEMMA 2. (i) *Let $\mathbf{B}$ be a $p \times p$ matrix and $\mathbf{w}$ be a $p \times 1$ vector. If $\mathbf{A} = \mathbf{B} + \mathbf{ww}'$ is nonsingular, then*

$$\mathbf{w}'\mathbf{A}^{-1}\mathbf{w} = (|\mathbf{A}| - |\mathbf{B}|)/|\mathbf{A}|.$$

(ii) *Let $\mathbf{w}_1, \mathbf{w}_2, \cdots$ be $p \times 1$ vectors and let $\mathbf{A}_n = \sum_1^n \mathbf{w}_i\mathbf{w}_i'$. Let λ_n^* denote the maximum eigenvalue of $\mathbf{A}_n$. Assume that $\mathbf{A}_N$ is nonsingular for some N. Then λ_n^* is nondecreasing* *and* $\mathbf{A}_n$ *is nonsingular for all* $n \geq N$. *Moreover, if* $\lim_{n\to\infty} \lambda_n^* < \infty$, *then* $\sum_{k=N}^{\infty} \mathbf{w}_k' \mathbf{A}_k^{-1}\mathbf{w}_k < \infty$. *On the other hand, if* $\lim_{n\to\infty} \lambda_n^* = \infty$, *then*

$$\sum_{k=N}^{n} \mathbf{w}_k'\mathbf{A}_k^{-1}\mathbf{w}_k = O(\log \lambda_n^*). \tag{2.6}$$

(iii) *Let* $\{\varepsilon_n\}$ *be a martingale difference sequence with respect to an increasing sequence of σ-fields* $\{\mathscr{F}_n\}$ *such that* $\sup_n E(\varepsilon_n^2 \mid \mathscr{F}_{n-1}) < \infty$ a.s. *Let* u_n *be an* $\mathscr{F}_{n-1}$*-measurable random variable for every n. Then*

$$\textstyle\sum_1^n u_i\varepsilon_i \text{ converges a.s. on } \{\sum_1^\infty u_i^2 < \infty\}, \tag{2.7}$$

$$\textstyle(\sum_1^n u_i\varepsilon_i)/\{(\sum_1^n u_i^2)^{1/2}[\log(\sum_1^n u_i^2)]^\eta\} \to 0 \text{ a.s. on } \{\sum_1^\infty u_i^2 = \infty\} \tag{2.8}$$

for every $\eta > 1/2$, *and consequently with probability* 1

$$\textstyle\sum_1^n u_i\varepsilon_i = o(\sum_1^n u_i^2) + O(1). \tag{2.9}$$

Moreover,

$$\textstyle\sum_1^\infty |u_i|\, \varepsilon_i^2 < \infty \text{ a.s. on } \{\sum_1^\infty |u_i| < \infty\}, \tag{2.10}$$

$$\textstyle(\sum_1^n |u_i|\varepsilon_i^2)/(\sum_1^n |u_i|)^\rho \to 0 \text{ a.s. on } \{\sum_1^\infty |u_i| = \infty\} \text{ for every } \rho > 1. \tag{2.11}$$

If (1.6) also holds, then (2.11) can be strengthened into

$$\textstyle\limsup_{n\to\infty} (\sum_1^n |u_i|\varepsilon_i^2)/(\sum_1^n |u_i|) < \infty \text{ a.s. on } \{\sup_n |u_n| < \infty, \sum_1^\infty |u_n| = \infty\}. \tag{2.12}$$

PROOF. (i) follows from the determinantal relation

$$|\mathbf{B}| = |\mathbf{A} - \mathbf{w}\mathbf{w}'| = |\mathbf{A}|(1 - \mathbf{w}'\mathbf{A}^{-1}\mathbf{w}).$$

To prove (ii), let λ_n denote the minimum eigenvalue of $\mathbf{A}_n$. Since $\mathbf{A}_n - \mathbf{A}_{n-1} = \mathbf{w}_n\mathbf{w}_n'$ is nonnegative definite, $\lambda_n^* \geq \lambda_{n-1}^*$, $\lambda_n \geq \lambda_{n-1}$, and $\mathbf{A}_n$ is nonsingular for $n \geq N$. By (i),

$$\textstyle\sum_{k=N}^{n} \mathbf{w}_k'\mathbf{A}_k^{-1}\mathbf{w}_k = \sum_{k=N}^{n} (|\mathbf{A}_k| - |\mathbf{A}_{k-1}|)/|\mathbf{A}_k|. \tag{2.13}$$

If $\lim_{n\to\infty} \lambda_n^* = \infty$, then for $n \geq N$, $|\mathbf{A}_n| \geq \lambda_N^{p-1} \lambda_n^* \to \infty$, and it follows from (2.13) that

$$\textstyle\sum_{k=N}^{n} \mathbf{w}_k'\mathbf{A}_k^{-1}\mathbf{w}_k = O(\log|\mathbf{A}_n|) = O(\log \lambda_n^*).$$

On the other hand, if $\lim_{n\to\infty} \lambda_n^* < \infty$, then it follows from (2.13) that

$$\textstyle\sum_{k=N}^{\infty} \mathbf{w}_k'\mathbf{A}_k^{-1}\mathbf{w}_k \leq \lambda_N^{-p} \sum_{k=N}^{\infty} (|\mathbf{A}_k| - |\mathbf{A}_{k-1}|) < \infty.$$

In (iii), $\{\sum_1^n u_i\varepsilon_i, \mathscr{F}_n, n \geq 1\}$ is a martingale transform although it need not be a martingale since $E(u_n\varepsilon_n)$ may be undefined. However, by choosing sufficiently large constants a_n such that $P[|u_n| > a_n] \leq n^{-2}$, we obtain that $P[u_n = u_n^*$ for all large $n] = 1$, where $u_n^* = u_n I_{[|u_n| \leq a_n]}$. Moreover, $E|u_n^*\varepsilon_n| < \infty$ and so $\sum_1^n u_i^*\varepsilon_i$ is a martingale. In view of this truncation argument, (2.7) and (2.8) follow from the local convergence theorem and the strong law for martingales (Chow, 1965); (2.10) follows from a theorem of Freedman (1973, page 919); while (2.11) follows by applying the Kronecker lemma and (2.10), since $\sum_1^\infty |u_i|/(\sum_1^i |u_j|)^\rho < \infty$. If (1.6) holds, then since $\sup_n E(|\varepsilon_n^2 - E(\varepsilon_n^2 \mid \mathscr{F}_{n-1})|^r \mid \mathscr{F}_{n-1}) < \infty$ a.s. for some $1 < r \leq 2$,

$$\textstyle\sum_1^n |u_i|\{\varepsilon_i^2 - E(\varepsilon_i^2 \mid \mathscr{F}_{i-1})\} = o(\sum_1^n |u_i|^r) = o(\sum_1^n |u_i|) \quad \text{a.s. on } \{\sup_n |u_n| < \infty, \sum_1^\infty |u_n| = \infty\}, \tag{2.14}$$

cf. Chow (1965). Since $\sum_1^n |u_i| E(\varepsilon_i^2 \mid \mathscr{F}_{i-1}) = O(\sum_1^n |u_i|^r)$ a.s. by (1.6), (2.12) follows from (2.14). □

PROOF OF LEMMA 1. Let $\mathbf{x}_n = (x_{n1}, \cdots, x_{np})'$. For $n \geq N$, let $\mathbf{V}_n = (\mathbf{X}_n'\mathbf{X}_n)^{-1} = (\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i')^{-1}$. Then $\mathbf{V}_{n+1}$ satisfies (1.4b), by the matrix inversion lemma (Goodwin and

Payne, 1977). By (1.4b), for $k > N$,

$$\mathbf{x}_k'\mathbf{V}_k = \mathbf{x}_k'\mathbf{V}_{k-1} - (1 + \mathbf{x}_k'\mathbf{V}_{k-1}\mathbf{x}_k)^{-1}(\mathbf{x}_k'\mathbf{V}_{k-1}\mathbf{x}_k)\mathbf{x}_k'\mathbf{V}_{k-1} = \mathbf{x}_k'\mathbf{V}_{k-1}/(1 + \mathbf{x}_k'\mathbf{V}_{k-1}\mathbf{x}_k). \tag{2.15}$$

This leads to the following recursion for Q_k. For $k > N$, by (2.2),

$$\begin{aligned} Q_k &= (\textstyle\sum_{i=1}^k \mathbf{x}_i'\varepsilon_i)\mathbf{V}_k(\sum_{i=1}^k \mathbf{x}_i\varepsilon_i) \\ &= (\textstyle\sum_{i=1}^{k-1} \mathbf{x}_i'\varepsilon_i)\mathbf{V}_k(\sum_{i=1}^{k-1}\mathbf{x}_i\varepsilon_i) + \mathbf{x}_k'\mathbf{V}_k\mathbf{x}_k\varepsilon_k^2 + 2(\mathbf{x}_k'\mathbf{V}_k \sum_{i=1}^{k-1} \mathbf{x}_i\varepsilon_i)\varepsilon_k \\ &= Q_{k-1} - (\mathbf{x}_k'\mathbf{V}_{k-1} \textstyle\sum_{i=1}^{k-1} \mathbf{x}_i\varepsilon_i)^2/(1 + \mathbf{x}_k'\mathbf{V}_{k-1}\mathbf{x}_k) + \mathbf{x}_k'\mathbf{V}_k\mathbf{x}_k\varepsilon_k^2 \\ &\quad + 2\{\mathbf{x}_k'\mathbf{V}_{k-1}(\textstyle\sum_{i=1}^{k-1} \mathbf{x}_i\varepsilon_i)\varepsilon_k\}/(1 + \mathbf{x}_k'\mathbf{V}_{k-1}\mathbf{x}_k), \end{aligned} \tag{2.16}$$

by (1.4b) and (2.15). Summing the recursion (2.16), we obtain that for $n > N$

$$\begin{aligned} &Q_n - Q_N + \textstyle\sum_{k=N+1}^n (\mathbf{x}_k'\mathbf{V}_{k-1} \sum_{i=1}^{k-1} \mathbf{x}_i\varepsilon_i)^2/(1 + \mathbf{x}_k'\mathbf{V}_{k-1}\mathbf{x}_k) \\ &\quad = \textstyle\sum_{k=N+1}^n \mathbf{x}_k'\mathbf{V}_k\mathbf{x}_k\varepsilon_k^2 + 2 \sum_{k=N+1}^n \{\mathbf{x}_k'\mathbf{V}_{k-1}(\sum_{i=1}^{k-1} \mathbf{x}_i\varepsilon_i)\varepsilon_k\}/(1 + \mathbf{x}_k'\mathbf{V}_{k-1}\mathbf{x}_k). \end{aligned} \tag{2.17}$$

Let $u_k = \mathbf{x}_k'\mathbf{V}_{k-1}(\sum_{i=1}^{k-1} \mathbf{x}_i\varepsilon_i)/(1 + \mathbf{x}_k'\mathbf{V}_{k-1}\mathbf{x}_k)$ if $k > N$ and set $u_k = 0$ if $k \le N$. Since u_k is $\mathscr{F}_{k-1}$-measurable, it follows from (2.9) and Lemma 2(iii) that with probability 1

$$\begin{aligned} &\textstyle\sum_{k=N+1}^n \{\mathbf{x}_k'\mathbf{V}_{k-1}(\sum_{i=1}^{k-1} \mathbf{x}_i\varepsilon_i)\varepsilon_k\}/(1 + \mathbf{x}_k'\mathbf{V}_{k-1}\mathbf{x}_k) \\ &\quad = o(\textstyle\sum_{k=N+1}^n (\mathbf{x}_k'\mathbf{V}_{k-1} \sum_{i=1}^{k-1} \mathbf{x}_i\varepsilon_i)^2/(1 + \mathbf{x}_k'\mathbf{V}_{k-1}\mathbf{x}_k)^2) + O(1). \end{aligned} \tag{2.18}$$

On $\{\lim_{n\to\infty} \lambda_{\max}(n) < \infty\}$ we obtain by Lemma 2(ii) that $\sum_{N+1}^\infty \mathbf{x}_k'\mathbf{V}_k\mathbf{x}_k < \infty$ a.s., and therefore by (2.10) of Lemma 2(iii), $\sum_{N+1}^\infty \mathbf{x}_k'\mathbf{V}_k\mathbf{x}_k\varepsilon_k^2 < \infty$. This together with (2.17) and (2.18) implies that

$$Q_n = O(1) \text{ a.s.} \quad \text{and} \tag{2.19}$$

$$\textstyle\sum_{k=N+1}^\infty (\mathbf{x}_k'\mathbf{V}_{k-1} \sum_{i=1}^{k-1} \mathbf{x}_i\varepsilon_i)^2/(1 + \mathbf{x}_k'\mathbf{V}_{k-1}\mathbf{x}_k) < \infty \quad \text{a.s.} \tag{2.20}$$

Now consider the event $\{\lim_{n\to\infty}\lambda_{\max}(n) = \infty\}$. Here we obtain by Lemma 2(ii) and (2.11) of Lemma 2(iii) that

$$\textstyle\sum_{k=N+1}^n \mathbf{x}_k'\mathbf{V}_k\mathbf{x}_k\varepsilon_k^2 = O((\sum_{N+1}^n \mathbf{x}_k'\mathbf{V}_k\mathbf{x}_k)^{1+\delta}) = O((\log \lambda_{\max}(n))^{1+\delta}) \quad \text{a.s.} \tag{2.21}$$

From (2.17), (2.18) and (2.21), it then follows that on $\{\lim_{n\to\infty}\lambda_{\max}(n) = \infty\}$

$$Q_n = O((\log \lambda_{\max}(n))^{1+\delta}) \quad \text{a.s., and} \tag{2.22}$$

$$\textstyle\sum_{k=N+1}^n (\mathbf{x}_k'\mathbf{V}_{k-1} \sum_{i=1}^{k-1} \mathbf{x}_i\varepsilon_i)^2/(1 + \mathbf{x}_k'\mathbf{V}_{k-1}\mathbf{x}_k) = O((\log \lambda_{\max}(n))^{1+\delta}) \quad \text{a.s.} \tag{2.23}$$

If (1.6) also holds, then noting that $\mathbf{x}_k'\mathbf{V}_k\mathbf{x}_k \le 1$ by Lemma 2(i), and using (2.12) of Lemma 2(iii) instead of (2.11), we can take $\delta = 0$ in (2.21), (2.22) and (2.23) and thereby obtain (2.4) in the event that $\lim_{n\to\infty}\lambda_{\max}(n) = \infty$. □

It is worth noting that in the preceding proof, we have also established the result in Corollary 1. This result is applied in the analysis by Lai and Wei (1981a) of the asymptotic properties of certain projections which are useful for studying the order of magnitude of $\lambda_{\min}(n)$ in many applications and which also provide an alternative proof and a refinement of Theorem 1; cf. Lai and Wei (1981a).

COROLLARY 1. *Let $\{\varepsilon_n\}$ be a martingale difference sequence with respect to an increasing sequence of σ-fields $\{\mathscr{F}_n\}$ such that* (2.1) *holds. Let $\mathbf{x}_n = (x_{n1}, \ldots, x_{np})'$ be an $\mathscr{F}_{n-1}$-measurable random vector for every n. Let $\mathbf{V}_n$ be the Moore-Penrose generalized inverse of $\sum_1^n \mathbf{x}_i\mathbf{x}_i'$, and let $\lambda_{\max}(n)$ denote the maximum eigenvalue of $\sum_1^n \mathbf{x}_i\mathbf{x}_i'$. Let $N = \inf\{n: \sum_1^n \mathbf{x}_i\mathbf{x}_i'$ is nonsingular$\}$;* $\inf \phi = \infty$.

(i) *On* $\{N < \infty$ *and* $\lim_{n\to\infty} \lambda_{\max}(n) < \infty\}$ (2.20) *holds.*

(ii) *On* $\{N < \infty$ *and* $\lim_{n\to\infty} \lambda_{\max}(n) = \infty\}$ (2.23) *holds for every* $\delta > 0$.

(iii) *If the assumption* (2.1) *is replaced by the stronger assumption* (1.6), *then* (2.23) *also holds with* $\delta = 0$ *on* $\{N < \infty, \lim_{n\to\infty}\lambda_{\max}(n) = \infty\}$.

The special case $p = 1$ in parts (i) and (ii) of Lemma 1 reduces to the convergence property (2.7) and the strong law (2.8) of martingale transforms in Lemma 2(iii). More importantly, Lemma 1(iii) in the special case $p = 1$ provides an improvement of (2.8) under the assumption (1.6). This is the content of

COROLLARY 2. *Let* $\{\varepsilon_n\}$ *be a martingale difference with respect to an increasing sequence of σ-fields* $\{\mathscr{F}_n\}$ *such that* (1.6) *holds. Let* u_n *be an* $\mathscr{F}_{n-1}$*-measurable random variable for every n. Then in the event* $\{\sum_1^\infty u_i^2 = \infty\}$,

$$\sum_1^n u_i\varepsilon_i = O(\{(\textstyle\sum_1^n u_i^2)\log(\sum_1^n u_i^2)\}^{1/2}) \quad \text{a.s.}$$

PROOF. Setting $\mathbf{X}_n = (u_i)_{1\le i\le n}$ in Lemma 1, we obtain that $\lambda_{\max}(n) = \sum_1^n u_i^2$ and

$$\mathbf{e}_n'\mathbf{X}_n(\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{X}_n'\mathbf{e}_n = (\textstyle\sum_1^n u_i\varepsilon_i)^2/(\sum_1^n u_i^2).$$

Hence the desired conclusion follows from Lemma 1(iii). □

PROOF OF THEOREM 1. Since $\lambda_{\min}(n) \to \infty$ a.s., $\mathbf{X}_n'\mathbf{X}_n$ is nonsingular for all large n with probability 1. Therefore

$$\begin{aligned}\|\mathbf{b}_n - \boldsymbol{\beta}\|^2 &= \|(\mathbf{X}_n'\mathbf{X}_n)^{-1}\textstyle\sum_{i=1}^n \mathbf{x}_i\varepsilon_i\|^2 \\ &\le \|(\mathbf{X}_n'\mathbf{X}_n)^{-1/2}\|^2\|(\mathbf{X}_n'\mathbf{X}_n)^{-1/2}\textstyle\sum_{i=1}^n \mathbf{x}_i\varepsilon_i\|^2 \\ &= (\lambda_{\min}(n))^{-1}\{\mathbf{e}_n'\mathbf{X}_n(\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{X}_n'\mathbf{e}_n\}.\end{aligned}$$

Hence from (1.7) and Lemma 1(iii), (1.8) follows. □

By applying Lemma 1(ii) instead of Lemma 1(iii) in the preceding proof, we can weaken the assumption (1.6) of Theorem 1 into (2.1) but at the expense of slightly strengthening the condition (1.7). This is the content of

COROLLARY 3. *Suppose that in the regression model* (1.1), $\{\varepsilon_n\}$ *is a martingale difference sequence with respect to an increasing sequence of σ-fields* $\{\mathscr{F}_n\}$ *such that* (2.1) *holds. Moreover, assume that the design levels* $x_{n1}, \dots, x_{np}$ *at stage n are* $\mathscr{F}_{n-1}$*-measurable random variables such that with probability* 1

$$\lambda_{\min}(n) \to \infty \quad \textit{and} \quad \{\log \lambda_{\max}(n)\}^{1+\delta} = o(\lambda_{\min}(n)) \quad \textit{for some} \quad \delta > 0. \tag{2.24}$$

Then the least squares estimate $\mathbf{b}_n$ *converges* a.s. *to* $\boldsymbol{\beta}$.

The following example of Lai and Robbins (1981) shows that the condition (1.7) of Theorem 1 is in some sense weakest possible.

EXAMPLE 1. Let $c \ne 0$ be a real constant and let $\varepsilon_1, \varepsilon_2, \cdots$ be i.i.d. random variables with $E\varepsilon_1 = 0$ and $E\varepsilon_1^2 = 1$. Let $\mathscr{F}_n$ be the σ-field generated by $\varepsilon_1, \dots, \varepsilon_n$. Consider the simple linear model

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i,$$

where the stochastic regressors x_i are defined inductively by

$$x_1 = 0, \qquad x_{n+1} = \bar{x}_n + c\bar{\varepsilon}_n, \qquad n \ge 1. \tag{2.25}$$

(The notation $\bar{a}_n$ denotes the arithmetic mean of n numbers $a_1, \dots, a_n$.) The random variable x_n is clearly $\mathscr{F}_{n-1}$-measurable. As shown by Lai and Robbins (1981), the least squares estimate b_{n2} of β_2 converges to $\beta_2 - c^{-1}(\neq\beta_2)$ a.s. Moreover, it can be shown by induction that

$$x_{n+1} = c \sum_1^n i^{-1}\varepsilon_i, \qquad n \geq 1. \tag{2.26}$$

By Kolmogorov's convergence theorem, $\sum_1^\infty i^{-1}\varepsilon_i$ converges a.s., and it therefore follows from (2.25) and (2.26) that

$$\bar{x}_n = x_{n+1} - c\bar{\varepsilon}_n \to c \sum_1^\infty i^{-1}\varepsilon_i \quad \text{a.s.} \tag{2.27}$$

Hence the least squares estimate

$$b_{n1} = \bar{y}_n - b_{n2}\bar{x}_n = \beta_1 + (\beta_2 - b_{n2})\bar{x}_n + \bar{\varepsilon}_n$$

converges to $\beta_1 + \sum_1^\infty i^{-1}\varepsilon_i$ a.s. We therefore have

$$b_{n1} \to \beta_1 + \sum_1^\infty i^{-1}\varepsilon_i \quad \text{a.s.}, \qquad b_{n2} \to \beta_2 - c^{-1} \quad \text{a.s.}; \tag{2.28}$$

i.e., the least squares estimates b_{n1}, b_{n2} of β_1, β_2 are both inconsistent. As shown by Lai and Robbins (1981),

$$\sum_1^n (x_i - \bar{x}_n)^2 \sim c^2 \log n \quad \text{a.s.} \tag{2.29}$$

Moreover, by (2.26),

$$\sum_1^n x_i^2 \sim nc^2 \left(\sum_1^\infty i^{-1}\varepsilon_i\right)^2 \quad \text{a.s.} \tag{2.30}$$

Making use of (2.29) and (2.30), it can be shown that the eigenvalues $\lambda_{\max}(n)$ and $\lambda_{\min}(n)$ of $\mathbf{X}_n'\mathbf{X}_n$ satisfy the asymptotic relations

$$\begin{aligned} \lambda_{\max}(n) &\sim n\{1 + c^2 (\textstyle\sum_1^\infty i^{-1}\varepsilon_i)^2\} \quad \text{a.s.}, \\ \lambda_{\min}(n) &\sim c^2(\log n)/\{1 + c^2 (\textstyle\sum_1^\infty i^{-1}\varepsilon_i)^2\} \quad \text{a.s.} \end{aligned} \tag{2.31}$$

Hence

$$\log \lambda_{\max}(n)/\lambda_{\min}(n) \to \{1 + c^2 (\textstyle\sum_1^\infty i^{-1}\varepsilon_i)^2\}/c^2 \quad \text{a.s.}$$

and condition (1.7) is only marginally violated.

3. Applications to identification and control of dynamic systems. In this section we consider the input-output dynamic model

$$y_n = \alpha_1 y_{n-1} + \cdots + \alpha_k y_{n-k} + \gamma_0'\mathbf{u}_n + \cdots + \gamma_h'\mathbf{u}_{n-h} + \varepsilon_n, \tag{3.1}$$

where the errors ε_n form a martingale difference sequence satisfying (2.1), $\mathbf{u}_n$ is the input vector and y_n is the output of the system at stage n. The column vectors γ_i and $\mathbf{u}_i$ in (3.1) are m-dimensional, while the y_i, α_i and ε_i are scalars. In a feedback system, the input $\mathbf{u}_n$ depends on the previous inputs and outputs $\mathbf{u}_j$, y_j, $j \leq n - 1$. Let

$$\boldsymbol{\beta} = (\alpha_1, \ldots, \alpha_k, \gamma_0', \ldots, \gamma_h')', \qquad \mathbf{x}_n = (y_{n-1}, \ldots, y_{n-k}, \mathbf{u}_n', \ldots, \mathbf{u}_{n-h}')'. \tag{3.2}$$

The least squares estimate $\mathbf{b}_n$ of $\boldsymbol{\beta}$ is commonly used in its recursive form (1.4) for the recursive on-line identification of the system (3.1). Motivated by systems with time-varying and adaptive feedback, Ljung (1976) examined the strong consistency problem of $\mathbf{b}_n$ under the assumption

$$\limsup_{n\to\infty} n^{-1} \sum_1^n (y_i^2 + \|\mathbf{u}_i\|^2) < \infty \quad \text{a.s.} \tag{3.3}$$

This assumption stems from the important case where the inputs satisfy $\|\mathbf{u}_n\|^2 = O(n)$ a.s. and the roots of the characteristic polynomial

$$\varphi(z) = z^k - \alpha_1 z^{k-1} - \cdots - \alpha_k \tag{3.4}$$

lie inside the unit circle. Example 1 in Section 2 shows, however, that the assumption (3.3) fails to ensure the strong consistency of $\mathbf{b}_n$. First note that Example 1 is a special case of (3.1) with $k = h = 0$, $\gamma_0 = (\beta_1, \beta_2)'$, $\mathbf{u}_n = (1, x_n)'$. Moreover, (2.30) and the fact that $n^{-1} \sum_1^n \varepsilon_i^2 \to 1$ a.s. imply that (3.3) holds. However, $\mathbf{b}_n$ in Example 1 is not strongly consistent. Ljung (1976, page 781) pointed out that $\mathbf{b}_n$ is strongly consistent in models satisfying (3.3) and the additional assumption that for some positive definite matrix Γ:

$$n^{-1} \textstyle\sum_1^n \mathbf{x}_i \mathbf{x}_i' \to \mathbf{\Gamma} \quad \text{a.s.} \tag{3.5}$$

The following theorem generalizes this result of Ljung by considerably weakening (3.3) and (3.5).

THEOREM 2. *Suppose that in the dynamic system* (3.1), $\{\varepsilon_n\}$ *is a martingale difference sequence with respect to an increasing sequence of σ-fields* $\{\mathscr{F}_n\}$ *such that* (2.1) *holds. Suppose that at every stage n the output* y_n *is* $\mathscr{F}_n$*-measurable, the input vector* $\mathbf{u}_n$ *is* $\mathscr{F}_{n-1}$*-measurable, and*

$$\|\mathbf{u}_n\| = O(n^r) \quad \text{a.s.} \quad \textit{for some } r > 0. \tag{3.6}$$

Define $\boldsymbol{\beta}$ *and* $\mathbf{x}_n$ *as in* (3.2). *Let* λ_n *denote the minimum eigenvalue of* $\sum_1^n \mathbf{x}_i \mathbf{x}_i'$.

(i) *If the roots* z_i *of the characteristic polynomial* $\varphi(z)$ *defined in* (3.4) *satisfy* $|z_i| \le 1$ *for* $i = 1, \cdots, k$, *then*

$$y_n = O(n^a) \quad \text{a.s.} \quad \textit{for some } a > 0. \tag{3.7}$$

(ii) *Assume that* (3.7) *holds and*

$$\lim_{n\to\infty} \lambda_n / (\log n)^\rho = \infty \quad \text{a.s.} \quad \textit{for some } \rho > 1, \tag{3.8}$$

then the least squares estimate $\mathbf{b}_n$ *of* $\boldsymbol{\beta}$ *is strongly consistent; in fact*

$$\|\mathbf{b}_n - \boldsymbol{\beta}\| = O(\{(\log n)^\rho / \lambda_n\}^{1/2}) \quad \text{a.s.} \tag{3.9}$$

(iii) *Assume that* (1.6) *and* (3.7) *hold, and replace* (3.8) *by the weaker assumption*

$$\lim_{n\to\infty} \lambda_n / (\log n) = \infty \quad \text{a.s.} \tag{3.10}$$

Then $\mathbf{b}_n$ *is still strongly consistent, and* (3.9) *can be strengthened into*

$$\|\mathbf{b}_n - \boldsymbol{\beta}\| = O(\{(\log n) / \lambda_n\}^{1/2}) \quad \text{a.s.} \tag{3.11}$$

PROOF. (ii) and (iii) follow from Theorem 1 and Corollary 2 since the maximum eigenvalue of $\sum_1^n \mathbf{x}_i \mathbf{x}_i'$ is majorized by $\operatorname{tr}(\sum_1^n \mathbf{x}_i \mathbf{x}_i')$, and (3.6) and (3.7) imply that $\operatorname{tr}(\sum_1^n \mathbf{x}_i \mathbf{x}_i') = O(n^n)$ a.s. for some $b > 0$. To prove (i), we note that

$$\textstyle\sum_1^\infty P[|\varepsilon_n| > n \mid \mathscr{F}_{n-1}] \le \sum_1^\infty n^{-2} E(\varepsilon_n^2 \mid \mathscr{F}_{n-1}) < \infty \quad \text{a.s.,} \qquad \text{by (2.1).}$$

Therefore by the conditional Borel-Cantelli lemma (Freedman, 1973),

$$\varepsilon_n = O(n) \quad \text{a.s.} \tag{3.12}$$

If $k = 0$, then (3.7) follows trivially from (3.1), (3.6) and (3.12). Now assume that $k \ge 1$. Define the k-dimensional vectors

$$\mathbf{y}_n = (y_n, \cdots, y_{n-k+1})', \qquad \mathbf{e}_n = (\gamma_0' \mathbf{u}_n + \cdots + \gamma_h' \mathbf{u}_{n-h} + \varepsilon_n, 0, \cdots, 0)'. \tag{3.13}$$

By (3.6) and (3.12),

$$\|\mathbf{e}_n\| = O(n^s) \quad \text{a.s.} \quad \text{for some } s > 0. \tag{3.14}$$

Consider the $k \times k$ matrix

$$\mathbf{A} = \begin{pmatrix} \alpha_1 \cdots \alpha_{k-1} & \alpha_k \\ \mathbf{I}_{k-1} & \mathbf{0} \end{pmatrix},$$

where $\mathbf{I}_p$ denotes the $p \times p$ identity matrix. By (3.1) and (3.13), $\mathbf{y}_n = \mathbf{A}\mathbf{y}_{n-1} + \mathbf{e}_n$ and therefore

(3.15) $$\mathbf{y}_n = \mathbf{A}^n \mathbf{y}_0 + \sum_{i=1}^n \mathbf{A}^{n-i} \mathbf{e}_i.$$

Express $\mathbf{A}$ in its Jordan form

(3.16) $$\mathbf{A} = \mathbf{C}\mathbf{D}\mathbf{C}^{-1},$$

where $\mathbf{D} = \operatorname{diag}\{\mathbf{D}_1, \cdots, \mathbf{D}_q\}$,

$$\mathbf{D}_j = \begin{pmatrix} z_j & 1 & 0 \cdots 0 \\ 0 & z_j & 1 \cdots 0 \\ \cdots & \cdots & \cdots \\ 0 & 0 & \cdots\cdots z_j \end{pmatrix} \text{ is an } m_j \times m_j \text{ matrix,}$$

z_j is a root of $\varphi(z)$ with multiplicity m_j, and $\mathbf{C}$ is a nonsingular matrix. Define

$$f_j(n, \nu) = \binom{n}{\nu} z_j^{n-\nu}, \qquad \nu = 0, 1, \cdots, \quad n = 1, 2, \cdots.$$

Then, since $|z_j| \le 1$,

(3.17) $$\|\mathbf{D}_j^n\| = \left\| \begin{pmatrix} f_j(n, 0) & f_j(n, 1) & \cdots f_j(n, m_j - 1) \\ 0 & f_j(n, 0) & \cdots f_j(n, m_j - 2) \\ \cdots & \cdots & \cdots \\ 0 & 0 & \cdots\cdots f_j(n, 0) \end{pmatrix} \right\| = O(n^{m_j - 1}).$$

Let $M = \max_j m_j$. By (3.16) and (3.17),

(3.18) $$\|\mathbf{A}^n\| = \|\mathbf{C}\mathbf{D}^n\mathbf{C}^{-1}\| \le \|\mathbf{C}\|\,\|\mathbf{C}^{-1}\| \sum_{j=1}^q \|\mathbf{D}_j^n\| = O(n^{M-1}).$$

From (3.14), (3.15) and (3.18), (3.7) follows. □

Some specific applications of Theorem 2 to system identification problems are given by Lai and Wei (1981a). They show in particular that if

(i) in the dynamic system (3.1), the roots of the characteristic polynomial $z^k - \alpha_1 z^{k-1} - \cdots - \alpha_k$ lie on or inside the unit circle,

(ii) the noise $\{\varepsilon_n\}$ is a martingale difference sequence such that (1.6) holds and $\liminf_{n\to\infty} E(\varepsilon_n^2 \mid \mathscr{F}_{n-1}) > 0$ a.s., and

(iii) the set of inputs $\{u_{n1}, \cdots, u_{nm}; n \ge 1\}$ is a set of independent random variables which are independent of the noise sequence $\{\varepsilon_n\}$ such that $Eu_{nj} \equiv 0$, $\inf_{j,n} Eu_{nj}^2 > 0$ and $\sup_{j,n} E|u_{nj}|^\alpha < \infty$ for some $\alpha > 2$,

then (3.6) and (3.8) both hold and therefore the least squares estimates of $\alpha_1, \cdots, \alpha_k, \gamma_0, \cdots, \gamma_h$ are strongly consistent by Theorem 2.

If the roots of the characteristic polynomial should all lie strictly inside the unit circle, then with the independent white noise inputs $\mathbf{u}_n$ as before, the system (3.1) is stable, and the aforementioned results of Ljung (1976), Anderson and Taylor (1979), and Christopeit and Helmes (1980) are related to such stable systems. By allowing the roots of the characteristic polynomial to lie on the unit circle as well, we can formulate unstable but non-explosive systems related to the ARIMA models of Box and Jenkins (1970). The following example shows that while the results of Anderson and Taylor (1979), Christopeit and Helmes (1980), Ljung (1976) are not applicable to the least squares identification method for such unstable systems, the conditions (3.6) and (3.8) of Theorem 2 are weak enough for these systems.

EXAMPLE 2. Consider the dynamic system

(3.19) $$y_n = \alpha y_{n-1} + \gamma u_n + \varepsilon_n, \qquad n \ge 1,$$

where the errors ε_n are i.i.d. random variables with mean 0 and variance $\sigma_1^2 > 0$, and the inputs u_n are also i.i.d. random variables independent of $\{\varepsilon_n, n \ge 1\}$ such that $Eu_n = 0$, $Eu_n^2 = \sigma_2^2 > 0$. Letting $\mathbf{x}_n = (y_{n-1}, u_n)'$, we have

$$\sum_1^n \mathbf{x}_i\mathbf{x}_i' = \begin{pmatrix} \sum_1^n y_{i-1}^2 & \sum_1^n y_{i-1}u_i \\ \sum_1^n y_{i-1}u_i & \sum_1^n u_i^2 \end{pmatrix}. \tag{3.20}$$

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

$$\lambda_{\max}(n) + \lambda_{\min}(n) = \sum_1^n y_{i-1}^2 + \sum_1^n u_i^2, \qquad \lambda_{\max}(n)\lambda_{\min}(n) = \det(\sum_1^n \mathbf{x}_i\mathbf{x}_i'). \tag{3.21}$$

Suppose that the unknown parameters α, γ actually take on the values $\alpha = 1$, $\gamma = 0$, so that the dynamic system (3.19) reduces to $y_n = y_{n-1} + \varepsilon_n$, or equivalently,

$$y_n = y_0 + S_n, \qquad \text{where} \quad S_n = \sum_1^n \varepsilon_i. \tag{3.22}$$

Making use of a theorem of Donsker and Varadhan (1977, page 751), it can be shown that

$$\liminf_{n\to\infty} n^{-2}(\log\log n) \sum_1^n y_{i-1}^2 = \liminf_{n\to\infty} n^{-2}(\log\log n) \sum_1^n S_{i-1}^2 = \sigma_1^2/4 \quad \text{a.s.} \tag{3.23}$$

On the other hand, by the law of the iterated logarithm,

$$\sum_1^n y_{i-1}^2 = O(n^2 \log\log n) \quad \text{a.s.} \tag{3.24}$$

Moreover, by (2.8) of Lemma 2(iii) and (3.24),

$$\sum_1^n y_{i-1}u_i = o((\sum_1^n y_i^2)^{1/2}\log(\sum_1^n y_i^2)) = o(n(\log n)^2) \quad \text{a.s.} \tag{3.25}$$

By the strong law of large numbers,

$$\sum_1^n u_i^2 \sim n\sigma_2^2 \quad \text{a.s.} \tag{3.26}$$

From (3.21) and (3.23)-(3.26), it then follows that

$$\lambda_{\max}(n) \sim \sum_1^n y_{i-1}^2 \quad \text{a.s.}, \qquad \lambda_{\min}(n) \sim n\sigma_2^2 \quad \text{a.s.} \tag{3.27}$$

In view of (3.26) and (3.27), conditions (3.6) and (3.8) of Theorem 3 are satisfied, and therefore the least squares estimates of α and γ are strongly consistent by Theorem 3. On the other hand, (3.23) shows that the condition (3.3) of Ljung (1976) is violated. Moreover, (3.23) and (3.27) imply that

$$\liminf_{n\to\infty} \{\lambda_{\max}(n)\log\log\lambda_{\max}(n)\}^{1/2}/\lambda_{\min}(n) = \sigma_1/2\sigma_2^2 \quad \text{a.s.},$$

violating the assumptions of Anderson and Taylor (1979, Theorem 1) and of Christopeit and Helmes (1980, Corollary 1).

Lai and Wei (1981b) apply Theorem 2 to the problem of adaptive control of the dynamic system (3.1) with scalar inputs when the errors ε_n are i.i.d. random variables with $E\varepsilon_n = 0$, $E\varepsilon_n^2 = \sigma^2 > 0$, and all roots of the characteristic polynomials $z^k - \alpha_1 z^{k-1} - \cdots - \alpha_k$ and $\gamma_0 z^h + \gamma_1 z^{h-1} + \cdots + \gamma_h$ lie inside the unit circle. To focus on the main ideas, we assume here that $\gamma_0 \neq 0$, and the details for the general case without such an assumption are given by Lai and Wei (1981b). If the parameters $\alpha_1, \cdots, \alpha_k, \gamma_0, \cdots, \gamma_h$ are known, then the input u_n which minimizes Ey_n^2 and which is based on the previous inputs and outputs $u_j, y_j, j \leq n$, is given by

$$u_n = -\{\alpha_1 y_{n-1} + \cdots + \alpha_k y_{n-k} + \gamma_1 u_{n-1} + \cdots + \gamma_h u_{n-h}\}/\gamma_0. \tag{3.28}$$

With this minimum variance controller u_n, $y_n = \varepsilon_n$ (Goodwin and Payne, 1977), and therefore

$$n^{-1}\sum_1^n y_i^2 \to \sigma^2 \quad \text{a.s.} \tag{3.29}$$

Moreover, (3.6) holds for such inputs u_n.

In ignorance of the parameters $\alpha_1, \cdots, \alpha_k, \gamma_0, \cdots, \gamma_h$, it is natural to try replacing them in (3.28) at each stage n by their least squares estimates based on past data. It is also natural to expect that (3.29) would still hold if the least squares estimates should in fact converge a.s. to the true parameters. On the other hand, if u_n should in fact be given by the minimum variance controller (3.28) with the parameters known, then the design matrix $\sum_1^n \mathbf{x}_i\mathbf{x}_i'$ would be singular since (3.28) implies that u_i is a linear combination of the other components of $\mathbf{x}_i = (y_{i-1}, \cdots, y_{i-k}, u_i, \cdots, u_{i-h})'$. However, Theorem 2 says that to have strong consistency of the least squares estimates of $\alpha_1, \cdots, \alpha_k, \gamma_0, \cdots, \gamma_h$, not only should $\sum_1^n \mathbf{x}_i\mathbf{x}_i'$ be nonsingular, but its minimum eigenvalue should grow faster than log n. This difficulty can be resolved by occasionally perturbing the system with white noise inputs to ensure that (3.8) holds, and Theorem 2 suggests that such white noise inputs need only be introduced in a negligible proportion, say $O((\log n)^4/n)$, of the first n stages as $n \to \infty$. Making use of this idea, Lai and Wei (1981b) obtain a class of adaptive controllers that satisfy (3.29).

4. Asymptotic normality of least squares estimates in stochastic regression models and related confidence ellipsoids. In this section we assume that the errors ε_n in the regression model (1.1) form a martingale difference sequence with respect to the σ-fields $\mathscr{F}_n$ such that (1.6) holds and

$$\lim_{n\to\infty} E(\varepsilon_n^2 \mid \mathscr{F}_{n-1}) = \sigma^2 \quad \text{a.s.} \quad \text{for some constant } \sigma. \tag{4.1}$$

An important special case is where the ε_n are independent random variables with zero means, variance σ^2, and $\sup_n E|\varepsilon_n|^\alpha < \infty$ for some $\alpha > 2$. The following theorem gives conditions on the stochastic regressors $\mathbf{x}_n$ that would ensure the asymptotic normality of the least squares estimate $\mathbf{b}_n$.

THEOREM 3. *Suppose that in the regression model* (1.1), $\{\varepsilon_n\}$ *is a martingale difference sequence with respect to an increasing sequence of σ-fields* $\{\mathscr{F}_n\}$ *such that* (1.6) *and* (4.1) *hold. Moreover, assume for each n that the design vector* $\mathbf{x}_n = (x_{n1}, \cdots, x_{np})'$ *at stage n is* $\mathscr{F}_{n-1}$*-measurable and that there exists a non-random positive definite symmetric matrix* $\mathbf{B}_n$ *for which*

$$\mathbf{B}_n^{-1}(\textstyle\sum_1^n \mathbf{x}_i\mathbf{x}_i')^{1/2} \to_P \mathbf{I}_p, \quad \text{and} \tag{4.2}$$

$$\max_{1\le i\le n} \| \mathbf{B}_n^{-1}\mathbf{x}_i \| \to_P 0. \tag{4.3}$$

Then the least squares estimate $\mathbf{b}_n$ *of* $\boldsymbol{\beta}$ *has an asymptotically normal distribution in the sense that*

$$(\textstyle\sum_1^n \mathbf{x}_i\mathbf{x}_i')^{1/2}(\mathbf{b}_n - \boldsymbol{\beta}) \to_D \mathbf{N}(\mathbf{0}, \sigma^2\mathbf{I}_p), \tag{4.4}$$

where $\to_D$ *denotes convergence in distribution.*

PROOF. In view of (1.3), (4.2) and the Cramér-Wold theorem, we need only show that for any non-random $p \times 1$ vector $\mathbf{c}$,

$$\mathbf{c}'\mathbf{B}_n^{-1}(\textstyle\sum_1^n \mathbf{x}_i\varepsilon_i) \to_D N(0, \sigma^2\|\mathbf{c}\|^2). \tag{4.5}$$

The desired conclusion (4.5) can be obtained by making use of a martingale central limit theorem of Dvoretzky (1972, Theorem 2.2) and an argument similar to that in Theorem 5 of Lai and Robbins (1981). □

The existence of a non-random matrix $\mathbf{B}_n$ satisfying condition (4.2) in Theorem 3 can be regarded as a stability assumption on the matrix $(\sum_1^n \mathbf{x}_i\mathbf{x}_i')^{1/2}$. Without making this assumption, we can replace condition (4.3) by

$$\max_{1\le i\le n} \mathbf{x}_i'(\textstyle\sum_{j=1}^n \mathbf{x}_j\mathbf{x}_j')^{-1}\mathbf{x}_i \to_P 0. \tag{4.6}$$

However, the following example shows that the asymptotic normality property (4.4) of $\mathbf{b}_n$ may fail to hold under (4.6) in the absence of the stability assumption (4.2).

EXAMPLE 3. Consider the autoregressive AR(1) process

$$y_n = \beta y_{n-1} + \varepsilon_n, \qquad n \ge 1, \tag{4.7}$$

where the errors ε_n are i.i.d. random variables with mean 0 and variance 1. The least squares estimate of β is

$$b_n = (\textstyle\sum_1^n y_i y_{i-1})/(\sum_1^n y_{i-1}^2) = \beta + (\sum_1^n y_{i-1}\varepsilon_i)/(\sum_1^n y_{i-1}^2). \tag{4.8}$$

Suppose that the unknown parameter β actually assumes the value 1, so that (4.7) reduces to $y_n = y_0 + S_n$, where $S_n = \sum_1^n \varepsilon_i$. Then $(\max_{1\le i\le n} y_i^2)/(\sum_1^n y_i^2) \to 0$ a.s. by (3.23) and the law of the iterated logarithm for S_n. On the other hand, by Donsker's invariance principle,

$$n^{-2} \textstyle\sum_1^n y_{i-1}^2 \to_D \displaystyle\int_0^1 w^2(t)\,dt, \tag{4.9}$$

and

$$(\textstyle\sum_1^n y_{i-1}^2)^{1/2}(b_n - \beta) \to_D \dfrac{1}{2}\{w^2(1) - 1\}\Big/\left\{\displaystyle\int_0^1 w^2(t)\,dt\right\}^{1/2}, \tag{4.10}$$

where $w(t)$ denotes the standard Wiener process (White, 1958, page 1196). Hence the asymptotic normality property (4.4) fails to hold for b_n.

The limiting normal distribution in (4.4) involves σ^2 which is usually unknown. Consistent estimation of σ^2 using the residual sum of squares is considered in the following.

LEMMA 3. *Suppose that in the regression model* (1.1), $\{\varepsilon_n\}$ *is a martingale difference sequence with respect to an increasing sequence of* σ-*fields* $\{\mathscr{F}_n\}$ *such that* (1.6) *and* (4.1) *hold. Moreover, assume that the design levels* $x_{n1}, \cdots, x_{np}$ *at stage n are* $\mathscr{F}_{n-1}$-*measurable random variables. Let* $\mathbf{X}_n = (x_{ij})_{1\le i\le n, 1\le j\le p}$. *Assume that* $N = \inf\{n \ge p : \mathbf{X}_n$ *is of rank* $p\}$ $< \infty$ a.s. *For* $n \ge N$, *let* $\mathbf{b}_n$ *be the least squares estimate of* $\boldsymbol{\beta}$ *defined by* (1.2), *and let*

$$\hat\sigma_n^2 = n^{-1} \textstyle\sum_1^n (y_i - b_{n1}x_{i1} - \cdots - b_{np}x_{ip})^2. \tag{4.11}$$

Let $\lambda_{\max}(n)$ *denote the maximum eigenvalue of* $\mathbf{X}_n'\mathbf{X}_n$. *If*

$$\lim_{n\to\infty}(\log \lambda_{\max}(n))/n = 0 \quad \text{a.s.}, \tag{4.12}$$

then $\hat\sigma_n^2 \to \sigma^2$ a.s.

PROOF. Let $\mathbf{e}_n = (\varepsilon_1, \cdots, \varepsilon_n)'$. Standard analysis of variance computations give the following expression for the residual sum of squares: For $n \ge N$,

$$\textstyle\sum_1^n (y_i - b_{n1}x_{i1} - \cdots - b_{np}x_{ip})^2 = \mathbf{e}_n'\mathbf{e}_n - \mathbf{e}_n'\mathbf{X}_n(\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{X}_n'\mathbf{e}_n. \tag{4.13}$$

By Lemma 1 (iii),

$$\mathbf{e}_n'\mathbf{X}_n(\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{X}_n'\mathbf{e}_n = O(\log \lambda_{\max}(n)) \quad \text{a.s.} \tag{4.14}$$

It follows from (1.6) that $\sum_1^n \{\varepsilon_i^2 - E(\varepsilon_i^2 \mid \mathscr{F}_{i-1})\} = o(n)$ a.s. (Chow, 1965), so

$$\mathbf{e}_n'\mathbf{e}_n = \textstyle\sum_1^n \varepsilon_i^2 = \sum_1^n E(\varepsilon_i^2 \mid \mathscr{F}_{i-1}) + o(n) \sim n\sigma^2 \quad \text{a.s.} \tag{4.15}$$

In view of (4.12), (4.13), (4.14) and (4.15), $\hat\sigma_n^2 \to \sigma^2$ a.s. □

Under the assumptions of Theorem 3 and the additional assumption (4.12), we note that (4.2) implies that

$$P\{\text{rank } (\textstyle\sum_1^n \mathbf{x}_i\mathbf{x}_i') = p \text{ for all large } n\} = 1,$$

and therefore by Theorem 3 and Lemma 3,

$$\lim_{n\to\infty} P\{\hat\sigma_n^{-2}(\mathbf{b}_n - \boldsymbol{\beta})'(\textstyle\sum_1^n \mathbf{x}_i\mathbf{x}_i')(\mathbf{b}_n - \boldsymbol{\beta}) \le u\} = P\{\chi^2(p) \le u\}, \qquad u > 0, \tag{4.16}$$

where $\chi^2(p)$ denotes the Chi squared distribution with p degrees of freedom. For $0 < \alpha < 1$, choosing u such that $P[\chi^2(p) \le u] = \alpha$, the relation (4.16) then provides an approximate α-level confidence ellipsoid for $\boldsymbol{\beta}$.

Making use of Theorem 3 to obtain the asymptotic normality of $\mathbf{b}_n$ and Lemma 3 to obtain the strong consistency of $\hat{\sigma}_n^2$, we can also extend the results of Albert (1966) and of Gleser (1965) on fixed size confidence ellipsoids for $\boldsymbol{\beta}$ in the case of non-random regressors $\mathbf{x}_n$ to the case where $\mathbf{x}_n$ is random but is $\mathscr{F}_{n-1}$-measurable.

REFERENCES

ALBERT, A. (1966). Fixed size confidence ellipsoids for linear regression parameters. *Ann. Math. Statist.* **37** 1602–1630.

ANDERSON, T. W. and TAYLOR, J. (1979). Strong consistency of least squares estimators in dynamic models. *Ann. Statist.* **7** 484–489.

ÅSTRÖM, K. J. (1970). *Introduction to Stochastic Control Theory.* Academic, New York.

BOX, G. E. P. and JENKINS, G. (1970). *Time Series Analysis, Forecasting and Control.* Holden-Day, San Francisco.

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

CHRISTOPEIT, N. and HELMES, K. (1980). Strong consistency of least squares estimators in linear regression models. *Ann. Statist.* **8** 778–788.

DONSKER, M. D. and VARADHAN, S. R. S. (1977). On laws of the iterated logarithm for local times. *Comm. Pure Appl. Math.* **30** 707–753.

DRYGAS, H. (1976). Weak and strong consistency of the least squares estimators in regression models. *Z. Wahrsch. verw. Gebiete* **34** 119–127.

DVORETZKY, A. (1972). Asymptotic normality for sums of dependent random variables. *Proc. Sixth Berkeley Symp. Math. Statist. Probability* **2** 513–535. Univ. California Press.

FREEDMAN, D. (1973). Another note on the Borel-Cantelli lemma and the strong law with the Poisson approximation as a by-product. *Ann. Probability* **1** 910–925.

GLESER, L. J. (1965). On the asymptotic theory of fixed-size confidence bounds for linear regression parameters. *Ann. Math. Statist.* **36** 463–467.

GOODWIN, G. C. and PAYNE, R. L. (1977). *Dynamic System Identification.* Academic, New York.

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., ROBBINS, H. and WEI, C. Z. (1978). Strong consistency of least squares estimates in multiple regression. *Proc. Nat. Acad. Sci. USA* **75** 3034–3036.

LAI, T. L., ROBBINS, H. and WEI, C. Z. (1979). Strong consistency of least squares estimates in multiple regression II. *J. Multivariate Anal.* **9** 343–361.

LAI, T. L. and WEI, C. Z. (1981a). Asymptotic properties of projections with applications to stochastic regression problems. *J. Multivariate Anal.*, to appear.

LAI, T. L. and WEI, C. Z. (1981b). Adaptive control of linear dynamic systems. Columbia Univ. Tech. Report.

LJUNG, L. (1976). Consistency of the least squares identification method. *IEEE Trans. Aut. Control* **AC-21** 779–781.

LJUNG, L. (1977). Analysis of recursive stochastic algorithms. *IEEE Trans. Aut. Control* **AC-22** 551–575.

MOORE, J. B. (1978). On strong consistency of least squares identification algorithms. *Automatica* **14** 505–509.

RAO, C. R. (1973). *Linear Statistical Inference and Its Applications, Second Edition.* Wiley, New York.

WHITE, J. S. (1958). The limiting distribution of the serial correlation coefficient in the explosive case. *Ann. Math. Statist.* **29** 1188–1197.

DEPARTMENT OF MATHEMATICAL STATISTICS
COLUMBIA UNIVERSITY
MATHEMATICS BUILDING
NEW YORK, N.Y. 10027

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF MARYLAND
COLLEGE PARK, MD. 20742