The Annals of Statistics
1988, Vol. 16, No. 1, 367–401

LIMITING DISTRIBUTIONS OF LEAST SQUARES ESTIMATES OF UNSTABLE AUTOREGRESSIVE PROCESSES

BY N. H. CHAN AND C. Z. WEI[1]

Indiana University and University of Maryland

An autoregressive process $y_n = \beta_1 y_{n-1} + \cdots + \beta_p y_{n-p} + \varepsilon_n$ is said to be unstable if the characteristic polynomial $\phi(z) = 1 - \beta_1 z - \cdots - \beta_p z^p$ has all roots on or outside the unit circle. The limiting distribution of the least squares estimate of $(\beta_1, \dots, \beta_p)$ is derived and characterized as a functional of stochastic integrals under a $2 + \delta$ moment assumption on ε_n. Up to the present, distributional results were available only with substantial restrictions on the possible roots which did not suggest the form of the distribution for the general case. To establish the limiting distribution, a result concerning the weak convergence of a sequence of random variables to a stochastic integral, which is of independent interest, is also developed.

1. Introduction. Consider the autoregressive AR(p) model

$$y_n = \beta_1 y_{n-1} + \cdots + \beta_p y_{n-p} + \varepsilon_n, \tag{1.1}$$

where y_n is the observation, ε_n is the (unobservable) random disturbance (noise) at time n, p is the order of the process and $\beta_1, \dots, \beta_p$ are the parameters of the model. Throughout the sequel, we shall assume that the initial conditions $y_0, \dots, y_{1-p}$ are fixed and $\{\varepsilon_n\}$ 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 of $\{\varepsilon_n\}$ is a sequence of independent random variables with zero means. We shall also assume $y_0, \dots, y_{1-p}$ are $\mathscr{F}_0$-measurable so that y_n is $\mathscr{F}_n$-measurable. Let

$$\phi(z) = 1 - \beta_1 z - \cdots - \beta_p z^p \tag{1.2}$$

denote the characteristic polynomial of the autoregressive model (1.1). When all roots of $\phi(z)$ are outside the unit circle, the model is said to be *asymptotically stationary*. If all roots are on or outside the unit circle, the model is said to be *unstable*. A commonly used estimate of the parameter vector $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$ is the least squares estimate

$$\mathbf{b}_n = \left(\sum_{t=0}^{n-1} \mathbf{y}_t \mathbf{y}_t' \right)^{-1} \sum_{t=1}^{n} \mathbf{y}_{t-1} y_t, \qquad n > p, \tag{1.3}$$

where

$$\mathbf{y}_t = (y_t, \dots, y_{t-p+1})'. \tag{1.4}$$

Received January 1986; revised April 1987.
[1]Supported by the National Science Foundation under Grant DMS-84-04081.
AMS 1980 *subject classifications*. Primary 62M10, 62E20; secondary 60F17.
Key words and phrases. Unstable autoregressive process, least squares, stochastic integral, limiting distribution.

The statistical properties of $\mathbf{b}_n$ have been well studied in the literature for the asymptotically stationary case. Recently, due to their importance in system identification and control [Goodwin and Payne (1977), Ljung (1976) and Graupe (1980)] and in modeling business and economic data [Anderson (1985), Phillips (1985) and Box and Jenkins (1976)], research in nonstationary autoregressive processes has been receiving considerable attention.

The consistency problem for $\mathbf{b}_n$ was recently solved by Lai and Wei (1983). They showed that irrespective of the location of the roots, $\mathbf{b}_n$ is always strongly consistent when $E(\varepsilon_n^2|\mathscr{F}_{n-1}) = \sigma^2$ a.s. and $\sup E(|\varepsilon_n|^{2+\delta}|\mathscr{F}_{n-1}) < \infty$ a.s. for some $\delta > 0$. However, the problem of finding the asymptotic distribution of $\mathbf{b}_n$ under the general nonstationarity assumption still remained open.

Mann and Wald (1943) first obtained the asymptotic distribution of $\mathbf{b}_n$ when y_n is asymptotically stationary. Under the assumption that the ε_n's are i.i.d. with all moments, they showed that

$$\sqrt{n}(\mathbf{b}_n - \boldsymbol{\beta}) \to_{\mathscr{L}} N(0, \Sigma), \tag{1.5}$$

where Σ is a positive definite matrix. The higher moment assumption on the $\{\varepsilon_n\}$ was subsequently removed by Anderson (1959). For the unstable case, White (1958) considered the AR(1) model with i.i.d. $N(0, \sigma^2)$ random errors ε_n and $\beta_1 = 1$. He obtained the limiting moment generating function of $n(\mathbf{b}_n - 1)$ and claimed that

$$n(\mathbf{b}_n - 1) \to_{\mathscr{L}} \tau = \tfrac{1}{2}\left(W^2(1) - 1\right)\Big/\int_0^1 W^2(t)\,dt, \tag{1.6}$$

where $W(t)$ is a standard Brownian motion. Rao (1978) expressed τ through complicated special function expressions which he described as "formidable." Dickey and Fuller (1979) proved that

$$n(\mathbf{b}_n - 1) \to_{\mathscr{L}} \nu = \tfrac{1}{2}(T^2 - 1)/\Gamma, \tag{1.7}$$

where

$$T = \sum_{k=1}^{\infty} \sqrt{2}\,\gamma_k Z_k, \qquad \Gamma = \sum_{k=1}^{\infty} \gamma_k^2 Z_k^2, \qquad \gamma_k = 2(-1)^{k+1}/(2k-1)\pi$$

and the Z_k's are i.i.d. $N(0,1)$ random variables. Utilizing this formula, they tabulated the percentiles of τ by the Monte Carlo method. An analytic method which involves the calculation of complex numbers was proposed by Evans and Savin (1981). Extensions and applications of (1.7) have been carried out by Phillips (1985), Dickey and Fuller (1981), Fuller and Hasza (1980) and Hasza and Fuller (1979). Ahtola and Tiao (1987) considered the AR(2) model where $\phi(z)$ has a pair of complex roots $e^{i\theta}$ and $e^{-i\theta}$. Under the assumption that θ is a rational multiple of 2π, they characterized the limiting distribution of $\mathbf{b}_n$.

However, there are important time series models, e.g., seasonal models, where $\phi(z)$ may have several roots on the unit circle. In such cases, not only the multiplicity of each root has to be handled, but also the relationship between different roots has to be considered. Apparently, the results stated previously are not adequate for such generality and a unified theory for the general unstable model is required. It is the objective of this paper to provide such a unified theory under a $2 + \delta$ moment condition (2.10) on ε_n. For related results under the second moment assumption, see Anderson (1959), Stigum (1974) and Solo (1984).

In the following, we shall use the functional central limit theorem to derive the limiting distribution of $\mathbf{b}_n$. Apart from the moment generating function method used by White (1958), the limiting distribution of $\mathbf{b}_n$ is traditionally obtained through quadratic form considerations. This approach can be easily explained by an AR(1) model. In this case,

$$(1.8) \qquad \begin{aligned}(\mathbf{b}_n - \beta_1) &= \sum_{t=1}^{n} y_{t-1}\varepsilon_t \Big/ \sum_{t=1}^{n} y_{t-1}^2 \\ &= \mathbf{e}_n' A_n \mathbf{e}_n / \mathbf{e}_n' B_n \mathbf{e}_n,\end{aligned}$$

where $\mathbf{e}_n = (\varepsilon_1, \dots, \varepsilon_n)'$ and A_n and B_n are $n \times n$ matrices defined accordingly. By finding the eigenvalues λ_{in} and eigenvectors $\mathbf{u}_{in}$ of B_n, one can express $\sum_{t=1}^{n} y_{t-1}^2$ as $\sum_{i=1}^{n} \lambda_{in} Z_{in}^2$, where $Z_{in} = \mathbf{e}_n' \mathbf{u}_{in}$, which are i.i.d. $N(0,1)$ random variables when the ε_n's are standard normal random errors. Reexpressing A_n in terms of the eigenvectors of B_n, $\sum_{t=1}^{n} y_{t-1}\varepsilon_t$ can then be represented as a weighted sum of Z_{in}. Dickey and Fuller (1979), instead of establishing a direct connection between τ and ν, used this transformation to derive (1.7). A similar transformation was also used by Ahtola and Tiao (1987) to obtain their result. However, instead of finding the eigenvalues and eigenvectors of B_n in the first place, it is equally plausible to obtain the eigenvalues and eigenvectors of A_n first and reexpress B_n in terms of the eigenvectors of A_n. It is unclear a priori which of A_n and B_n should be chosen so that a convenient expansion can be achieved. This ambiguity becomes more severe in the AR(p) case where more quadratic forms are involved.

On the other hand, the functional central limit theorem approach enables us to express the limiting distribution of $\mathbf{b}_n$ in terms of functionals of standard Brownian motions. Through series expansions of the underlying Brownian motions, the limiting distribution in turn can be expressed as a form similar to ν. This observation (see Corollary 3.1.3) establishes a direct connection between τ and ν, which answers the question raised recently by Solo (1984). Moreover, it also provides a "neater" series representation (see Corollary 3.3.8 and the remark following it) for the distribution considered by Ahtola and Tiao (1987). Above all, we believe the method developed herein can also be used to derive the limiting distributions of the related test statistics for unstable AR(p) models discussed in the literature. For an overview of the hypothesis testing problems and related issues, see Fuller (1985).

Note that the functional central limit theorem is a common tool for deriving limiting distributions in the statistical literature. For the nonstationary AR(1) model, Solo (1984) and Phillips (1985) applied it even to cases with moving average or mixing errors. However, their statistics (as in many other situations in the literature) have to be expressed as a continuous functional on some function spaces so that the continuous mapping theorem can be applied to obtain the limiting distributions. This is not the case for the AR(p) model. The functionals involved in $\mathbf{b}_n$ may not be continuous, so that the continuous mapping theorem is no longer applicable. In Section 2, we shall develop some auxiliary probability theorems to overcome this difficulty. In particular, Theorem 2.4 is of independent interest since it gives a sufficient condition for a sequence of random variables to converge in distribution to a stochastic integral. Section 3, the main body of this paper, consists of applying such stochastic integrals to characterize the limiting distribution of $\mathbf{b}_n$. A further discussion on some related topics is briefly given in Section 4.

2. Auxiliary probability theorems.

Theorem 2.1. *Let* $\{X_n\}$ *be a sequence of random variables such that*

(2.1)
(i) $E|X_n| = O(n^\alpha)$ *for some* $\alpha > 0$;
(ii) *there exist random variables* $A(n, m)$, $B(n, m)$ *and constants* $\gamma \geq 0$, $\delta \geq 0$ *and* $c > 0$ *with* $|X_n - X_m| \leq A(n, m)B(n, m)$;

$$EA^2(n, m) \leq cn^\gamma \quad \text{and} \quad EB^2(n, m) \leq cn^\delta(n - m) \text{ for } n \geq m. \tag{2.2}$$

If $2\alpha > \gamma + \delta$ *and* $e^{i\theta} \neq 1$, *then*

$$\sup_{1 \leq j \leq n} \left| \sum_{t=1}^{j} e^{it\theta} X_t \right| = o_p(n^{\alpha+1}). \tag{2.3}$$

Proof. Since $2\alpha - (\gamma + \delta) > 0$, for each n we can always choose $N(n) + 1$ integers n_k with $1 = n_0 < n_1 < \cdots < n_N = n$ such that

$$\max\{n_{k+1} - n_k \colon 0 \leq k \leq N - 1\} = o(n^{2\alpha - (\gamma+\delta)}) \tag{2.4}$$

and

$$N = N(n) \uparrow \infty, \qquad N(n) = o(n). \tag{2.5}$$

Let

$$Y_k = Y_k(n) = \sum_{t=n_{k-1}}^{n_k - 1} e^{it\theta}(X_t - X_{n_{k-1}}) \tag{2.6}$$

and

$$s = s(j) = \sup\{k \colon n_k \leq j\}. \tag{2.7}$$

Then

$$\sum_{t=1}^{j} e^{it\theta} X_t = \sum_{t=n_s}^{j} e^{it\theta}(X_t - X_{n_s}) + \sum_{k=1}^{s} Y_k + \left(\sum_{t=n_s}^{j} e^{it\theta} \right) X_{n_s} + \sum_{k=1}^{s} \left(\sum_{t=n_{k-1}}^{n_k - 1} e^{it\theta} \right) X_{n_{k-1}}.$$

Since for any integers $l \geq m \geq 1$,

$$\left|\sum_{t=m}^{l} e^{it\theta}\right| = |(e^{im\theta} - e^{i(l+1)\theta})/(1 - e^{i\theta})| \leq 2/(1 - \cos\theta),$$

we have that

$$\begin{aligned}\left|\sum_{t=1}^{j} e^{it\theta}X_t\right| &\leq \sum_{n_s}^{j}|X_t - X_{n_s}| + \sum_{k=1}^{s}\sum_{n_{k-1}}^{n_k-1}|X_t - X_{n_{k-1}}| \\ &\quad + (2/(1-\cos\theta))\left(|X_{n_s}| + \sum_{k=1}^{s}|X_{n_{k-1}}|\right) \\ &\leq \sum_{k=1}^{N}\sum_{n_{k-1}}^{n_k-1}|X_t - X_{n_{k-1}}| + (2/(1-\cos\theta))\left(\sum_{k=1}^{N}|X_{n_k}|\right).\end{aligned} \tag{2.8}$$

Note that the right-hand side of (2.8) is independent of j. Consequently, by (2.1), (2.2) and the Cauchy–Schwarz inequality, we obtain

$$\begin{aligned}E\sup_{1\leq j\leq n}\left|\sum_{t=1}^{j} e^{it\theta}X_t\right| &\leq E\sum_{k=1}^{N}\sum_{n_{k-1}}^{n_k-1}A(t, n_{k-1})B(t, n_{k-1}) \\ &\quad + (2/(1-\cos\theta))\left(\sum_{k=1}^{N}E|X_{n_k}|\right) \\ &\leq \left(\sum_{k=1}^{N}\sum_{n_{k-1}}^{n_k-1}EA(t, n_{k-1})^2\right)^{1/2}\left(\sum_{k=1}^{N}\sum_{n_{k-1}}^{n_k-1}EB(t, n_{k-1})^2\right)^{1/2} \\ &\quad + O(Nn^{\alpha}) \\ &\leq (cn^{\gamma+1})^{1/2}\left(c\sum_{k=1}^{N}n^{\delta}(n_k - n_{k-1})^2\right)^{1/2} + o(n^{\alpha+1}) \\ &\leq cn^{(\gamma+\delta+2)/2}\max_{1\leq k\leq N}(n_k - n_{k-1})^{1/2} + o(n^{\alpha+1}) \\ &= o(n^{\alpha+1}).\end{aligned} \tag{2.9}$$

By Markov's inequality, (2.3) is proved. □

REMARK. Suppose $e^{i\theta} \neq 1$ and (i) holds. Then, by similar arguments, Theorem 2.1 still holds if (ii) is replaced by:

> (ii′) There exist random variables $A_j(n, m)$, $B_j(n, m)$ and constants $\gamma_j \geq 0$, $\delta_j \geq 0$ for $j = 1, \dots, q$ such that $|X_n - X_m| \leq \sum_{j=1}^{q} A_j(n, m)B_j(n, m)$, $EA_j^2(n, m) \leq cn^{\gamma_j}$, $EB_j^2(n, m) \leq cn^{\delta_j}(n - m)$ for $n \geq m$ and $\gamma_j + \delta_j < 2\alpha$ for $j = 1, \dots, q$.

In the rest of this section, we shall consider several results related to the functional central limit theorem. For definitions and background facts, we refer the reader to Billingsley (1968).

Throughout the sequel, we use $D = D[0,1]$ to designate the space of functions $x(t)$ on $[0,1]$ which are right continuous and have left-hand limits. D will be equipped with the Skorokhod topology. The weak convergence (or the convergence in distribution) of a sequence of random elements X_n in D to a random element X in D will be denoted by $X_n \to_{\mathscr{L}} X$.

THEOREM 2.2. *Let $\{\varepsilon_n\}$ be a sequence of martingale differences with respect to an increasing sequence of σ-fields $\{\mathscr{F}_n\}$ such that*

$$E\{\varepsilon_n^2 | \mathscr{F}_{n-1}\} = 1 \quad a.s., \tag{2.10}$$
$$\sup_n E\{|\varepsilon_n|^{2+\alpha} | \mathscr{F}_{n-1}\} < \infty \quad a.s. \text{ for some } \alpha > 0.$$

Suppose $\{\mathbf{z}_n\}$ is a sequence of random vectors in R^q which satisfies

$$\mathbf{z}_n = A\mathbf{z}_{n-1} + \boldsymbol{\varepsilon}_n, \tag{2.11}$$

where $\mathbf{z}_0 = 0$, $\boldsymbol{\varepsilon}_n' = (\varepsilon_n, 0, \dots, 0)$ and A is a $q \times q$ constant matrix with all eigenvalues lying inside the unit circle. Let $\theta_k \in (0, \pi)$ such that $\theta_k \neq \theta_j$ if $k \neq j$ for $k, j = 1, 2, \dots, l$. Consider the following random elements in $(\prod_{i=1}^{2l+2} D) \times R^q$:

$$X_n'(u, v, t_1, \dots, t_{2l}) = \frac{1}{\sqrt{n}} \left(\sum_1^{[nu]} \varepsilon_k, \sum_1^{[nv]} (-1)^k \varepsilon_k, \sum_1^{[nt_1]} \sqrt{2} \sin k\theta_1 \varepsilon_k, \right. \tag{2.12}$$
$$\left. \sum_1^{[nt_2]} \sqrt{2} \cos k\theta_1 \varepsilon_k, \dots, \sum_1^{[nt_{2l}]} \sqrt{2} \cos k\theta_l \varepsilon_k, \sum_1^n \mathbf{z}_{k-1}' \varepsilon_k \right).$$

We have

$$X_n \to_{\mathscr{L}} (W, N),$$

where W is a standard Brownian motion of dimension $2l + 2$ and N a normal random vector which is independent of W and has zero mean and covariance matrix $\Sigma = \sum_{k=0}^{\infty} A^k \mathbf{e}\mathbf{e}'(A')^k$, where $\mathbf{e}' = (1, 0, \dots, 0)$.

REMARK. The elements of (2.12) will be the basic processes corresponding to various components with different unit roots.

PROOF. It is known [Lai and Wei (1985)] that under (2.10),

$$\lim_{n \to \infty} \frac{1}{n} \sum_1^n \mathbf{z}_t \mathbf{z}_t' = \Sigma \quad \text{a.s.} \tag{2.13}$$

Note that for $\theta, \delta \in [0, 2\pi)$,

$$\lim_{n \to \infty} \frac{1}{n} \sum_1^{[nt]} \cos k\theta \sin k\delta = 0$$

and

$$\lim_{n \to \infty} \frac{1}{n} \sum_1^{[nt]} \cos k\theta \cos k\delta = \lim_{n \to \infty} \frac{1}{n} \sum_1^{[nt]} \sin k\theta \sin k\delta = 0, \quad \text{if } \theta \neq \delta.$$

Furthermore, by (2.13), $\lim_{n\to\infty}(1/n)\sum_{k=1}^{n}\|z_k\|^2 = c$ a.s. Hence,

$$\lim_{n\to\infty} \max_{1\le k\le n} \|z_k\|^2 \to 0 \text{ a.s.}$$

In view of this, (2.11) and the strong law of large numbers,

$$\begin{aligned}\frac{1}{n}\sum_{k=1}^{j} e^{ik\theta}\mathbf{z}_k &= \frac{1}{n}\sum_{k=1}^{j} e^{ik\theta}A\mathbf{z}_{k-1} + \frac{1}{n}\sum_{k=1}^{j} e^{ik\theta}\boldsymbol{\varepsilon}_k \\ &= e^{i\theta}A\left\{\frac{1}{n}\sum_{k=1}^{j} e^{ik\theta}\mathbf{z}_k\right\} - \frac{1}{n}e^{i(j+1)\theta}A\mathbf{z}_j + \frac{1}{n}\sum_{k=1}^{j} e^{ik\theta}\boldsymbol{\varepsilon}_k \\ &= e^{i\theta}A\left\{\frac{1}{n}\sum_{k=1}^{j} e^{ik\theta}\mathbf{z}_k\right\} + o(1) \quad \text{a.s.}\end{aligned}$$

Hence, with probability 1, every limit point of $\{(1/n)\sum_1^{[nt]} e^{ik\theta}\mathbf{z}_k \colon n \ge 1\}$ satisfies $\mathbf{z} = e^{i\theta}(A\mathbf{z})$. Since $e^{-i\theta}$ is not an eigenvalue of A, $\mathbf{z} = \mathbf{0}$. Consequently,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{1}^{[nt]} e^{ik\theta}\mathbf{z}_k = \mathbf{0} \quad \text{a.s.}$$

Moreover, by the conditional Markov inequality and (2.10), for all $\delta > 0$,

$$\begin{aligned}&\sum_{k=1}^{n} E\left(\|\mathbf{z}_k\varepsilon_{k+1}\|^2 I\left[\|\mathbf{z}_k\varepsilon_{k+1}\| > \sqrt{n}\,\delta\right] | \varepsilon_1, \dots, \varepsilon_k\right) \\ &\quad\le \left(\sum_{k=1}^{n}\|\mathbf{z}_k\|^{2+\alpha}\right) n^{-\alpha/2}\delta^{-\alpha}K \quad (\text{for some } K > 0) \\ &\quad\le \left(\sum_{k=1}^{n}\|\mathbf{z}_k\|^2\right)\left(\frac{1}{n}\max_{1\le k\le n}\|\mathbf{z}_k\|^2\right)^{\alpha/2}\delta^{-\alpha}K \\ &\quad= o(n) \quad \text{a.s.}\end{aligned}$$

Now, applying the standard functional central limit theorem [Helland (1982), Theorem 3.3], the proof is complete. □

THEOREM 2.3. *Let $\{X_n\}$ and X be random elements taking values in $\prod_{i=1}^{m} D$. For each $u \in [0,1]$ and any continuous function $f\colon R^m \to R$, define*

$$Y(u) = \left(X^1(u), \dots, X^m(u)\right) \quad \textit{and} \quad Z(t) = \int_0^t f(Y(u))\,du.$$

Similarly, we can define Y_n and Z_n. If $X_n \to_{\mathscr{L}} X$, then $(X_n, Z_n) \to_{\mathscr{L}} (X, Z)$.

PROOF. It is clear that the processes defined in Theorem 2.3 are taking values in D. Now consider the mapping $g\colon \prod_{i=1}^{m} D \to \prod_{i=1}^{m+1} D$, defined by

$$g(x)(s,t) = \left(x(s), \int_0^t f(y(u))\,du\right),$$

$$\forall\, x \in \prod_{i=1}^{m} D,\ \forall\, s \in [0,1]^m \text{ and } \forall\, t \in [0,1].$$

Since f is continuous, g is continuous. By the continuous mapping theorem [Billingsley (1968), Theorem 5.1], the result is established. □

THEOREM 2.4. *Let $\{X_n\}$ and $\{Y_n\}$ be two sequences of random variables. Let $U_n(t) = \sum_{k=1}^{[nt]} X_k$, $V_n(t) = \sum_{k=1}^{[nt]} Y_k$ and $S_n(t) = X_{[nt]}$. Assume one of the following two conditions holds*:

(i) *There is $a_n \uparrow \infty$ such that*

$$\left(a_n^{-1} S_n, n^{-1/2} V_n\right) \to_{\mathscr{L}} (S, W),$$

where W is the standard Brownian motion with respect to an increasing sequence of σ-fields $\mathscr{G}_t$ and S is $\mathscr{G}_t$-adaptive.

(ii) *There are increasing σ-fields $\mathscr{F}_n$ such that $(X_n, Y_n)'$ is a sequence of martingale differences with respect to $\mathscr{F}_n$. Moreover,*

$$E\left(X_n^2 + Y_n^2 \mid \mathscr{F}_{n-1}\right) \le c \quad a.s. \text{ for some constant } c > 0 \tag{2.14}$$

and

$$n^{-1/2}(U_n, V_n) \to_{\mathscr{L}} (B, W), \tag{2.15}$$

where B and W are two Brownian motions with respect to an increasing sequence of σ-fields $\mathscr{G}_t$.

Then

$$\left(b_n^{-1} U_n, n^{-1/2} V_n, \left(n^{1/2} b_n\right)^{-1} \sum_{k=1}^{n-1} U_n\left(\frac{k}{n}\right) Y_{k+1}\right) \to_{\mathscr{L}} \left(H, W, \int_0^1 H\, dW\right), \tag{2.16}$$

where $b_n = n a_n$ and $H(t) = \int_0^t S(u)\, du$ in (i) *and $b_n = n^{1/2}$ and $H = B$ in* (ii).

REMARK. In Section 3, Theorem 2.4(i) will be applied to the case where S is a functional of W. Hence, we can choose $\mathscr{G}_t = \sigma\{W(s)\colon s \le t\}$. Theorem 2.4(ii) will be applied to the cases where either $H = W$ or H and W are independent. In both cases, we can choose $\mathscr{G}_t = \sigma\{H(s), W(s)\colon s \le t\}$.

PROOF. Suppose condition (i) holds.

Note that $U_n(k/n) = \sum_1^k X_i$. Summation by parts gives

$$\sum_{k=1}^{n-1} U_n\left(\frac{k}{n}\right) Y_{k+1} = -\sum_{k=1}^{n-1} V_n\left(\frac{k}{n}\right) X_k + \left(\sum_1^{n-1} X_k\right) V_n(1). \tag{2.17}$$

Let $F(t) = \int_0^t f(s)\, ds$. Since the functional

$$K(f, g) = \left(F, g, -\int_0^1 g(s) f(s)\, ds + g(1) F(1)\right)$$

is a continuous functional from $D \times D$ into $D \times D \times R$, by the continuous mapping theorem and Theorem 2.3,

$$\left(b_n^{-1} U_n, n^{-1/2} V_n, \left(n^{1/2} b_n\right)^{-1} \sum_{k=1}^{n-1} U_n\left(\frac{k}{n}\right) Y_{k+1}\right) \to_{\mathscr{L}} (H, W, L),$$

where $H(t) = \int_0^t S(u)\,du$ and

$$L = -\int_0^1 W(t)S(t)\,dt + H(1)W(1).$$

Since $S \in D$, $\|S\|_\infty < \infty$ a.s. This implies that H is continuous and of bounded variation. Consequently, H is a semimartingale with respect to G_t and

$$L = -\int_0^1 W\,dH + H(1)W(1)$$
$$= \int_0^1 H\,dW,$$

by integration by parts [Elliott (1982), page 141].

Now assume condition (ii) holds.

Since with probability 1, B and W have continuous paths, the convergence in the Skorokhod topology is equivalent to uniform convergence [Billingsley (1968), page 112]. Furthermore, we can equip D with a complete metric so that the induced topology is equivalent to the Skorokhod topology [Billingsley (1968), pages 114–115]. By the Skorokhod representation theorem [Skorokhod (1956)], there are a probability space Ω and random elements U^n, V^n in $D[0,1]$ such that

$$\|(U^n, V^n) - (B, W)\|_\infty \to 0 \quad \text{a.s.} \tag{2.18}$$

and

$$(U^n, V^n) =_{\mathscr{L}} n^{-1/2}(U_n, V_n).$$

Let

$$G^n = \sum_{k=1}^{n-1} U^n\left(\frac{k}{n}\right)\left(V^n\left(\frac{k+1}{n}\right) - V^n\left(\frac{k}{n}\right)\right) \tag{2.19}$$

and

$$G_n = \frac{1}{n}\sum_{k=1}^{n-1} U_n\left(\frac{k}{n}\right)Y_{k+1}.$$

Then

$$(U^n, V^n, G^n) =_{\mathscr{L}} \left(n^{-1/2}U_n, n^{-1/2}V_n, G_n\right).$$

In order to show (2.16), it is sufficient to show that

$$G^n \to_p \int_0^1 B\,dW. \tag{2.20}$$

By (2.18) and Egorov's theorem, given $\varepsilon > 0$, there is an event $\Omega_\varepsilon \subset \Omega$ such that $P(\Omega_\varepsilon) \ge 1 - \varepsilon$ and

$$\sup\{\|(U^n(\omega), V^n(\omega)) - (B(\omega), W(\omega))\|_\infty : \omega \in \Omega_\varepsilon\} = \delta_n \to 0. \tag{2.21}$$

Note that δ_n is a sequence of constants. We can choose integers $N(n) \uparrow \infty$ such that

$$N(n)\,\delta_n^2 \to 0 \quad \text{and} \quad N(n)/n \to 0. \tag{2.22}$$

For each n, we can further choose a partition $\{t_0, \dots, t_{N(n)}\}$ of $[0,1]$ such that

$$(2.23)\quad 0 = t_0 < t_1(n) = \frac{n_1}{n} < t_2(n) = \frac{n_2}{n} < \cdots < t_{N(n)} = \frac{n_{N(n)}}{n} = 1,$$

$$(2.24)\qquad \max\{|t_{i+1} - t_i| : 0 \le i \le N(n) - 1\} = o(1).$$

We first claim that

$$(2.25)\qquad G^n = \sum_{k=1}^{N(n)} U^n(t_{k-1})(V^n(t_k) - V^n(t_{k-1})) + o_p(1).$$

Clearly,

$$(2.26)\qquad \begin{aligned} J_n &= G^n - \sum_{k=1}^{N(n)} U^n(t_{k-1})(V^n(t_k) - V^n(t_{k-1})) \\ &= \sum_{k=1}^{N(n)} \left[\sum_{i=n_{k-1}}^{n_k - 1} \left(U^n\left(\frac{i}{n}\right) - U^n(t_{k-1}) \right) \left(V^n\left(\frac{i+1}{n}\right) - V^n\left(\frac{i}{n}\right) \right) \right]. \end{aligned}$$

Using the fact that (X_k, Y_k) are martingale differences and (2.14), we have

$$\begin{aligned} EJ_n^2 &= \sum_{k=1}^{N(n)} \sum_{i=n_{k-1}}^{n_k-1} E\left(U^n\left(\frac{i}{n}\right) - U^n(t_{k-1}) \right)^2 \left(V^n\left(\frac{i+1}{n}\right) - V^n\left(\frac{i}{n}\right) \right)^2 \\ &\le \sum_{k=1}^{N(n)} c^2 \sum_{i=n_{k-1}}^{n_k-1} \left(\frac{i}{n} - \frac{n_{k-1}}{n} \right) \frac{1}{n} \\ &\le \frac{c^2}{n^2} \sum_{k=1}^{N(n)} (n_k - n_{k-1})^2 \\ &\le c^2 \max_{1 \le k \le N(n)} \left\{ \frac{n_k}{n} - \frac{n_{k-1}}{n} \right\} \to 0, \quad \text{by (2.24).} \end{aligned}$$

By the Markov inequality, (2.25) is proved.

Next, we show that

$$(2.27)\qquad \begin{aligned} &I_{\Omega_\varepsilon} \sum_1^{N(n)} U^n(t_{k-1})(V^n(t_k) - V^n(t_{k-1})) \\ &\quad = I_{\Omega_\varepsilon} \sum_1^{N(n)} B(t_{k-1})(V^n(t_k) - V^n(t_{k-1})) + o_p(1). \end{aligned}$$

Note that by the Cauchy–Schwarz inequality and (2.21),

$$\begin{aligned} &\left| \sum_1^{N(n)} [U^n(t_{k-1}) - B(t_{k-1})] I_{\Omega_\varepsilon}(V^n(t_k) - V^n(t_{k-1})) \right|^2 \\ &\quad \le \sum_1^{N(n)} [U^n(t_{k-1}) - B(t_{k-1})]^2 I_{\Omega_\varepsilon} \sum_1^{N(n)} (V^n(t_k) - V^n(t_{k-1}))^2 \\ &\quad \le N(n)\, \delta_n^2 \sum_1^{N(n)} (V^n(t_k) - V^n(t_{k-1}))^2. \end{aligned}$$

By (2.22),

$$E\left\{N(n)\,\delta_n^2\sum_1^{N(n)}\left(V^n(t_k)-V^n(t_{k-1})\right)^2\right\}=N(n)\,\delta_n^2\sum_1^{N(n)}E\left(V^n(t_k)-V^n(t_{k-1})\right)^2$$
$$=N(n)\,\delta_n^2\sum_1^{N(n)}(t_k-t_{k-1})$$
$$=N(n)\,\delta_n^2\to 0.$$

This shows (2.27).

Now, summation by parts gives

$$\sum_1^{N(n)}B(t_{k-1})(V^n(t_k)-V^n(t_{k-1}))$$
$$=-\sum_1^{N(n)}V^n(t_k)(B(t_k)-B(t_{k-1}))+B(1)V^n(1).$$

By a similar argument, we can replace $V^n(t_k)$ by $W(t_k)$ and obtain

$$I_{\Omega_\varepsilon}\sum_1^{N(n)}B(t_{k-1})(V^n(t_k)-V^n(t_{k-1}))$$
$$=-I_{\Omega_\varepsilon}\sum_1^{N(n)}W(t_k)(B(t_k)-B(t_{k-1}))+I_{\Omega_\varepsilon}B(1)W(1)+o_p(1)$$
$$=I_{\Omega_\varepsilon}\sum_1^{N(n)}B(t_{k-1})(W(t_k)-W(t_{k-1}))+o_p(1) \tag{2.28}$$
$$=I_{\Omega_\varepsilon}\int_0^1 B(t)\,dW(t)+o_p(1).$$

The last identity is insured by the fact that

$$E\left(\sum_1^{N(n)}B(t_{k-1})(W(t_k)-W(t_{k-1}))-\int_0^1 B(t)\,dW(t)\right)^2$$
$$=E\left(\sum_1^{N(n)}\int_{t_{k-1}}^{t_k}(B(t_{k-1})-B(t))\,dW(t)\right)^2$$
$$=\sum_1^{N(n)}\int_{t_{k-1}}^{t_k}(t_k-t)\,dt$$
$$\le\max_{1\le k\le N(n)}(t_k-t_{k-1})=o(1).$$

Combining (2.28), (2.25) and (2.27), our theorem is proved. □

REMARK. By a similar argument, (2.16) can be generalized to the high dimension case. More precisely, assume that there exist increasing σ-fields $\mathscr{F}_n$

such that $\mathbf{X}_n = (X_n(1), \dots, X_n(l))'$ is a sequence of martingale differences with respect to $\mathscr{F}_n$. Define

$$\mathbf{U}_n(\mathbf{t}) = n^{-1/2}\left(\sum_{k=1}^{[nt_1]} X_k(1), \dots, \sum_{k=1}^{[nt_l]} X_k(l)\right)'$$

and

$$\mathbf{G}_n(i, j) = n^{-1} \sum_{k=1}^{n-1} \left\{\left[\sum_{u=1}^{k} X_u(i)\right] X_{k+1}(j)\right\}.$$

Assume that $\mathbf{U}_n \to_{\mathscr{L}} \mathbf{W} = (W_1, \dots, W_l)$, where $W_1, \dots, W_l$ are l Brownian motions with respect to increasing σ-fields $\mathscr{G}_t$. Then

$$(\mathbf{U}_n, \mathbf{G}_n) \to_{\mathscr{L}} (\mathbf{W}, \mathbf{G}),$$

where $\mathbf{G}(i, j) = \int_0^1 W_i \, dW_j$.

3. Main results. The distribution of $\mathbf{b}_n$ will be obtained through several stages. First, since time series with different characteristic roots are expected to behave differently, we transform the original time series into several components so that each component has its own characteristic root. More precisely, we express the characteristic polynomial $\phi(z)$ in (1.2) as

$$\phi(z) = (1 - z)^a (1 + z)^b \prod_{k=1}^{l} \left(1 - 2\cos\theta_k z + z^2\right)^{d_k} \psi(z), \tag{3.1}$$

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

$$u_t = \left[\phi(\mathbb{B})(1 - \mathbb{B})^{-a}\right] y_t,$$
$$v_t = \left[\phi(\mathbb{B})(1 + \mathbb{B})^{-b}\right] y_t,$$
$$z_t = \left[\phi(\mathbb{B})(\psi(\mathbb{B}))^{-1}\right] y_t$$

and for $k = 1, \dots, l$,

$$x_t(k) = \left[\phi(\mathbb{B})\left(1 - 2\cos\theta_k \mathbb{B} + \mathbb{B}^2\right)^{-d_k}\right] y_t,$$

where $\mathbb{B}$ is the backshift operator. Then $(1 - \mathbb{B})^a u_t = \varepsilon_t$, $(1 + \mathbb{B})^b v_t = \varepsilon_t$, $(1 - 2\cos\theta_k \mathbb{B} + \mathbb{B}^2)^{d_k} x_t(k) = \varepsilon_t$ and $\psi(\mathbb{B}) z_t = \varepsilon_t$. Define

$$\mathbf{u}_t = (u_t, \dots, u_{t-a+1})',$$
$$\mathbf{v}_t = (v_t, \dots, v_{t-b+1})',$$
$$\mathbf{x}_t(k) = \left(x_t(k), \dots, x_{t-2d_k+1}(k)\right)'$$

and

$$\mathbf{z}_t = (z_t, \dots, z_{t-q+1})'.$$

It is not difficult to see that there is a $p \times p$ matrix Q (see Appendix 1) such that

$$Q\mathbf{y}_t = (\mathbf{u}'_t, \mathbf{v}'_t, \mathbf{x}'_t(1), \dots, \mathbf{x}'_t(l), \mathbf{z}'_t)', \tag{3.2}$$

where $\mathbf{y}_t$ is defined in (1.4). Next, we introduce block diagonal matrices (see Section 3.4)

$$G_n = \begin{pmatrix} J_n & 0 & 0 & & 0 & 0 \\ 0 & K_n & \cdot & & \cdot & \cdot \\ & & L_n(1) & & & \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ & & & & L_n(l) & \\ 0 & 0 & 0 & & 0 & M_n \end{pmatrix}, \tag{3.3}$$

where J_n, K_n, $L_n(1), \dots, L_n(l)$ and M_n are $a \times a$, $b \times b$, $2d_1 \times 2d_1, \dots, 2d_l \times 2d_l$ and $q \times q$ matrices to be chosen later. In Section 3.4, we shall show that if we normalize $Q\Sigma_1^n \mathbf{y}_t\mathbf{y}_t'Q'$ by G_n, then the cross product terms vanish as $n \to \infty$. Specifically,

$$G_n Q \sum_1^n \mathbf{y}_t\mathbf{y}_t' Q' G_n' \sim_p \begin{pmatrix} J_n \sum_1^n \mathbf{u}_t\mathbf{u}_t' J_n' & 0 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & & & M_n \sum_1^n \mathbf{z}_t\mathbf{z}_t' M_n' \end{pmatrix}. \tag{3.4}$$

This indicates that different components are asymptotically uncorrelated. Furthermore, in Section 3.5, we shall also show that (3.4) converges in distribution to a random matrix which is positive definite a.s. This not only implies that Q and G_n are invertible, but also reduces our problem to a componentwise problem. More precisely,

$$(Q'G_n')^{-1}(\mathbf{b}_n - \boldsymbol{\beta}) \sim_p \begin{pmatrix} (J_n')^{-1}\left(\sum_1^{n-1} \mathbf{u}_t\mathbf{u}_t'\right)^{-1} \sum_1^n \mathbf{u}_{t-1}\varepsilon_t \\ \vdots \\ (M_n')^{-1}\left(\sum_1^{n-1} \mathbf{z}_t\mathbf{z}_t'\right)^{-1} \sum_1^n \mathbf{z}_{t-1}\varepsilon_t \end{pmatrix}. \tag{3.5}$$

If we observe that

$$\mathbf{b}_n - \boldsymbol{\beta} = \left(\sum_1^{n-1} \mathbf{y}_t\mathbf{y}_t'\right)^{-1} \sum_1^n \mathbf{y}_{t-1}\varepsilon_t, \tag{3.6}$$

it is clear that each component in (3.5) is a least squares estimate after normalization. From Sections 3.1–3.3, we will consider each component separately. In fact, each component will be shown to be a functional of the corresponding component of the random element X_n in (2.12). Since the limiting process of X_n has independent components, the componentwise limiting distribution of (3.5) characterizes the whole distribution of $\mathbf{b}_n$. (See Appendix 3.)

In the sequel, for the sake of convenience, we will assume $\mathbf{y}_0 = \mathbf{0}$. This in turn implies $\mathbf{u}_0 = \mathbf{0}$, $\mathbf{v}_0 = \mathbf{0}$, $\mathbf{x}_0(1) = \mathbf{0}, \dots, \mathbf{x}_0(k) = \mathbf{0}$ and $\mathbf{z}_0 = \mathbf{0}$. We will also assume $\{\varepsilon_n\}$ is a sequence of martingale differences satisfying (2.10). From time to time, we also denote $X_n \to_{\mathscr{L}} X$ by $X_n(t) \to_{\mathscr{L}} X(t)$ in D to indicate the time variable t of the random elements.

3.1. *Roots equal to* 1. In this section, we consider the model

$$(1 - \mathbb{B})^a u_t = \varepsilon_t, \qquad t = 1, 2, \dots, \tag{3.1.1}$$

with initial condition $\mathbf{u}_0 = \mathbf{0}$. Let

$$u_t(j) = (1 - \mathbb{B})^{a-j} u_t \tag{3.1.2}$$

and

$$U_t = (u_t(a), \dots, u_t(1))'.$$

Then

$$M\mathbf{u}_t = U_t, \tag{3.1.3}$$

where M is an $a \times a$ matrix defined by

$$M = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & -1 & 0 & & \cdot \\ 1 & -2 & 1 & & \cdot \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & (-1)\binom{a-1}{1} & & \cdots & (-1)^{a-1} \end{pmatrix}.$$

From (3.1.2), it is obvious that $u_t(j) = (1 - \mathbb{B}) u_t(j+1)$ for $j = 0, 1, \dots, a-1$. Hence,

$$u_t(j+1) = \sum_{k=1}^{t} u_k(j) \quad \text{for } j = 0, 1, \dots, a-1. \tag{3.1.4}$$

Note that

$$u_t(1) = \sum_{k=1}^{t} u_k(0) = \sum_{k=1}^{t} \varepsilon_k. \tag{3.1.5}$$

The continuous analogues of these recursions are

$$\begin{aligned} F_0(t) &= W(t), \\ F_j(t) &= \int_0^t F_{j-1}(s)\, ds, \end{aligned} \tag{3.1.6}$$

where $W(t)$ is a standard Brownian motion. In order to state our theorem, we introduce the following definitions. Let

$$\begin{aligned} F = (\sigma_{jl}), \qquad \sigma_{jl} &= \int_0^1 F_{j-1}(t) F_{l-1}(t)\, dt \quad \text{for } j, l = 1, \dots, a, \\ \xi &= \left(\int_0^1 F_{a-1}(t)\, dW(t), \dots, \int_0^1 F_0(t)\, dW(t) \right)' \end{aligned} \tag{3.1.7}$$

and

$$(3.1.8)\qquad J_n = N_n^{-1}M, \quad \text{where } N_n = \begin{pmatrix} n^a & 0 & \cdots & 0 \\ 0 & n^{a-1} & & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & n \end{pmatrix}.$$

LEMMA 3.1.1. *F is nonsingular a.s.*

This lemma will be proved in Appendix 2. Now we are ready to state the theorem.

THEOREM 3.1.2.

$$(3.1.9)\qquad J_n \sum_1^{n-1} \mathbf{u}_j \mathbf{u}_j' J_n' \to_{\mathscr{L}} F$$

and

$$(3.1.10)\qquad (J_n')^{-1}\left[\sum_1^{n-1} \mathbf{u}_j \mathbf{u}_j'\right]^{-1} \sum_1^n \mathbf{u}_{j-1}\varepsilon_j \to_{\mathscr{L}} F^{-1}\xi.$$

PROOF. By (3.1.5) and Theorem 2.2, the process

$$(3.1.11)\qquad \frac{1}{\sqrt{n}} u_{[nt]}(1) \to_{\mathscr{L}} W(t) \quad \text{in } D.$$

Applying Theorem 2.3, (3.1.11), (3.1.4) and (3.1.6), we have

$$(3.1.12)\qquad (n)^{1/2-j} u_{[nt]}(j) \to_{\mathscr{L}} F_{j-1}(t) \quad \text{in } D \text{ for } j = 1, 2, \dots, a.$$

This in turn implies

$$(3.1.13)\qquad n^{-(j+l)} \sum_{k=1}^{n-1} u_k(j) u_k(l) \to_{\mathscr{L}} \sigma_{jl}.$$

Now observe that by (3.1.3) and (3.1.8),

$$J_n\left(\sum_1^{n-1} \mathbf{u}_j \mathbf{u}_j'\right) J_n' = N_n^{-1}\left(\sum_1^{n-1} U_j U_j'\right) N_n^{-1}.$$

The (j, l) element of the preceding matrix is $(1/n^{j+l})\sum_{k=1}^{n-1} u_k(j) u_k(l)$. Hence, (3.1.9) is proved by (3.1.13). Now express (3.1.10) as

$$\left[J_n \sum_1^{n-1} \mathbf{u}_j \mathbf{u}_j' J_n'\right]^{-1} J_n \sum_1^n \mathbf{u}_{j-1}\varepsilon_j.$$

Since all quantities involved in $J_n \sum_1^n \mathbf{u}_{j-1}\varepsilon_j$ will be shown to be functionals of $(1/\sqrt{n}) u_{[nt]}(1) = (1/\sqrt{n}) \sum_1^{[nt]} \varepsilon_k$, by Lemma 3.1.1, (3.1.13) and the continuous mapping theorem, (3.1.10) holds if

$$(3.1.14)\qquad J_n \sum_1^n \mathbf{u}_{j-1}\varepsilon_j \to_{\mathscr{L}} \xi.$$

But the $(a-k+1)$st element of $J_n\Sigma_1^n \mathbf{u}_{j-1}\varepsilon_j$ is $n^{-k}\Sigma_{t=1}^n u_{t-1}(k)\varepsilon_t$ which converges in distribution to $\int_0^1 F_{k-1}(t)\,dF_0(t)$ by Theorem 2.4. This completes the proof of (3.1.14). □

COROLLARY 3.1.3. *Assume* $u_n = bu_{n-1} + \varepsilon_n$ *for* $n = 1, 2, \ldots$. *Then under the assumption* $b = 1$,

$$\begin{aligned} n(\hat{b}_n - 1) &\to_{\mathscr{L}} \tfrac{1}{2}(W^2(1) - 1)\Big/\int_0^1 W^2(t)\,dt \\ &=_{\mathscr{L}} \tfrac{1}{2}(T^2 - 1)/\Gamma, \end{aligned} \tag{3.1.15}$$

where W *is a standard Brownian motion*, $T = \Sigma_{k=1}^\infty \sqrt{2}\,\gamma_k Z_k$, $\Gamma = \Sigma_{k=1}^\infty \gamma_k^2 Z_k^2$, $\gamma_k = 2(-1)^k/(2k-1)\pi$ *and* Z_k *are i.i.d.* $N(0,1)$ *random variables.*

PROOF. Apply Itô's formula and Theorem 3.1.2 with $a = 1$. We have that

$$\begin{aligned} n(\hat{b}_n - 1) &\to_{\mathscr{L}} \int_0^1 W(t)\,dW(t)\Big/\int_0^1 W^2(t)\,dt \\ &= \tfrac{1}{2}(W^2(1) - 1)\Big/\int_0^1 W^2(t)\,dt. \end{aligned}$$

Now expand the Brownian motion by the reproducing kernel method [Kac (1980)]:

$$W(t) = \sum_{n=0}^{\infty} \frac{2\sqrt{2}}{(2n+1)\pi} \sin\left(n + \frac{1}{2}\right)\pi t Z_n, \tag{3.1.16}$$

where Z_n are i.i.d. $N(0,1)$ random variables. Then

$$\int_0^1 W^2(t)\,dt = \sum_{n=0}^{\infty} (2/(2n+1)\pi)^2 Z_n^2 = \Gamma$$

and

$$W(1) = \sum_{n=0}^{\infty} \frac{2\sqrt{2}}{(2n+1)\pi}(-1)^n Z_n = T.$$

Hence,

$$\tfrac{1}{2}(W^2(1) - 1)\Big/\int_0^1 W^2(t)\,dt = \tfrac{1}{2}(T^2 - 1)/\Gamma.$$

This completes the proof. □

REMARK 1. The second representation in (3.1.15) was proved by Dickey and Fuller (1979) by the quadratic form approach. Since their method is different from the functional central limit theorem approach, Solo (1984) asked what is the relation between this representation and the other representation in (3.1.15), its stochastic integral counterpart. The proof of Corollary 3.1.3 gives an answer to his question. Note that Solo (1984) also mentioned the possibility of using a series expansion of the Brownian motion to establish this relation. However, it is not obvious that such a direct relation can be derived by using the example (the Walsh series expansion) cited there.

REMARK 2. The series expansion also provides some light on the efficient calculation of the distribution. Since there are infinitely many possible ways to express a Brownian motion [Loève (1978)] as an infinite series, faster convergent series may lead to numerical improvement over slower ones. Further research along this line may be interesting.

3.2. *Roots equal to* -1. In this section, we consider the model

$$(1+\mathbb{B})^b v_t = \varepsilon_t, \qquad t = 1,2,\ldots, \tag{3.2.1}$$

with initial condition $\mathbf{v}_0 = \mathbf{0}$. The result of this section is essentially similar to that of Section 3.1 except that every quantity involved here is shown to be a functional of $\Sigma_{k=1}^{[nt]}(-1)^k\varepsilon_k$ instead of being a functional of $\Sigma_{k=1}^{[nt]}\varepsilon_k$. As a consequence of this, the distribution derived here is a mirror image of the case where roots are equal to 1 (see the remark following Theorem 3.2.1). More precisely, let us define

$$\tilde{M} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & & \cdot \\ 1 & 2 & 1 & & \cdot \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \binom{b-1}{1} & \cdot & \cdots & 1 \end{pmatrix}, \tag{3.2.2}$$

$$K_n = \tilde{N}_n^{-1}\tilde{M}, \quad \text{where } \tilde{N}_n = \begin{pmatrix} n^b & 0 & \cdots & 0 \\ 0 & n^{b-1} & & \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & n \end{pmatrix}.$$

We also define F_j as in (3.1.6) and $\tilde{F} = (\sigma_{jl})$, for $j, l = 1,\ldots, b$, as in (3.1.7) and

$$\boldsymbol{\eta} = -\left(\int_0^1 F_{b-1}(t)\,dW(t),\ldots,\int_0^1 F_0(t)\,dW(t)\right)'. \tag{3.2.3}$$

THEOREM 3.2.1.

$$K_n\sum_{k=1}^{n-1}\mathbf{v}_k\mathbf{v}_k'K_n' \to_{\mathscr{L}} \tilde{F} \tag{3.2.4}$$

and

$$(K_n')^{-1}\left(\sum_1^{n-1}\mathbf{v}_k\mathbf{v}_k'\right)^{-1}\sum_1^n\mathbf{v}_{k-1}\varepsilon_k \to_{\mathscr{L}} \tilde{F}^{-1}\boldsymbol{\eta}. \tag{3.2.5}$$

PROOF. The proof follows essentially the same line as the proof of Theorem 3.1.2. The only difference is that any typical element of $K_n\Sigma_{k=1}^{n-1}\mathbf{v}_k\mathbf{v}_k'K_n'$ or

$K_n \Sigma_1^n \mathbf{v}_{k-1}\varepsilon_k$ is a functional of $\Sigma_1^{[nt]}(-1)^k\varepsilon_k$ instead of $\Sigma_1^{[nt]}\varepsilon_k$. To see this, let

$$\tilde{M}\mathbf{v}_t = V_t = (v_t(b),\dots,v_t(1))' \quad \text{and} \quad v_t(0) = \varepsilon_t.$$

Then

$$v_t(j) = (1+\mathbb{B})v_t(j+1) \quad \text{for } j = 0,1,\dots,b-1$$

or

$$v_t(j+1) = \sum_{k=1}^{t} (-1)^{t-k} v_k(j) \quad \text{for } j = 0,\dots,b-1.$$

Consequently,

$$(-1)^t v_t(j+1) = \sum_{k=1}^{t} (-1)^k v_k(j) \quad \text{for } j = 0,\dots,b-1.$$

Note that $(-1)^t v_t(1) = \Sigma_{k=1}^t(-1)^k\varepsilon_k$. This implies that $(-1)^t v_t(j)$ is a functional of $\Sigma_{k=1}^t(-1)^k\varepsilon_k$ for $j = 1,2,\dots,b$. Now observe that the (k,j)th element of $\Sigma_{t=1}^{n-1}\mathbf{v}_t\mathbf{v}_t'$ is equal to $\Sigma_{t=1}^{n-1}[(-1)^t v_t(k)][(-1)^t v_t(j)]$ for $k, j \geq 1$ and the kth element of $\tilde{M}\Sigma_1^n \mathbf{v}_{t-1}\varepsilon_t$ is equal to

$$\sum_{t=1}^{n-1} v_{t-1}(k)\varepsilon_t = \sum_{t=1}^{n-1} (-1)^t v_t(k)\big((-1)^t v_t(1) - (-1)^{t-1} v_{t-1}(1)\big).$$

This completes the proof. □

REMARK. From (3.1.7) and (3.2.3), under the assumption that $a = b$, we have that $(\tilde{F}, \boldsymbol{\eta}) =_{\mathscr{L}} (F, -\boldsymbol{\xi})$. Hence, $(\tilde{F})^{-1}\boldsymbol{\eta} =_{\mathscr{L}} -F^{-1}\boldsymbol{\xi}$. This is observed by Fuller (1976) for the case $b = 1$. The importance of this observation is that we only have to tabulate one distribution instead of two. Theorem 3.1.2 shows this observation still holds for the multiple root case.

3.3. *Roots equal to $e^{i\theta}$ or $e^{-i\theta}$.* In this section, we consider the model

$$(1 - 2\cos\theta\mathbb{B} + \mathbb{B}^2)^d x_t = \varepsilon_t, \qquad t = 1,2,\dots,$$

with initial condition $\mathbf{x}_0 = \mathbf{0}$, where here and in the sequel $\mathbf{x}_t = (x_t,\dots,x_{t-2d+1})'$. We also assume $\theta \in (0,\pi)$. Let

$$y_t(j) = (1 - 2\cos\theta\mathbb{B} + \mathbb{B}^2)^{d-j} x_t \quad \text{for } j = 0,1,\dots,d \tag{3.3.1}$$

and

$$Y_t = (y_t(1), y_{t-1}(1),\dots, y_t(d), y_{t-1}(d))'.$$

It is not difficult to see that there exists a $2d \times 2d$ matrix C (see Appendix 1) such that

$$C\mathbf{x}_t = Y_t. \tag{3.3.2}$$

By (3.3.1),

$$(1 - 2\cos\theta\mathbb{B} + \mathbb{B}^2)y_t(j+1) = y_t(j) \qquad \text{for } j = 0,1,\dots,d-1.$$

Note that $\mathbf{x}_0 = \mathbf{0}$ implies $\mathbf{Y}_0 = \mathbf{0}$. This in turn implies

$$y_t(j+1) = \frac{1}{\sin\theta} \sum_{k=1}^{t} \sin(t-k+1)\theta y_k(j) \tag{3.3.3}$$

for $j = 0, 1, \dots, d-1$.

Now, for $j = 0, \dots, d$, let

$$S_t(j) = \sum_{k=1}^{t} \cos k\theta y_k(j), \qquad T_t(j) = \sum_{k=1}^{t} \sin k\theta y_k(j),$$

$$S_t = S_t(0) \quad \text{and} \quad T_t = T_t(0).$$

LEMMA 3.3.1. *For* $j = 1, \dots, d$,

$$\sin\theta y_t(j) = S_t(j-1)\sin(t+1)\theta - T_t(j-1)\cos(t+1)\theta; \tag{3.3.4}$$

$$\begin{aligned} 2\sin\theta S_t(j) = \sum_{k=1}^{t} \{&\sin(2k+1)\theta S_k(j-1) + \sin\theta S_k(j-1) \\ &-\cos(2k+1)\theta T_k(j-1) - \cos\theta T_k(j-1)\}; \end{aligned} \tag{3.3.5}$$

$$\begin{aligned} 2\sin\theta T_t(j) = \sum_{k=1}^{t} \{&\cos\theta S_k(j-1) - \cos(2k+1)\theta S_k(j-1) \\ &+\sin\theta T_k(j-1) - \sin(2k+1)\theta T_k(j-1)\}. \end{aligned} \tag{3.3.6}$$

This lemma implies that $y_t(j)$, $S_t(j)$ and $T_t(j)$ can all be expressed as functionals of S_t and T_t. Thus, to show a random variable can be expressed as a functional of S_t and T_t, it is sufficient to show that it can be expressed as a functional of $S_t(j)$, $T_t(j)$ and $y_t(j)$.

LEMMA 3.3.2. *For* $j = 0, \dots, d$ *and* $k = 0, \dots, d$,

$$\begin{aligned} &2\sin^2\theta \sum_{t=1}^{n} y_t(k) y_t(j) \\ &= \sum_{t=1}^{n} (S_t(k-1)S_t(j-1) + T_t(k-1)T_t(j-1)) \\ &\quad - \sum_{t=1}^{n} (S_t(k-1)T_t(j-1) + T_t(k-1)S_t(j-1))\sin(2t+2)\theta \\ &\quad - \sum_{t=1}^{n} (S_t(k-1)S_t(j-1) - T_t(k-1)T_t(j-1))\cos(2t+2)\theta, \end{aligned}$$

$$\begin{aligned}
2\sin^2\theta \sum_{t=1}^{n} y_{t-1}(k)y_t(j)
&= \cos\theta \sum_{t=1}^{n} (T_{t-1}(k-1)T_t(j-1) + S_{t-1}(k-1)S_t(j-1)) \\
&\quad -\sin\theta \sum_{t=1}^{n} (T_{t-1}(k-1)S_t(j-1) - S_{t-1}(k-1)T_t(j-1)) \\
&\quad - \sum_{t=1}^{n} (S_{t-1}(k-1)S_t(j-1) - T_{t-1}(k-1)T_t(j-1))\cos(2t+1)\theta \\
&\quad - \sum_{t=1}^{n} (S_{t-1}(k-1)T_t(j-1) + S_t(j-1)T_{t-1}(k-1))\sin(2t+1)\theta,
\end{aligned}$$

$$\sin\theta \sum_{t=1}^{n} y_t(j)\varepsilon_{t+1} = \sum_{t=1}^{n} S_t(j-1)(T_{t+1} - T_t) - \sum_{t=1}^{n} T_t(j-1)(S_{t+1} - S_t),$$

$$\begin{aligned}
\sin\theta \sum_{t=1}^{n} y_{t-1}(j)\varepsilon_{t+1}
&= \cos\theta \sum_{t=1}^{n} (S_{t-1}(j-1)(T_{t+1} - T_t) - T_{t-1}(j-1)(S_{t+1} - S_t)) \\
&\quad -\sin\theta \sum_{t=1}^{n} (S_{t-1}(j-1)(S_{t+1} - S_t) + T_{t-1}(j-1)(T_{t+1} - T_t)).
\end{aligned}$$

The proofs of Lemmas 3.3.1 and 3.3.2 are straightforward applications of trigonometric identities. Details are omitted. Lemma 3.3.2 provides recursive forms for those terms involved in the least squares estimates. Moreover, it also leads to the definitions of the following continuous analogues. Let $W_1(t)$ and $W_2(t)$ be two independent Brownian motions. Define

$$f_0(t) = W_1(t), \qquad g_0(t) = W_2(t),$$

$$f_j(t) = \frac{1}{2\sin\theta}\left(\sin\theta \int_0^t f_{j-1}(s)\,ds - \cos\theta \int_0^t g_{j-1}(s)\,ds\right),$$

$$g_j(t) = \frac{1}{2\sin\theta}\left(\cos\theta \int_0^t f_{j-1}(s)\,ds + \sin\theta \int_0^t g_{j-1}(s)\,ds\right),$$

$$\zeta_{2j-1} = \frac{1}{2\sin\theta}\left(\int_0^1 f_{j-1}(s)\,dW_2(s) - \int_0^1 g_{j-1}(s)\,dW_1(s)\right),$$

$$\begin{aligned}
\zeta_{2j} = \frac{1}{2\sin\theta}\Bigg\{&\cos\theta\left(\int_0^1 f_{j-1}(s)\,dW_2(s) - \int_0^1 g_{j-1}(s)\,dW_1(s)\right) \\
&-\sin\theta\left(\int_0^1 f_{j-1}(s)\,dW_1(s) + \int_0^1 g_{j-1}(s)\,dW_2(s)\right)\Bigg\},
\end{aligned}$$

$$\sigma_{2k-1,2j-1} = \sigma_{2k,2j}$$
$$= \frac{1}{4\sin^2\theta}\left(\int_0^1 f_{k-1}(s) f_{j-1}(s)\, ds + \int_0^1 g_{k-1}(s) g_{j-1}(s)\, ds\right),$$
$$\sigma_{2k-1,2j} = \sigma_{2j,2k-1}$$
$$= \frac{1}{4\sin^2\theta}\left\{\cos\theta\left(\int_0^1 f_{k-1}(s) f_{j-1}(s)\, ds + \int_0^1 g_{k-1}(s) g_{j-1}(s)\, ds\right)\right.$$
$$\left. -\sin\theta\left(\int_0^1 f_{j-1}(s) g_{k-1}(s)\, ds - \int_0^1 g_{j-1}(s) f_{k-1}(s)\, ds\right)\right\},$$

$\zeta = (\zeta_1, \dots, \zeta_{2d})'$ and $H = (\sigma_{ij})$, a $2d \times 2d$ random matrix.

LEMMA 3.3.3. *H is nonsingular a.s.*

This lemma will be proved in Appendix 2. Now let $L_n = N_n^{-1}C$, where C is defined in (3.3.2) and N_n is a $2d \times 2d$ diagonal matrix satisfying

$$N_n = \begin{pmatrix} nI & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & n^d I \end{pmatrix} \quad \text{and} \quad I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

THEOREM 3.3.4.

$$L_n\left(\sum_1^{n-1} \mathbf{x}_t \mathbf{x}_t'\right) L_n' \to_{\mathscr{L}} H \tag{3.3.7}$$

and

$$(L_n')^{-1}\left(\sum_1^{n-1} \mathbf{x}_t \mathbf{x}_t'\right)^{-1} \sum_1^n \mathbf{x}_{t-1}\varepsilon_t \to_{\mathscr{L}} H^{-1}\zeta, \tag{3.3.8}$$

where $\{\varepsilon_n\}$ satisfies (2.10).

Before proving this theorem, we first derive some auxiliary lemmas.

LEMMA 3.3.5. *For $k = 0, \dots, d-1$,*

$$ES_n^2(k) = O(n^{2k+1}) \quad \text{and} \quad ET_n^2(k) = O(n^{2k+1}). \tag{3.3.9}$$

PROOF. We shall prove this by induction on k. For $k = 0$,

$$ES_n^2(0) = \sum_{j=1}^n (\cos(j\theta))^2 \le n.$$

Similarly, $ET_n^2(0) \le n$. Now assume (3.3.9) is true for $k = l$. By (3.3.5) and the

Cauchy–Schwarz inequality,

$$\begin{aligned} ES_n^2(l+1) &\le \frac{1}{4\sin^2\theta}4\left\{4n\sum_{j=1}^{n}E\big(S_j^2(l)\big)+4n\sum_{j=1}^{n}E\big(T_j^2(l)\big)\right\} \\ &= O\left(n\sum_{j=1}^{n}j^{2l+1}\right) \\ &= O(n^{2(l+1)+1}). \end{aligned}$$

Similarly, $ET_n^2(l+1)=O(n^{2l+3})$. This completes the proof. □

LEMMA 3.3.6. *For* $k, l = 0,\dots,d-1$,

(3.3.10) $$\sup_{1\le j\le n}\left|\sum_{t=1}^{j}e^{it\theta}S_t(k)\right| = o_p(n^{k+3/2}),$$

(3.3.11) $$\sup_{1\le j\le n}\left|\sum_{t=1}^{j}e^{it\theta}S_t^2(k)\right| = o_p(n^{2k+2}),$$

(3.3.12) $$\sup_{1\le j\le n}\left|\sum_{t=1}^{j}e^{it\theta}S_t(l)T_t(k)\right| = o_p(n^{l+k+2}).$$

PROOF. We are going to prove Lemma 3.3.6 by Theorem 2.1. For (3.3.10),

$$E|S_n(k)| \le \big(E|S_n^2(k)|\big)^{1/2} = O(n^{k+1/2}),$$

by (3.3.9). Now assume $n\ge m$. By the definition of $S_n(k)$, (3.3.4) and the Cauchy–Schwarz inequality,

$$|S_n(k)-S_m(k)|$$

(3.3.13) $$\begin{aligned} &\le \left\{\sum_{t=m+1}^{n}\cos^2(t\theta)\right\}^{1/2}\left\{\sum_{t=m+1}^{n}y_t^2(k)\right\}^{1/2} \\ &\le (n-m)^{1/2}\left\{\frac{4}{\sin^2\theta}\sum_{t=m+1}^{n}\big(S_t^2(k-1)+T_t^2(k-1)\big)\right\}^{1/2}. \end{aligned}$$

By Lemma 3.3.5,

(3.3.14) $$E\sum_{t=m+1}^{n}\{S_t^2(k-1)+T_t^2(k-1)\} = O\left(\sum_{t=1}^{n}t^{2k-1}\right) = O(n^{2k}).$$

Since $2(k+\frac{1}{2}) > 2k+0$, (3.3.10) is proved. For (3.3.11),

$$E|S_n^2(k)| = O(n^{2k+1}),$$

by (3.3.9). Now assume $n\ge m$. We have

$$|S_n^2(k)-S_m^2(k)| = |S_n(k)+S_m(k)|\,|S_n(k)-S_m(k)|.$$

By (3.3.9), (3.3.13) and (3.3.14),

$$
\begin{aligned}
&E|S_n(k) + S_m(k)|^2 \le 2\big(ES_n^2(k) + ES_m^2(k)\big) = O(n^{2k+1}),\\
&E|S_n(k) - S_m(k)|^2 \le O\big((n-m)n^{2k}\big).
\end{aligned} \tag{3.3.15}
$$

Choose $\alpha = \gamma = 2k+1$ and $\delta = 2k$ in Theorem 2.1. Since $2\alpha = 4k+2 > 4k+1 = \gamma + \delta$, (3.3.11) is proved. For (3.3.12), by (3.3.9),

$$E|S_n(l)T_n(k)| \le \{ES_n^2(l)\}^{1/2}\{ET_n^2(k)\}^{1/2} = O(n^{l+k+1}).$$

Moreover, for $n \ge m$,

$$|S_n(l)T_n(k) - S_m(l)T_m(k)| \le |S_n(l)|\,|T_n(k) - T_m(k)| + |T_n(k)|\,|S_n(l) - S_m(l)|.$$

Now,

$$ES_n^2(l) = O(n^{2l+1}) \quad \text{and} \quad ET_n^2(k) = O(n^{2k+1}).$$

As in (3.3.15),

$$E|T_n(k) - T_m(k)|^2 = O\big((n-m)n^{2k}\big)$$

and

$$E|S_n(l) - S_m(l)|^2 = O\big((n-m)n^{2l}\big).$$

Using the remark following Theorem 2.1, (3.3.12) is proved. □

REMARK 1. Analogous results hold in Lemma 3.3.6 when $S_t(k)$ is replaced by $S_{t-1}(k)$ or $T_t(k)$.

LEMMA 3.3.7. *For $d \ge j \ge 0$,*

$$\sqrt{2}\, n^{-j-1/2}\big(S_{[nt]}(j), T_{[ns]}(j)\big) \to_{\mathscr{L}} \big(f_j(t), g_j(s)\big) \quad \text{in } D \times D. \tag{3.3.16}$$

PROOF. We prove (3.3.16) by induction. By (3.3.1) and the definitions of $S_t(0)$ and $T_t(0)$,

$$S_t(0) = \sum_{k=1}^{t} \cos k\theta\varepsilon_k, \qquad T_t(0) = \sum_{k=1}^{t} \sin k\theta\varepsilon_k.$$

In view of Theorem 2.2, (3.3.16) holds for $j = 0$. Now suppose (3.3.17) holds for $j-1$. By (3.3.5),

$$
\begin{aligned}
\sqrt{2}\, n^{-j-1/2} S_{[nt]}(j) &= (\sqrt{2}\, n^{-j-1/2}) \frac{1}{2\sin\theta} \sum_{k=1}^{[nt]} \{\sin\theta S_k(j-1) - \cos\theta T_k(j-1)\}\\
&\quad + (\sqrt{2}\, n^{-j-1/2}) \frac{1}{2\sin\theta} \sum_{k=1}^{[nt]} \{\sin(2k+1)\theta S_k(j-1)\\
&\qquad\qquad - \cos(2k+1)\theta T_k(j-1)\}\\
&\equiv I_1(t) + I_2(t).
\end{aligned}
$$

By Lemma 3.3.6, $\|I_2\|_\infty = o_p(1)$. Similarly,

$$\sqrt{2}\, n^{-j-1/2} T_{[ns]}(j) = J_1(s) + J_2(s),$$

where $\|J_2\|_\infty = o_p(1)$ and

$$J_1(s) = (\sqrt{2}\, n^{-j-1/2}) \frac{1}{2\sin\theta} \sum_{k=1}^{[ns]} \{\cos\theta S_k(j-1) + \sin\theta T_k(j-1)\}.$$

Hence, to prove (3.3.16), it suffices to show

$$(I_1, J_1) \to_{\mathscr{L}} (f_j, g_j). \tag{3.3.17}$$

Now consider the continuous functional H from $D \times D$ to $D \times D$ defined by

$$H(f, g)(t, s) = \frac{1}{2\sin\theta}\left(\int_0^t [\sin\theta f(u) - \cos\theta g(u)]\, du,\right.$$
$$\left.\int_0^s [\cos\theta f(u) + \sin\theta g(u)]\, du\right).$$

In view of the induction hypothesis and the continuous mapping theorem,

$$(I_1, J_1) \to_{\mathscr{L}} H(f_{j-1}, g_{j-1}) = (f_j, g_j). \qquad \square$$

REMARK 2. All of the limiting results stated in Lemma 3.3.7 are jointly convergent. This can be shown by applying Theorem 2.3 repeatedly. Also see Appendix 3.

PROOF OF THEOREM 3.3.4. By Lemmas 3.3.2 and 3.3.6,

$$n^{-k-j} \sum_{t=1}^{n} y_t(k) y_t(j)$$
$$= (2\sin^2\theta)^{-1} n^{-k-j} \sum_{t=1}^{n} (S_t(k-1)S_t(j-1) + T_t(k-1)T_t(j-1)) + o_p(1).$$

By Lemma 3.3.7 and the continuous mapping theorem,

$$n^{-k-j} \sum_{t=1}^{n} y_t(k) y_t(j) \to_{\mathscr{L}} \sigma_{2k,2j}. \tag{3.3.18}$$

Similarly,

$$n^{-k-j} \sum_{t=1}^{n} y_{t-1}(k) y_t(j) \to_{\mathscr{L}} \sigma_{2k-1,2j}. \tag{3.3.19}$$

Since a typical element in $L_n(\Sigma_1^n \mathbf{x}_t \mathbf{x}_t')L_n'$ is either of the form (3.3.18) or (3.3.19), which is a functional of (S_t, T_t), the joint law follows. Consequently,

$$L_n\left(\sum_1^n \mathbf{x}_t \mathbf{x}_t'\right) L_n' \to_{\mathscr{L}} H.$$

For (3.3.8), as argued in Theorem 3.1.2, we only have to show that

$$L_n \sum_1^n \mathbf{x}_{t-1}\varepsilon_t = N_n^{-1} \sum_1^n Y_{t-1}\varepsilon_t \to_{\mathscr{L}} \zeta. \tag{3.3.20}$$

From Lemmas 3.3.2 and 3.3.7 and Theorem 2.4,

$$n^{-j} \sum_{t=1}^{n} y_t(j)\varepsilon_{t+1} \to_{\mathscr{L}} \zeta_{2j-1}.$$

Similarly,

$$n^{-j}\sum_{t=1}^{n} y_{t-1}(j)\varepsilon_{t+1} \to_{\mathscr{L}} \zeta_{2j}.$$

These are typical forms of $N_n^{-1}\Sigma_1^n Y_{t-1}\varepsilon_t$. Hence, (3.3.20) is shown and the proof of Theorem 3.3.4 is complete. □

Corollary 3.3.8. *Assume* $x_t = \beta_1 x_{t-1} + \beta_2 x_{t-2} + \varepsilon_t$, *for* $t = 1, 2, \dots$. *Let* $\hat{\beta}_{n,2}$ *be the least squares estimate of* β_2 *at stage* n. *If the characteristic polynomial* $1 - \beta_1 z - \beta_2 z^2$ *has roots* $e^{i\theta}$ *and* $e^{-i\theta}$ *with* $\theta \in (0, \pi)$, *then*

$$(3.3.21)\quad \begin{aligned} n(\hat{\beta}_{n,2} + 1) &\to_{\mathscr{L}} (2 - W_1^2(1) - W_2^2(1)) \Big/ \int_0^1 (W_1^2(t) + W_2^2(t))\,dt \\ &=_{\mathscr{L}} (2 - T - \tilde{T})/(\Gamma + \tilde{\Gamma}), \end{aligned}$$

where W_1, W_2 *are two independent standard Brownian motions,* T, Γ *are defined as in* (3.1.15), $\tilde{T} = \Sigma_{k=1}^{\infty}\sqrt{2}\,\gamma_k \tilde{Z}_k$, $\tilde{\Gamma} = \Sigma_{k=1}^{\infty}\gamma_k^2 \tilde{Z}_k^2$, *and* $\tilde{Z}_k$ *are i.i.d.* $N(0,1)$ *random variables which are independent of* $\{Z_k\}$.

Proof. Under the assumption, $\beta_2 = -1$. By Theorem 3.3.4,

$$(3.3.22)\quad n(\hat{\beta}_{n,2} + 1) \to_{\mathscr{L}} (-\zeta_1\sigma_{12} + \zeta_2\sigma_{11})/(\sigma_{11}\sigma_{22} - \sigma_{12}\sigma_{21}).$$

By the definition of σ_{ij},

$$\begin{aligned} \sigma_{2j,2j-1} = \sigma_{2j-1,2j} &= \frac{\cos\theta}{4\sin^2\theta}\left(\int_0^1 f_{j-1}^2(s)\,ds + \int_0^1 g_{j-1}^2(s)\,ds\right) \\ &= \cos\theta\,\sigma_{2j,2j} = \cos\theta\,\sigma_{2j-1,2j-1}. \end{aligned}$$

Substituting this identity into (3.3.22), we obtain

$$\begin{aligned} &(-\zeta_1\sigma_{12} + \zeta_2\sigma_{11})/(\sigma_{11}\sigma_{22} - \sigma_{12}\sigma_{21}) \\ &\quad= \sigma_{11}(\zeta_2 - \zeta_1\cos\theta)/\sigma_{11}^2(1 - \cos^2\theta) \\ &\quad= (\zeta_2 - \zeta_1\cos\theta)/\sigma_{11}\sin^2\theta \\ &\quad= -2\left(\int_0^1 W_1\,dW_1 + \int_0^1 W_2\,dW_2\right)\Big/\int_0^1 (W_1^2 + W_2^2)\,ds \\ &\quad= (2 - W_1^2(1) - W_2^2(1))\Big/\int_0^1 (W_1^2 + W_2^2)\,ds, \quad \text{by Itô's formula.} \end{aligned}$$

Using the expansions

$$W_1(t) = \sum_{n=0}^{\infty}\frac{2\sqrt{2}}{(2n+1)\pi}\sin\left(n + \frac{1}{2}\right)\pi t Z_n,$$

$$W_2(t) = \sum_{n=0}^{\infty}\frac{2\sqrt{2}}{(2n+1)\pi}\sin\left(n + \frac{1}{2}\right)\pi t \tilde{Z}_n,$$

we obtain (3.3.21) by an argument similar to that in Corollary 3.1.3. □

REMARK 3. The limiting distribution of $n(\hat{\beta}_{n,2} + 1)$ is independent of θ. This is observed by Ahtola and Tiao (1987) who, using the quadratic form approach, represented the limiting distribution as $U^{-1}\{\frac{1}{2} - V_1^2 - V_2^2\}$, where

$$U = \tfrac{1}{4}(V_1^2 + V_2^2) + (V_3V_2 - V_1V_4) + \tfrac{1}{4}(V_5 + V_6),$$
$$V_1 = \sum_j \delta_j N_j, \qquad V_2 = \sum_j \delta_j W_j,$$
$$V_3 = \sum_j \delta_j^2 N_j, \qquad V_4 = \sum_j \delta_j^2 W_j,$$
$$V_5 = \sum_j \delta_j^2 N_j^2, \qquad V_6 = \sum_j \delta_j^2 W_j^2,$$
$$\delta_j = (-1)/(2j - 1)\pi,$$

N_j, W_j are i.i.d. $N(0,1)$ random variables and $j = 0, \pm 1, \pm 2, \dots$. Apparently, (3.3.21) is "neater" than Ahtola and Tiao's representation and provides a simpler expansion for tabulating the distribution.

REMARK 4. Note that the series representation of (3.3.21) is the negative of the series representation of the limiting distribution of $n(\hat{\alpha}_2 - 1)$ given in Dickey, Hasza and Fuller (1984). These two distributions arise from different situations, and it seems difficult to reveal this relation through Ahtola and Tiao's result.

3.4. *Cross product terms.* Let J_n and K_n be matrices defined as in (3.1.8) and (3.2.2), respectively. Let $M_n = n^{-1/2}I_q$, where I_q is the $q \times q$ identity matrix. Moreover, for each $\mathbf{x}_t(j)$ in (3.2), define $L_n(j)$ similar to the L_n defined in Theorem 3.3.4. Hence, G_n in (3.3) is well defined now. In this section, we shall show that (3.4) holds. First, let us consider the products between unstable components and then products between unstable and asymptotically stationary components.

THEOREM 3.4.1.

(i) $$J_n\sum_1^n \mathbf{u}_t\mathbf{v}_t'K_n' \to_p \mathbf{0};$$

(ii) $$J_n\sum_1^n \mathbf{u}_t\mathbf{x}_t'(j)L_n'(j) \to_p \mathbf{0} \quad \textit{for } 1 \le j \le l;$$

(iii) $$K_n\sum_1^n \mathbf{v}_t\mathbf{x}_t'(j)L_n'(j) \to_p \mathbf{0} \quad \textit{for } 1 \le j \le l;$$

(iv) $$L_n(h)\sum_1^n \mathbf{x}_t(h)\mathbf{x}_t'(j)L_n'(j) \to_p \mathbf{0} \quad \textit{for } 1 \le h \ne j \le l.$$

PROOF. We are going to show that each element in (i)–(iv) can be expressed as a sum of the forms $n^{-j}\sum_1^n \sin(t\theta)A_tB_t$ and $n^{-j}\sum_1^n \cos(t\theta)A_tB_t$ for some positive number j, random variables A_t, B_t and some θ such that $e^{i\theta} \neq 1$. By means of Theorem 2.1, we can then derive the desired results. To verify the assumptions of Theorem 2.1, similar arguments used in proving (3.3.12) of Lemma 3.3.6 can be applied and details are omitted. Typical elements of (i)–(iv) are

$$n^{-j-h} \sum_{t=1}^{n} u_t(j)v_t(h) = n^{-j-h} \sum_{t=1}^{n} \cos(t\pi) u_t(j)\big((-1)^t v_t(h)\big),$$

$$n^{-m-h} \sum_{t=1}^{n} u_t(h) y_t(j,m) = \big(n^{m+h} \sin\theta_j\big)^{-1} \sum_{t=1}^{n} \{\sin(t+1)\theta_j u_t(h) S_t(j, m-1) - \cos(t+1)\theta_j u_t(h) T_t(j, m-1)\},$$

$$n^{-m-h} \sum_{t=1}^{n} v_t(h) y_t(j,m) = \big(n^{m+h} \sin\theta_j\big)^{-1} \sum_{t=1}^{n} \Big\{\sin(t+1)\theta_j \cos t\pi S_t(j, m-1)\big((-1)^t v_t(h)\big) - \cos(t+1)\theta_j \cos t\pi T_t(j, m-1)\big((-1)^t v_t(h)\big)\Big\},$$

$$n^{-s-m} \sum_{t=1}^{n} y_t(h,s) y_t(j,m) = \big(n^{s+m} \sin\theta_j \sin\theta_h\big)^{-1} \sum_{t=1}^{n} \big\{\big(S_t(h, s-1)\sin(t+1)\theta_h - T_t(h, s-1)\cos(t+1)\theta_h\big) \times \big(S_t(j, m-1)\sin(t+1)\theta_j - T_t(j, m-1)\cos(t+1)\theta_j\big)\big\},$$

where θ_j, $y_t(j,m)$, $S_t(j,m)$ and $T_t(j,m)$ are defined in Section 3.3 by replacing x_t in (3.3.1) with $x_t(j)$. Using the trigonometric identities, we can reexpress these forms as the sums of $n^{-j}\sum_1^n \sin(t\theta) A_t B_t$ and $n^{-j}\sum_1^n \cos(t\theta) A_t B_t$. Note that $\theta_j \neq \theta_h$ and $\theta_j, \theta_h \in (0, \pi)$ imply that $e^{i(\theta_j \pm \pi)} \neq 1$ and $e^{i(\theta_j \pm \theta_h)} \neq 1$. This completes our proof. □

THEOREM 3.4.2.

(i) $$J_n \sum_1^n \mathbf{u}_t \mathbf{z}_t' M_n' \to_p \mathbf{0};$$

(ii) $$K_n \sum_1^n \mathbf{v}_t \mathbf{z}_t' M_n' \to_p \mathbf{0};$$

(iii) $$L_n(j) \sum_1^n \mathbf{x}_t(j) \mathbf{z}_t' M_n' \to_p \mathbf{0} \quad \text{for } 1 \leq j \leq l,$$

where $\mathbf{z}_t$ *was defined earlier* [*after* (3.1)].

Before we prove Theorem 3.4.2, we need a lemma.

LEMMA 3.4.3. *Assume* $\{\mathbf{g}_t\}$ *and* $\{\mathbf{h}_t\}$ *are two sequences of random vectors in* R^s *such that* $\mathbf{g}_t$ *and* $\mathbf{h}_t$ *are* $\mathscr{F}_t$*-measurable. Suppose there exists an* $s \times s$ *constant matrix* M *such that* $\mathbf{g}_t = M\mathbf{g}_{t-1} + \mathbf{h}_t$, *where* $\mathbf{g}_0 = \mathbf{h}_0 = \mathbf{0}$. *If*

$$(3.4.1)\quad E\sum_{t=1}^{n}\|\mathbf{g}_t\|^2 = O(n^\alpha) \quad \text{and} \quad E\sum_{t=1}^{n}\|\mathbf{h}_t\|^2 = o(n^\alpha) \quad \text{for some } \alpha > 0,$$

then for any fixed integer j,

$$(3.4.2)\qquad E\left\|\sum_{t=1}^{n}\mathbf{g}_t\varepsilon_{t+j}\right\| = o(n^{(\alpha+1)/2})$$

and

$$(3.4.3)\qquad E\left\|\sum_{t=1}^{n}\mathbf{g}_t\mathbf{z}_t'\right\| = o(n^{(\alpha+1)/2}),$$

where $\{\varepsilon_t\}$ *satisfies* (2.10) *and* $\mathbf{z}_t$ *is defined in* (3.2).

PROOF. Assume $j \ge 1$. Then $\sum_{t=1}^{n}\mathbf{g}_t\varepsilon_{t+j}$ is a martingale. Hence,

$$(3.4.4)\qquad E\left\|\sum_{t=1}^{n}\mathbf{g}_t\varepsilon_{t+j}\right\|^2 = E\sum_{t=1}^{n}\|\mathbf{g}_t\|^2 = O(n^\alpha) = o(n^{\alpha+1}).$$

Assume $j = -k \le 0$. Since

$$(3.4.5)\qquad \mathbf{g}_t = M^{k+1}\mathbf{g}_{t-(k+1)} + \sum_{m=0}^{k}M^m\mathbf{h}_{t-m},$$

$$(3.4.6)\qquad \sum_{t=1}^{n}\mathbf{g}_t\varepsilon_{t+j} = M^{k+1}\sum_{t=1}^{n}\mathbf{g}_{t-(k+1)}\varepsilon_{t-k} + \sum_{m=0}^{k}M^m\sum_{t=1}^{n}\mathbf{h}_{t-m}\varepsilon_{t-k}.$$

By (3.4.4),

$$(3.4.7)\qquad \begin{aligned} E\left\|M^{k+1}\sum_{t=1}^{n}\mathbf{g}_{t-(k+1)}\varepsilon_{t-k}\right\| &\le \|M^{k+1}\|\left\{E\left\|\sum_{t=1}^{n}\mathbf{g}_{t-(k+1)}\varepsilon_{t-k}\right\|^2\right\}^{1/2} \\ &= o(n^{(\alpha+1)/2}). \end{aligned}$$

Now, by (3.4.1),

$$(3.4.8)\qquad \begin{aligned} E\left\|M^m\sum_{t=1}^{n}\mathbf{h}_{t-m}\varepsilon_{t-k}\right\| &\le \|M^m\|\left\{\sum_{t=1}^{n}E\|\mathbf{h}_{t-m}\|^2\right\}^{1/2}\left\{\sum_{t=1}^{n}E(\varepsilon_{t-k}^2)\right\}^{1/2} \\ &= o(n^{\alpha/2})n^{1/2} = o(n^{(\alpha+1)/2}). \end{aligned}$$

Combining (3.4.6)–(3.4.8), we have

$$E\left\|\sum_{t=1}^{n}\mathbf{g}_t\varepsilon_{t+j}\right\| = o(n^{(\alpha+1)/2}).$$

This together with (3.4.4) implies (3.4.2). For (3.4.3), first observe that if $\psi(z) = 1 + a_1 z + \cdots + a_p z^p$, then

$$\mathbf{z}_t = A\mathbf{z}_{t-1} + \boldsymbol{\varepsilon}_t, \qquad \mathbf{z}_0 = \mathbf{0}, \tag{3.4.9}$$

where

$$A = \begin{pmatrix} a_1 & \cdots & a_p \\ I_{p-1} & & \mathbf{0} \end{pmatrix},$$

I_{p-1} is the $(p-1)\times(p-1)$ identity matrix and $\boldsymbol{\varepsilon}_t = (\varepsilon_t, 0, \dots, 0)'$. A direct computation shows that

$$E\mathbf{z}_n\mathbf{z}_n' = \sum_0^{n-1} A^t\mathbf{e}\mathbf{e}'(A^t)', \quad \text{where } \mathbf{e} = (\sigma^2, 0, \dots, 0)'.$$

Hence,

$$\lim_{n\to\infty} \frac{1}{n}\sum_1^n E\|\mathbf{z}_t\|^2 = c < \infty. \tag{3.4.10}$$

Fix k. Then, by (3.4.9),

$$\mathbf{z}_t = A^k\mathbf{z}_{t-k} + \sum_{j=1}^{k-1} A^j\boldsymbol{\varepsilon}_{t-j}$$

and, consequently,

$$\sum_{t=1}^n \mathbf{g}_t\mathbf{z}_t' = \sum_{t=1}^n \mathbf{g}_t\mathbf{z}_{t-k}'(A')^k + \sum_{j=0}^{k-1}\sum_{t=1}^n \mathbf{g}_t\boldsymbol{\varepsilon}_{t-j}'(A')^j. \tag{3.4.11}$$

By (3.4.2),

$$\begin{aligned} &\limsup_{n\to\infty} n^{-(\alpha+1)/2} E\left\| \sum_{t=1}^n \mathbf{g}_t\boldsymbol{\varepsilon}_{t-j}'(A')^j \right\| \\ &\quad\le \limsup_{n\to\infty} \left\|(A')^j\right\| n^{-(\alpha+1)/2} E\left\| \sum_{t=1}^n \mathbf{g}_t\varepsilon_{t-j} \right\| = 0. \end{aligned} \tag{3.4.12}$$

By (3.4.1), there exists a constant $\gamma > 0$ such that $E\sum_{t=1}^n \|\mathbf{g}_t\|^2 \le \gamma n^\alpha$. By (3.4.10)–(3.4.12),

$$\begin{aligned} &\limsup_{n\to\infty} n^{-(\alpha+1)/2} E\left\| \sum_{t=1}^n \mathbf{g}_t\mathbf{z}_t' \right\| \\ &\quad= \limsup_{n\to\infty} n^{-(\alpha+1)/2} E\left\| \sum_{t=1}^n \mathbf{g}_t\mathbf{z}_{t-k}'(A')^k \right\| \\ &\quad\le \|(A')^k\| \limsup_{n\to\infty} n^{-(\alpha+1)/2}\left(E\sum_{t=1}^n \|\mathbf{g}_t\|^2 \right)^{1/2}\left(E\sum_{t=1}^n \|\mathbf{z}_{t-k}\|^2 \right)^{1/2} \\ &\quad\le \|(A')^k\|\gamma^{1/2}c \to 0 \quad \text{as } k\to\infty, \end{aligned}$$

since A has all eigenvalues inside the unit circle. This completes the proof. □

PROOF OF THEOREM 3.4.2. The jth row of $K_n\Sigma_1^n \mathbf{v}_t\mathbf{z}_t'M_n'$ is $n^{-j-1/2}\Sigma_1^n v_t(j)\mathbf{z}_t'$. But $v_t(j) = (-1)v_{t-1}(j) + v_t(j-1)$ and $E\Sigma_{t=1}^n v_t^2(j) = O(n^{2j})$. Hence, Lemma 3.4.3 is satisfied with $s = 1$, $M = (-1)$ and $\alpha = 2j$. Consequently, (ii) holds. Similarly, (i) holds. For (iii), observe that

$$\begin{pmatrix} y_t(j,m) \\ y_{t-1}(j,m) \end{pmatrix} = \begin{pmatrix} 2\cos\theta_j & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} y_{t-1}(j,m) \\ y_{t-2}(j,m) \end{pmatrix} + \begin{pmatrix} y_t(j,m-1) \\ 0 \end{pmatrix}$$

and $E\Sigma_{t=1}^n y_t^2(j,m) = O(n^{2j-1})$, by (3.3.4) and (3.3.9). It is not difficult to see that (iii) holds by using Lemma 3.4.3, with $\alpha = 2j - 1$ and

$$M = \begin{pmatrix} 2\cos\theta_j & -1 \\ 1 & 0 \end{pmatrix}. \qquad \square$$

3.5. *General model.* We are now ready to return to the general model (1.1) with the characteristic polynomial (3.1). Define J_n, K_n and $L_n(k)$ as in Section 3.4. Also, define F, ξ, $\tilde{F}$ and η as in Sections 3.1 and 3.2. For each k, define H_k and ζ_k by replacing d by d_k in Section 3.3.

THEOREM 3.5.1.

$$G_nQ\sum_1^{n-1}\mathbf{y}_t\mathbf{y}_t'Q'G_n^{-1} \to_{\mathscr{L}} \Delta \tag{3.5.1}$$

and

(3.5.2)

$$(Q'G_n')^{-1}(\mathbf{b}_n - \beta) \to_{\mathscr{L}} ((F^{-1}\xi)', (\tilde{F}^{-1}\eta)', (H_1^{-1}\zeta_1)', \dots, (H_l^{-1}\zeta_l)', N')',$$

where

$$\Delta = \begin{pmatrix} F & 0 & & \cdots & & 0 \\ 0 & \tilde{F} & & & & \\ & & H_1 & & & \\ \vdots & & \vdots & \ddots & & \vdots \\ & & & & H_l & \\ 0 & & & \cdots & & \Sigma \end{pmatrix} \quad \textit{is nonsingular a.s.,}$$

$(F, \xi), (\tilde{F}, \eta), (H_1, \zeta_1), \dots, (H_l, \zeta_l)$, N *are independent and* G_n, Q *and* N *are defined in* (3.2), (3.3) *and Theorem* 2.2, *respectively.*

PROOF. By Theorems 3.4.1 and 3.4.2, (3.4) is established. In view of (3.4), (3.1.9), (3.2.4), (3.3.7), (2.12) and Appendix 3, (3.5.1) is proved. Since each random matrix involved in Δ is nonsingular a.s., Δ is nonsingular a.s. Now, by (3.5.1), (3.5) is achieved. In view of (3.5), (3.1.10), (3.2.5) and (3.3.8), (3.5.2) is proved. Note that the random matrices and vectors involved in (3.4) and (3.5) are functionals of the corresponding processes of (2.12). The independence of $(F, \xi), \dots, N$ follows from Theorem 2.2. $\square$

4. Further discussion. 1. As shown in Section 3, the limiting distribution of $\mathbf{b}_n$ is represented by ratios of certain Brownian functionals. The closed form of the distribution functions of these functionals may be extremely complicated, as Rao's (1978) result indicates, and we have not attempted to derive such expressions in this paper.

2. Since the closed forms are not known, we may wonder how to use the functional expressions presented in Section 3 to calculate the percentiles of the limiting distributions. In the literature, these percentiles are tabulated by Monte Carlo simulations on some series representations which are derived from the quadratic form approach. By similar reasoning, we may expect that the Monte Carlo method can also be applied directly on the functional representations. In fact, a stochastic integral can be approximated by a partial sum of random variables, which in turn can be simulated easily. Since an approximation (or trunction) has to be made in the series expansion approach as well, it will be interesting to see which method produces a better result for the same amount of computing time.

3. In applications of nonstationary autoregressions, we may encounter the model where a constant drift μ is added to (1.1). It is known that [Dickey and Fuller (1979)] for an AR(1) model with $\mu \neq 0$, the least squares estimate for β_1 has normal limiting distribution regardless of whether $|\beta_1| < 1$ or $\beta_1 = 1$. Such phenomenon is also true for the multiple roots situation when all unit roots are equal to 1. However, for the general unstable AR(p) model, the limiting distributions of those components associated with roots $e^{i\theta}$, $\theta \in (0, 2\pi)$, will not be affected by the presence of a nonzero μ. For the AR(1) model where μ can be interpreted as the slope of the mean of y_t when $\beta_1 = 1$, the least squares estimate of μ has a normal limiting distribution as long as $|\beta_1| \leq 1$. For further details, see Chan (1987).

4. In order to apply the distributional results of this paper, some knowledge about the location of the roots is required. For example, the $\hat{\beta}_{n,2}$ in Corollary 3.3.8 is used by Ahtola and Tiao (1987) to test whether an AR(2) model possesses a pair of complex roots. In this case, even though the exact knowledge for the location of the roots is not required, the existence of a pair of complex roots has to be assumed a priori in order to compute the significance level of the associated test. As another interesting application, using the distribution results of this paper, Wei (1987) obtains a consistent estimate of the number of differencings in an integrated autoregressive model (IAR) without knowing the order of the model. However, as is well known, the IAR model does not have any unit roots besides 1.

APPENDIX 1.

Matrices Q, C and $L_n(j)$ were previously defined implicitly in Section 3.

We shall only give a detailed definition of the matrix Q here. The matrices C and $L_n(j)$ can be defined in exactly the same manner.

Let

$$\phi(z)(1-z)^{-a} = 1 + f_1 z + \cdots + f_{p-a} z^{p-a}.$$

Then

$$u_t = y_t + f_1 y_{t-1} + \cdots + f_{p-a} y_{t-p+a}.$$

Define the $a \times p$ matrix Q_1 by

$$Q_1 = \begin{bmatrix} 1 & f_1 & & \cdots & f_{p-a} & 0 & \cdots & 0 \\ & 1 & f_1 & \cdots & & f_{p-a} & \cdots & 0 \\ \vdots & & & & & & & \vdots \\ 0 & \cdots & & 1 & f_1 & & \cdots & f_{p-a} \end{bmatrix}.$$

Then $Q_1 \mathbf{y}_t = \mathbf{u}_t$. Similarly, we can define a $b \times p$ matrix Q_2, a $q \times p$ matrix Q_3 and a $2d_k \times p$ matrix $Q_4(k)$ such that

$$Q_2 \mathbf{y}_t = \mathbf{v}_t, \qquad Q_3 \mathbf{y}_t = \mathbf{z}_t \quad \text{and} \quad Q_4(k)\mathbf{y}_t = \mathbf{x}_t(k).$$

Then

$$Q = (Q_1', Q_2', Q_4'(1), \ldots, Q_4'(l), Q_3')'.$$

APPENDIX 2.

PROOF OF LEMMA 3.1.1. Let $\Omega = \{\omega\colon W(\omega, t)$ is continuous and nondifferentiable for $0 \le t \le 1\}$. It is well known that $P(\Omega) = 1$. We will show that $F(\omega)$ is nonsingular for any $\omega \in \Omega$. If not, then $\exists\ \omega \in \Omega$ and $\mathbf{c} = (c_1, \ldots, c_a)' \neq \mathbf{0}$ such that $\mathbf{c}'F(\omega)\mathbf{c} = 0$, i.e.,

$$\text{(A1)} \qquad \int_0^1 \left(\sum_{j=1}^{a} c_j F_{j-1}(\omega, s) \right)^2 ds = 0.$$

By the choice of ω, $\sum_{j=1}^{a} c_j F_{j-1}(\omega, s)$ is a continuous function in s. Hence, (A1) implies

$$\sum_{j=1}^{a} c_j F_{j-1}(\omega, s) = 0 \quad \text{for } 0 \le s \le 1.$$

Let $l = \inf\{j\colon c_j \neq 0\}$. Since $\mathbf{c} \neq 0$, $1 \le l \le a$. Consequently,

$$\text{(A2)} \qquad F_{l-1}(\omega, s) = \frac{1}{c_l} \sum_{j=l+1}^{a} c_j F_{j-1}(\omega, s) \quad \text{for } 0 \le s \le 1.$$

Since $F_j(\omega, s) = \int_0^s F_{j-1}(\omega, t)\, dt$ for $j \ge 1$, differentiating (A2) $(l-1)$ times, we obtain

$$W(\omega, s) = \frac{1}{c_l} \sum_{j=l+1}^{a} c_j F_{j-l}(\omega, s) \quad \text{for } 0 \le s \le 1.$$

This is a contradiction, as $F_1, \ldots, F_{a-l}$ are differentiable. □

PROOF OF LEMMA 3.3.3. Let

$$\mathbf{h} = \left(f_0 + ig_0, e^{-i\theta}(f_0 + ig_0), \ldots, e^{-i\theta}(f_{d-1} + ig_{d-1}) \right)$$

and $\bar{\mathbf{h}}$ be its conjugate pair, i.e.,

$$\bar{\mathbf{h}} = \left(f_0 - ig_0, e^{i\theta}(f_0 - ig_0), \dots, e^{i\theta}(f_{d-1} - ig_{d-1})\right).$$

Set

$$\text{(A3)} \qquad \mathbf{h}\bar{\mathbf{h}}' = \Gamma + iM,$$

where Γ and M are $2d \times 2d$ real matrices. It is not difficult to see that

$$\text{(A4)} \qquad H = \frac{1}{4\sin^2\theta}\int_0^1 \Gamma(s)\, ds.$$

Since for any $\mathbf{c} \in R^{2d\times 2d}$,

$$0 \le \mathbf{c}'\mathbf{h}\bar{\mathbf{h}}'\mathbf{c} = \mathbf{c}'\Gamma\mathbf{c} + i\mathbf{c}'M\mathbf{c} = \mathbf{c}'\Gamma\mathbf{c},$$

we have

$$\mathbf{c}'H\mathbf{c} = \frac{1}{4\sin^2\theta}\int_0^1 |\mathbf{c}'\mathbf{h}(s)|^2\, ds.$$

Now,

$$\mathbf{c}'\mathbf{h} = \left(c_1 + c_2 e^{-i\theta}\right)\left(f_0 + ig_0\right) + \cdots + \left(c_{2d-1} + c_{2d} e^{-i\theta}\right)\left(f_{d-1} + ig_{d-1}\right).$$

By assumption, $e^{i\theta} \neq 1$ or -1. Since c_{2j-1} and c_{2j} are real numbers, we have that

$$c_{2j-1} + c_{2j} e^{-i\theta} = 0 \quad \text{iff} \quad c_{2j-1} = c_{2j} = 0.$$

Hence, if $\mathbf{c} \neq \mathbf{0}$, then there exists $1 \le j \le d$ such that $c_{2j-1} + c_{2j} e^{-i\theta} \neq 0$. Using this fact and a similar argument as in the proof of Lemma 3.1.1, we can show that H is nonsingular everywhere on

$$\Omega = \{\omega\colon W_1(\omega) \text{ and } W_2(\omega) \text{ are continuous but nondifferentiable}\}.$$

Since $P(\Omega) = 1$, the proof is complete. □

APPENDIX 3.

In general, the weak convergence of the marginal distributions may not imply the joint convergence. The following proposition, which gives such a result, and Theorems 2.2–2.4 imply all claims of joint convergence in this paper.

PROPOSITION. *Assume that* $X_n = (X_n(1), \dots, X_n(m))$ *and* $X = (X(1), \dots, X(m))$ *are random elements taking values in* $\prod_{i=1}^m D$ *such that*

$$\text{(A5)} \qquad X_n \to_{\mathscr{L}} X.$$

For each i, $1 \le i \le l$, *let* I_i *be a nonempty subset of* $\{1, \dots, m\}$ *and assume that* $G_n(i)$ *is a sequence of random variables such that*

$$\text{(A6)} \qquad \{X_n(k),\, k \in I_i,\, G_n(i)\} \to_{\mathscr{L}} \{X(k),\, k \in I_i,\, G_i\},$$

where G_i *is measurable with respect to* $\sigma\{X(k),\, k \in I_i\}$. *Then*

$$\text{(A7)} \qquad \{X_n, G_n(i), 1 \le i \le l\} \to_{\mathscr{L}} \{X, G_i, 1 \le i \le l\}.$$

PROOF. By (A5) and (A6), $\{X_n, G_n(i), 1 \le i \le l\}$ is tight. In order to show (A7), it is sufficient [Billingsley (1968), page 16] to prove that for any subsequence n_j such that

$$\{X_{n_j}, G_{n_j}(i), 1 \le i \le l\} \to_{\mathscr{L}} \{\tilde{X}, \tilde{G}_i, 1 \le i \le l\}, \tag{A8}$$

we have that

$$\{\tilde{X}, \tilde{G}_i, 1 \le i \le l\} =_{\mathscr{L}} \{X, G_i, 1 \le i \le l\}. \tag{A9}$$

In view of (A6) and (A8), for each i,

$$\{\tilde{X}(k),\ k \in I_i, \tilde{G}_i\} =_{\mathscr{L}} \{X(k),\ k \in I_i, G_i\}. \tag{A10}$$

Since G_i is $\sigma\{X(k),\ k \in I_i\}$-measurable, by a lemma in Gihman and Skorokhod [(1974), page 8] there is a measurable function g_i: $\prod_{k \in I_i} D \to R$ such that $G_i = g_i(X(k),\ k \in I_i)$. (A10) implies that

$$\tilde{G}_i = g_i(\tilde{X}(k),\ k \in I_i) \quad \text{a.s.} \tag{A11}$$

But from (A5) and (A8) we have that $\tilde{X} =_{\mathscr{L}} X$. Now (A9) is a consequence of this and (A11). □

Acknowledgments. The results of this paper were obtained in June 1985. While the second author attended the ISI conference in August, he also learned that Nyblom (1985) had independently obtained the results for the AR(2) case. We would also like to thank an Associate Editor and two referees for their useful comments.

REFERENCES

AHTOLA, J. A. and TIAO, G. C. (1987). Distributions of least squares estimators of autoregressive parameters for a process with complex roots on the unit circle. *J. Time Ser. Anal.* **8** 1–14.

ANDERSON, O. D. (1985). Practical examples of ARUMA modeling. In *Time Series Analysis: Theory and Practice* 7 (O. D. Anderson, ed.) 15–21. North-Holland, Amsterdam.

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

BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

BOX, G. E. P. and JENKINS, G. M. (1976). *Time Series Analysis: Forecasting and Control.* Holden-Day, San Francisco.

CHAN, N. H. (1987). Limiting distributions of least square estimates of unstable autoregressive models with drifts. Technical Report, Dept. Mathematics, Indiana Univ.

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.

DICKEY, D. A. and FULLER, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica* **49** 1057–1072.

DICKEY, D. A., HASZA, D. P. and FULLER, W. A. (1984). Testing for units roots in seasonal time series. *J. Amer. Statist. Assoc.* **79** 355–367.

ELLIOTT, R. J. (1982). *Stochastic Calculus and Applications.* Springer, New York.

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.

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

FULLER, W. A. (1985). Nonstationary autoregressive time series. In *Handbook of Statistics: Time Series in the Time Domain* (E. J. Hannan, P. R. Krishnaiah and M. M. Rao, eds.) **5** 1–23. North-Holland, Amsterdam.

FULLER, W. A. and HASZA, D. P. (1980). Predictors for the first order autoregressive process. *J. Econometrics* **13** 139–157.

GIHMAN, I. I. and SKOROKHOD, A. V. (1974). *The Theory of Stochastic Processes* **1**. Springer, New York.

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

GRAUPE, D. (1980). On convergence of least-squares identifiers of autoregressive models having stable and unstable roots. *IEEE Trans. Automat. Control* **AC-25** 999–1002.

HASZA, D. P. and FULLER, W. A. (1979). Estimation for autoregressive processes with unit roots. *Ann. Statist.* **7** 1106–1120.

HELLAND, I. S. (1982). Central limit theorems for martingales with discrete or continuous time. *Scand. J. Statist.* **9** 79–94.

KAC, M. (1980). *Integration in Function Spaces and Some of Its Applications*. Academia Nationale Dei Lincei, Pisa.

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.

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

LOÈVE, M. (1978). *Probability Theory* **2**, 4th ed. Springer, Berlin.

MANN, H. B. and WALD, A. (1943). On the statistical treatment of linear stochastic difference equations. *Econometrica* **11** 173–220.

NYBLOM, J. (1985). On the least squares estimators for the unstable AR(2) models. *Contributed Papers*, 45th Session, ISI. 385–386.

PHILLIPS, P. C. B. (1985). Time series regression with unit roots. Cowles Foundation Discussion Paper 740.

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

SKOROKHOD, A. V. (1956). Limit theorems for stochastic processes. *Theory Probab. Appl.* **1** 261–290.

SOLO, V. (1984). The order of differencing in ARIMA models. *J. Amer. Statist. Assoc.* **79** 916–921.

STIGUM, R. P. (1974). Asymptotic properties of dynamic parameter estimates, III. *J. Multivariate Anal.* **4** 351–381.

WEI, C. Z. (1987). Adaptive prediction by least squares predictors in stochastic regression models with applications to time series. *Ann. Statist.* **15** 1667–1682.

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

DEPARTMENT OF MATHEMATICS
INDIANA UNIVERSITY
BLOOMINGTON, INDIANA 47405

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF MARYLAND
COLLEGE PARK, MARYLAND 20742