ADVANCES IN APPLIED MATHEMATICS 3, 50–73 (1982)

Iterated Least Squares in Multiperiod Control*

T. L. LAI AND HERBERT ROBBINS

Department of Mathematical Statistics, Columbia University, New York, New York 10027

1. INTRODUCTION

Consider the linear regression model

$$y_i = \alpha + \beta x_i + \epsilon_i, \qquad i = 1, 2, \ldots, \tag{1.1}$$

where α and β are unknown parameters and the errors $\epsilon_1, \epsilon_2, \ldots$ are independent and identically distributed (i.i.d.) random variables with mean 0 and variance σ^2. In the econometrics literature, the "multiperiod control problem" is to choose successive levels $x_1, \ldots, x_n$ in the model (1.1) so that the outputs $y_1, \ldots, y_n$ are as close as possible to a given target value y^*. Several authors have approached this problem from a Bayesian point of view, formulating it as the problem of minimizing

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E_{\alpha,\beta}\left[\sum_{i=1}^{n}(y_i - y^*)^2\right] d\pi(\alpha,\beta)$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left\{n\sigma^2 + \beta^2 E_{\alpha,\beta}\left[\sum_{i=1}^{n}(x_i - \theta)^2\right]\right\} d\pi(\alpha,\beta), \tag{1.2}$$

where π is a prior distribution of the unknown parameters α and β (cf. [15, 17]). However, because of the computational complexities in the numerical solution of the dynamic programming problems and the analytical difficulties in studying the properties of the Bayes rules, not much is known about the performance of these rules and it is difficult to implement them in practice.

A recent departure from the Bayesian approach is due to Anderson and Taylor [1]. Noting that the optimal level is $x = (y^* - \alpha)/\beta$ when α and

*Research supported by the National Science Foundation and the National Institutes of Health.

0196-8858/82/010050-24$05.00/0

$\beta \neq 0$ are known, they assume for the case of unknown α and β prior knowledge of bounds K_1 and K_2 such that

$$-\infty < K_1 \leq (y^* - \alpha)/\beta \leq K_2 < \infty, \tag{1.3}$$

and propose the rule

$$x_{i+1} = K_2 \wedge \{\hat{\beta}_i^{-1}(y^* - \hat{\alpha}_i) \vee K_1\}, \qquad i \geq 2, \tag{1.4}$$

where $\vee$ and $\wedge$ denote maximum and minimum, respectively, and

$$\hat{\beta}_i = \left\{ \sum_{r=1}^{i} (x_r - \bar{x}_i) y_r \right\} \Big/ \sum_{r=1}^{i} (x_r - \bar{x}_i)^2, \qquad \hat{\alpha}_i = \bar{y}_i - \hat{\beta}_i \bar{x}_i, \tag{1.5}$$

are the least-squares estimates of β and α at stage i. (Here and in the sequel we use the notation $\bar{a}_i$ for the arithmetic mean of $a_1, \ldots, a_i$.) The initial values x_1, x_2 of the recursion (1.4) are distinct but otherwise arbitrary numbers between K_1 and K_2. Anderson and Taylor call this rule the "least-squares certainty equivalence" (LSCE) rule and, assuming the errors ϵ_i to be normally distributed, they carry out some Monte Carlo simulations of its performance. Based on the results of these simulations, they conjecture that for the LSCE rule (1.4), x_n converges to θ with probability 1, where $\theta = (y^* - \alpha)/\beta$, and that $n^{1/2}(x_n - \theta)$ converges in distribution to a normal random variable with mean 0 and variance σ^2/β^2. They also raise the question whether the least-squares estimates $\hat{\alpha}_i$ and $\hat{\beta}_i$ are strongly consistent. In Section 2 we disprove the conjecture and give a negative answer to the question.

Another suggestion for treating the multiperiod control problem is due to Aoki [2]. He assumes that the sign of β is known, say $\beta > 0$, and proposes the use of a Robbins–Monro stochastic approximation scheme

$$x_{i+1} = x_i - c_i(y_i - y^*), \tag{1.6}$$

where $\{c_i\}$ is a sequence of positive constants such that

$$\sum_1^\infty c_i^2 < \infty, \qquad \sum_1^\infty c_i = \infty. \tag{1.7}$$

(If $\beta < 0$, then (1.6) is replaced by $x_{i+1} = x_i + c_i(y_i - y^*)$.) The condition (1.7) ensures (in the case $\beta > 0$) that the stochastic approximation scheme (1.6) converges to θ with probability 1 (cf. [3, 16]). As shown by Chung [6], the choice $c_i = (i\beta)^{-1}$ leads to an asymptotically normal distribution of x_i with the smallest asymptotic variance. For this optimal Robbins–Monro stochastic approximation scheme

$$x_{i+1} = x_i - (y_i - y^*)/(i\beta), \tag{1.8}$$

the following properties hold (cf. [10]):

$$n^{1/2}(x_n - \theta) \xrightarrow{\mathfrak{D}} N(0, \sigma^2/\beta^2), \tag{1.9}$$

$$\limsup_{n\to\infty} (n/2\log\log n)^{1/2} |x_n - \theta| = \sigma/\beta \qquad \text{a.s.}, \tag{1.10}$$

$$\lim_{n\to\infty} \sum_1^n (x_i - \theta)^2/\log n = \sigma^2/\beta^2 \qquad \text{a.s.} \tag{1.11}$$

Here and in the sequel, the notation $\xrightarrow{\mathfrak{D}}$ denotes convergence in distribution, "a.s." means "almost surely" (with probability 1), and $N(\mu, \sigma^2)$ denotes the normal distribution with mean μ and variance σ^2.

In the case of *known* β, the least-squares estimate of θ based on the observations $x_1, y_1, \ldots, x_i, y_i$ is $\bar{x}_i - \beta^{-1}(\bar{y}_i - y^*)$, and therefore the iterated least-squares procedure for choosing the level x_i amounts to the recursive scheme

$$x_{i+1} = \bar{x}_i - (\bar{y}_i - y^*)/\beta. \tag{1.12}$$

This recursion turns out to be equivalent to the stochastic approximation scheme (1.8); in fact, for every constant c and positive integer n, we have the equivalence

$$\begin{aligned} & x_{i+1} = \bar{x}_i - c(\bar{y}_i - y^*) \qquad \text{for all } i = 1,\ldots,n \\ \Leftrightarrow\ & x_{i+1} = x_i - c(y_i - y^*)/i \qquad \text{for all } i = 1,\ldots,n \end{aligned} \tag{1.13}$$

(cf. [10]).

When β is unknown, it is natural to replace β in (1.8) or (1.12) by some estimate $b_i = b_i(x_1, y_1, \ldots, x_i, y_i)$ of β based on the data already observed. Such a modification of (1.8) leads to the adaptive stochastic approximation scheme

$$x_{i+1} = x_i - (y_i - y^*)/(ib_i). \tag{1.14}$$

Modifying the iterated least-squares procedure (1.12) likewise leads to

$$x_{i+1} = \bar{x}_i - (\bar{y}_i - y^*)/b_i. \tag{1.15}$$

In spite of the equivalence between (1.8) and (1.12), the recursions (1.14) and (1.15) are no longer equivalent when the b_i are changing with i. In Section 3 we obtain a general representation theorem for (1.15) and compare it with the corresponding result for the stochastic approximation scheme (1.14).

We have recently developed in [10–12] an asymptotic theory of adaptive stochastic approximation schemes of the form (1.14). In this paper we extend the theory to recursive schemes of the type (1.15). Note that if we let $b_i = \hat{\beta}_i$, where $\hat{\beta}_i$ is the least-squares estimate of β in (1.5), then the recursive scheme (1.15) reduces to the LSCE rule (1.4) with infinite truncation points $K_1 = -\infty$, $K_2 = \infty$. In the counterexample of Section 2 on the LSCE rule, we exhibit an event with positive probability in which the sign of $\hat{\beta}_i$ differs from that of β for all i. In practice, although the value of β is unknown, its sign is often known. Making this assumption and therefore choosing b_i in (1.15) to have the same sign as β, Theorem 2 of Section 4 shows that the recursive scheme (1.15) converges a.s. to θ. The requirement that b_i should have the same sign as β also plays a vital role in establishing the a.s. convergence of the stochastic approximation scheme (1.14) (cf. [3, 10]). Estimates of the rate of convergence of the recursive scheme (1.15) under various general assumptions on b_i are also obtained in Section 4.

As in [10], we call the cumulative squared difference $\Sigma_1^n(x_i - \theta)^2$ of the design levels $x_1, \ldots, x_n$ from the optimal level θ the *cost* of the design at stage n. The relevance of this quantity to the multiperiod control problem is shown by (1.2). In Section 5 we obtain estimates of the cost $\Sigma_1^n(x_i - \theta)^2$ for the recursive scheme (1.15). In particular, we show that if $b_n \to \beta$ a.s., then the cost $\Sigma_1^n(x_i - \theta)^2$ of (1.15) also satisfies the asymptotic relation (1.11) for the optimal Robbins–Monro stochastic approximation scheme (1.8).

We have recently shown in [12] that if bounds B_1 and B_2 for β are known such that $0 < B_1 < \beta < B_2 < \infty$ and we let $b_i = B_2 \wedge (\hat{\beta}_i \vee B_1)$, then the stochastic approximation scheme (1.14) with this choice of b_i has the asymptotic properties (1.9), (1.10), and (1.11) of the optimal Robbins–Monro stochastic approximation scheme (1.8). In Section 6, by setting b_i in the recursive scheme (1.15) equal to a similar truncated least-squares estimate of β, we obtain a modified version of the LSCE rule which also has the asymptotic properties (1.9), (1.10), and (1.11). Thus, although the natural idea of using the least-squares estimates $\hat{\alpha}_i$, $\hat{\beta}_i$ iteratively to replace the unknown parameters α, β in the optimal level $(y^* - \alpha)/\beta$ does not lead to an a.s. convergent rule, a suitable modification of this idea does have the desirable convergence properties conjectured by Anderson and Taylor.

2. Counterexample to the Anderson–Taylor Conjecture

Consider the linear regression model (1.1) in which the errors ϵ_i are i.i.d. $N(0, \sigma^2)$ random variables with $\sigma > 0$ and the levels x_i are defined recursively by the LSCE rule (1.4). Note that in this case of normal errors, the maximum likelihood estimate of $\theta = (y^* - \alpha)/\beta$, subject to the bounds

(1.3), based on the observations $x_1, y_1, \ldots, x_i, y_i$ is $K_2 \wedge \{\hat{\beta}_i^{-1}(y^* - \hat{\alpha}_i) \vee K_1\}$. Therefore the LSCE rule (1.4) simply uses the maximum likelihood estimate of θ as the choice of the next level x_{i+1}. Based on Monte Carlo simulations involving normal errors, Anderson and Taylor [1] conjecture that the LSCE rule converges a.s. to θ and that $n^{1/2}(x_n - \theta) \xrightarrow{\mathfrak{D}} N(0, \sigma^2/\beta^2)$. In this section we give a negative answer to this conjecture by exhibiting an event with positive probability in which x_n does not converge to θ.

Without loss of generality we shall assume that $\beta > 0$, $\theta = 0$, and $K_2 = K = -K_1$ with $K > 0$. Consider the LSCE rule (1.4) with initial values $x_1 = 0$ and $x_2 = K$. Letting

$$A = \left\{ -\frac{25}{16}K\beta < \epsilon_2 - \epsilon_1 < -\frac{3}{2}K\beta, \ \frac{5}{16}K\beta < \bar{\epsilon}_2 < \frac{21}{64}K\beta, \text{ and} \right.$$
$$\left. -\frac{n+40}{64}K\beta < \sum_{i=3}^{n} \epsilon_i < \frac{n-42}{64}K\beta \text{ for all } n \geq 3 \right\}, \tag{2.1}$$

it follows from the strong law of large numbers, the independence between $\epsilon_2 - \epsilon_1$ and $\bar{\epsilon}_2$, and their independence of $\{\sum_{i=3}^{n}\epsilon_i, n \geq 3\}$, that $P(A) > 0$. We now show that

$$x_n = K \qquad \text{for all } n \geq 2 \text{ on } A. \tag{2.2}$$

The proof of (2.2) is by induction and makes repeated use of the following algebraic identities: For $n \geq 3$,

$$\sum_{i=3}^{n} i^{-1}(\epsilon_i - \bar{\epsilon}_{i-1}) = \sum_{i=3}^{n} i^{-1}\epsilon_i - \sum_{i=3}^{n} \{i(i-1)\}^{-1}\left\{2\bar{\epsilon}_2 + \sum_{j=3}^{i-1}\epsilon_j\right\}$$
$$= n^{-1}\sum_{j=3}^{n-1}\epsilon_j - 2\bar{\epsilon}_2\left(\frac{1}{2} - \frac{1}{n}\right) = \bar{\epsilon}_n - \bar{\epsilon}_2, \tag{2.3}$$

while for $n \geq 2$,

$$\hat{\beta}_n = \beta + \frac{\sum_1^n (x_i - \bar{x}_n)\epsilon_i}{\sum_1^n (x_i - \bar{x}_n)^2} = \beta + \frac{\sum_2^n i^{-1}(i-1)(x_i - \bar{x}_{i-1})(\epsilon_i - \bar{\epsilon}_{i-1})}{\sum_2^n i^{-1}(i-1)(x_i - \bar{x}_{i-1})^2} \tag{2.4}$$

(cf. [8]).

Since $K_2 = K = -K_1$ and $\theta = 0$, the LSCE rule (1.4) can be written as

$$x_{i+1} = K \wedge \left\{\left[\left(1 - \hat{\beta}_i^{-1}\beta\right)\bar{x}_i - \hat{\beta}_i^{-1}\bar{\epsilon}_i\right] \vee (-K)\right\}, \qquad i \geq 2. \quad (2.5)$$

Since $x_1 = 0$ and $x_2 = K$, it follows from (2.4) that $\hat{\beta}_2 = \beta + K^{-1}(\epsilon_2 - \epsilon_1)$, and therefore

$$-\tfrac{9}{16}\beta < \hat{\beta}_2 < -\tfrac{1}{2}\beta \qquad \text{on } A. \quad (2.6)$$

Noting that $\bar{x}_2 = \frac{1}{2}K$ and that $\bar{\epsilon}_2 > 0$ on A, we then obtain that on A

$$\left(1 - \hat{\beta}_2^{-1}\beta\right)\bar{x}_2 - \hat{\beta}_2^{-1}\bar{\epsilon}_2 > \tfrac{1}{2}K\left(1 - \hat{\beta}_2^{-1}\beta\right) > K,$$

and therefore $x_3 = K$ on A by (2.5).

Let $n \geq 3$ and assume that $x_i = K$ for all $i = 2, \ldots, n$ on A. Then for $n \geq i \geq 2$, $\bar{x}_i = i^{-1}(i-1)K$ and $x_i - \bar{x}_{i-1} = K/(i-1)$ on A. Therefore on A,

$$\sum_2^n i^{-1}(i-1)(x_i - \bar{x}_{i-1})^2 = K^2 \sum_2^n \{i(i-1)\}^{-1} = K^2(1 - n^{-1}), \quad (2.7)$$

$$\begin{aligned}\sum_2^n i^{-1}(i-1)(x_i - \bar{x}_{i-1})(\epsilon_i - \bar{\epsilon}_{i-1}) &= K \sum_2^n i^{-1}(\epsilon_i - \bar{\epsilon}_{i-1}) \\ &= K\left\{\tfrac{1}{2}(\epsilon_2 - \epsilon_1) + (\bar{\epsilon}_n - \bar{\epsilon}_2)\right\}, \\ &\qquad \text{by (2.3)}. \end{aligned} \quad (2.8)$$

From (2.1), it follows that on A

$$-\tfrac{50}{64}K\beta < \tfrac{1}{2}(\epsilon_2 - \epsilon_1) < -\tfrac{48}{64}K\beta, \qquad -\tfrac{22}{64}K\beta < \bar{\epsilon}_n - \bar{\epsilon}_2 < -\tfrac{19}{64}K\beta. \quad (2.9)$$

By (2.4), (2.7), (2.8), and (2.9), we obtain that on A

$$\begin{aligned}\hat{\beta}_n &> \beta - \tfrac{72}{64}\beta/(1 - n^{-1}) \geq -\tfrac{11}{16}\beta, \\ \hat{\beta}_n &< \beta - \tfrac{67}{64}\beta/(1 - n^{-1}) < -\tfrac{3}{64}\beta. \end{aligned} \quad (2.10)$$

Since $\bar{x}_n = n^{-1}(n-1)K \geq \frac{2}{3}K$ and $\bar{\epsilon}_n > -\frac{1}{64}K\beta$ on A, we obtain from (2.10) that on A,

$$\left(1 - \hat{\beta}_n^{-1}\beta\right)\bar{x}_n - \hat{\beta}_n^{-1}\bar{\epsilon}_n > \left(1 + \tfrac{16}{11}\right)\tfrac{2}{3}K - \tfrac{1}{3}K > K,$$

and therefore $x_{n+1} = K$ by (2.5), completing the induction argument.

3. A Representation Theorem for the Recursion (1.15)

For any real sequence $\{a_n\}$, let $\Sigma_{n=i}^{k} a_n = 0$ if $i > k$. In view of (1.1) and the fact $y^* = \alpha + \beta\theta$, the recursion (1.15) can be written as

$$x_{i+1} - \theta = \left(1 - \beta b_i^{-1}\right)(\bar{x}_i - \theta) - b_i^{-1}\bar{\epsilon}_i. \tag{3.1}$$

The following representation theorem for the recursion (3.1) provides a useful tool for analyzing the recursive scheme (1.15).

THEOREM 1. *Let m be a positive integer, and let $\{x_n\}, \{\epsilon_n\}, \{a_n\}, \{c_n\}$, $n \geq m$, be sequences of real numbers such that*

$$x_{n+1} = (1 - a_n)\bar{x}_n - c_n\bar{\epsilon}_n, \qquad n \geq m. \tag{3.2}$$

Then for $n \geq m$,

$$x_{n+1} = \beta_{m-1,n}\bar{x}_m - \sum_{j=m}^{n-1} \beta_{jn} c_j \bar{\epsilon}_j / (j+1) - c_n\bar{\epsilon}_n, \tag{3.3}$$

where

$$\beta_{nn} = 1, \beta_{n-1,n} = 1 - a_n,$$
$$\beta_{jn} = (1 - a_n) \prod_{k=j+2}^{n} (1 - a_{k-1}/k), \qquad n \geq j + 2. \tag{3.4}$$

We preface the proof of Theorem 1 by the following

LEMMA 1. *Let N, m be positive integers such that $N > m$, and let $\{a_n\}, \{d_n\}$, $m \leq n \leq N$, be two sequences of real numbers. Suppose that $d_m = 1$. Then the following statements are equivalent*:

$$d_{n+1} = (1 - \alpha_n)n^{-1} \sum_{i=m}^{n} d_i, \qquad N - 1 \geq n \geq m; \tag{3.5}$$

$$n^{-1} \sum_{i=m}^{n} d_i = m^{-1} \prod_{k=m+1}^{n} (1 - \alpha_{k-1}/k), \qquad N \geq n > m; \tag{3.6}$$

$$d_{m+1} = m^{-1}(1 - \alpha_m), d_n = m^{-1}(1 - \alpha_{n-1}) \prod_{k=m+1}^{n-1} (1 - \alpha_{k-1}/k)$$
$$\textit{for } N \geq n > m + 1. \tag{3.7}$$

Proof. Simple algebra shows (3.6) ⇒ (3.7), and both the implications (3.5) ⇒ (3.6) and (3.7) ⇒ (3.5) can easily be proved by induction on N. □

Proof of Theorem 1. We prove (3.3) by induction on n. Since $\beta_{m-1,m} = 1 - a_m$, (3.3) obviously holds for $n = m$. Assume that (3.3) holds for all n with $m \le n \le N-1$. Then by (3.2),

$$\begin{aligned} x_{N+1} &= (1-a_N)\left(m\bar{x}_m + \sum_{i=m}^{N-1} x_{i+1}\right)/N - c_N\bar{\epsilon}_N \\ &= N^{-1}(1-a_N)\left\{\left(m + \sum_{i=m}^{N-1}\beta_{m-1,i}\right)\bar{x}_m - \sum_{i=m}^{N-1}\sum_{j=m}^{i-1}\beta_{ji}c_j\bar{\epsilon}_j/(j+1)\right. \\ &\qquad \left. - \sum_{i=m}^{N-1} c_i\bar{\epsilon}_i\right\} - c_N\bar{\epsilon}_N, \qquad \text{by induction hypothesis,} \\ &= N^{-1}(1-a_N)\left\{\left(m + \sum_{i=m+1}^{N}\beta_{m-1,i-1}\right)\bar{x}_m - c_{N-1}\bar{\epsilon}_{N-1}\right. \\ &\qquad \left. - \sum_{j=m}^{N-2}\left[\sum_{i=j+1}^{N-1}\beta_{ji} + (j+1)\right]c_j\bar{\epsilon}_j/(j+1)\right\} - c_N\bar{\epsilon}_N. \end{aligned} \tag{3.8}$$

Put $d_i = m^{-1}\beta_{m-1,i-1}$ for $i > m$ and $d_m = 1$ in Lemma 1 and note that (3.4) implies that (3.7) holds with $\alpha_i = a_i$. Hence we obtain from (3.5) that

$$\begin{aligned} N^{-1}(1-a_N)\left(m + \sum_{i=m+1}^{N}\beta_{m-1,i-1}\right) &= m(1-a_N)N^{-1}\sum_{i=m}^{N} d_i \\ &= md_{N+1} = \beta_{m-1,N}. \end{aligned} \tag{3.9}$$

Likewise, putting $d_i' = (j+1)^{-1}\beta_{j,i-1}$ for $i \ge j+2$ and $d_{j+1}' = 1$ in Lemma 1, we obtain from (3.5) that

$$\begin{aligned} N^{-1}(1-a_N)\left[\sum_{i=j+2}^{N}\beta_{j,i-1} + (j+1)\right] &= (j+1)(1-a_N)N^{-1}\sum_{i=j+1}^{N} d_i' \\ &= (j+1)d_{N+1}' = \beta_{jN}. \end{aligned} \tag{3.10}$$

Moreover, by (3.4),

$$N^{-1}(1-a_N) = \beta_{N-1,N}/\{(N-1)+1\}. \tag{3.11}$$

From (3.8)–(3.11) it follows that (3.3) also holds for $n = N$, completing the induction proof. □

It is of interest to compare Theorem 1 with the corresponding result for the stochastic approximation scheme (1.14) which, in view of (1.1), can be rewritten as

$$x_{i+1} - \theta = \left(1 - \beta b_i^{-1}/i\right)(x_i - \theta) - b_i^{-1}\epsilon_i/i. \tag{3.12}$$

The following lemma (cf. [10, p. 1202]) provides the analog of the representation (3.3) for the recursion (3.12).

LEMMA 2. *Let m be a positive integer, and let $\{x_n\}, \{\epsilon_n\}, \{a_n\}, \{c_n\}$, $n \geq m$, be sequences of real numbers such that*

$$x_{n+1} = (1 - a_n/n)x_n - c_n\epsilon_n/n. \tag{3.13}$$

Then for $n \geq m$

$$x_{n+1} = \beta'_{m-1,n}x_m - \sum_{j=m}^{n} \beta'_{jn}c_j\epsilon_j/j, \tag{3.14}$$

where

$$\beta'_{nn} = 1, \beta'_{jn} = \prod_{k=j+1}^{n} (1 - a_k/k) \qquad \text{for } n \geq j+1. \tag{3.15}$$

For the special case $c_n = c$ and $a_n = \beta c$ for all n, it follows from (3.15) that for $n \geq j+1$,

$$\begin{aligned}\beta'_{jn}c_j/j - \beta'_{j+1,n}c_{j+1}/(j+1) &= c(1-\beta c)\beta'_{j+1,n}/\{j(j+1)\}\\ &= c\beta_{jn}/\{j(j+1)\},\end{aligned} \tag{3.16}$$

where β_{jn} is as defined in (3.4). In view of (3.16) and the fact that $\beta'_{0n} = \beta_{0n}$, application of partial summation to (3.14) in the case $m = 1$ then reduces it to the representation (3.3). This shows the equivalence of (3.3) and (3.14) in the special case $m = 1$ and $c_n = c$, $a_n = \beta c$. However, when a_n and c_n are changing with n, (3.3) and (3.14) are no longer equivalent.

4. CONVERGENCE PROPERTIES OF THE RECURSIVE SCHEME $x_{i+1} = \bar{x}_i - (\bar{y}_i - y^*)/b_i$

In the counterexample of Section 2 on the LSCE rule, (2.6) and (2.10) show that $\hat{\beta}_n$ and β are of different signs on the event A. When the sign of β is known, we should therefore choose b_n in the recursive scheme (1.15) to be of the same sign as β. Throughout the sequel we shall assume that $\beta > 0$ and that $b_n > 0$ for all n. The following theorem shows that the recursive scheme (1.15) converges a.s. to θ under very weak assumptions on b_n.

THEOREM 2. *Let $\epsilon, \epsilon_1, \epsilon_2, \ldots$ be i.i.d. random variables with $E\epsilon = 0$ and $E\epsilon^2 = \sigma^2 < \infty$, and let $\{b_n\}$ be a sequence of positive random variables.*

Consider the linear regression model

$$y_n = y^* + \beta(x_n - \theta) + \epsilon_n, \tag{4.1}$$

where $\beta > 0$, y^* *and* θ *are constants, and* x_n *are random variables defined recursively by* (1.15).

(i) *On* $\{\inf b_n > 0$ *and* $\sum_1^\infty (nb_n)^{-1} = \infty\}$, $x_n \to \theta$ *a.s.*

(ii) *Suppose that there exist positive random variables* U_n *such that with probability* 1

$$\lim_{n\to\infty} U_n = \infty, \qquad \sum_1^\infty (nU_n)^{-1} = \infty, \tag{4.2}$$

$$U_n \ge b_n \qquad \text{for all large } n, \tag{4.3}$$

and

$$\limsup_{n\to\infty} \left(\log b_n^{-1}\right) \Big/ \sum_1^n (iU_i)^{-1} < \beta. \tag{4.4}$$

Then $x_n \to \theta$ a.s. *In particular,* $x_n \to \theta$ a.s. *if there exist* $\rho > 0$ *and* $0 < \delta < 1$ *such that with probability* 1

$$(\log n)^{-\rho} \le b_n \le (\log n)^{\delta} \qquad \text{for all large } n. \tag{4.5}$$

Proof. From (3.1) and Theorem 1, it follows that for $n \ge m$

$$x_{n+1} - \theta = \beta_{m-1,n}(\bar{x}_m - \theta) - \sum_{j=m}^{n-1} \beta_{jn}\bar{\epsilon}_j / \left\{(j+1)b_j\right\} - \bar{\epsilon}_n / b_n, \tag{4.6}$$

where β_{jn} is as defined in (3.4) with $a_k = \beta b_k^{-1}$. To prove (ii), since $\sum_1^n (iU_i)^{-1} = o(\log n)$ a.s. by (4.2), it follows from (4.4) that $\lim_{n\to\infty} nb_{n-1} = \infty$ a.s. In view of this and (4.3), with probability 1 we can choose m sufficiently large such that

$$1 - \beta/(nb_{n-1}) \ge \tfrac{1}{2} \qquad \text{and} \qquad U_n \ge b_n \qquad \text{for all } n \ge m. \tag{4.7}$$

From (3.4), (4.7), and the inequality $1 - x < e^{-x}$ for $x > 0$, it follows that with probability 1, for $n > j \ge m$,

$$\begin{aligned} |\beta_{jn}| &\le (1 + \beta/b_n) \exp\left\{-\beta \sum_{i=j+2}^{n} (iU_{i-1})^{-1}\right\} \\ &\le \left\{(1 + \beta/b_n) \exp\left(-\beta \sum_{i=m+1}^{n} (iU_{i-1})^{-1}\right)\right\} \exp\left\{\beta \sum_{i=m+1}^{j+1} (iU_{i-1})^{-1}\right\}. \end{aligned} \tag{4.8}$$

Since $\Sigma_{m+1}^{n}(iU_{i-1})^{-1} \sim \Sigma_{m}^{n-1}(iU_i)^{-1}$, we obtain from (4.4) that

$$b_n^{-1}\exp\left(-\beta\sum_{i=m+1}^{n}(iU_{i-1})^{-1}\right) \to 0 \text{ a.s.} \tag{4.9}$$

By the law of the iterated logarithm,

$$\bar{\epsilon}_j = 0\left(j^{-1/2}(\log\log j)^{1/2}\right) \qquad \text{a.s.} \tag{4.10}$$

Since $\Sigma_{i=m+1}^{j+1}(iU_{i-1})^{-1} = o(\log j)$ a.s., it follows from (4.9) and (4.10) that

$$\bar{\epsilon}_n/b_n \to 0 \qquad \text{and}$$

$$\sum_{j=m}^{\infty}\{(j+1)b_j\}^{-1}|\bar{\epsilon}_j|\exp\left\{\beta\sum_{i=m+1}^{j+1}(iU_{i-1})^{-1}\right\} < \infty \qquad \text{a.s.} \tag{4.11}$$

From (4.6), (4.8), (4.9), and (4.11), we obtain that $x_n \to \theta$ a.s. A similar argument proves (i). □

We now study the rate of convergence of x_n to θ in the following

THEOREM 3. *With the same notations and assumptions as in Theorem* 2, *let* $b^* = \limsup_{n\to\infty} b_n$.

(i) *On* $\{\inf b_n > 0, b^* < 2\beta\}$, $x_n - \theta = 0(n^{-1/2}(\log\log n)^{1/2})$ *a.s.*

(ii) *For* $\lambda > 2\beta$, $x_n - \theta = o(n^{-\beta/\lambda})$ *a.s. on* $\{\inf b_n > 0, b^* < \lambda\}$.

Proof. To prove (ii), let $\lambda > \tilde{\lambda} > 2\beta$ and let $A_{\tilde{\lambda}} = \{\inf b_n > 0 \text{ and } b_n \le \tilde{\lambda}$ for all large $n\}$. On $A_{\tilde{\lambda}}$, we have for $n > j \ge m$ (sufficiently large),

$$|\beta_{jn}| \le (1+\beta/b_n)\exp\left\{-(\beta/\tilde{\lambda})\sum_{i=j+2}^{n} i^{-1}\right\}. \tag{4.12}$$

Since $\inf b_n > 0$ on $A_{\tilde{\lambda}}$ and $\beta/\tilde{\lambda} < \frac{1}{2}$, it then follows from (4.6), (4.10), and (4.12) that with probability 1, $x_n - \theta = 0(n^{-\beta/\tilde{\lambda}}) = o(n^{-\beta/\lambda})$ on $A_{\tilde{\lambda}}$. Part (i) is an immediate corollary of Theorem 4 below. □

The following theorem, which is a refinement of Theorem 2(i), says that with probability 1, a sufficiently long string of b_n not exceeding $(2-\eta)\beta$ leads to a corresponding string of x_n differing from θ by less than a constant times $n^{-1/2}(\log\log n)^{1/2}$. An analogous result for the stochastic approximation scheme (1.14) was recently established in [11] under additional assumptions on b_i.

THEOREM 4. *With the same notations and assumptions as in Theorem* 2, *assume that* $\inf b_n > 0$ a.s. *Then there exists an event* Ω_0 *with* $P(\Omega_0) = 1$ *such that all sample points* $\omega \in \Omega_0$ *have the following property: For every given* $0 < \eta < 2$, *there exist* $C > 0$ *and positive integers* N, k (*depending on* ω *and*

η) such that at ω, for all $m \geq N$ and $l \geq m^k$,

$$\max_{m \leq n \leq l} b_n \leq (2-\eta)\beta$$
$$\Rightarrow |x_n - \theta| \leq Cn^{-1/2}(\log\log n)^{1/2} \qquad \textit{for all } m^k \leq n \leq l, \textit{ and}$$
$$|\bar{x}_n - \theta| \leq Cn^{-1/2}(\log\log n)^{1/2} \qquad \textit{for all } m^k \leq n \leq l + l^{1/2}. \tag{4.13}$$

Proof. The assumption $\inf b_n > 0$ a.s. implies that $\sup_{1 \leq j \leq n < \infty} |\beta_{jn}| < \infty$ a.s., and therefore in view of (4.6) and (4.10), $x_n = 0(1)$ a.s. This in turn implies that with probability 1

$$\sup_m |\bar{x}_m| < \infty, \quad \sum_{l \leq i \leq l + l^{1/2}} |x_i - \theta| = 0(l^{1/2}). \tag{4.14}$$

Let Ω_0 be the event in which (4.14) holds and

$$b_* = \inf_n b_n > 0, \qquad |\bar{\epsilon}_j| = 0\big(j^{-1/2}(\log\log j)^{1/2}\big). \tag{4.15}$$

Let $\omega \in \Omega_0$ and let $0 < \eta < 2$. Choosing m_0 large enough such that $\beta/(ib_{i-1}) < 1$ for $i \geq m_0$, we have at ω

$$\max_{m \leq n \leq l} b_n \leq (2-\eta)\beta \qquad \text{and} \qquad m \geq m_0$$
$$\Rightarrow |\beta_{jn}| \leq (1 + \beta/b_*) \prod_{i=j+2}^{n} \left(1 - \frac{1}{(2-\eta)i}\right) = (1 + \beta/b_*)\gamma_n/\gamma_{j+1}$$
$$\text{for } l \geq n > j \geq m, \tag{4.16}$$

where

$$\gamma_n = \prod_{i=m_0}^{n} \left(1 - \frac{1}{(2-\eta)i}\right) \sim Dn^{-1/(2-\eta)}$$

for some $D > 0$. Letting $k \geq 2$ such that $(1 - k^{-1})/(2-\eta) > \frac{1}{2}$, we obtain from (4.16) that at ω, for $m \geq m_1$ (sufficiently large) and $l \geq m^k$,

$$\max_{m \leq i \leq l} b_i \leq (2-\eta)\beta \Rightarrow |\beta_{m-1,n}| \leq n^{-1/2} \qquad \text{for } m^k \leq n \leq l. \tag{4.17}$$

Making use of (4.6) and (4.14)–(4.17), we obtain the desired conclusion (4.13) on $x_n - \theta$ by choosing C and N sufficiently large; this and (4.14) then provide the desired conclusion on $\bar{x}_n - \theta$ by choosing k sufficiently large. □

The estimates of the rate of convergence of x_n to θ on the event $\{\inf b_n > 0, \sup b_n < \infty\}$ given by Theorem 3 are sharp, in view of the following precise estimates on the events

$$E = \{b_n \text{ converges to a finite positive limit}\},$$
$$E_1 = \left\{\inf_i b_i > 0, \sup_{i>n} |b_i - b_n| = 0\left((\log n)^{-\rho}\right) \text{ for some } \rho > 1\right\} \subset E,$$
$$E_2 = \left\{\inf_i b_i > 0, \sup_{i>n} |b_i - b_n| = 0\left((\log n)^{-\rho}\right) \text{ for some } \rho > \tfrac{3}{2}\right\} \subset E_1. \tag{4.18}$$

THEOREM 5. *With the same notations and assumptions as in Theorem 2, define the events E, E_1, E_2 by* (4.18) *and let $b = \lim_{n\to\infty} b_n$ on E.*

(i) *On $E \cap \{b < 2\beta\}$,*

$$\limsup_{n\to\infty} (n/2 \log\log n)^{1/2} |x_n - \theta| = (\sigma/\beta) f^{1/2}(b/\beta) \quad a.s., \tag{4.19}$$

where

$$f(t) = 1/\{t(2-t)\}, \qquad 0 < t \neq 2. \tag{4.20}$$

(ii) *On $E_2 \cap \{b = 2\beta\}$,*

$$\limsup_{n\to\infty} n^{1/2} |x_n - \theta| / \{2(\log n)(\log\log\log n)\}^{1/2} = \sigma/2\beta \quad a.s. \tag{4.21}$$

(iii) *On $E_1 \cap \{b > 2\beta\}$, $n^{\beta/b}(x_n - \theta)$ converges a.s. Moreover, on $E_1 \cap \{b > 2\beta\} \cap \{\lim_{n\to\infty} n^{\beta/b}(x_n - \theta) = 0\}$,*

$$\limsup_{n\to\infty} (n/2 \log\log n)^{1/2} |x_n - \theta| = (\sigma/\beta) |f(b/\beta)|^{1/2} \quad a.s., \tag{4.22}$$

where f is as defined in (4.20).

To prove Theorem 5, we make use of the properties of slowly varying sequences; a sequence of positive numbers $L(n)$ is said to be slowly varying if $\lim_{n\to\infty} L([cn])/L(n) = 1$ for all $c > 0$ (cf. [4]). We also make use of the following uniform law of the iterated logarithm for certain integral transforms of Brownian motion.

LEMMA 3. *Let $w(t)$, $t \geq 0$, be a standard Brownian motion. Then*

(i) $P[\limsup_{t\to\infty} t^{1/2} | (1-\alpha)t^{-\alpha}\int_0^t s^{\alpha-2}w(s)\,ds + t^{-1}w(t) | / (2\log\log t)^{1/2} = (2\alpha-1)^{-1}$ *for all* $\alpha > \frac{1}{2}] = 1$;
(ii) $P[\limsup_{t\to\infty} t^{1/2} | (1-\alpha)t^{-\alpha}\int_t^\infty s^{\alpha-2}w(s)\,ds - t^{-1}w(t) | / (2\log\log t)^{1/2} = (1-2\alpha)^{-1}$ *for all* $\alpha < \frac{1}{2}] = 1$.

Proof. To prove (ii), let

$$X_\alpha(t) = (1-\alpha)t^{1/2-\alpha}\int_t^\infty s^{\alpha-2}w(s)\,ds = (1-\alpha)\int_t^\infty (s/t)^{\alpha-1/2}s^{-3/2}w(s)\,ds, \tag{4.23}$$

$$\Omega_\alpha = \left\{ \limsup_{t\to\infty} | X_\alpha(t) - t^{-1/2}w(t) | / (2\log\log t)^{1/2} = (1-2\alpha)^{-1} \right\}. \tag{4.24}$$

For every *fixed* $\alpha < \frac{1}{2}$,

$$t^{\alpha-1/2}\left(X_\alpha(t) - t^{-1/2}w(t)\right) = (1-\alpha)\int_t^\infty s^{\alpha-2}w(s)\,ds - t^{\alpha-1}w(t) = \int_t^\infty s^{\alpha-1}\,dw(s) = (1-2\alpha)^{-1}\tilde{w}(t^{-(1-2\alpha)}), \tag{4.25}$$

in which $\tilde{w}(t)$, $t \geq 0$, is a standard Brownian motion. By the law of the iterated logarithm,

$$\limsup_{s\to\infty} | \tilde{w}(s) | / (2s\log\log s)^{1/2} = 1 \quad \text{a.s.} \tag{4.26}$$

From (4.24), (4.25), and (4.26), it follows that $P(\Omega_\alpha) = 1$ for every $\alpha < \frac{1}{2}$. Therefore,

$$P\left(\cap \left\{\Omega_\alpha \colon \alpha < \tfrac{1}{2}, \alpha \text{ is rational}\right\}\right) = 1. \tag{4.27}$$

For fixed $c < d < \frac{1}{2}$ with $d - c < 1$, we obtain from (4.23) that

$$\sup_{c\leq\alpha\leq d} | X_\alpha(t) - X_c(t) | \leq \left\{(1-c)\left[(d-c)^{-(d-c)} - 1\right] + d - c\right\} \times \int_t^{t/(d-c)} (s/t)^{c-1/2}s^{-3/2} | w(s) | \,ds + 2(1-c)\int_{t/(d-c)}^\infty (s/t)^{d-1/2}s^{-3/2} | w(s) | \,ds,$$

and therefore by the law of the iterated logarithm (4.26) for $w(s)$,

$$\limsup_{t\to\infty}\left\{\sup_{c\le\alpha\le d}|X_\alpha(t)-X_d(t)|/(\log\log t)^{1/2}\right\}\le K(c,d) \qquad \text{a.s.,} \tag{4.28}$$

where $\lim_{r\downarrow 0}K(d-r,d)=0$ uniformly for d belonging to compact subsets of $(-\infty,\frac{1}{2})$. From (4.27) and (4.28), (ii) follows. Part (i) can be proved by a similar argument. □

Proof of Theorem 5. On E, we can choose m sufficiently large such that $1-\beta/(ib_{i-1})\ge\frac{1}{2}$ for all $i\ge m$. Letting $\gamma_n=\prod_{i=m}^{n}(1-\beta/ib_{i-1})$ for $n\ge m$, we note that on E, $\gamma_n=n^{-\beta/b}\tau_n$ where $\{\tau_n\}$ is a slowly varying sequence of positive numbers (cf. [10, p. 1202]). Since $\beta_{jn}=(1-\beta b_n^{-1})\gamma_n/\gamma_{j+1}$ for $n>j\ge m-1$, it then follows that on E

$$\beta_{jn}=(1-\beta b_n^{-1})\{(j+1)/n\}^{\beta/g}\tau_b\tau_{j+1}^{-1} \qquad \text{for } n>j\ge m-1. \tag{4.29}$$

in the event $E_1\subset E$, since

$$\beta b_n^{-1}=\beta b^{-1}+0((\log n)^{-\rho}) \tag{4.30}$$

for some $\rho>1$, we have furthermore that

$$\sup_{n>j}|\tau_n\tau_{j+1}^{-1}-1|=0((\log j)^{-(\rho-1)}), \text{ and}$$
$$\tau=\lim_{n\to\infty}\tau_n \text{ exists and is positive on } E_1. \tag{4.31}$$

From (4.6) and (4.29), it follows that on E

$$x_{n+1}-\theta=(1-\beta b_n^{-1})(m/n)^{\beta/b}\tau_n\tau_m^{-1}(\bar{x}_m-\theta)-\bar{\epsilon}_n/b_n$$
$$-(1-\beta b_n^{-1})n^{-\beta/b}\tau_n\sum_{j=m}^{n-1}(j+1)^{\beta/b-1}\bar{\epsilon}_j/(\tau_{j+1}b_j). \tag{4.32}$$

To prove (i), letting $S(0)=0$ and $S(t)=j\bar{\epsilon}_j$ for $j-1<t\le j$, and redefining the random variables on a new probability space if necessary, there exists by the strong invariance principle (cf. [7]) a standard Brownian motion $w(t)$ such that

$$S(t)-\sigma w(t)=o((t\log\log t)^{1/2}) \qquad \text{a.s.} \tag{4.33}$$

On $E\cap\{b<2\beta\}$, since (4.26) holds (with $\beta b_n^{-1}\to\beta/b>\frac{1}{2}$ and $\{\tau_n\}$ slowly varying), we obtain (4.19) by using (4.33) and Lemma 3(i) together with an argument similar to that in Theorem 7 of [9] for adaptive stochastic approximation schemes.

To prove (iii), we note from (4.10), (4.31), and (4.32) that on $E_1 \cap \{\frac{1}{2} > \beta/b\}$, $\Sigma_{j=m}^{\infty}(j+1)^{\beta/b-1}|\bar{\epsilon}_j|/(\tau_{j+1}b_j) < \infty$ a.s. and therefore

$$n^{\beta/b}(x_{n+1}-\theta) \to \tau(1-\beta/b)\left\{m^{\beta/b}\tau_m^{-1}(\bar{x}_m-\theta) - \sum_{j=m}^{\infty}(j+1)^{\beta/b-1}\bar{\epsilon}_j/\left(\tau_{j+1}b_j\right)\right\} \quad \text{a.s.}$$

Moreover, it follows from (4.26) that on $E_1 \cap \{\frac{1}{2} > \beta/b\} \cap \{m^{\beta/b}\tau_m^{-1}(\bar{x}_m - \theta) = \Sigma_{j=m}^{\infty}(j+1)^{\beta/b-1}\bar{\epsilon}_j/(\tau_{j+1}b_j)\}$,

$$x_{n+1}-\theta = \left(1-\beta b_n^{-1}\right)n^{-\beta/b}\tau_n\sum_{j=n}^{\infty}(j+1)^{\beta/b-1}\bar{\epsilon}_j/\left(\tau_{j+1}b_j\right) - \bar{\epsilon}_n/b_n. \tag{4.34}$$

From (4.31), (4.33), (4.34), and Lemma 3(ii), it then follows that (4.22) holds on $E_1 \cap \{\frac{1}{2} > \beta/b\} \cap \{m^{\beta/b}\tau_m^{-1}(\bar{x}_m - \theta) = \Sigma_{j=m}^{\infty}(j+1)^{\beta/b-1}\bar{\epsilon}_j/(\tau_{j+1}b_j)\}$.

To prove (ii), we note from (4.30), (4.31), and (4.32) that on $E_2 \cap \{b = 2\beta\}$,

$$x_{n-1}-\theta = 0(n^{-1/2}) - \left\{1+0\left((\log n)^{-\rho}\right)\right\}\bar{\epsilon}_n/b - \left\{\tfrac{1}{2}+0\left((\log n)^{-\rho}\right)\right\}b^{-1}n^{-1/2}\sum_{j=m}^{n-1}(j+1)^{-1/2}\bar{\epsilon}_j\left\{1+0\left((\log j)^{-(\rho-1)}\right)\right\} \tag{4.35}$$

with $\rho > 3/2$. Let $\frac{1}{2} > \delta > 2-\rho$. Making use of the law of the iterated logarithm (4.10) and partial summation, we then obtain from (4.35) that on $E_2 \cap \{b = 2\beta\}$,

$$x_{n+1}-\theta = -(2\beta)^{-1}n^{-1/2}\left\{\sum_{j=m}^{n}j^{-1/2}\epsilon_j + o\left((\log n)^{\delta}\right)\right\} \quad \text{a.s.} \tag{4.36}$$

By the law of the iterated logarithm for the weighted sum $\Sigma_{j=m}^{n} j^{-1/2}\epsilon_j$ whose variance is $\sigma^2\Sigma_{j=m}^{n} j^{-1} \sim \sigma^2 \log n$,

$$\limsup_{n\to\infty}\left|\sum_{j=m}^{n}j^{-1/2}\epsilon_j\right| / \{2(\log n)(\log\log\log n)\}^{1/2} = \sigma \quad \text{a.s.} \tag{4.37}$$

From (4.36) and (4.37), (4.21) follows. □

While the law of the iterated logarithm (4.19) for x_n follows from the representation (4.32) and the strong invariance principle (4.33), an application of Donsker's invariance principle (cf. [5]) and (4.32) gives the following result on the limiting distribution of x_n.

THEOREM 6. *With the same notations and assumptions as in Theorem* 2, *suppose that there exists a postive constant b such that* $b < 2\beta$ *and* $b_n \to b$ a.s. *Then*

$$n^{1/2}(x_n - \theta) \xrightarrow{\mathfrak{D}} N(0, (\sigma^2/\beta^2)f(b/\beta)),$$

where f is defined in (4.20).

5. Some Asymptotic Properties of the Cost $\sum_1^n(x_i - \theta)^2$

In this section we prove the following theorem on the order of magnitude of the cost $\sum_1^n (x_i - \theta)^2$ of the recursive scheme $x_{i+1} = \bar{x}_i - (\bar{y}_i - y^*)/b_i$ in the event $E = \{b_n$ converges to a finite positive limit$\}$.

THEOREM 7. *Under the same assumptions and notations as in Theorem* 5,

(i) $\sum_1^n(x_i - \theta)^2/\log n \to (\sigma^2/\beta^2)f(b/\beta)$ a.s. *on* $E \cap \{b < 2\beta\}$;

(ii) $n^{-(1-2\beta/b)}\sum_1^n(x_i - \theta)^2$ *converges* a.s. *on* $E_1 \cap \{b > 2\beta\}$; *moreover, on* $E_1 \cap \{b > 2\beta\} \cap \{\lim_{n\to\infty} n^{-(1-2\beta/b)}\sum_1^n(x_i - \theta)^2 = 0\}$,

$$\sum_1^n (x_i - \theta)^2/\log n \to (\sigma^2/\beta^2)\,|f(b/\beta)| \qquad a.s.; \tag{5.1}$$

(iii) *on* $E_2 \cap \{b = 2\beta\}$,

$$\limsup_{n\to\infty} \sum_1^n (x_i - \theta)^2/\left\{(\log n)^2(\log\log\log n)\right\} = \frac{2}{\pi^2}\frac{\sigma^2}{\beta^2} \qquad a.s. \tag{5.2}$$

$$\liminf_{n\to\infty} \sum_1^n (x_i - \theta)^2/\left\{(\log n)^2/(\log\log\log n)\right\} = (4\beta)^{-2}\sigma^2 \qquad a.s. \tag{5.3}$$

The proof of Theorem 6 makes use of the following result of [14].

LEMMA 4. *Let* $\epsilon, \epsilon_1, \epsilon_2, \ldots$ *be i.i.d. random variables with* $E\epsilon = 0$ *and* $E\epsilon^2 = \sigma^2 < \infty$. *Let* $\tilde{\epsilon}_n(\alpha) = n^{-\alpha}\sum_{j=1}^n j^{\alpha-1}\epsilon_j$ *for* $\alpha \geq \frac{1}{2}$, *and let* $\hat{\epsilon}_n(\alpha) = n^{-\alpha}\sum_{j=n}^\infty j^{\alpha-1}\epsilon_j$ *for* $\alpha < \frac{1}{2}$.

(i) *For every* $b > a > \frac{1}{2}$,

$$P\left[\sum_1^n \tilde{\epsilon}_i^2(\alpha)/\log n \to \sigma^2/(2\alpha - 1) \text{ uniformly in } a \leq \alpha \leq b\right] = 1. \quad (5.4)$$

(ii) *For every* $c < d < \frac{1}{2}$,

$$P\left[\sum_1^n \hat{\epsilon}_i^2(\alpha)/\log n \to \sigma^2/(1 - 2\alpha) \text{ uniformly in } c \leq \alpha \leq d\right] = 1. \quad (5.5)$$

(iii) *For the case* $\alpha = \frac{1}{2}$,

$$\limsup_{n\to\infty} \sum_1^n \tilde{\epsilon}_i^2(\alpha)/\left\{(\log n)^2(\log\log\log n)\right\} = 8\sigma^2/\pi^2 \quad a.s., \quad (5.6)$$

$$\liminf_{n\to\infty} \sum_1^n \tilde{\epsilon}_i^2(\alpha)/\left\{(\log n)^2/(\log\log\log n)\right\} = \sigma^2/4 \quad a.s. \quad (5.7)$$

(iv) *For every* $r > s > \frac{1}{2}$ *and* $0 < \lambda \leq 1$,

$$\limsup_{n\to\infty} \left\{\max_{s\leq\alpha\leq r} \sum_{k=1}^n \left(k^{-\alpha} \sum_{j\leq\lambda k} j^{\alpha-1}|\bar{\epsilon}_j|\right)^2\right\}/(\log n)$$
$$\leq \left(s - \tfrac{1}{2}\right)^{-2}\sigma^2\lambda^{2s-1} \quad \text{a.s.} \quad (5.8)$$

Proof of Theorem 7. To prove (i), we first note that Lemma 4(i) implies

$$P\left[\sum_{k=1}^n \left(k^{-\alpha}\sum_{j=1}^k j^{\alpha-1}\epsilon_j\right)^2 \Big/ \log n \to \sigma^2/(2\alpha - 1) \quad \text{for all } \alpha > \tfrac{1}{2}\right] = 1.$$
$$(5.9)$$

By partial summation and the law of the iterated logarithm (4.10), we obtain that on $E \cap \{b < 2\beta\}$

$$(1 - \beta b^{-1})k^{-\beta/b}\sum_{j=1}^{k-1} j^{\beta/b-1}\bar{\epsilon}_j + \bar{\epsilon}_k = k^{-\beta/b}\sum_{j=1}^k j^{\beta/b-1}\epsilon_j + o(k^{-1/2})$$

a.s.,

and therefore by (5.9),

$$\sum_{k=m}^{n}\left\{(1-\beta b^{-1})k^{-\beta/b}\sum_{j=m}^{k-1} j^{\beta/b-1}\bar{\epsilon}_j+\bar{\epsilon}_k\right\}^2/\log n \to \sigma^2/(2\beta b^{-1}-1) \quad \text{a.s.} \tag{5.10}$$

Making use of Lemma 4(iv), we then obtain from (5.10) that on $E \cap \{b < 2\beta\}$,

$$\sum_{k=m}^{n}\left\{(1-\beta b_k^{-1})k^{-\beta/b}\sum_{j=m}^{k-1} b^{-1}j^{\beta/b-1}\bar{\epsilon}_j+\bar{\epsilon}_k/b_k\right\}^2/\log n$$

$$\to (\sigma^2/\beta^2)f(b/\beta) \qquad \text{a.s.} \tag{5.11}$$

We note from (4.32) that on $E \cap \{b < 2\beta\}$,

$$\sum_{k=m}^{n}(x_{k+1}-\theta)^2=\sum_{k=m}^{n}\left\{(1-\beta b_k^{-1})k^{-\beta/b}\sum_{j=m}^{k-1} b_j^{-1}(j+1)^{\beta/b-1}\bar{\epsilon}_j\tau_k/\tau_{j+1}\right.$$

$$\left.+\bar{\epsilon}_k/b_k+0(k^{-\beta/b}\tau_k)\right\}^2 \tag{5.12}$$

In view of (5.11) and (5.12), it therefore suffices for the proof of (i) to show that

$$\sum_{k=m}^{n}\left\{k^{-\beta/b}\sum_{j=m}^{k-1}|b_j^{-1}(j+1)^{\beta/b-1}\tau_k/\tau_{j+1}-b^{-1}j^{\beta/b-1}||\bar{\epsilon}_j|\right\}^2$$

$$=o(\log n) \qquad \text{a.s.} \quad \text{on } E\cap\{b<2\beta\}. \tag{5.13}$$

To prove (5.13), let $\delta > 0$ be a random variable such that $\beta/b - \delta > \frac{1}{2}$ on $\{b < 2\beta\}$. As indicated in the proof of Theorem 5, $\{\tau_n\}$ is slowly varying on E. Therefore on E, we can choose m so large that

$$\tau_k/\tau_{j+1} \le (k/(j+1))^{\delta} \qquad \text{for } m \le j < k \tag{5.14}$$

(cf. [4]). Let $0 < \lambda < 1$. Since $\inf b_j > 0$ on E, it then follows from (5.14) and Lemma 4(iv) that on $E \cap \{b < 2\beta\}$,

$$\sum_{k=m}^{n} \left\{ k^{-\beta/b} \sum_{m \le j \le \lambda k} |\, b_j^{-1}(j+1)^{\beta/b-1} \tau_k/\tau_{j+1} - b^{-1} j^{\beta/b-1} |\, |\bar{\epsilon}_j| \right\}^2$$

$$\le \sum_{k=m}^{n} \left[\left\{ k^{-\beta/b+\delta} \sum_{j \le \lambda k} b_j^{-1}(j+1)^{\beta/b-\delta-1} |\bar{\epsilon}_j| \right\}^2 + \left\{ k^{-\beta/b} \sum_{j \le \lambda k} b^{-1} j^{\beta/b-1} |\bar{\epsilon}_j| \right\}^2 \right]$$

$$\le (z_\lambda + o(1)) \log n \qquad \text{a. s., where } z_\lambda \to 0 \quad \text{as} \quad \lambda \to 0. \tag{5.15}$$

Since $\{\tau_n\}$ is slowly varying on E, we also obtain that on $E \cap \{b < 2\beta\}$,

$$\sum_{k=m}^{n} \left\{ k^{-\beta/b} \sup_{\lambda k \le j < k} |\, b_j^{-1}((j+1)/j)^{\beta/b-1} \tau_k/\tau_{j+1} - b^{-1} |\, \sum_{j=m}^{k-1} j^{\beta/b-1} |\bar{\epsilon}_j| \right\}^2$$

$$= \sum_{k=m}^{n} o\left(\left\{ k^{-\beta/b} \sum_{j=m}^{k-1} j^{\beta/b-1} |\bar{\epsilon}_j| \right\}^2 \right) = o(\log n) \qquad \text{a.s., by Lemma 4(iv).} \tag{5.16}$$

From (5.15) and (5.16), (5.13) follows.

To prove (ii), we note by Theorem 5(iii) that on $E_1 \cap \{b > 2\beta\}$, $n^{\beta/b}(x_n - \theta)$ converges a.s. to some random variable z, and therefore

$$\sum_{1}^{n} (x_i - \theta)^2 \sim (1 - 2\beta/b)^{-1} n^{1-2\beta/b} z^2 \quad \text{a.s.} \tag{5.17}$$

Moreover, on $E_1 \cap \{b > 2\beta\} \cap \{z = 0\}$, (4.34) holds, and therefore

$$x_{k+1} - \theta = (1 - \beta b_k^{-1}) k^{-\beta/b} \sum_{j=k}^{\infty} b_j^{-1}(j+1)^{\beta/b-1} \bar{\epsilon}_j (\tau_k/\tau_{j+1}) - \bar{\epsilon}_k/b_k, \qquad \text{by (4.34),}$$

$$= (1 - \beta b^{-1}) b^{-1} k^{-\beta/b} \sum_{j=k}^{\infty} j^{\beta/b-1} \bar{\epsilon}_j - \bar{\epsilon}_k/b$$

$$+ 0\left(k^{-1/2} (\log k)^{-(\rho-1)} (\log\log k)^{1/2}\right) \qquad \text{a.s., by (4.10), (4.30), (4.31),}$$

$$= b^{-1} k^{-\beta/b} \sum_{j=k}^{\infty} j^{\beta/b-1} \epsilon_j + o(k^{-1/2}) \qquad \text{a.s.} \tag{5.18}$$

The last equality above follows from partial summation and (4.10). By Lemma 4(ii),

$$P\left[\sum_{k=m}^{n}\left(k^{-\alpha}\sum_{j=k}^{\infty} j^{\alpha-1}\epsilon_j\right)^2 / \log n \to \sigma^2/(1-2\alpha) \quad \text{for all } \alpha < \tfrac{1}{2}\right] = 1. \tag{5.19}$$

From (5.18) and (5.19), it then follows that on $E_1 \cap \{b > 2\beta\} \cap \{z = 0\}$,

$$\sum_1^n (x_i - \theta)^2/\log n \to b^{-2}\sigma^2/(1 - 2\beta b^{-1}) = (\sigma^2/\beta^2)\,|f(b/\beta)| \qquad \text{a.s.}$$

To prove (iii), since (4.36) holds on $E_2 \cap \{b = 2\beta\}$, we have on $E_2 \cap \{b = 2\beta\}$

$$\sum_{k=m}^{n} (x_{k+1} - \theta)^2 = (2\beta)^{-2} \sum_{k=m}^{n} k^{-1}\left\{\sum_{j=m}^{k} j^{-1/2}\epsilon_j + o\left((\log k)^{\delta}\right)\right\}^2 \qquad \text{a.s.,} \tag{5.20}$$

where $\delta < \frac{1}{2}$. From (5.20) and Lemma 4(iii), the desired conclusion follows. □

Noting that $\Sigma_1^n(x_i - \bar{x}_n)^2 = \Sigma_1^n(x_i - \theta)^2 - n(\bar{x}_n - \theta)^2$, we can combine Theorem 7 with Theorem 5 to obtain

COROLLARY 1. *Under the same assumptions and notations as in Theorem 5,*

(i) $\Sigma_1^n(x_i - \bar{x}_n)^2/\log n \to (\sigma^2/\beta^2)f(b/\beta)$ a.s. *on* $E \cap \{b < 2\beta\}$;

(ii) *on* $E_1 \cap \{b > 2\beta\}$, $n^{-(1-2\beta/b)}\sum_1^n(x_i - \bar{x}_n)^2$ *converges a.s. to* $(\beta/b)^2z^2/\{(1 - 2\beta/b)(1 - \beta/b)^2\}$, *where* $z = \lim_{n\to\infty} n^{\beta/b}(x_n - \theta)$; *moreover, on* $E_1 \cap \{b > 2\beta\} \cap \{z = 0\}$,

$$\sum_1^n (x_i - \bar{x}_n)^2/\log n \to (\sigma^2/\beta^2)\,|f(b/\beta)| \qquad a.s.; \tag{5.21}$$

(iii) *on* $E_2 \cap \{b = 2\beta\}$,

$$\limsup_{n\to\infty} \sum_1^n (x_i - \bar{x}_n)^2/\left\{(\log n)^2(\log\log\log n)\right\} = \frac{2}{\pi^2}\frac{\sigma^2}{\beta^2} \qquad a.s. \tag{5.22}$$

$$\liminf_{n\to\infty} \sum_1^n (x_i - \bar{x}_n)^2/\left\{(\log n)^2/(\log\log\log n)\right\} = (4\beta)^{-2}\sigma^2 \qquad a.s. \tag{5.23}$$

6. An Asymptotically Efficient Modification of the LSCE Rule

In this section we assume that positive lower and upper bounds B_1 and B_2 for the slope β in the linear regression model (3.1) are known. We do not, however, assume the knowledge of bounds K_1, K_2 on θ, as assumed by Anderson and Taylor [1]. In ignorance of bounds on θ, we have to set $K_1 = -\infty$ and $K_2 = \infty$ in the LSCE rule (1.4), and this amounts to the recursive scheme (1.15) with $b_i = \hat{\beta}_i$. On the other hand, since upper and lower bounds B_2 and B_1 on β are known, it is natural to truncate the least squares estimate $\hat{\beta}_i$ by these bounds and therefore to take $b_i = B_2 \wedge (\hat{\beta}_i \vee B_1)$ in (1.15).

For the case of a fixed design in which $x_1, x_2, \ldots$ are nonrandom constants, $\hat{\beta}_n$ is an unbiased estimate of β and has variance $\sigma^2/\Sigma_1^n(x_i - \bar{x}_n)^2$, and the strong consistency of $\hat{\beta}_n$ under the sole condition that $\Sigma_1^n(x_i - \bar{x}_n)^2 \to \infty$ was recently established in [8]. This condition, however, is not sufficient to ensure the strong consistency of $\hat{\beta}_n$ when the x_i are sequentially determined random variables (cf. [12]). For the recursive scheme (1.15), we obtain from Corollary 1(i) that on $\{\lim_{n\to\infty} b_n = \beta\}$,

$$\sum_1^n (x_i - \bar{x}_n)^2 \sim (\sigma^2/\beta^2)\log n \qquad \text{a.s.,} \tag{6.1}$$

and therefore

$$n^{-1}\sum_1^n (y_i - \bar{y}_n)^2 = n^{-1}\sum_1^n \{\beta(x_i - \bar{x}_n) + (\epsilon_i - \bar{\epsilon}_n)\}^2 \to \sigma^2 \qquad \text{a.s.} \tag{6.2}$$

Let $s_n^2 = n^{-1}\Sigma_1^n(y_i - \bar{y}_n)^2$. Since $\beta \le B_2$, it then follows from (6.1) and (6.2) that

$$\liminf_{n\to\infty} \sum_1^n (x_i - \bar{x}_n)^2/\log n \ge s_n^2/B_2^2 \qquad \text{a.s. on } \left\{\lim_{n\to\infty} b_n = \beta\right\}. \tag{6.3}$$

Let $\{c_n\}$ be any sequence of positive constants such that

$$\liminf_{n\to\infty} c_n > 0, \qquad \limsup_{n\to\infty} c_n < 1. \tag{6.4}$$

From (6.3), it follows that on $\{\lim_{n\to\infty} b_n = \beta\}$,

$$\sum_1^n (x_i - \bar{x}_n)^2 > (c_n s_n^2/B_2^2)\log n \tag{6.5}$$

for all large n, with probability 1. Noting also that the accuracy of the least-squares estimate $\hat{\beta}_n$ of β is closely related to the magnitude of $\Sigma_1^n(x_i - \bar{x}_n)^2$, we therefore define b_n for the recursive scheme (1.15) as follows:

$$b_n = B_2 \wedge (\hat{\beta}_n \vee B_1) \qquad \text{if (6.5) holds,}$$
$$= b_{n-1} \qquad \text{otherwise,} \tag{6.6}$$

where b_1 is any constant between B_1 and B_2. We shall call the recursive scheme (1.15) with b_n defined by (6.6) the *modified LSCE rule*.

Making use of the local convergence properties in Corollary 1 and Theorem 4 for recursive schemes of the form (1.15) and a general theorem on the strong consistency of $\hat{\beta}_n$ in stochastic designs, it can be shown that $b_n \to \beta$ a.s. in the modified LSCE rule. The details of the proof are given in [13]. It then follows from Theorems 5(i), 6, and 7(i) that the asymptotic properties (1.9), (1.10), and (1.11) for the asymptotically optimal Robbins–Monro stochastic approximation scheme (1.8) (or its least-squares equivalent (1.12)) assuming known β still hold for the modified LSCE rule in the present case of unknown β. Hence this modification of the LSCE rule has the desirable convergence properties of the Anderson–Taylor conjecture.

References

1. T. W. Anderson and J. Taylor, Some experimental results on the statistical properties of least squares estimates in control problems, *Econometrica* **44** (1976), 1289–1302.
2. M. Aoki, On some price adjustment schemes, *Ann. Econ. Soc. Measure.* **3** (1974), 95–116.
3. J. R. Blum, Approximation methods which converge with probability one, *Ann. Math. Statist.* **25** (1954), 382–386.
4. R. Bojanic and E. Seneta, A unified theory of regularly varying sequences, *Math. Z.* **134** (1973), 91–106.
5. L. Breiman, "Probability," Addison-Wesley, Reading, Mass., 1968.
6. K. L. Chung, On a stochastic approximation method, *Ann. Math. Statist.* **25** (1954), 463–483.
7. N. C. Jain, K. Jogdeo, and W. F. Stout, Upper and lower functions for martingales and mixing processes, *Ann. Probab.* **3** (1975), 119–145.
8. T. L. Lai and H. Robbins, Strong consistency of least squares estimates in regression models, *Proc. Nat. Acad. Sci. USA* **74** (1977), 2667–2669.
9. T. L. Lai and H. Robbins, Limit theorems for weighted sums and stochastic approximation processes, *Proc. Nat. Acad. Sci. USA* **75** (1978), 1068–1070.
10. T. L. Lai and H. Robbins, Adaptive design and stochastic approximation, *Ann. Statist.* **7** (1979), 1196–1221.
11. T. L. Lai and H. Robbins, Local convergence theorems for adaptive stochastic approximation schemes, *Proc. Nat. Acad. Sci. USA* **76** (1979), 3065–3067.
12. T. L. Lai and H. Robbins, Consistency and asymptotic efficiency of slope estimates in stochastic approximation schemes, *Z. Wahrsch. Verw. Gebiete* **56** (1981), 329–360.
13. T. L. Lai and H. Robbins, Adaptive design and the multiperiod control problem, *in*

"Proceedings, Third Purdue Symposium on Statistical Decision Theory and Related Topics" (S. Gupta, Ed.), in press.

14. T. L. LAI, On Cesàro and other weighted means of independent random variables, to appear.
15. E. C. PRESCOTT, The multiperiod control problem under uncertainty, *Econometrica* **40** (1972), 1043–1058.
16. H. ROBBINS AND S. MONRO, A stochastic approximation method, *Ann. Math. Statist.* **22** (1951), 400–407.
17. A. ZELLNER, "An Introduction to Bayesian Inference in Econometrics," Wiley, New York, 1971.