The Annals of Statistics
1983, Vol. 11, No. 2, 478–485

FIXED ACCURACY ESTIMATION OF AN AUTOREGRESSIVE PARAMETER

BY T. L. LAI[1] AND D. SIEGMUND[2]

Columbia University and Stanford University

For a first order non-explosive autoregressive process with unknown parameter $\beta \in [-1, 1]$, it is shown that if data are collected according to a particular stopping rule, the least squares estimator of β is asymptotically normally distributed uniformly in β. In the case of normal residuals, the stopping rule may be interpreted as sampling until the observed Fisher information reaches a preassigned level. The situation is contrasted with the fixed sample size case, where the estimator has a non-normal unconditional limiting distribution when $|\beta| = 1$.

1. Introduction and summary. Consider the first order, non-explosive, autoregressive model

$$x_n = \beta x_{n-1} + \varepsilon_n, \quad n = 1, 2, \cdots, \tag{1.1}$$

where $\varepsilon_1, \varepsilon_2, \cdots$ are independent, identically distributed random variables with $E\varepsilon_1 = 0$ and $0 < E\varepsilon_1^2 = \sigma^2 < \infty$. The initial state x_0 is a random variable (not depending on β) which is stochastically independent of $\{\varepsilon_n\}$. The constant $\beta \in [-1, 1]$ is an unknown parameter, which at stage n is customarily estimated by the least squares estimate,

$$b_n = (\textstyle\sum_{i=1}^n x_{i-1}x_i)/(\sum_{i=1}^n x_{i-1}^2) = \beta + (\sum_{i=1}^n x_{i-1}\varepsilon_i)/(\sum_1^n x_{i-1}^2). \tag{1.2}$$

If the ε's are normally distributed, then b_n is also the maximum likelihood estimator of β, and the observed Fisher information about β contained in $x_0, x_1, \cdots, x_n$ is

$$I_n = -\frac{d^2}{d\beta^2}\left(\beta \textstyle\sum_{i=1}^n x_{i-1}x_i - \frac{1}{2}\beta^2 \sum_{i=1}^n x_{i-1}^2\right) = \sum_{i=1}^n x_{i-1}^2. \tag{1.3}$$

It is well-known and easy to prove that for fixed $\beta \in (-1, 1)$, as $n \to \infty$

$$I_n^{1/2}(b_n - \beta) \to_{\mathscr{L}} N(0, \sigma^2). \tag{1.4}$$

Here $N(\mu, \sigma^2)$ denotes a normal random variable with mean μ and variance σ^2, and $\to_{\mathscr{L}}$ indicates convergence in law. See, for example, Anderson (1959). However, for $\beta = \pm 1$, an entirely different limiting distribution occurs (White, 1958, Rao, 1978). For example, if $\beta = 1$, $x_n - x_0$ is the sum of i.i.d. random variables $\varepsilon_1 + \cdots + \varepsilon_n$, and summation by parts in (1.2) yields

$$(\textstyle\sum_{i=1}^n x_{i-1}^2)^{1/2}(b_n - 1) = \frac{1}{2}n^{-1}(x_n^2 - x_0^2 - \sum_1^n \varepsilon_i^2)/(n^{-2}\sum_{i=1}^n x_{i-1}^2)^{1/2}.$$

By Donsker's theorem this converges in law to

$$\frac{\sigma}{2}\{W^2(1) - 1\} \Big/ \left\{\int_0^1 W^2(t)\, dt\right\}^{1/2}. \tag{1.5}$$

where $\{W(t), 0 \le t \le 1\}$ is a standard Brownian motion process. Of course, this result indicates that the asymptotic normality of (1.4) breaks down for β in a neighborhood of

Received May 1982; revised November 1982.
[1] Research supported by National Institutes of Health and the National Science Foundation.
[2] Research supported by ONR Contract N00014-77-C-0306 and by NSF Grant MCS 80-24649.
AMS 1980 *subject classification.* Primary 62L10.
Key words and phrases. Stopping rule, fixed width confidence interval, uniform asymptotic normality.

± 1, in the sense that for any given n, no matter how large, there will be a neighborhood of ± 1 in which one should not expect (1.4) to yield reasonable approximations.

Examples in econometrics having values of β close to 1 are cited by Evans and Savin (1981) and by Dickey and Fuller (1979).

In this paper we consider the asymptotic behavior of $\{b_n\}$ under a sequential sampling scheme which measures time in terms of accumulated (observed) Fisher information. Define

$$N_c = \text{first } n \geq 1 \quad \text{such that} \quad I_n \geq c\sigma^2. \tag{1.6}$$

Our principal result (Theorem 2.1) is that as $c \to \infty$

$$I_{N_c}^{1/2}(b_{N_c} - \beta) \to_{\mathscr{L}} N(0, \sigma^2) \tag{1.7}$$

uniformly in β for $-1 \leq \beta \leq 1$.

The sampling rule (1.6) is motivated by the theory of fixed width confidence intervals, cf. Anscombe (1953), Chow and Robbins (1965), or Grambsch (1982). Some precedent for the rather surprising uniformity in the convergence of (1.7) is found in the work of Siegmund (1981), although the underlying reasons are quite different in the present case. A discussion of this problem in the comparatively trivial continuous time case is given by Lipster and Shiryaev (1978).

When it is feasible to use the sampling rule (1.6), its advantages appear to be threefold: (i) the accuracy of b_{N_c} as an estimator of β (as measured by the variance of its asymptotic distribution) is approximately a small constant c^{-1}, rather than an uncontrolled random variable, (ii) the appropriate asymptotic distribution theory does not depend on the value of the unknown parameter β; and (iii) the convergence to asymptotic normality is much more rapid, even when β is not near the values ± 1. (See Section 3.)

The remainder of this paper is arranged as follows. Section 2 contains a proof of (1.7). In Section 3 we give the results of some simulations comparing confidence intervals obtained by indiscriminate use of (1.4) for fixed sample sizes with (1.7) for sequentially determined sample sizes. Section 4 is concerned with some related asymptotic results, and in particular the appropriate modification of (1.6) when σ is unknown.

Partial results for the model $x_n = \alpha + \beta x_{n-1} + \varepsilon_n$ indicate that the multiparameter case can be appreciably more complicated, because if α is close to 0, one must recognize this and use a stopping rule equivalent to (1.6); otherwise a substantially different stopping rule is required.

2. Uniform asymptotic normality of b_{N_c}. The main result of this section is the proof of (1.7) (Theorem 2.1 below). Our approach is motivated by the observation that $b_n - \beta = \sum_{i=1}^n x_{i-1}\varepsilon_i / \sum_{i=1}^n x_{i-1}^2$ is of the form of a martingale divided by the sum of the conditional variances of its increments. The martingale central limit theorem (e.g. Dvoretzky, 1972) almost immediately implies (1.7) for each *fixed* $\beta \in [-1, 1]$, but the requirement of uniformity demands a substantially more complicated argument. The novelty of our approach lies in systematic exploitation of the condition (2.6) and the simple identity (2.16). We begin with some preliminary probabilistic results.

PROPOSITION 2.1. *Let $x_n, \varepsilon_n, n = 0, 1, \cdots$ be random variables adapted to the increasing sequence of σ-algebras $\mathscr{F}_n$, $n = 0, 1, \cdots$.* Let $\{P_\theta, \theta \in \Xi\}$ *be a family of probability measures such that under every P_θ*

$$\varepsilon_1, \varepsilon_2, \cdots \quad \textit{are } \text{i.i.d.} \textit{ with} \quad E_\theta\varepsilon_1 = 0, \quad E_\theta\varepsilon_1^2 = 1; \tag{2.1}$$

$$\sup_\theta E_\theta\{\varepsilon_1^2; |\varepsilon_1| > a\} \to 0 \quad \textit{as} \quad a \to \infty; \tag{2.2}$$

$$\varepsilon_n \quad \textit{is independent of} \quad \mathscr{F}_{n-1} \quad \textit{for each} \quad n \geq 1; \tag{2.3}$$

$$P_\theta\{\textstyle\sum_{i=0}^\infty x_i^2 = \infty\} = 1; \tag{2.4}$$

(2.5) $$\sup_\theta P_\theta\{x_n^2 > a\} \to 0 \quad \text{as} \quad a \to \infty \quad \text{for each} \quad n \geq 0;$$

and for each $\delta > 0$

(2.6) $$\lim_{m\to\infty}[\sup_\theta P_\theta\{x_n^2 \geq \delta \textstyle\sum_{i=0}^{n-1} x_i^2 \quad \text{for some} \quad n \geq m\}] = 0.$$

For $c > 0$ *let*

(2.7) $$T_c = \inf\{n: \textstyle\sum_1^n x_{i-1}^2 \geq c\} \ (\inf \phi = +\infty).$$

Then uniformly in $\theta \in \Xi$ *and* $-\infty < t < \infty$

(2.8) $$P_\theta\{c^{-1/2} \textstyle\sum_{i=1}^{T_c} x_{i-1}\varepsilon_i \leq t\} \to \Phi(t) \quad \text{as} \quad c \to \infty,$$

where Φ *is the standard normal distribution function.*

REMARK. For the autoregressive model (1.1), if we identify θ with the autoregressive parameter β, all the conditions of Proposition 2.1 are trivially verified, with the single exception of (2.6).

Proposition 2.1 is proved by reducing it to the following convenient martingale central limit theorem (cf. Freedman, 1971, pages 90–92).

LEMMA 2.1. *Let* $0 < \delta < 1$ *and* $r > 0$. *Assume that* $\{u_n, \mathscr{F}_n, n \geq 0\}$ *is a martingale difference sequence satisfying*

(2.9) $$|u_n| \leq \delta \quad \text{for all} \quad n$$

and

(2.10) $$\textstyle\sum_1^\infty E(u_n^2 \mid \mathscr{F}_{n-1}) > r \quad \text{a.s.}$$

Let

$$\tau = \inf\{n: \textstyle\sum_1^n E(u_i^2 \mid \mathscr{F}_{i-1}) \geq r\}.$$

There exists a function ρ: $(0, \infty) \to [0, 2]$, *not depending on the distribution of the martingale difference sequence, such that* $\lim_{x\to 0}\rho(x) = 0$ *and*

$$\sup_x |P\{\textstyle\sum_1^\tau u_i \leq x\} - \Phi(x/r^{1/2})| \leq \rho(\delta/r^{1/2}).$$

PROOF OF PROPOSITION 2.1. By (2.4) $P_\theta\{T_c < \infty\} = 1$, and obviously $P_\theta\{\lim_{c\to\infty} T_c = \infty\} = 1$ for all θ. Let $0 < \delta < 1$ and define $\tilde{x}_n = x_n$ if $x_n^2 \leq \delta^2 c$ and $\tilde{x}_n = \delta c^{1/2}$ otherwise. Then for all θ

$$\begin{aligned} P_\theta\{x_n \neq \tilde{x}_n \quad &\text{for some} \quad n < T_c\} \\ &\leq \textstyle\sum_{i=1}^m P_\theta\{x_{i-1}^2 > \delta^2 c\} + P_\theta\{T_c > m, x_n \neq \tilde{x}_n \quad \text{for some} \quad m \leq n < T\} \\ &\leq \textstyle\sum_{i=1}^m P_\theta\{x_{i-1}^2 > \delta^2 c\} + P_\theta\{x_n^2 \geq \delta^2 \textstyle\sum_0^{n-1} x_i^2 \quad \text{for some} \quad n \geq m\}, \end{aligned}$$

which is bounded above by 2δ if we first choose m large enough and use (2.6), then choose c large and use (2.5). Hence if

$$\Omega_c = \{x_n = \tilde{x}_n \quad \text{for all} \quad n < T_c\},$$

then for all large c and all θ

(2.11) $$P_\theta(\Omega_c) \geq 1 - 2\delta.$$

Define $\tilde{\varepsilon}_n = \varepsilon_n I\{|\varepsilon_n| \leq \delta^{-1/2}\}$ and $\tilde{\tilde{\varepsilon}}_n = \varepsilon_n - \tilde{\varepsilon}_n$. Then under P_θ, $\{c^{-1/2}\tilde{x}_{n-1}(\tilde{\varepsilon}_n - E_\theta\tilde{\varepsilon}_n), \mathscr{F}_n, 0 \leq n < \infty\}$ is a martingale difference sequence satisfying

$$|c^{-1/2}\tilde{x}_{n-1}(\tilde{\varepsilon}_n - E_\theta\tilde{\varepsilon}_n)| \leq 2\delta^{1/2},$$

and by (2.4)

$$P_\theta\{\sum_0^\infty \tilde{x}_n^2 = \infty\} = 1.$$

Also by (2.2), as $\delta \to 0$

$$(2.12) \qquad v_\theta(\delta) = \operatorname{var}_\theta \tilde{\varepsilon}_1 \to 1$$

uniformly in θ. Hence, if

$$\tau_c = \inf\{n: \sum_1^n \tilde{x}_{i-1}^2 \ge c\},$$

then by Lemma 2.1

$$(2.13) \qquad |P_\theta\{c^{-1/2}\sum_1^\tau \tilde{x}_{i-1}(\tilde{\varepsilon}_i - E_\theta\tilde{\varepsilon}_i) \le t\} - \Phi(t/(v_\theta(\delta))^{1/2})| \le \rho[2(\delta/v_\theta(\delta))^{1/2}].$$

On Ω_c, $\tau_c = T_c$ and

$$(2.14) \qquad |\sum_1^\tau \tilde{x}_{i-1}(\tilde{\varepsilon}_i - E_\theta\tilde{\varepsilon}_i) - \sum_1^T x_{i-1}\varepsilon_i| = |\sum_1^\tau \tilde{x}_{i-1}(\tilde{\tilde{\varepsilon}}_i - E_\theta\tilde{\tilde{\varepsilon}}_i)|.$$

By Wald's identity (cf. Chow, Robbins, and Siegmund, 1971, page 23)

$$(2.15) \qquad E_\theta[c^{-1/2}\sum_1^\tau \tilde{x}_{i-1}(\tilde{\tilde{\varepsilon}}_i - E_\theta\tilde{\tilde{\varepsilon}}_i)]^2 = c^{-1}(1 - v_\theta(\delta))E_\theta(\sum_1^\tau \tilde{x}_{i-1}^2) \le (1 - v_\theta(\delta))(1 + \delta^2).$$

The Proposition follows from (2.11) – (2.15) by letting $\delta \to 0$.

For ease of reference we state without proof the following lemma. It is related to the strong law of large numbers given e.g. by Neveu (1965, page 148), whose method of proof can be adapted to the present purpose. Alternatively it follows easily from Proposition 2 of Robbins and Siegmund (1971) and a straightforward calculation along the lines of their Lemma 1.

LEMMA 2.2. *Suppose that the measurability conditions of Proposition* 2.1, (2.1), *and* (2.3) *are satisfied. For each* $\gamma > \frac{1}{2}$, $\delta > 0$, *and increasing sequence of positive constants* $c_n \to \infty$,

$$\sup_\theta P_\theta\{|\sum_1^n x_{i-1}\varepsilon_i| \ge \delta \max(c_n, (\sum_1^n x_{i-1}^2)^\gamma) \text{ for some } n \ge m\} \to 0$$

as $m \to \infty$.

We now return to the autoregressive model (1.1) and write P_β to denote dependence of probabilities on the parameter β. (The joint distribution of $x_0, \varepsilon_1, \varepsilon_2, \cdots$, however, is assumed not to depend on β.) The principal result of this section is

THEOREM 2.1. *Define* $\{b_n, n = 2, 3, \cdots\}$ *by* (1.2) *and* N_c *by* (1.6). *If* $\varepsilon_1, \varepsilon_2, \cdots$ *are* i.i.d. *with mean* 0 *and variance* σ^2, *and are independent of* x_0, *then*

$$\lim_{c\to\infty} P_\beta\{I_{N_c}^{1/2}(b_{N_c} - \beta) \le t\} = \Phi(t/\sigma)$$

uniformly for $-1 \le \beta \le 1$ *and* $-\infty < t < \infty$.

PROOF. Theorem 2.1 follows immediately from Proposition 2.1 once we have verified (2.6). To this end, note that squaring (1.1) and summing yields

$$(2.16) \qquad x_n^2 + (1 - \beta^2)\sum_{i=0}^{n-1} x_i^2 - x_0^2 = \sum_1^n \varepsilon_i^2 + 2\beta\sum_{i=1}^n x_{i-1}\varepsilon_i.$$

Let $|\beta| \le 1$ and $0 < \lambda < \sigma^2/4$, and define

$$(2.17) \qquad \Omega_{m,\lambda} = \{|n^{-1}\sum_1^n \varepsilon_i^2 - \sigma^2| < \lambda, |\sum_1^n x_{i-1}\varepsilon_i| < \max(\lambda n, (\sum_1^n x_{i-1}^2)^{2/3}) \text{ for all } n \ge m\}.$$

On $\Omega_{m,\lambda}$ if $n \ge m$ and $x_n^2 \le \lambda n$, then (2.16) implies

$$\sum_0^{n-1} x_i^2 \ge (\sigma^2 - \lambda)n - \lambda n - 2\max\{\lambda n, (\sum_1^n x_{i-1}^2)^{2/3}\}$$

and hence for all m sufficiently large

$$2\sum_0^{n-1} x_i^2 \geq (\sigma^2 - 4\lambda)n \geq (\sigma^2\lambda^{-1} - 4)x_n^2. \tag{2.18}$$

On the other hand, since $|x_{n-1}| \geq |\beta x_{n-1}| \geq |x_n| - |\varepsilon_n|$, it follows that

$$\min_{1\leq j\leq k} |x_{n-j}| \geq |x_n| - \sum_{j=0}^{k-1} |\varepsilon_{n-j}| \geq |x_n|\{1 - \sum_{j=0}^{k-1} |\varepsilon_{n-j}|/(\lambda n)^{1/2}\}$$

if $x_n^2 \geq \lambda n$. Hence $n \geq k$ and $x_n^2 \geq \lambda n$ imply

$$\sum_0^{n-1} x_i^2/x_n^2 \geq k\{1 - \sum_{j=0}^{k-1} |\varepsilon_{n-j}|/(\lambda n)^{1/2}\}^2; \tag{2.19}$$

and since $\sum_{j=0}^{k-1} |\varepsilon_{n-j}|$, $n = k, k+1, \cdots$ are identically distributed with finite second moment, the right hand side of (2.19) converges to k with probability one, at a rate which does not involve β. Hence by choosing λ small and k so that $2/(\sigma^2\lambda^{-1} - 4) < \delta$ and $k > 1/\delta$, we see by (2.18) and (2.19) that

$$\lim_{m\to\infty}\sup_{|\beta|\leq 1} P_\beta\{x_n^2 \geq \delta \sum_0^{n-1} x_i^2 \quad \text{for some} \quad n \geq m\} \leq \lim_{m\to\infty}\sup_{|\beta|\leq 1} P_\beta(\Omega_{m,\lambda}^c),$$

which equals 0 by Lemma 2.2 and the strong law of large numbers. This establishes (2.6) and hence the theorem.

3. Monte Carlo results. In this section we report the results of a Monte Carlo experiment to compare the fixed sample size and sequential asymptotic distributions. The basic experiment to assess the accuracy of the normal approximation indicated by (1.4) and (1.7) consisted of a frequency count of the number of times the normalized estimator of β exceeded z or was less than $-z$ for commonly used quantiles z of the standard normal distribution. Since very similar results were obtained for various z, we report here only for $z = 1.28$, for which the (one-tailed) probabilities are nominally $p_r = p_\ell = 0.10$. For simplicity $x_0 = 0$, and $\varepsilon_1, \varepsilon_2, \cdots$ were taken to be $N(0, 1)$. In the fixed sample experiment, $n = 50$ observations were taken; and the results are reported in Table 1. In Table 2 observations are taken sequentially with c chosen so that $E_\beta N_c \cong 50$ for all β; in Table 3, $c = 50$ and $E_\beta N_c$ varies with β. There were 1600 replications.

For normally distributed ε_i, $b_n(b_{N_c})$ fails to be normally distributed only because of variability in $\sum_1^n x_{i-1}^2$ $(\sum_1^{N_c} x_{i-1}^2)$. (See Dvoretzky, 1972, page 520.) Of course, the sequential experiment is designed to reduce this variability. The columns of Tables 1–3 with the headings $E(I)$ and $SD(I)$ report the observed average and standard deviation of $\sum x_{i-1}^2$ respectively. Variability for $\sum x_{i-1}^2$ in the fixed sample experiment goes into variability in N_c for the sequential experiment, so Tables 2 and 3 also report estimates of EN and $SD(N)$. Note that variability in $\sum x_{i-1}^2$ for the fixed sample size case and $SD(N)/EN$ for the sequential case increase dramatically as β approaches one.

The columns headed $\hat{p}_r$ and $\hat{p}_\ell$ report the percentage of excesses in the right and left tails of the distributions.

The figures in Tables 1 and 2 indicate that the fixed sample size asymptotic theory is not especially good when $n = 50$, even for small $|\beta|$, and it deteriorates quite noticeably for β near 1. In the sequential case the asymptotic theory is much better and shows no dependence upon the value of β.

4. Additional asymptotic theory. Here we give some additional asymptotic results, which follow from the techniques developed in Section 2. Theorem 4.1 describes the asymptotic behavior of N_c, and to some extent explains the rather surprising differences between the fixed sample and sequential cases. Theorem 4.2 is concerned with the uniform strong consistency of b_n and $\hat{\sigma}_n^2 = n^{-1}\sum_{i=1}^n (x_i - b_n x_{i-1})^2$. It provides the foundation for consideration of the case of unknown σ.

THEOREM 4.1. *Under the conditions of Theorem* 2.1,

(i) *for each* $\beta \in (-1, 1)$, $P_\beta\{\lim_{c\to\infty} c^{-1}N_c = (1 - \beta^2)\} = 1$,

TABLE 1
Fixed Sample Case
$n = 50$, $z = 1.28$, p_r *and* p_ℓ *are nominally* 0.10

β	$\hat{p}_r$	$\hat{p}_\ell$	$E(I)$	$SD(I)$
0.1	.084	.108	49	10
0.5	.086	.118	65	18
0.9	.052	.136	232	136
1.0	.048	.166	1272	1505

TABLE 2
Sequential Case
$z = 1.28$, p_r *and* p_ℓ *are nominally* 0.10

β	$\hat{p}_r$	$\hat{p}_\ell$	c	$E(I)$	$SD(I)$	EN	$SD(N)$
0.1	.094	.105	50	52	1.7	52	10
0.5	.103	.096	65	67	2.2	52	13
0.9	.103	.099	200	207	8.0	51	21
1.0	.098	.101	500	524	22	48	24

TABLE 3
Sequential Case
$c = 50$, $z = 1.28$, p_r *and* p_ℓ *are nominally* 0.10

β	$\hat{p}_r$	$\hat{p}_\ell$	$E(I)$	$SD(I)$	EN	$SD(N)$
.5	.097	.099	52	2.4	41	11
.9	.092	.102	55	5.4	20	9.0
1.0	.108	.107	58	8.2	16	7.6
1.1	.096	.093	61	10	13	5.9

and

(ii) *for* $|\beta| = 1$, $c^{-1/2}N_c \to_{\mathscr{L}} \inf\left\{t: \int_0^t W^2(s)\,ds = 1\right\}$,

where $W(t)$, $0 \le t < \infty$, *is a standard Brownian motion process.*

REMARK. In addition to containing information on the sample size of our sequential procedure, Theorem 4.1 has interesting connections with Theorem 2.1 and a theorem of Anscombe (1952). As generalized by Mogyoródi (1962), Anscombe's theorem says that if $Y_n \to_{\mathscr{L}} Y$, some additional technical conditions are satisfied, and if $\nu(c)$ are integer-valued random variables which can be normalized by constants $k(c) \to +\infty$ in such a way that $\nu(c)/k(c)$ converges *in probability* to a positive random variable, then $Y_{\nu(c)} \to_{\mathscr{L}} Y$. Hence (1.4), Theorem 4.1 (i), and Anscombe's theorem show that $I_{N_c}^{1/2}(b_{N_c} - \beta) \to_{\mathscr{L}} N(0, 1)$ for each fixed $\beta \in (-1, 1)$. However, because the convergence of Theorem 4.1 (ii) is in law and not in probability, Anscombe's theorem is not applicable for $|\beta| = 1$, and in fact its conclusion would be incorrect.

PROOF OF THEOREM 4.1. On the event $\Omega_{m,\lambda}$ defined in (2.17), (2.16) implies for all $n \ge m$

$$|x_n^2 - x_0^2 + (1 - \beta^2)\textstyle\sum_1^n x_{i-1}^2 - n\sigma^2| \le 3\lambda n + 2(\sum_1^n x_{i-1}^2)^{2/3}. \tag{4.1}$$

If $|\beta|$ is bounded away from 1, say $|\beta| \le \rho < 1$, so $1 - \beta^2 \ge 1 - \rho^2 > 0$, then (2.6), (4.1), and $P_\beta(\Omega_{m,\lambda}) \to 1$ (uniformly in β) imply that

$$\lim_{m\to\infty}\sup_{|\beta|\le\rho} P_\beta\{|(1 - \beta^2)n^{-1}\textstyle\sum_1^n x_{i-1}^2 - \sigma^2| \ge \delta \quad \text{for some} \quad n \ge m\} = 0. \tag{4.2}$$

Theorem 4.1 (i) follows easily from (4.2).

If $\beta = 1$, then $x_n = x_0 + S_n$, where $S_n = \varepsilon_1 + \cdots + \varepsilon_n$. Thus Theorem 4.1 (ii) follows from Donsker's Theorem, which implies that

$$\{c^{-1}\textstyle\sum_{i=1}^{[c^{1/2}t]} S_i^2,\ t \geq 0\} \to_{\mathscr{L}} \left\{\sigma^2 \int_0^t W^2(s)\, ds, \quad t \geq 0\right\}.$$

A similar but slightly more complicated argument handles the case $\beta = -1$.

THEOREM 4.2. *Define the least squares estimate for* β *as in* (1.2), *and in the case of unknown* σ, *estimate* σ^2 *at stage n by*

$$\hat{\sigma}_n^2 = n^{-1} \textstyle\sum_{i=1}^n (x_i - b_n x_{i-1})^2. \tag{4.3}$$

Then b_n *and* $\hat{\sigma}_n^2$ *are uniformly consistent for* $|\beta| \leq 1$ *in the sense that for all* $\delta > 0$

$$\lim_{m\to\infty} \sup_{|\beta|\leq 1} P_\beta\{|b_n - \beta| \geq \delta \quad \textit{for some} \quad n \geq m\} = 0 \tag{4.4}$$

and

$$\lim_{m\to\infty} \sup_{|\beta|\leq 1} P_\beta\{|\hat{\sigma}_n^2 - \sigma^2| \geq \delta \quad \textit{for some} \quad n \geq m\} = 0. \tag{4.5}$$

PROOF. From (2.6), (2.16), and (4.2) it is easy to see that for all $0 < \lambda < 1$

$$\sup_{|\beta|\leq 1} P_\beta\{n^{-1} \textstyle\sum_1^n x_{i-1}^2 \leq \lambda\sigma^2 \quad \textit{for some} \quad n \geq m\} \to 0, \quad (m \to \infty),$$

which together with (1.2) and Lemma 2.2 proves (4.4).

To prove (4.5) note that

$$\hat{\sigma}_n^2 = n^{-1} \textstyle\sum_{i=1}^n (\varepsilon_i + (\beta - b_n)x_{i-1})^2 = n^{-1} \sum_1^n \varepsilon_i^2 - (\sum_{i=1}^n x_{i-1}\varepsilon_i)^2/n \sum_{i=1}^n x_{i-1}^2. \tag{4.6}$$

For $|\beta| \leq 1$, $|x_n| \leq |x_0| + \sum_1^n |\varepsilon_i| = U_n$, say, where the distribution of U_n does not depend on β. Moreover, $\sum_1^n x_{i-1}^2 \leq \sum_0^{n-1} U_k^2 = o(n^4)$ with probability one. Hence by Lemma 2.2

$$\lim_{m\to\infty} \sup_{|\beta|\leq 1} P_\beta\{(\textstyle\sum_1^n x_{i-1}\varepsilon_i)^2 \geq \delta n \sum_1^n x_{i-1}^2 \quad \textit{for some} \quad n \geq m\} = 0,$$

which along with (4.6) and the strong law of large numbers implies (4.5).

With the help of (4.5) it is possible to modify the definition of N_c to handle the case of unknown σ. The following result can be proved along the lines of Theorem 2.1, although the details are considerably more complicated. The proof is omitted.

THEOREM 4.3. *Define* b_n *by* (1.2) *and* $\hat{\sigma}_n^2$ *by* (4.3). *Let* $\{\delta_n\}$ *be a sequence of positive constants with* $\delta_n \to 0$. *Let* $s_n^2 = \max(\delta_n, \hat{\sigma}_N^2)$ *and for* $c < 0$ *define*

$$\hat{N}_c = \inf\{n\colon n \geq 2, \textstyle\sum_1^n x_{i-1}^2 \geq c s_n^2\}. \tag{4.7}$$

Then as $c \to \infty$

$$P_\beta\{(\textstyle\sum_1^{\hat{N}_c} x_{i-1}^2)^{1/2}(b_{\hat{N}_c} - \beta)/\hat{\sigma}_{\hat{N}_c} \leq t\} \to \Phi(t)$$

uniformly in $|\beta| \leq 1$ *and* $-\infty < t < \infty$.

REFERENCES

ANSCOMBE, F. J. (1952). Large sample theory of sequential estimation. *Proc. Camb. Phil. Soc.* **48** 600–607.

ANSCOMBE, F. J. (1953). Sequential estimation. *J. Roy. Statist. Soc. Ser. B* **15** 1–21.

ANDERSON, T. W. (1959). On asymptotic distributions of estimates of parameters of stochastic difference equations. *Ann. Math. Statist.* **30** 676–687.

CHOW, Y. S. and ROBBINS, H. E. (1965). On the asymptotic theory of fixed width confidence intervals for the mean. *Ann. Math. Statist.* **36** 457–462.

CHOW, Y. S., ROBBINS, H. E. and SIEGMUND, D. (1971). *Great Expectations: the Theory of Optimal Stopping*. Houghton Mifflin, Boston.

DICKEY, D. A. and FULLER, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *J. Amer. Statist. Assoc.* **74** 427-431.

DVORETZKY, A. (1972). Asymptotic normality for sums of dependent random variables. *Proc. Sixth Berk. Symp. Math. Statist. and Probab.*, vol. II. Univ. of California Press, Berkeley, 513-535.

EVANS, G. B. A. and SAVIN, N. E. (1981). The calculation of the limiting distribution of the least squares estimator of the parameter in a random walk model. *Ann. Statist.* **9** 1114-1118.

FREEDMAN, D. (1971). *Brownian Motion and Diffusion.* Holden Day, San Francisco.

GRAMBSCH, P. (1983). Sequential sampling based on the observed Fisher information to guarantee the accuracy of the maximum likelihood estimator. *Ann. Statist.* **11** 68-77.

LIPSTER, R. S. and SHIRYAEV, A. N. (1978). *Statistics of Random Processes II.* Springer-Verlag, Berlin.

MOGYORÓDI, J. (1962). A central limit theorem for the sum of a random number of independent random variables. *Publ. Math. Inst. Hung. Acad. Sci.* **7** 409-424.

NEVEU, J. (1965). *Mathematical Foundations of the Calculus of Probabilities.* Holden Day, San Francisco.

RAO, M. M. (1978). Asymptotic distribution of an estimator of the boundary parameter of an unstable process. *Ann. Statist.* **6** 185-190.

ROBBINS, H. and SIEGMUND, D. (1971). A convergence theorem for nonnegative almost supermartingales and some applications. In *Optimizing Methods in Statistics*, ed. by J. S. Rustagi. Academic, New York, 223-258.

SIEGMUND, D. (1981). A sequential confidence interval for the odds ratio. To appear in *Probability and Mathematical Statistics.*

WHITE, J. S. (1958). The limiting distribution of the serial correlation coefficient in the explosive case. *Ann. Math. Statist.* **29** 188-197.

DEPARTMENT OF MATHEMATICAL STATISTICS
COLUMBIA UNIVERSITY
NEW YORK, NEW YORK 10027

DEPARTMENT OF STATISTICS
SEQUOIA HALL
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305