The Annals of Statistics
1987, Vol. 15, No. 4, 1667–1682

ADAPTIVE PREDICTION BY LEAST SQUARES PREDICTORS IN STOCHASTIC REGRESSION MODELS WITH APPLICATIONS TO TIME SERIES[1]

BY C. Z. WEI

University of Maryland

Herein we consider the asymptotic performance of the least squares predictors $\hat{y}_n$ of the stochastic regression model $y_n = \beta_1 x_{n1} + \cdots + \beta_p x_{np} + \varepsilon_n$. In particular, the accumulated cost function $\sum_{k=1}^n (y_k - \hat{y}_k - \varepsilon_k)^2$ is studied. The results are then applied to nonstationary autoregressive time series. A statistic is also constructed to show how many times one should difference a nonstationary time series in order to obtain a stationary series.

1. Introduction. Consider the multiple regression model

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

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

$$\mathbf{b}_n = (b_{n1}, \ldots, b_{np})' = \left(\sum_{k=1}^n \mathbf{x}_k \mathbf{x}_k' \right)^{-1} \sum_{k=1}^n \mathbf{x}_k y_k$$

denotes the least squares estimate of $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ based on the observations $\mathbf{x}_1, y_1, \ldots, \mathbf{x}_n, y_n$, assuming that $\sum_1^n \mathbf{x}_i \mathbf{x}_i'$ is nonsingular. Throughout the sequel we shall assume that $\{\varepsilon_n\}$ is a martingale difference sequence with respect to an increasing sequence of σ-fields $\{\mathscr{F}_n\}$; i.e., ε_n is $\mathscr{F}_n$-measurable and $E(\varepsilon_n | \mathscr{F}_{n-1}) = 0$ a.s. for every n. An important example is the case where ε_n are i.i.d. random variables with $E(\varepsilon_n) = 0$. We shall also assume that the design vector at stage n depends on the previous observations $\mathbf{x}_1, y_1, \ldots, \mathbf{x}_{n-1}, y_{n-1}$; i.e., $\mathbf{x}_n$ is $\mathscr{F}_{n-1}$-measurable. The asymptotic properties of the least squares estimates were recently studied by Lai and Wei (1982) and Wei (1985). Strong consistency and asymptotic normality of $\mathbf{b}_n$ were established under general assumptions. In this paper we shall study the asymptotic performance of $\{\mathbf{b}_{n-1}' \mathbf{x}_n\}$ as a sequence of predictors of $\{y_n\}$.

Let $\{\hat{y}_n\}$ be a sequence of predictors of $\{y_n\}$. Since we cannot foresee the future, $\hat{y}_n$ is assumed to be $\mathscr{F}_{n-1}$-measurable. If one is interested in one period prediction, $(y_n - \hat{y}_n)^2$ would be the "cost" to be minimized. However, in the sequential prediction case [see Goodwin and Sin (1984) for some examples], the predictors are going to be updated adaptively and used repeatedly over many

Received May 1986; revised February 1987.

[1]Research supported by the National Science Foundation under grant DMS-84-04081.

AMS 1980 *subject classifications*. Primary 62J05; secondary 62M10, 62M20.

Key words and phrases. Adaptive prediction, least squares, stochastic regression, nonstationary autoregressive models, order selection.

periods. In this situation the cumulative "cost" $\sum_{k=1}^{n}(y_k - \hat{y}_k)^2$ would be more appropriate. Under the assumption that

$$E(\varepsilon_n^2 | \mathscr{F}_{n-1}) = \sigma^2 \quad \text{a.s. for all } n, \tag{1.2}$$

for a reasonable sequence of predictors, one may expect that

$$\frac{1}{n}\sum_{k=1}^{n}(y_k - \hat{y}_k)^2 \to \sigma^2 \quad \text{a.s.} \tag{1.3}$$

However, by Chow's theorem (1965), it is not difficult to see

$$\sum_{k=1}^{n}(y_k - \hat{y}_k)^2 = \sum_{k=1}^{n}\varepsilon_k^2 + C_n(1 + o(1)) \quad \text{a.s.}$$

on the set $\{C_n \to \infty\}$ and

$$\sum_{k=1}^{n}(y_k - \hat{y}_k)^2 = \sum_{k=1}^{n}\varepsilon_k^2 + C_n(1 + O(1)) \quad \text{a.s.}$$

on the set $\{\lim_{n\to\infty} C_n < \infty\}$, where

$$C_n = \sum_{k=1}^{n}(y_k - \hat{y}_k - \varepsilon_k)^2.$$

Hence, C_n is of essential importance. In fact, the quantity $(y_n - \hat{y}_n - \varepsilon_n)^2$ was considered in the time series literature [Fuller and Hasza (1981)] and C_n can be viewed as a second order quantity in comparison with (1.3).

For the least squares predictors

$$C_n = \sum_{k=1}^{n}(\boldsymbol{\beta}'\mathbf{x}_k - \mathbf{b}'_{k-1}\mathbf{x}_k)^2. \tag{1.4}$$

In Section 2 sufficient conditions are imposed on the design vector $\mathbf{x}_k$ to show that

$$C_n \sim \sigma^2 \log\det\left(\sum_1^n \mathbf{x}_k\mathbf{x}'_k\right). \tag{1.5}$$

These results are then applied to the autoregressive models in Section 3. Since our results reveal some deeper properties of nonstationary autoregressive time series, a statistic, which is based on $\det(\sum_{k=1}^{n}\mathbf{x}_i\mathbf{x}'_i)$, is then constructed to show how many times we should difference an integrated autoregressive time series in order to obtain a stationary one when the exact order of the underlying time series is unknown; see Theorems 5 and 6. The result (1.5) provides a statistical interpretation for our statistic. It seems difficult for one period predictors to reveal the similar property (see the remark following Theorem 4). For detailed discussion see Section 3.

2. Main theorems.

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

$$\sup_n E\{|\varepsilon_n|^\alpha | \mathscr{F}_{n-1}\} < \infty \quad a.s. \text{ for some } \alpha > 2, \tag{2.1}$$

and

$$\mathbf{x}_n'\left(\sum_{k=1}^{n} \mathbf{x}_k\mathbf{x}_k'\right)^{-1}\mathbf{x}_n \to v \quad a.s. \text{ as } n \to \infty, \tag{2.2}$$

where v is a nonnegative random variable. Then

$$(1-v)C_n + \sum_{k=1}^{n} [(\mathbf{b}_n - \boldsymbol{\beta})'\mathbf{x}_k]^2 \sim nv\sigma^2 \quad a.s. \tag{2.3}$$

on the set $\{1 > v > 0, C_n \to \infty\}$ and

$$C_n + \sum_{k=1}^{n} [(\mathbf{b}_n - \boldsymbol{\beta})'\mathbf{x}_k]^2 \sim \sigma^2 \log\det\left(\sum_{k=1}^{n} \mathbf{x}_k\mathbf{x}_k'\right) \quad a.s. \tag{2.4}$$

on the set $\{v = 0, C_n \to \infty, \lambda_n \to \infty\}$, where λ_n is the minimum eigenvalue of $\sum_{k=1}^{n}\mathbf{x}_k\mathbf{x}_k'$.

REMARK. Usually, one design vector will not dominate the whole design and $v = 0$. But for the explosive models, $v \neq 0$; see Theorem 4.

PROOF. Let $\mathbf{V}_n = (\sum_{k=1}^{n}\mathbf{x}_k\mathbf{x}_k')^{-1}$ and $Q_n = (\sum_{k=1}^{n}\mathbf{x}_k'\varepsilon_k)V_n(\sum_{k=1}^{n}\mathbf{x}_k\varepsilon_k)$. Then (2.16) of Lai and Wei (1982) gives

$$\begin{aligned} Q_n - Q_N &+ \sum_{k=N+1}^{n}\left(\mathbf{x}_k'V_{k-1}\sum_{j=1}^{k-1}\mathbf{x}_j\varepsilon_j\right)^2 \Big/ (1 + \mathbf{x}_k'V_{k-1}\mathbf{x}_k) \\ &= \sum_{k=N+1}^{n} \mathbf{x}_k'V_k\mathbf{x}_k\varepsilon_k^2 + 2\sum_{k=N+1}^{n}\left\{\mathbf{x}_k'V_{k-1}\left(\sum_{j=1}^{k-1}\mathbf{x}_j\varepsilon_j\right)\varepsilon_k\right\}\Big/(1 + \mathbf{x}_k'V_{k-1}\mathbf{x}_k). \end{aligned} \tag{2.5}$$

By (2.15) of Lai and Wei (1982),

$$\mathbf{x}_k'V_k\mathbf{x}_k = \mathbf{x}_k'V_{k-1}\mathbf{x}_k/(1 + \mathbf{x}_k'V_{k-1}\mathbf{x}_k).$$

Hence,

$$1 - \mathbf{x}_k'V_k\mathbf{x}_k = 1/(1 + \mathbf{x}_k'V_{k-1}\mathbf{x}_k). \tag{2.6}$$

Notice that

$$\mathbf{b}_k - \boldsymbol{\beta} = \mathbf{V}_k\sum_{j=1}^{k}\mathbf{x}_j\varepsilon_j. \tag{2.7}$$

This, (2.5) and (2.6) imply

$$\begin{aligned} Q_n - Q_N &+ \sum_{k=N+1}^{n} [\mathbf{x}_k'(\mathbf{b}_{k-1} - \boldsymbol{\beta})]^2(1 - \mathbf{x}_k'V_k\mathbf{x}_k) \\ &= \sum_{k=N+1}^{n} \mathbf{x}_k'V_k\mathbf{x}_k\varepsilon_k^2 + 2\sum_{k=N+1}^{n} [\mathbf{x}_k'(\mathbf{b}_{k-1} - \boldsymbol{\beta})]\varepsilon_k(1 - \mathbf{x}_k'V_k\mathbf{x}_k). \end{aligned} \tag{2.8}$$

On the set $\{1 > v \geq 0,\ C_n \to \infty\}$,

$$\sum_{k=N+1}^{n} [\mathbf{x}_k'(\mathbf{b}_{k-1} - \boldsymbol{\beta})]^2(1 - \mathbf{x}_k'V_k\mathbf{x}_k) \sim C_n(1 - v) \quad \text{a.s.} \tag{2.9}$$

Using the local martingale convergence theorem [Lai and Wei, (1982), Lemma 2(iii)] on the set $\{1 > v \geq 0,\ C_n \to \infty\}$,

$$\sum_{k=N+1}^{n} [\mathbf{x}_k'(\mathbf{b}_{k-1} - \boldsymbol{\beta})]\,\varepsilon_k(1 - \mathbf{x}_k'V_k\mathbf{x}_k) = o(C_n) \quad \text{a.s.} \tag{2.10}$$

Consequently, (2.8)–(2.10) imply

$$Q_n - Q_N + C_n(1 - v)(1 + o(1)) = \sum_{k=N+1}^{n} \mathbf{x}_k'V_k\mathbf{x}_k\varepsilon_k^2 \quad \text{a.s.} \tag{2.11}$$

on the set $\{1 > v \geq 0,\ C_n \to \infty\}$. By (2.1) and the local martingale convergence theorem [Chow (1965)], on the set $\{\sum_{k=N+1}^{n}\mathbf{x}_k'V_k\mathbf{x}_k \to \infty\}$,

$$\sum_{k=N+1}^{n} \mathbf{x}_k'V_k\mathbf{x}_k\{\varepsilon_k^2 - E(\varepsilon_k^2|\mathscr{F}_{k-1})\} = o\left\{\sum_{k=N+1}^{n} \mathbf{x}_kV_k\mathbf{x}_k\right\} \quad \text{a.s.}$$

Hence, by (1.2), on the set $\{\sum_{k=N+1}^{n}\mathbf{x}_k'V_k\mathbf{x}_k \to \infty\}$,

$$\sum_{k=N+1}^{n} \mathbf{x}_k'V_k\mathbf{x}_k\varepsilon_k^2 \sim \left(\sum_{k=N+1}^{n} \mathbf{x}_k'V_k\mathbf{x}_k\right)\sigma^2 \quad \text{a.s.} \tag{2.12}$$

On the set $\{1 > v > 0\}$,

$$\sum_{k=N+1}^{n} \mathbf{x}_k'V_k\mathbf{x}_k \sim nv \quad \text{a.s.} \tag{2.13}$$

On the set $\{v = 0,\ \lambda_n \to \infty\}$, by Lemma 2(i) and (ii) of Lai and Wei (1982),

$$\sum_{k=N+1}^{n} \mathbf{x}_k'V_k\mathbf{x}_k \sim \log\det\left(\sum_{k=1}^{n} \mathbf{x}_k\mathbf{x}_k'\right) \quad \text{a.s.} \tag{2.14}$$

Finally, notice that

$$Q_n = (\mathbf{b}_n - \boldsymbol{\beta})'V_n^{-1}(\mathbf{b}_n - \boldsymbol{\beta}) = \sum_{k=1}^{n} [(\mathbf{b}_n - \boldsymbol{\beta})'\mathbf{x}_k]^2. \tag{2.15}$$

Thus, (2.11)–(2.15) imply (2.3) and (2.4). □

The following theorems provide some simple conditions to ensure $C_n \to \infty$ a.s.

THEOREM 2. *Suppose that in the regression model* (1.1), (1.2) *and* (2.1) *hold. If*

$$\mathbf{x}_n'\left(\sum_{k=1}^{n} \mathbf{x}_k\mathbf{x}_k'\right)^{-1}\mathbf{x}_n \to 0 \quad a.s., \tag{2.16}$$

$$\lambda_n \to \infty \quad a.s. \tag{2.17}$$

and

$$\sum_{k=1}^{n} [(\mathbf{b}_n - \boldsymbol{\beta})'\mathbf{x}_k]^2 = O_p(1), \tag{2.18}$$

then

$$C_n \sim \sigma^2 \log\det\left(\sum_{k=1}^{n} \mathbf{x}_k\mathbf{x}_k'\right) \quad \textit{in probability}.$$

REMARK. Condition (2.18) is a natural condition. Usually, we may even expect that the quantity $\sum_{k=1}^{n}[(\mathbf{b}_n - \boldsymbol{\beta})'\mathbf{x}_k]^2$ has a limiting distribution. (It is a multiple of a χ^2-distribution in the classical regression theory.)

PROOF. By (2.4) of Theorem 1, we only have to show that

$$C_n \to \infty \quad \text{a.s.} \tag{2.19}$$

We are going to use (2.8), (2.12) and (2.14). First, notice that these results hold without the assumption (2.19). By (2.18) of Lai and Wei (1982)

$$\begin{aligned} &\sum_{k=N+1}^{n} [\mathbf{x}_k'(\mathbf{b}_{k-1} - \boldsymbol{\beta})]\,\varepsilon_k(1 - \mathbf{x}_k'V_k\mathbf{x}_k) \\ &\quad = o\left(\sum_{k=N+1}^{n} [\mathbf{x}_k'(\mathbf{b}_{k-1} - \boldsymbol{\beta})]^2(1 - \mathbf{x}_k'V_k\mathbf{x}_k)\right) + O(1) \quad \text{a.s.} \end{aligned} \tag{2.20}$$

By (2.8), (2.12), (2.14), (2.20) and (2.18),

$$\begin{aligned} &\left\{\sum_{k=N+1}^{n} [\mathbf{x}_k'(\mathbf{b}_{k-1} - \boldsymbol{\beta})]^2(1 - \mathbf{x}_k'V_k\mathbf{x}_k)\right\}(1 + O(1)) \\ &\quad \sim \sigma^2 \log\det\left(\sum_{k=1}^{n} \mathbf{x}_k\mathbf{x}_k'\right) \quad \text{in probability.} \end{aligned}$$

Hence,

$$\sum_{k=N+1}^{n} [\mathbf{x}_k'(\mathbf{b}_{k-1} - \boldsymbol{\beta})]^2(1 - \mathbf{x}_k'V_k\mathbf{x}_k) \to \infty \quad \text{in probability.} \tag{2.21}$$

But $\sum_{k=N+1}^{n}[\mathbf{x}_k'(\mathbf{b}_{k-1} - \boldsymbol{\beta})]^2(1 - \mathbf{x}_k'V_k\mathbf{x}_k)$ is a positive and increasing function of n. By (2.21)

$$\sum_{k=N+1}^{n} [\mathbf{x}_k'(\mathbf{b}_{k-1} - \boldsymbol{\beta})]^2(1 - \mathbf{x}_k'V_k\mathbf{x}_k) \to \infty \quad \text{a.s.} \tag{2.22}$$

Now (2.19) follows from the fact that

$$C_n \geq \sum_{k=N+1}^{n} [\mathbf{x}_k'(\mathbf{b}_{k-1} - \boldsymbol{\beta})]^2(1 - \mathbf{x}_k'V_k\mathbf{x}_k) \quad \text{a.s.}$$ □

When the original model can be reparametrized so that the new design vectors are weakly correlated, we have a better result.

THEOREM 3. *Suppose that in the regression model* (1.1), (1.2), (2.1) *and* (2.17) *hold. If there is a constant nonsingular matrix A such that $\mathbf{z}_n = A\mathbf{x}_n$ satisfy*

$$\mathbf{z}_n'\left(\sum_1^n \mathbf{z}_i\mathbf{z}_i'\right)^{-1}\mathbf{z}_n \to 0 \quad a.s. \tag{2.23}$$

and

$$\liminf \lambda_{\min}\left(D_n^{-1}\left(\sum_1^n \mathbf{z}_i\mathbf{z}_i'\right)D_n^{-1}\right) > 0 \quad a.s., \tag{2.24}$$

where $D_n = \{\operatorname{diag}(\Sigma_1^n \mathbf{z}_i\mathbf{z}_i')\}^{1/2}$ and $\lambda_{\min}(M)$ is the minimum eigenvalue of M, then

$$C_n \sim \sigma^2 \log\det\left(\sum_{k=1}^n \mathbf{x}_k\mathbf{x}_k'\right) \quad a.s. \tag{2.25}$$

REMARK. Usually, there is an increasing sequence of positive real numbers, such that $a_n \to \infty$ and

$$\left(\sum_{k=1}^n \mathbf{x}_k\mathbf{x}_k'\right)\Big/ a_n \to \Gamma \quad \text{a.s.}, \tag{2.26}$$

where Γ is a positive definite matrix. Then we can take A to be the identity matrix and (2.24) is satisfied automatically. But (2.26) would not be satisfied if the components of $\mathbf{x}_n$ are linear combinations of some "weak" and "strong" signals. This is the case when $x_n = \beta_1 x_{n-1} + \beta_2 x_{n-2} + \varepsilon_n$ and the characteristic polynomial $z^2 - \beta_1 z - \beta_2$ has one root equal to 1 and another root with magnitude less than 1 (see Appendix 2 for further details).

We shall prove Theorem 3 via the following strong law of a martingale transform, whose proof is given in Appendix 1.

LEMMA 1. *Let $\{\varepsilon_n, \mathscr{F}_n\}$ be a sequence of martingale differences such that $\sup_n E\{|\varepsilon_n|^\alpha | \mathscr{F}_{n-1}\} < \infty$ a.s. for some $\alpha \geq 2$. Let u_n be an $\mathscr{F}_{n-1}$-measurable random variable, $s_n^2 = \Sigma_{k=1}^n u_i^2$ and f be a nondecreasing function such that*

$$\int_1^\infty (xf^\alpha(x))^{-1}\,dx < \infty. \tag{2.27}$$

Then on the set $\{s_n^2 \to \infty\}$,

$$S_n = \sum_{k=1}^n u_k\varepsilon_k = o\left(s_n f(s_n^2)\right) \quad a.s. \tag{2.28}$$

REMARKS. (a) Lemma 1 is an improvement of Lemma 2 of Wei (1985). Under the assumption $\alpha > 2$, Wei (1985) shows that

$$\sum_{k=1}^n u_k\varepsilon_k = O\left(s_n(\log s_n)^\delta\right) \quad \text{a.s. for } \delta > (\min(\alpha, 4))^{-1}. \tag{2.29}$$

Lemma 1 implies

$$(2.30)\qquad \sum_{k=1}^{n} u_k \varepsilon_k = o\big(s_n (\log s_n)^{\delta}\big) \quad \text{a.s. for } \delta > \alpha^{-1}.$$

(b) Theorem 1 of Wei (1985) on the strong consistency of the least squares estimates can be improved by (2.30). For details, see Wei (1985).

PROOF OF THEOREM 3. First we claim that

$$(2.31)\qquad \sum_{k=1}^{n} [(\mathbf{b}_n - \boldsymbol{\beta})'\mathbf{x}_k]^2 = o\left(\log\det\left(\sum_{k=1}^{n} \mathbf{x}_k\mathbf{x}_k'\right)\right) \quad \text{a.s.}$$

Note that $\mathbf{z}_k = A\mathbf{x}_k$ and A nonsingular imply that

$$\begin{aligned}
\sum_{k=1}^{n} [(\mathbf{b}_n - \boldsymbol{\beta})'\mathbf{x}_k]^2 &= \left(\sum_{k=1}^{n} \mathbf{x}_k\varepsilon_k\right)'\left(\sum_{k=}^{n} \mathbf{x}_k\mathbf{x}_k'\right)^{-1}\left(\sum_{k=1}^{n} \mathbf{x}_k\varepsilon_k\right)\\
&= \left(\sum_{k=1}^{n} \mathbf{z}_k\varepsilon_k\right)'\left(\sum_{k=1}^{n} \mathbf{z}_k\mathbf{z}_k'\right)^{-1}\left(\sum_{k=1}^{n} \mathbf{z}_k\varepsilon_k\right)\\
(2.32)\qquad &= \left(D_n^{-1}\sum_{k=1}^{n} \mathbf{z}_k\varepsilon_k\right)'\left(D_n^{-1}\left(\sum_{k=1}^{n} \mathbf{z}_k\mathbf{z}_k'\right)D_n^{-1}\right)^{-1}\left(D_n^{-1}\sum_{k=1}^{n} \mathbf{z}_k\varepsilon_k\right)\\
&= O\left(\left\|D_n^{-1}\sum_{k=1}^{n} z_k\varepsilon_k\right\|^2\right), \quad \text{by (2.24)}.
\end{aligned}$$

Apply (2.30) to each component of $D_n^{-1}\sum_{k=1}^{n}\mathbf{z}_k\varepsilon_k$. We obtain

$$(2.33)\qquad \left\|D_n^{-1}\sum_{k=1}^{n} \mathbf{z}_k\varepsilon_k\right\|^2 = o\left(\left(\log\left(\sum_{k=1}^{n} \|\mathbf{z}_k\|^2\right)\right)^{\delta}\right) \quad \text{a.s. for } \delta > 2\alpha^{-1}.$$

Since $\mathbf{z}_k = A\mathbf{x}_k$,

$$\begin{aligned}
\log\left(\sum_{k=1}^{n} \|\mathbf{z}_k\|^2\right) &= O\left(\log\left(\sum_{k=1}^{n} \|\mathbf{x}_k\|^2\right)\right) \quad \text{a.s.}\\
(2.34)\qquad &= O\left(\log\det\left(\sum_{k=1}^{n} \mathbf{x}_k\mathbf{x}_k'\right)\right) \quad \text{a.s. by (2.17)}.
\end{aligned}$$

Combining (2.32)–(2.34) and the fact that $\alpha > 2$, (2.31) is proved. Now using (2.8), (2.12), (2.14), (2.20) and (2.31), we have

$$(2.35)\qquad C_n \to \infty \quad \text{a.s.}$$

By (2.31), (2.35) and (2.4) of Theorem 1, (2.25) is proved. □

3. Applications.

THEOREM 4. *Assume* $y_n = \rho y_{n-1} + \varepsilon_n$, *where* $\{\varepsilon_n\}$ *is a sequence of martingale differences satisfying* (1.2) *and* (2.1). *Let* y_0 *be an* $\mathscr{F}_0$*-measurable random variable*, $\hat{\rho}_n = \left(\sum_{k=1}^{n-1} y_k y_{k+1}\right)/\left(\sum_{k=1}^{n-1} y_k^2\right)$ *and* $C_n = \sum_{k=2}^{n}(\hat{\rho}_{k-1} - \rho)^2 y_k^2$. *Then*

$$|\rho| < 1 \quad \text{implies} \quad C_n \sim \sigma^2 \log n \quad a.s.; \tag{3.1}$$

$$|\rho| = 1 \quad \text{implies} \quad C_n \sim 2\sigma^2 \log n \quad a.s.; \tag{3.2}$$

$$|\rho| > 1 \quad \text{implies} \quad C_n \sim \sigma^2(\rho^2 - 1)n \quad a.s. \tag{3.3}$$

PROOF. By Theorem 3, Theorem 4 and (3.7) of Lai and Wei (1983), we have that for $|\rho| \le 1$,

$$y_n^2 \Big/ \left(\sum_{k=1}^{n} y_k^2 \right) \to 0 \quad \text{a.s.} \tag{3.4}$$

and

$$\log\left(\sum_{k=1}^{n} y_k^2 \right) \Big/ \log n \to c \quad \text{a.s.,} \tag{3.5}$$

where $c = 1$ if $|\rho| < 1$ and $c = 2$ if $|\rho| = 1$. Now apply Theorem 3 with $A = 1$ and observe that (2.24) is automatically satisfied. We obtain (3.1) and (3.2).

For the case $|\rho| > 1$, by Theorem 2 of Lai and Wei (1983),

$$\rho^{-n} y_n \to z \quad \text{a.s.,} \tag{3.6}$$

where $z = y_0 + \sum_{k=1}^{\infty} \rho^{-k} \varepsilon_k$ is nonzero a.s. This implies

$$\left(\sum_{k=1}^{n} y_k^2 \right)(\rho^2 - 1)/\rho^{2(n+1)} \to z^2 \quad \text{a.s.} \tag{3.7}$$

and

$$y_n^2 \Big/ \left(\sum_{k=1}^{n} y_k^2 \right) \to (\rho^2 - 1)/\rho^2 \quad \text{a.s.} \tag{3.8}$$

By Lemma 1 and (3.7),

$$\begin{aligned} \sum_{k=1}^{n-1} (\hat{\rho}_n - \rho)^2 y_k^2 &= \left[\left(\sum_{k=1}^{n-1} y_k \varepsilon_{k+1} \right)^2 \Big/ \left(\sum_{k=1}^{n-1} y_k^2 \right) \right] \\ &= o\left(\log\left(\sum_{k=1}^{n-1} y_k^2 \right) \right) \\ &= o(n) \quad \text{a.s.} \end{aligned} \tag{3.9}$$

Applying (2.3) of Theorem 1, we have

$$\begin{aligned} C_n &\sim n\sigma^2 \left[(\rho^2 - 1)/\rho^2 \right] / \left[1 - (\rho^2 - 1)/\rho^2 \right] \quad \text{a.s.} \\ &= n\sigma^2(\rho^2 - 1). \end{aligned}$$

This completes the proof of Theorem 4. □

REMARK. Under the assumption that $|\rho| \le 1$, Fuller and Hasza (1981) have shown that

$$(\hat{\rho}_{n-1} - \rho)^2 y_n^2 = O\left(\frac{1}{n}\right) \quad \text{in probability.}$$

It seems difficult for their arguments to show the "doubling effect" which appears in (3.2).

THEOREM 5. *Assume $y_n = \beta_1 y_{n-1} + \cdots + \beta_p y_{n-p} + \varepsilon_n$, where $\{\varepsilon_n\}$ satisfies* (1.2) *and* (2.1). *Let $y_0, \ldots, y_{1-p}$ be $\mathscr{F}_0$-measurable random variables and*

$$\begin{aligned}\phi(z) &= z^p - \beta_1 z^{p-1} - \cdots - \beta_p \\ &= (1-z)^a (1+z)^b \prod_{k=1}^{l} (1 - 2\cos\theta_k z + z^2)^{d_k} \psi(z),\end{aligned} \tag{3.10}$$

where a, b, l and d_k are nonnegative integers, $\theta_k \in (0, \pi)$, $\theta_k \ne \theta_j$, if $k \ne j$ and $\psi(z)$ is a polynomial of order $q = p - (a + b + 2d_1 + \cdots + 2d_l)$ which has all roots inside the unit circle. Then

$$\log\det\left(\sum_{k=1}^{n} \mathbf{y}_k \mathbf{y}_k'\right) \sim c \log n \quad \textit{in probability}, \tag{3.11}$$

where $\mathbf{y}_n' = (y_n, \ldots, y_{n-p+1})$, $c = q + a(a+1) + b(b+1) + 2\sum_{k=1}^{l} d_k(d_k+1)$ *and*

$$C_n \sim c\sigma^2 \log n \quad \textit{in probability}. \tag{3.12}$$

PROOF. By Chan and Wei (1986) there exist nonsingular $p \times p$ matrices G and H_n such that

$$H_n G\left(\sum_{k=1}^{n} \mathbf{y}_k \mathbf{y}_k'\right) G' H_n' \to_{\mathscr{L}} \Delta, \tag{3.13}$$

where Δ is positive definite a.s. and

$$(G'H_n')^{-1}(\mathbf{b}_n - \boldsymbol{\beta}) \to_{\mathscr{L}} \text{some random vector } \boldsymbol{\eta}. \tag{3.14}$$

Furthermore,

$$H_n^{-1} = \begin{pmatrix} A_n & 0 & 0 & \cdots & 0 & 0 \\ 0 & B_n & \cdot & & \cdot & \cdot \\ 0 & 0 & D_n(1) & & \cdot & \cdot \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \cdot & \cdot & \cdot & & D_n(l) & \cdot \\ 0 & \cdot & \cdot & \cdots & 0 & Q_n \end{pmatrix}, \tag{3.15}$$

where $Q_n = \sqrt{n} I_q$,

$$A_n = \begin{pmatrix} n^a & 0 & \cdots & 0 \\ 0 & n^{a-1} & & \cdot \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & n \end{pmatrix}, \qquad B_n = \begin{pmatrix} n^b & 0 & \cdots & 0 \\ 0 & n^{b-1} & & \cdot \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & n \end{pmatrix}$$

and

$$D_n(k) = \begin{pmatrix} nI & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & n^{d_k}I \end{pmatrix} \quad \text{with } I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

By (3.13) and (3.15), in probability,

$$\begin{aligned} \log\det\left(\sum_{k=1}^{n} \mathbf{y}_k\mathbf{y}_k'\right) &\sim 2\log\left(\det A_n \det B_n \prod_{k=1}^{l} \det D_n(k) \det Q_n\right) \\ &= \left(a(a+1) + b(b+1) + 2\sum_{k=1}^{l} d_k(d_k+1) + q\right)\log n \\ &= c\log n. \end{aligned}$$

Hence, (3.11) is proved. Now combining (3.13) and (3.14), we have

$$\sum_{k=1}^{n} [(\mathbf{b}_n - \boldsymbol{\beta})'\mathbf{y}_k]^2 \to_{\mathscr{L}} \|\Delta^{-1}\boldsymbol{\eta}\|^2. \tag{3.16}$$

By Theorems 3 and 4 of Lai and Wei (1983),

$$\mathbf{y}_n'\left(\sum_{k=1}^{n} \mathbf{y}_k\mathbf{y}_k'\right)^{-1}\mathbf{y}_n \to 0 \quad \text{a.s.} \tag{3.17}$$

and

$$\liminf(\lambda_n/n) > 0 \quad \text{a.s.} \tag{3.18}$$

In view of (3.16)–(3.18) and Theorem 2, (3.12) is shown. □

THEOREM 6. *Assume the autoregressive model in Theorem 5 has roots equal to 1 or less than 1 in magnitude only (i.e., $b = d_1 = \cdots = d_l = 0$). Suppose $\beta_p \neq 0$ and p is unknown, but $r \geq p$ is given. Let $\mathbf{y}_n = (y_n, \ldots, y_{n-r+1})'$. Then*

$$\hat{a}_n = \left[\log\det\left(\sum_{k=r}^{n} \mathbf{y}_k\mathbf{y}_k'\right)\Big/\log n - r\right]^{1/2} \to a, \quad \textit{in probability}.$$

PROOF. We can enlarge the model to be

$$y_n = \alpha_1 y_{n-1} + \cdots + \alpha_r y_{n-r} + \varepsilon_n.$$

By Theorem 5,

$$\begin{aligned} \hat{a}_n &\to (q + a(a+1) - r)^{1/2} \\ &= (r - a + a^2 + a - r)^{1/2} = a, \quad \text{in probability}. \end{aligned}$$ □

REMARKS. (a) Theorem 6 answers the question of how many times we should difference an integrated autoregressive model when the exact order is unknown.

(b) After differencing, the standard AIC and BIC criteria can be applied to estimate the exact order. Therefore, Theorem 5 also provides a two step order selection procedure. For one step procedures and background knowledge of AIC and BIC criteria, see Tsay (1984).

(c) Note that if only one root is suspected to be equal to 1 (i.e., $a = 1$ or 0), we may use $\hat{a}_n$ as a test statistic. See Fuller (1976) or Dickey and Fuller (1979) for other tests which are based on least squares estimates.

APPENDIX 1

In this appendix, we prove Lemma 1. First, we need another lemma.

LEMMA 2. *Let $\{\varepsilon_n, \mathscr{F}_n\}$ be a sequence of martingale differences such that for some $\alpha \geq 2$ and positive constant C,*

$$\sup_n E\{|\varepsilon_n|^\alpha | \mathscr{F}_{n-1}\} \leq C \quad a.s. \tag{A1.1}$$

Let u_n be $\mathscr{F}_{n-1}$-measurable random variables, $S_n = \sum_{k=1}^n u_k \varepsilon_k$ and $S_n^* = \sup_{1 \leq k \leq n} |S_k|$. Then

$$E(S_n^*)^\alpha \leq KE\left(\sum_{k=1}^n u_i^2\right)^{\alpha/2}, \tag{A1.2}$$

where K depends only on α and C.

PROOF. Let $d_k = u_k \varepsilon_k$ and $d_n^* = \sup_{1 \leq k \leq n} |d_k|$. Then by an inequality due to Burkholder, Davis and Gundy (1972) [see also, Chow and Teicher (1978), page 397] there exists a constant B which depends on α such that

$$E(S_n^*)^\alpha \leq B\left[E\left(\sum_{k=1}^n E(d_k^2 | \mathscr{F}_{k-1})\right)^{\alpha/2} + E(d_n^*)^\alpha\right]. \tag{A1.3}$$

Since

$$\begin{aligned} &E(|\varepsilon_k|^2 | \mathscr{F}_{k-1}) \leq E(|\varepsilon_k|^\alpha | \mathscr{F}_{k-1})^{2/\alpha} \leq C^{2/\alpha}, \\ &E\left[\sum_{k=1}^n E(d_k^2 | \mathscr{F}_{k-1})\right]^{\alpha/2} \leq CE\left(\sum_{k=1}^n u_k^2\right)^{\alpha/2}. \end{aligned} \tag{A1.4}$$

Now

$$\begin{aligned} E(d_n^*)^\alpha &\leq E\sum_{k=1}^n |d_k|^\alpha = E\sum_{k=1}^n E(|u_k\varepsilon_k|^\alpha | \mathscr{F}_{k-1}) \\ &\leq CE\left(\sum_{k=1}^n |u_k|^\alpha\right) \leq CE\left(\sum_{k=1}^n |u_k|^2\right)\left(\max_{1 \leq k \leq n} |u_k|^{\alpha-2}\right) \\ &\leq CE\left(\sum_{k=1}^n |u_k|^2\right)\left(\sum_{k=1}^n |u_k|^2\right)^{(\alpha-2)/2} \\ &= CE\left(\sum_{k=1}^n |u_k|^2\right)^{\alpha/2}. \end{aligned} \tag{A1.5}$$

Combining (A1.3)–(A1.5), (A1.2) is proved. □

REMARK. Taking the limits on both sides of (A1.2), Lemma 2 also holds for $n = \infty$.

PROOF OF LEMMA 1. Without loss of generality, we can assume that each u_k is a bounded random variable. Otherwise, choose a_k so large that $\sum_{k=1}^{\infty} P[|u_k| \leq a_k] < \infty$ and replace u_k by $u_k I_{[|u_k| \leq a_k]}$. First, let us assume that $\sup_n E(|\varepsilon_n|^\alpha | \mathscr{F}_{n-1}) \leq C$ a.s. for some constant $C > 0$. Fix a constant $e > 1$. Define

$$\tau_k = \inf\{n: s_n^2 \geq e^k\}.$$

Notice that on the set $\{s_n^2 \to \infty\}$, $\tau_k < \infty$ a.s. for all k. In order to prove (2.28), it is sufficient to show that on the set $\{s_n^2 \to \infty\}$,

$$S_{\tau_k} / \left(s_{\tau_k} f\left(s_{\tau_k}^2\right)\right) \to 0 \quad \text{a.s.} \tag{A1.6}$$

and

$$\sup_{\tau_k < i < \tau_{k+1}} |S_i - S_{\tau_k}| / e^{k/2} f(e^k) \to 0 \quad \text{a.s.} \tag{A1.7}$$

Since $S_{\tau_k} = S_{\tau_k - 1} + u_{\tau_k} \varepsilon_{\tau_k}$, (A1.6) can be shown by proving

$$u_{\tau_k} \varepsilon_{\tau_k} / s_{\tau_k} f\left(s_{\tau_k}^2\right) \to 0 \quad \text{a.s. on } \left\{s_n^2 \to \infty\right\} \tag{A1.8}$$

and

$$S_{\tau_k - 1} / e^{k/2} f(e^k) \to 0 \quad \text{a.s.} \tag{A1.9}$$

Now on the set $\{s_n^2 \to \infty\}$,

$$\sum_{j=1}^{\infty} E\left(|u_j \varepsilon_j / s_j f\left(s_j^2\right)|^\alpha | \mathscr{F}_{k-1}\right) \leq C \sum_{k=1}^{\infty} u_j^2 / \left(s_j^2 f^\alpha\left(s_j^2\right)\right) < \infty \quad \text{a.s. by (2.27).}$$

With a standard argument which uses the conditional Borel–Cantelli lemma [Stout (1974), page 55], on the set $\{s_n^2 \to \infty\}$,

$$u_n \varepsilon_n / \left(s_n^2 f\left(s_n^2\right)\right) \to 0 \quad \text{a.s.} \tag{A1.10}$$

Since on the set $\{s_n^2 \to \infty\}$, $\tau_k \to \infty$ a.s., (A1.8) follows from (A1.10). For (A1.9), let us rewrite $S_{\tau_k - 1}$ as

$$Y_k = \sum_{j=1}^{\infty} I_{[\tau_k > j]} u_j \varepsilon_j.$$

Note that s_n^2 is $\mathscr{F}_{n-1}$-measurable. This implies that $[\tau_k > n] = [s_n^2 < e^k]$ is $\mathscr{F}_{n-1}$-measurable and $\{I_{[\tau_k > n]} u_n \varepsilon_n, \mathscr{F}_n\}$ is a sequence of martingale differences. By Lemma 2,

$$\begin{aligned} E|Y_k|^\alpha &\leq KE\left\{\sum_{j=1}^{\infty} I_{[\tau_k > j]} u_j^2\right\}^{\alpha/2} \\ &= KE\left\{\sum_{j=1}^{\tau_k - 1} u_j^2\right\}^{\alpha/2} \\ &\leq Ke^{k\alpha/2}. \end{aligned}$$

Thus, for any $\delta > 0$,

$$\begin{aligned}\sum_{j=1}^{\infty} P\{|Y_j| \geq \delta e^{j/2} f(e^j)\} &\leq \sum_{j=1}^{\infty} E|Y_j|^{\alpha}/(\delta^{\alpha} e^{j\alpha/2} f^{\alpha}(e^j)) \\ &\leq (K/\delta^{\alpha}) \sum_{j=1}^{\infty} 1/f^{\alpha}(e^j) \\ &\leq (K/\delta^{\alpha}) \sum_{j=1}^{\infty} e(e-1)^{-1} \int_{e^{j-1}}^{e^j} (xf^{\alpha}(x))^{-1}\, dx \\ &< \infty, \quad \text{by (2.27)}.\end{aligned}$$

By the Borel–Cantelli lemma, (A1.9) is proved. Now let us show (A1.7). For $\tau_k < i < \tau_{k+1}$,

$$S_i - S_{\tau_k} = \sum_{j=1}^{i} I_{[\tau_k < j < \tau_{k+1}]} u_j \varepsilon_j.$$

Note that

$$[\tau_k < j < \tau_{k+1}] = [\tau_k < j] \cap [\tau_{k+1} \leq j]^c \in \mathscr{F}_{j-1}.$$

Hence, $\{I_{[\tau_k < j < \tau_{k+1}]} u_j \varepsilon_j, \mathscr{F}_j\}$ is a sequence of martingale differences. Observe that

$$Z_k = \sup_{\tau_k < i < \tau_{k+1}} |S_i - S_{\tau_k}| \leq \sup_{1 \leq i < \infty} \left| \sum_{j=1}^{i} I_{[\tau_k < j < \tau_{k+1}]} u_j \varepsilon_j \right|.$$

By Lemma 2,

$$\begin{aligned}EZ_k^{\alpha} &\leq KE\left(\sum_{j=1}^{\infty} I_{[\tau_k < j < \tau_{k+1}]} u_j^2 \right)^{\alpha/2} \\ &= KE\left(\sum_{\tau_k + 1}^{\tau_{k+1} - 1} u_j^2 \right)^{\alpha/2} \\ &\leq K(e^{k+1} - e^k)^{\alpha/2} = K(e-1)e^{\alpha k/2}.\end{aligned}$$

By the same arguments as we show for (A1.9), (A1.7) is proved. Finally, we have to remove the assumption $\sup_n E(|\varepsilon_n|^{\alpha} | \mathscr{F}_{n-1}) \leq C$ a.s. by $\sup_n E(|\varepsilon_n|^{\alpha} | \mathscr{F}_{n-1}) < \infty$ a.s. Let

$$A_m = \left\{ \sup_n E(|\varepsilon_n|^{\alpha} | \mathscr{F}_{n-1}) \leq m \right\}.$$

Then $P(\bigcup_{m=1}^{\infty} A_m) = 1$. Define

$$\varepsilon_n' = \varepsilon_n I_{[E(|\varepsilon_n|^{\alpha} | \mathscr{F}_{n-1}) \leq m]}.$$

Clearly, $\{\varepsilon_n', \mathscr{F}_n\}$ is a sequence of martingale differences such that on A_m, $\varepsilon_n = \varepsilon_n'$ for all n and

$$\sup_n E(|\varepsilon_n'| \mathscr{F}_{n-1}) \leq m \quad \text{a.s.}$$

Applying the case we had proved, (2.28) holds on the set $A_m \cap \{s_n^2 \to \infty\}$ for all m. Hence, (2.28) holds on the set $(\bigcup_{m=1}^{\infty} A_m) \cap \{s_n^2 \to \infty\}$. This completes our proof. □

APPENDIX 2

In this appendix we are going to justify the assertion given in the remark following Theorem 3. We restate this assertion as a proposition.

PROPOSITION. *Suppose that* $x_n = \beta_1 x_{n-1} + \beta_2 x_{n-2} + \varepsilon_n$, *where the initial values* $x_{-1} = x_0 = 0$ *and the* ε_n *are i.i.d. random variables with* $E(\varepsilon_n) = 0$, $0 < E(\varepsilon_n^2) < \infty$ *and* $E(|\varepsilon_n|^\alpha) < \infty$ *for some* $\alpha > 2$. *If* $\phi(z) = z^2 - \beta_1 z - \beta_2 = (z-1)(z-\gamma)$ *with* $|\gamma| < 1$, *then* (2.23) *and* (2.24) *hold but* (2.26) *is violated.*

PROOF. Let $t_n = x_n - \gamma x_{n-1}$ and $u_n = x_n - x_{n-1}$. Then for $n \geq 1$,

(A2.1) $$t_n = t_{n-1} + \varepsilon_n, \qquad u_n = \gamma u_{n-1} + \varepsilon_n$$

and

(A2.2) $$\mathbf{z}_n = A\mathbf{x}_n = (t_n, u_n)', \quad \text{where } A = \begin{pmatrix} 1 & -\gamma \\ 1 & -1 \end{pmatrix}.$$

Since $\gamma \neq 1$, A is nonsingular. This implies (2.23) holds iff

(A2.3) $$\mathbf{x}_n' \left(\sum_1^n \mathbf{x}_i \mathbf{x}_i' \right)^{-1} \mathbf{x}_n \to 0 \quad \text{a.s.}$$

But (A2.3) is a special case of (3.17). Hence, (2.23) is proved.

Now let $T_n^2 = \sum_1^n t_i^2$ and $U_n^2 = \sum_1^n u_i^2$. Then

$$\sum_1^n \mathbf{z}_i \mathbf{z}_i' = \begin{pmatrix} T_n^2 & \sum_1^n t_i u_i \\ \sum_1^n t_i u_i & U_n^2 \end{pmatrix}$$

and

$$D_n = \begin{pmatrix} T_n & 0 \\ 0 & U_n \end{pmatrix}.$$

We claim that

(A2.4) $$D_n^{-1} \left(\sum_1^n \mathbf{z}_i \mathbf{z}_i' \right) D_n^{-1} \to \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{a.s.}$$

Clearly, (2.24) follows from (A2.4) directly. We will prove (A2.4) later. Let us use it to show that (2.26) is violated first. Denote the maximum and minimum eigenvalues of $\sum_1^n \mathbf{z}_i \mathbf{z}_i'$ by δ_n and λ_n. Then (A2.4) implies that

(A2.5) $$\liminf_{n \to \infty} \delta_n / \lambda_{\max}(D_n) \geq 1 \quad \text{a.s.}$$

and

(A2.6) $$\limsup_{n \to \infty} \lambda_n / \lambda_{\min}(D_n) \leq 1 \quad \text{a.s.}$$

It is well known that [cf. Lai and Wei (1985)]

$$n^{-1}U_n^2 \to (1-\gamma^2)^{-1}\sigma^2 \quad \text{a.s.} \tag{A2.7}$$

It is also known that [Lai and Wei (1982), page 163]

$$T_n^2 = O(n^2 \log\log n) \quad \text{a.s.} \tag{A2.8}$$

and

$$\liminf_{n\to\infty} n^{-2}(\log\log n)T_n^2 = \sigma^2/4 \quad \text{a.s.} \tag{A2.9}$$

By (A2.5)–(A2.7) and (A2.9),

$$\lim_{n\to\infty} (\lambda_n/\delta_n) = 0 \quad \text{a.s.} \tag{A2.10}$$

If (2.26) were true, then (A2.2) would imply

$$\left(\sum_1^n \mathbf{z}_i\mathbf{z}_i'\right)\bigg/ a_n \to A\Gamma A' \quad \text{a.s.} \tag{A2.11}$$

Since $A\Gamma A'$ is positive definite, (A2.11) implies that $\liminf_{n\to\infty}(\lambda_n/\delta_n) > 0$ a.s. This contradicts (A2.10). Hence, (2.26) does not hold. Now, let us prove (A2.4). Observe that this is equivalent to showing that

$$e_n = \left(\sum_1^n t_i u_i\right)\bigg/(T_n U_n) \to 0 \quad \text{a.s.} \tag{A2.12}$$

In view of (3.17) (with $p = 1$) and (A2.1), we have that

$$(t_n u_n)/(T_n U_n) \to 0 \quad \text{a.s.,} \tag{A2.13}$$

$$T_n^{-1}T_{n-1} \to 1 \quad \text{a.s.} \quad \text{and} \quad U_n^{-1}U_{n-1} \to 1 \quad \text{a.s.} \tag{A2.14}$$

These imply that

$$e_n - e_{n-1} \to 0 \quad \text{a.s.} \tag{A2.15}$$

By (A2.1), (A2.7)–(A2.9), Lemma 1 and the strong law of large numbers, we obtain that

$$\begin{aligned}
\sum_1^n t_i u_i &= \gamma \sum_1^{n-1} t_i u_i + \sum_1^{n-1} t_i \varepsilon_{i+1} + \gamma \sum_1^{n-1} u_i \varepsilon_{i+1} + \sum_1^n \varepsilon_i^2 \\
&= \gamma \sum_1^{n-1} t_i u_i + o(T_{n-1}\log T_{n-1}) + o(U_{n-1}\log U_{n-1}) + O(n) \\
&= \gamma \sum_1^{n-1} t_i u_i + o(T_{n-1}\log n) + o(n\log n) \\
&= \gamma \sum_1^{n-1} t_i u_i + o(T_{n-1}U_{n-1}) \quad \text{a.s.}
\end{aligned}$$

This together with (A2.14) implies that

$$e_n = \gamma e_{n-1} + o(1) \quad \text{a.s.} \tag{A2.16}$$

Let Ω be the event that (A2.15) and (A2.16) hold. Then, $P(\Omega) = 1$. By the Cauchy–Schwarz inequality, $\sup_n |e_n(\omega)| \leq 1$, $\forall\, \omega \in \Omega$. In order to prove (A2.12), it is sufficient to show that for all $\omega \in \Omega$, the limit set of $C = \{e_n(\omega)\colon n \geq 1\}$ consists of 0 only. Let e be a limit point of C. By (A2.15) and (A2.16), $e = \gamma e$. Since $\gamma \neq 1$, $e = 0$. This completes our proof. □

REFERENCES

BURKHOLDER, D. L., DAVIS, B. J. and GUNDY, R. F. (1972). Inequalities for convex functions of operators on martingales. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **2** 223–240. Univ. California Press.

CHAN, N. H. and WEI, C. Z. (1986). Limiting distributions of least squares estimates of unstable autoregressive processes. *Ann. Statist.* To appear.

CHOW, Y. S. (1965). Local convergence of martingales and the law of large numbers. *Ann. Math. Statist.* **36** 552–58.

CHOW, Y. S. and TEICHER, H. (1978). *Probability Theory: Independence, Interchangeability, Martingales.* Springer, New York.

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.

FULLER, W. A. (1976). *Introduction to Statistical Time Series.* Wiley, New York.

FULLER, W. A. and HASZA, D. P. (1981). Properties of predictors for autoregressive time series. *J. Amer. Statist. Assoc.* **76** 155–161.

GOODWIN, G. C. and SIN, K. S. (1984). *Adaptive Filtering Prediction and Control.* Prentice-Hall, Englewood Cliffs, N.J.

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

LAI, T. L. and WEI, C. Z. (1983). Asymptotic properties of general autoregressive models and strong consistency of least squares estimates of their parameters. *J. Multivariate Anal.* **13** 1–23.

LAI, T. L. and WEI, C. Z. (1985). Asymptotic properties of multivariate weighted sums with applications to stochastic regression in linear dynamic systems. In *Multivariate Analysis VI* (P. R. Krishnaiah, ed.) 375–393. North-Holland, Amsterdam.

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

TSAY, R. S. (1984). Order selection in nonstationary autoregressive models. *Ann. Statist.* **12** 1425–1433.

WEI, C. Z. (1985). Asymptotic properties of least squares estimates in stochastic regression models. *Ann. Statist.* **13** 1498–1508.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF MARYLAND
COLLEGE PARK, MARYLAND 20742