The Annals of Statistics
1979, Vol. 7, No. 5, 1106–1120

ESTIMATION FOR AUTOREGRESSIVE PROCESSES WITH UNIT ROOTS

BY DAVID P. HASZA AND WAYNE A. FULLER

Kansas State University and Iowa State University

Let Y_t satisfy the stochastic difference equation $Y_t = \Sigma_{j=1}^p \eta_j Y_{t-j} + e_t$ for $t = 1, 2, \cdots$, where the e_t are independent identically distributed $(0, \sigma^2)$ random variables and the initial conditions $(Y_{-p+1}, Y_{-p+2}, \cdots, Y_0)$ are fixed constants. It is assumed the true, but unknown, roots $m_1, m_2, \cdots, m_p$ of $m^p - \Sigma_{j=1}^p \eta_j m^{p-j} = 0$ satisfy $m_1 = m_2 = 1$ and $|m_j| < 1$ for $j = 3, 4, \cdots, p$. Let $\hat{\boldsymbol{\eta}}$ denote the least squares estimator of $\boldsymbol{\eta} = (\eta_1, \eta_2, \cdots, \eta_p)'$ obtained by the least squares regression of Y_t on $Y_{t-1}, Y_{t-2}, \cdots, Y_{t-p}$ for $t = 1, 2, \cdots, n$. The asymptotic distributions of $\hat{\boldsymbol{\eta}}$ and of a test statistic designed to test the hypothesis that $m_1 = m_2 = 1$ are characterized. Analogous distributional results are obtained for models containing time trend and intercept terms. Estimated percentiles for these distributions are obtained by the Monte Carlo method.

1. Introduction. Let the time series $\{Y_t\}$ satisfy

$$Y_t = \Sigma_{j=1}^p \eta_j Y_{t-j} + e_t \qquad t = 1, 2, \cdots, \tag{1.1}$$

where $\{e_t\}_{t=1}^\infty$ is a sequence of independent random variables with mean zero and variance σ^2. It is assumed that the e_t are either identically distributed or that $E\{|e_t|^{2+\delta}\} < M$ for some $\delta > 0$ and all t. It is further assumed that the initial conditions $(Y_{-p+1}, Y_{-p+2}, \cdots, Y_0)$ are known constants. The time series Y_t is said to be an autoregressive process of order p. Let

$$m^p - \Sigma_{j=1}^p \eta_j m^{p-j} = 0 \tag{1.2}$$

be the characteristic equation of the process. The roots of (1.2), denoted by $m_1, m_2, \cdots, m_p$ are the characteristic roots of the process.

It is assumed that $\boldsymbol{\eta} = (\eta_1, \eta_2, \cdots, \eta_p)'$ and σ^2 are unknown. We shall consider estimation of $\boldsymbol{\eta}$ and tests of hypotheses concerning $\boldsymbol{\eta}$. Let the observations $(Y_1, Y_2, \cdots, Y_n)$ be available and let $\boldsymbol{\Phi}_t = (Y_{t-1}, Y_{t-2}, \cdots, Y_{t-p})'$. Define the least squares estimator of $\boldsymbol{\eta}$ by

$$\hat{\boldsymbol{\eta}} = (\Sigma_{t=1}^n \boldsymbol{\Phi}_t \boldsymbol{\Phi}_t')^{-1} \Sigma_{t=1}^n \boldsymbol{\Phi}_t Y_t. \tag{1.3}$$

The estimator $\hat{\boldsymbol{\eta}}$ is the maximum likelihood estimator of $\boldsymbol{\eta}$ if the e_t are normally distributed. If $|m_j| < 1$ for $j = 1, 2, \cdots, p$, Y_t is converging to a weakly stationary process as $t \to \infty$. In the stationary case the asymptotic distributions of $n^{\frac{1}{2}}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})$, and of related "$F$-type" likelihood ratio test statistics are well known. See Mann and Wald (1943), Anderson (1959) and Hannan and Heyde (1972).

Received September 1977; revised July 1978.

AMS 1970 *subject classifications.* Primary 62M10; secondary 62J05.

Key words and phrases. Time series, autoregression, nonstationary, differencing.

White (1958) characterized the asymptotic distribution of $\hat{\eta}_1$ when $p = 1$ and $\eta_1 = 1$. Dickey (1976) obtained an alternative representation of the distribution of $\hat{\eta}_1$ for $p = 1$ and $\eta_1 = 1$. He extended his results to models containing intercept and time terms. Fuller (1976, Section 8.5) discussed estimation of the pth order model when the true, but unknown, $\boldsymbol{\eta}$ is such that $|m_1| = 1$ and $|m_j| < 1$ for $j = 2, 3, \cdots, p$. Fuller presents tables of the percentage points of statistics that can be used to test the hypothesis that $m_1 = 1$ given that $|m_j| < 1$ for $j = 2, 3, \cdots, p$. Rao (1978) gave the density of the asymptotic distribution of $\hat{\eta}_1$ when $p = 1$ and $\eta_1 = 1$. Anderson (1959) and Rao (1961), among others, have considered processes with roots greater than one in modulus.

We shall study the asymptotic properties of $\hat{\boldsymbol{\eta}}$ and related statistics when the true but unknown roots of (1.2) satisfy $m_1 = m_2 = 1$ and $|m_j| < 1$ for $j = 3, 4, \cdots, p$. We shall also construct a test of the hypothesis $m_1 = m_2 = 1$. Models with intercept and slope terms are also considered. The case $p = 2$ is considered in Sections 2, 3, and 4 and the results are extended to higher order processes in Section 5. An example is given in Section 7.

2. Order results. Consider the second order model

$$Y_t = \eta_1 Y_{t-1} + \eta_2 Y_{t-2} + e_t \qquad t = 1, 2, \cdots, \tag{2.1}$$

and assume, for convenience, that $Y_{-1} = Y_0 = 0$. We reparametrize the model to obtain

$$Y_t = \alpha Y_{t-1} + \beta(Y_{t-1} - Y_{t-2}) + e_t. \tag{2.2}$$

The condition that the roots of (1.2) satisfy $m_1 = m_2 = 1$ is equivalent to $\alpha = \beta = 1$. Let

$$(\hat{\alpha}, \hat{\beta})' = (\Sigma_{t=1}^n \psi_t \psi_t')^{-1} \Sigma_{t=1}^n \psi_t Y_t, \tag{2.3}$$

where $\psi_t = (Y_{t-1}, Z_{t-1})'$ and $Z_t = Y_t - Y_{t-1}$. In this section we establish the order in probability of the differences $\hat{\alpha} - \alpha$ and $\hat{\beta} - \beta$. As preliminaries the order of statistics entering the definition of $(\hat{\alpha}, \hat{\beta})$ are derived in Lemmas 2.1 and 2.2.

LEMMA 2.1. *Let model* (2.1) *hold with* $m_1 = m_2 = 1$. *Let* $\{e_t\}_{t=1}^{\infty}$ *be a sequence of independent* $(0, \sigma^2)$ *random variables. Assume the* e_t *are identically distributed or* $E\{|e_t|^{2+\delta}\} < M$ *for all* t *and some* $\delta > 0$. *Let* $\varepsilon > 0$ *be given. Then there exist* $M_1 > 0$, $M_2 > 0$, *and* N *such that for* $n > N$

$$P\{M_1 n^3 < Y_n^2 < M_2 n^3\} > 1 - \varepsilon,$$

$$P\{M_1 n^4 < \Sigma_{t=1}^n Y_t^2 < M_2 n^4\} > 1 - \varepsilon.$$

PROOF. Solving the defining difference equation we obtain

$$Y_n = \Sigma_{t=1}^n t e_{n-t+1} \qquad n = 1, 2, \cdots.$$

It is readily verified that

$$3n^{-3} Y_n^2 \to_{\mathcal{L}} \sigma^2 \chi_1^2$$

and

$$20n^{-5}(\Sigma_{t=1}^{n} Y_t)^2 \to_{\mathcal{L}} \sigma^2 \chi_1^2,$$

where χ_1^2 denotes a chi-square random variable with one degree of freedom. Now

$$n^3 \Sigma_{t=1}^{n} t^{-3} Y_t^2 \geqslant \Sigma_{t=1}^{n} Y_t^2 \geqslant n^{-1}(\Sigma_{t=1}^{n} Y_t)^2$$

and the result follows because $E\{t^{-3}Y_t^2\}$ is uniformly bounded. □

LEMMA 2.2. *Let the model assumptions of Lemma 2.1 hold. Let $\varepsilon > 0$ be given and define*

$$b(n) = (\Sigma_{t=1}^{n} Y_{t-1}^2)^{-1} \Sigma_{t=1}^{n} Y_{t-1} Z_{t-1}. \tag{2.4}$$

Then there exist $M_1 > 0$, $M_2 > 0$, and N such that for all $n > N$

$$P\{M_1 n^{-1} < b(n) < M_2 n^{-1}\} > 1 - \varepsilon.$$

PROOF. Using $Y_t = 2Y_{t-1} - Y_{t-2} + e_t$, we obtain

$$2\Sigma_{t=1}^{n} Y_{t-1} Z_t = 2\Sigma_{t=1}^{n} Y_t Y_{t-1} - 2\Sigma_{t=1}^{n} Y_{t-1}^2 = Y_n Y_{n-1} + \Sigma_{t=1}^{n} Y_{t-1} e_t$$

and

$$b(n) = (\Sigma_{t=1}^{n} Y_{t-1}^2)^{-1}\left(\tfrac{1}{2} Y_n Y_{n-1} - \tfrac{1}{2}\Sigma_{t=1}^{n} Y_{t-1} e_t\right).$$

Because $Y_n(Y_n - Y_{n-1}) = O_p(n^2)$, $\Sigma_{t=1}^{n} Y_{t-1} e_t = O_p(n^2)$, and $(\Sigma_{t=1}^{n} Y_{t-1}^2)^{-1} = O_p(n^{-4})$ we have

$$b(n) = \tfrac{1}{2} Y_{n-1}^2 (\Sigma_{t=1}^{n} Y_{t-1}^2)^{-1} + O_p(n^{-2}).$$

The result follows by Lemma 2.1. □

We now give the main result of this section.

THEOREM 2.1. *Let the model assumptions of Lemma 2.1 hold. Then*

$$[n^2(\hat{\alpha} - 1), n(\hat{\beta} - 1)]' = O_p(1).$$

PROOF. The error in the least squares estimator is

$$(\hat{\alpha} - 1, \hat{\beta} - 1)' = (\Sigma_{t=1}^{n} \psi_t \psi_t')^{-1} \Sigma_{t=1}^{n} \psi_t e_t$$

$$= D^{-1}(N_1, N_2)',$$

where

$$N_1 = (\Sigma_{t=1}^{n} Z_{t-1}^2)(\Sigma_{t=1}^{n} Y_{t-1} e_t) - (\Sigma_{t=1}^{n} Y_{t-1} Z_{t-1})(\Sigma_{t=1}^{n} Z_{t-1} e_t),$$

$$N_2 = (\Sigma_{t=1}^{n} Y_{t-1}^2)(\Sigma_{t=1}^{n} Z_{t-1} e_t) - (\Sigma_{t=1}^{n} Y_{t-1} Z_{t-1})(\Sigma_{t=1}^{n} Y_{t-1} e_t),$$

$$D = (\Sigma_{t=1}^{n} Y_{t-1}^2)(\Sigma_{t=1}^{n} Z_{t-1}^2) - (\Sigma_{t=1}^{n} Y_{t-1} Z_{t-1})^2.$$

It can be established that

$$(\Sigma_{t=1}^{n} Z_{t-1}^2)(\Sigma_{t=1}^{n} Y_{t-1}e_t) = O_p(n^4),$$

$$(\Sigma_{t=1}^{n} Y_{t-1}Z_{t-1})(\Sigma_{t=1}^{n} Z_{t-1}e_t) = O_p(n^4).$$

From Lemma 2.1 $\Sigma_{t=1}^{n} Y_{t-1}^2 = O_p(n^4)$ and it follows that $N_1 = O_p(n^4)$ and $N_2 = O_p(n^5)$. To complete the proof we shall show that $D^{-1} = O_p(n^{-6})$. Now

$$D = (\Sigma_{t=1}^{n} Y_{t-1}^2)(\Sigma_{t=1}^{n} Z_{t-1}^2 - b^2(n)\Sigma_{t=1}^{n} Y_{t-1}^2),$$

where $b(n)$ was defined in Lemma 2.2. Let $\varepsilon > 0$ be given and choose $M_1 > 0$, $M_2 > 0$, and N as in Lemma 2.2 so that, for $n > N$

$$P\{M_1 n^{-1} < b(n) < M_2 n^{-1}\} > 1 - \varepsilon/3.$$

Let $n > 4M_1^{-1}M_2 N$ and $n^* = [2^{-1}M_2^{-1}M_1 n]$, where $[\cdot]$ denotes the greatest integer function. Note that $n^* > N$ and

$$P\{b(n^*) > 2M_2 n^{-1}\} > 1 - \varepsilon/3,$$

where

$$b(n^*) = (\Sigma_{t=1}^{n^*} Y_{t-1}Z_{t-1})(\Sigma_{t=1}^{n^*} Y_{t-1}^2)^{-1}.$$

Therefore $P\{[b(n^*) - b(n)]^2 > M_2^2 n^{-2}\} > 1 - 2\varepsilon/3$. We have

$$\begin{aligned} D &= (\Sigma_{t=1}^{n} Y_{t-1}^2)(\Sigma_{t=1}^{n}[Z_{t-1} - b(n)Y_{t-1}]^2) \\ &\geqslant (\Sigma_{t=1}^{n} Y_{t-1}^2)(\Sigma_{t=1}^{n^*}[Z_{t-1} - b(n^*)Y_{t-1} + \{b(n^*) - b(n)\}Y_{t-1}]^2) \\ &= (\Sigma_{t=1}^{n} Y_{t-1}^2)(\Sigma_{t=1}^{n^*}[Z_{t-1} - b(n^*)Y_{t-1}]^2 + [b(n^*) - b(n)]^2\Sigma_{t=1}^{n^*} Y_{t-1}^2) \\ &\geqslant [b(n^*) - b(n)]^2(\Sigma_{t=1}^{n^*} Y_{t-1}^2)^2. \end{aligned}$$

By Lemma 2.1 there exist $M_3 > 0$ and $N_0 > N$ such that for $n > N_0$

$$P\{\Sigma_{t=1}^{n} Y_{t-1}^2 < M_3 n^4\} < \varepsilon/3.$$

Therefore, for $n^* > N_0$

$$P\left\{D < \frac{M_2^2}{n^2} M_3^2 \left(\frac{M_1 n}{2M_2}\right)^8\right\} < \varepsilon$$

and $P\{D < Mn^6\} < \varepsilon$, where $M = (256M_2^6)^{-1}M_1^8 M_3^2$. Thus $D^{-1} = O_p(n^{-6})$. □

3. Asymptotic distributions. In this section we consider generalizations of model (2.1). To aid in differentiating among the models we add subscripts to the parameters appearing in the models. Thus we rewrite model (2.2) as

$$Y_t = \alpha_1 Y_{t-1} + \beta_1(Y_{t-1} - Y_{t-2}) + e_t. \tag{3.1}$$

The first generalization of model (3.1) is the model

$$Y_t = \mu_2 + \alpha_2 Y_{t-1} + \beta_2(Y_{t-1} - Y_{t-2}) + e_t \tag{3.2}$$

and the second generalization is the model

$$Y_t = \mu_3 + \theta_3 t + \alpha_3 Y_{t-1} + \beta_3(Y_{t-1} - Y_{t-2}) + e_t. \tag{3.3}$$

If the roots of the characteristic equation (1.2) satisfy $m_1 = 1$ and $|m_2| < 1$ and if $\mu_2 \neq 0$ the process (3.2) will display a linear time trend or drift. The process (3.3) will display somewhat similar behavior if $\theta_3 \neq 0$ and both roots of the characteristic equation are less than one in absolute value. If $m_1 = m_2 = 1$ and $\mu_2 \neq 0$ the process (3.2) will contain a quadratic trend. Therefore the three models (3.1)–(3.3) form an interesting class of time series models.

Let $\xi_1 = (\alpha_1, \beta_1)'$, $\xi_2 = (\mu_2, \alpha_2, \beta_2)'$, $\xi_3 = (\mu_3, \theta_3, \alpha_3, \beta_3)'$, $\psi_{1t} = (Y_{t-1}, Z_{t-1})'$, $\psi_{2t} = (1, Y_{t-1}, Z_{t-1})'$, and $\psi_{3t} = (1, t, Y_{t-1}, Z_{t-1})'$. Then the least squares estimator of the parameters of the model indexed by i is

$$\hat{\xi}_i = (\Sigma_{t=1}^n \psi_{it}\psi_{it}')^{-1}\Sigma_{t=1}^n \psi_{it} Y_t.$$

For normal e_t, the likelihood ratio tests of hypotheses about elements of ξ_i are monotone functions of the usual "F-statistics" of normal regression theory. For example, to test the hypothesis $H_1 : (\alpha_1, \beta_1) = (1, 1)$ under the maintained model (3.1) with normal e_t and fixed (Y_0, Y_{-1}), the test statistic is

$$\Phi_1(2) = (2s_1^2)^{-1}(\hat{\xi}_1 - \mathbf{1})'\Sigma_{t=1}^n \psi_{1t}(Y_t - 2Y_{t-1} + Y_{t-2}),$$

where $\mathbf{1} = (1, 1)'$ and

$$s_1^2 = (n-2)^{-1}\Sigma_{t=1}^n [Y_t - \psi_{1t}'\hat{\xi}_1]^2.$$

Let $\Phi_2(2)$ and $\Phi_2(3)$ denote the analogous test statistics for the hypotheses $(\alpha_2, \beta_2) = (1, 1)$ and $(\mu_2, \alpha_2, \beta_2) = (0, 1, 1)$ for model (3.2) and let $\Phi_3(2)$ and $\Phi_3(4)$ denote the test statistics for the hypotheses $(\alpha_3, \beta_3) = (1, 1)$ and $(\mu_3, \theta_3, \alpha_3, \beta_3) = (0, 0, 1, 1)$ for model (3.3).

We shall obtain representations of the limiting distributions of the Φ-statistics in two steps. We first demonstrate that each statistic can be expressed approximately as a continuous differentiable function of a vector $\mathbf{W}_n$ of six elementary statistics defined in Lemma 3.1. The limiting distribution of the vector of elementary statistics is obtained in Theorem 3.1. The limiting distributions of the $\hat{\xi}_i$ are presented in Theorem 3.2 and the limiting distributions of the Φ-statistics in Corollary 3.1.

LEMMA 3.1. *Let model* (2.1) *hold with* $m_1 = m_2 = 1$. *Let* $\{e_t\}_{t=1}^\infty$ *be a sequence of independent* $(0, \sigma^2)$ *random variables. Assume the* e_t *are identically distributed or* $E\{|e_t|^{2+\delta}\} < M$ *for all* t *and some* $\delta > 0$. *Let* $\mathbf{W}_n' = (W_{1n}, W_{2n}, W_{3n}, W_{4n}, W_{5n}, W_{6n})$, *where*

$$(W_{1n}, W_{2n}) = n^{-\frac{1}{2}}(Z_{n-1}, n^{-1}Y_{n-1}),$$

$$(W_{3n}, W_{4n}) = n^{-\frac{1}{2}}(n^{-2}\Sigma_{t=1}^n Y_{t-1}, n^{-3}\Sigma_{t=1}^n tY_{t-1}),$$

$$(W_{5n}, W_{6n}) = (n^{-2}\mathbf{e}_n'\mathbf{A}_n\mathbf{e}_n, n^{-4}\mathbf{e}_n'\mathbf{A}_n^2\mathbf{e}_n),$$

$\mathbf{e}_n = (e_1, e_2, \cdots, e_{n-1})'$, $\mathbf{A}_n = \mathbf{B}_n'\mathbf{B}_n$ *and* $\mathbf{B}_n$ *is the lower triangular matrix of dimen-*

sion $n-1$ *with* $B_{ij}=1$ *for* $i \geqslant j$. *Then*

(i) $n^{-1}\Sigma_{t=1}^{n} Z_{t-1}e_t = \frac{1}{2}W_{1n}^2 - \frac{\sigma^2}{2} + o_p(1)$,

(ii) $n^{-\frac{3}{2}}\Sigma_{t=1}^{n} te_t = W_{1n} - W_{2n} + O_p(n^{-\frac{1}{2}})$,

(iii) $n^{-2}\Sigma_{t=1}^{n} Y_{t-1}e_t = W_{1n}W_{2n} - W_{5n} + O_p(n^{-\frac{1}{2}})$,

(iv) $n^{-\frac{5}{2}}\Sigma_{t=1}^{n} tZ_{t-1} = W_{2n} - W_{3n} + O_p(n^{-1})$,

(v) $n^{-3}\Sigma_{t=1}^{n} Y_{t-1}Z_{t-1} = \frac{1}{2}W_{2n}^2 + O_p(n^{-1})$,

(vi) $n^{-4}\Sigma_{t=1}^{n} Y_{t-1}^2 = W_{6n} - W_{2n}^2 + 2W_{2n}W_{3n}$.

PROOF. Result (i) is given by Anderson (1959). We present a proof only for result (vi) as similar techniques apply to the other results.

The tth element of $\mathbf{A}_n\mathbf{e}_n$ is $Y_{n-1} - Y_{t-1}$. Therefore

$$n^4W_{6n} = \Sigma_{t=1}^{n-1}(Y_{n-1} - Y_{t-1})^2$$
$$= (n-1)Y_{n-1}^2 - 2Y_{n-1}\Sigma_{t=1}^{n-1}Y_{t-1} + \Sigma_{t=1}^{n-1}Y_{t-1}^2$$

and

$$\Sigma_{t=1}^{n} Y_{t-1}^2 = n^4W_{6n} - nY_{n-1}^2 + 2Y_{n-1}\Sigma_{t=1}^{n}Y_{t-1}.$$

Therefore

$$n^{-4}\Sigma_{t=1}^{n} Y_{t-1}^2 = W_{6n} - W_{2n}^2 + 2W_{2n}W_{3n}. \qquad \square$$

Let $\boldsymbol{\lambda}_n = (\lambda_{1n}, \lambda_{2n}, \cdots, \lambda_{n-1,n})'$ denote the vector of eigenvalues of $\mathbf{A}_n$, where $\lambda_{1n} > \lambda_{2n} > \cdots > \lambda_{n-1,n}$. Furthermore let $\mathbf{x}_{in} = (x_{in1}, x_{in2}, \cdots, x_{in,n-1})'$ denote the eigenvector associated with λ_{in} and define the orthogonal transformation of $\mathbf{e}_n$ into $\mathbf{u}_n = (u_{1n}, u_{2n}, \cdots, u_{n-1,n})'$ by

$$u_{in} = \Sigma_{t=1}^{n-1}x_{int}\, e_t,$$

where

$$\lambda_{in} = \tfrac{1}{4}\sec^2(2n-1)^{-1}(n-i)\pi$$

and

$$x_{int} = 2(2n-1)^{-\frac{1}{2}}\cos\left[(2n-1)^{-1}(2i-1)\pi\left(t-\tfrac{1}{2}\right)\right].$$

See Dickey (1976) for a derivation of the roots and vectors of $\mathbf{A}_n$.

LEMMA 3.2. *Let the assumptions of Lemma* 3.1 *hold and let* $k \geqslant 1$ *be fixed. Then*

$$(u_{1n}, u_{2n}, \cdots, u_{kn})' \rightarrow_{\mathcal{L}} N_k(\mathbf{0}, \sigma^2\mathbf{I}_k).$$

PROOF. Because the cosine function is bounded we have

$$\lim_{n\to\infty}\sup_{1<t<n-1}x_{int}^2 = 0.$$

Therefore we may apply the Lindeberg central limit theorem for triangular arrays, Chung (1974, page 205) to obtain

$$u_{in} \rightarrow_{\mathcal{L}} N(0, \sigma^2).$$

Because the covariance matrix of $(u_{1n}, u_{2n}, \cdots, u_{kn})'$ is $\sigma^2 \mathbf{I}_k$ for all $n > k$, $(u_{1n}, u_{2n}, \cdots, u_{kn})' \rightarrow_{\mathcal{L}} N_k(\mathbf{0}, \sigma^2 \mathbf{I}_k)$. □

In obtaining the asymptotic properties of our estimators we require the asymptotic joint distribution (as $n \rightarrow \infty$) of the W_{jn}. Lemma 3.3 will be used to obtain the joint distribution.

LEMMA 3.3. *Let the assumptions of Lemma* 3.1 *hold. Then, for fixed* $i \geqslant 1$,

$$\lim_{n\rightarrow\infty}(\delta_{1in}, \delta_{2in}) = 2^{\frac{1}{2}}(\gamma_i, \gamma_i^2),$$

$$\lim_{n\rightarrow\infty}(\delta_{3in}, \delta_{4in}) = 2^{\frac{1}{2}}(\gamma_i^2 - \gamma_i^3, \tfrac{1}{2}\gamma_i^2 - \gamma_i^3 + \gamma_i^4),$$

$$\lim_{n\rightarrow\infty}\lambda_{in} = \gamma_i^2,$$

where

$$\gamma_i = 2[(2i-1)\pi]^{-1}(-1)^{i+1} \qquad i = 1, 2, \cdots,$$

and

$$\delta_{jin} = \sigma^{-2}\operatorname{Cov}(u_{in}, W_{jn}) \qquad j = 1, \cdots, 4; i = 1, 2, \cdots, n-1.$$

PROOF. The results for δ_{1in}, δ_{2in}, and λ_{in} are proved in Dickey (1976). We prove only the result for δ_{4in} as the other proofs are similar. We have

$$\operatorname{Cov}(u_{in}, W_{4n}) = \operatorname{Cov}\left(\Sigma_{t=1}^{n-1}x_{\text{int}}e_t, n^{-\frac{7}{2}}\Sigma_{t=1}^{n} tY_{t-1}\right).$$

Therefore

$$\sigma^2\delta_{4in} = n^{-\frac{7}{2}}\operatorname{Cov}\left(\Sigma_{t=1}^{n-1}x_{\text{int}}e_t, \Sigma_{t=1}^{n-1}\left[\frac{(n-t)^3}{3} + \frac{t(n-t)^2}{2}\right]e_t\right) + O(n^{-1})$$

$$= \sigma^2 n^{-\frac{7}{2}}\left(\Sigma_{t=1}^{n-1}\left[\frac{(n-t)^3}{3} + \frac{t(n-t)^2}{2}\right]\cos\left[\frac{2i-1}{2n-1}\left(t-\tfrac{1}{2}\right)\pi\right]\right) + O(n^{-1}).$$

For fixed $r \in \{1, 2, \cdots\}$ we have

$$\lim_{n\rightarrow\infty}n^{-r}\Sigma_{t=1}^{n-1}t^{r-1}\cos\left[\frac{2i-1}{2n-1}\left(t-\tfrac{1}{2}\right)\pi\right] = \int_0^1 x^{r-1}\cos\left[\frac{(2i-1)\pi}{2}x\right]dx.$$

It follows that

$$\lim_{n\rightarrow\infty}\delta_{4in} = 2^{\frac{1}{2}}\left(\tfrac{1}{2}\gamma_i^2 - \gamma_i^3 + \gamma_i^4\right). \qquad \square$$

We are now in a position to obtain the limiting distribution of $\mathbf{W}_n$.

THEOREM 3.1. *Let the assumptions of Lemma* 3.1 *hold. Let the elements of* $\mathbf{W} = (W_1, W_2, \cdots, W_6)'$ *be defined by*

$$(W_1, W_2) = 2^{\frac{1}{2}}(\Sigma_{i=1}^{\infty}\gamma_i V_i, \Sigma_{i=1}^{\infty}\gamma_i^2 V_i),$$

$$(W_3, W_4) = 2^{\frac{1}{2}}\left(\Sigma_{i=1}^{\infty}(\gamma_i^2 - \gamma_i^3)V_i, \Sigma_{i=1}^{\infty}\left(\tfrac{1}{2}\gamma_i^2 - \gamma_i^3 + \gamma_i^4\right)V_i\right),$$

$$(W_5, W_6) = (\Sigma_{i=1}^{\infty}\gamma_i^2 V_i^2, \Sigma_{i=1}^{\infty}\gamma_i^4 V_i^2),$$

where $\{V_i\}_{i=1}^{\infty}$ *is a sequence of normal independent* $(0, \sigma^2)$ *random variables. Then*

$$\mathbf{W}_n \to_{\mathcal{L}} \mathbf{W}.$$

PROOF. From the definition of u_{in} and the fact that the u_{in} are uncorrelated $(0, \sigma^2)$ random variables it follows that

$$W_{jn} = \Sigma_{i=1}^{n-1}\delta_{jin}u_{in} \qquad j = 1, \cdots, 4,$$

$$W_{5n} = n^{-2}\Sigma_{i=1}^{n-1}\lambda_{in}u_{in}^2,$$

$$W_{6n} = n^{-4}\Sigma_{i=1}^{n-1}\lambda_{in}^2u_{in}^2.$$

Define W_{jnk} for $j = 1, 2, \cdots, 6$ and $k < n - 1$ analogous to W_{jn} with the summations truncated at k. We have, for example,

$$W_{1nk} = \Sigma_{i=1}^{k}\delta_{1in}u_{in}.$$

Let $W_j^{(k)}$ for $j = 1, 2, \cdots, 6$ be defined by

$$\left(W_1^{(k)}, W_2^{(k)}\right) = 2^{\frac{1}{2}}\left(\Sigma_{i=1}^{k}\gamma_i V_i, \Sigma_{i=1}^{k}\gamma_i^2 V_i\right),$$

$$\left(W_3^{(k)}, W_4^{(k)}\right) = 2^{\frac{1}{2}}\left(\Sigma_{i=1}^{k}\left(\gamma_i^2 - \gamma_i^3\right)V_i, \Sigma_{i=1}^{k}\left(\tfrac{1}{2}\gamma_i^2 - \gamma_i^3 + \gamma_i^4\right)V_i\right),$$

$$\left(W_5^{(k)}, W_6^{(k)}\right) = \left(\Sigma_{i=1}^{k}\gamma_i^2 V_i^2, \Sigma_{i=1}^{k}\gamma_i^4 V_i^2\right).$$

By Lemma 3.2, for fixed k, the limiting joint distribution (as $n \to \infty$) of $(W_{1nk}, W_{2nk}, \cdots, W_{6nk})'$ is that of $\mathbf{W}^{(k)} = (W_1^{(k)}, W_2^{(k)}, \cdots, W_6^{(k)})'$. The vector $\mathbf{W}$ is well defined as a limit a.e. and $\mathbf{W}^{(k)} \to_{\mathcal{L}} \mathbf{W}$ as $k \to \infty$.

We wish to show that $\mathbf{W}_n \to_{\mathcal{L}} \mathbf{W}$ (as $n \to \infty$). Let $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \cdots, \lambda_6)'$ be an arbitrary vector with $\boldsymbol{\lambda}'\boldsymbol{\lambda} \neq 0$. Using Lemma 6.3.1 of Fuller (1976) (see also Dianada 1953) the result will be obtained by showing that

$$\operatorname{plim}_{k\to\infty}\Sigma_{j=1}^{6}\lambda_j\left(W_{jn} - W_{jnk}\right) = 0$$

uniformly in n. Let $\varepsilon > 0$ be given. For $k < n - 1$ we have

$$W_{1n} - W_{1nk} = \Sigma_{i=k+1}^{n-1}\delta_{1in}u_{in},$$

and

$$\Sigma_{i=1}^{n-1}\delta_{1in}^2 = \sigma^{-2}\operatorname{Var}\left(n^{-\frac{1}{2}}\Sigma_{t=1}^{n-1}e_t\right) = \frac{n-1}{n}.$$

Furthermore $\Sigma_{i=1}^{\infty}2\gamma_i^2 = 1$ and we may choose N_0 so that

$$\Sigma_{i=1}^{N_0}2\gamma_i^2 > 1 - \tfrac{1}{2}\varepsilon.$$

Because $\lim_{n\to\infty}\delta_{1in}^2 = 2\gamma_i^2$ it follows that, for k sufficiently large,

$$\Sigma_{i=k+1}^{n-1}\delta_{1in}^2 < \varepsilon \qquad \text{for all} \quad n > k + 1.$$

Therefore $\operatorname{plim}_{k\to\infty}(W_{1n} - W_{1nk}) = 0$ uniformly in n. Similar methods can be used to establish the analogous result for $W_{2n}, W_{3n}, \cdots, W_{6n}$. Therefore $\Sigma_{j=1}^{6}\lambda_j(W_{jn} - W_{jnk})$ converges to zero in probability uniformly in n and $\mathbf{W}_n \to_{\mathcal{L}} \mathbf{W}$. □

We now express the least squares estimators as functions of $\mathbf{W}_n$ and terms of smaller order in probability. Define

$$\mathbf{H}_3 = \begin{bmatrix} 1 & \frac{1}{2} & W_3 & W_2 \\ \frac{1}{2} & \frac{1}{3} & W_4 & W_2 - W_3 \\ W_3 & W_4 & W_6 - W_2^2 + 2W_2W_3 & \frac{1}{2}W_2^2 \\ W_2 & W_2 - W_3 & \frac{1}{2}W_2^2 & W_5 \end{bmatrix}.$$

Let $\mathbf{H}_1$ be the submatrix of $\mathbf{H}_3$ obtained by deleting the first two rows and columns and let $\mathbf{H}_2$ be the submatrix of $\mathbf{H}_3$ obtained by deleting the second row and column. Define

$$\mathbf{h}_1 = \left(W_1W_2 - W_5, \tfrac{1}{2}W_1^2 - \frac{\sigma^2}{2}\right)',$$

$$\mathbf{h}_2 = \left(W_1, W_1W_2 - W_5, \tfrac{1}{2}W_1^2 - \frac{\sigma^2}{2}\right)',$$

$$\mathbf{h}_3 = \left(W_1, W_1 - W_2, W_1W_2 - W_5, \tfrac{1}{2}W_1^2 - \frac{\sigma^2}{2}\right)'.$$

Let $\mathbf{H}_{1n}$, $\mathbf{h}_{1n}$, $\mathbf{H}_{2n}$, $\mathbf{h}_{2n}$, $\mathbf{H}_{3n}$, and $\mathbf{h}_{3n}$ be defined by analogy to $\mathbf{H}_1$, $\mathbf{h}_1$, $\mathbf{H}_2$, $\mathbf{h}_2$, $\mathbf{H}_3$, and $\mathbf{h}_3$ with W_j replaced by W_{jn}.

THEOREM 3.2. *Let the assumptions of Lemma* 3.1 *hold. Then*
(i) $[n^2(\hat{\alpha}_1 - 1), n(\hat{\beta}_1 - 1)]' \to_{\mathcal{L}} \mathbf{H}_1^{-1}\mathbf{h}$,
(ii) $[n^{\frac{1}{2}}\hat{\mu}_2, n^2(\hat{\alpha}_2 - 1), n(\hat{\beta}_2 - 1)]' \to_{\mathcal{L}} \mathbf{H}_2^{-1}\mathbf{h}_2$,
(iii) $[n^{\frac{1}{2}}\hat{\mu}_3, n^{\frac{3}{2}}\hat{\theta}_3, n^2(\hat{\alpha}_3 - 1), n(\hat{\beta}_3 - 1)]' \to_{\mathcal{L}} \mathbf{H}_3^{-1}\mathbf{h}_3$.

PROOF. By Lemma 3.1

$$\left[n^2(\hat{\alpha}_1 - 1), n(\hat{\beta}_1 - 1)\right]' = \left[\mathbf{H}_{1n} + O_p(n^{-\frac{1}{2}})\right]^{-1}\left[\mathbf{h}_{1n} + o_p(1)\right],$$

$$\left[n^{\frac{1}{2}}\hat{\mu}_2, n^2(\hat{\alpha}_2 - 1), n(\hat{\beta}_2 - 1)\right]' = \left[\mathbf{H}_{2n} + O_p(n^{-\frac{1}{2}})\right]^{-1}\left[\mathbf{h}_{2n} + o_p(1)\right],$$

$$\left[n^{\frac{1}{2}}\hat{\mu}_3, n^{\frac{3}{2}}\hat{\theta}_3, n^2(\hat{\alpha}_3 - 1), n(\hat{\beta}_3 - 1)\right]' = \left[\mathbf{H}_{3n} + O_p(n^{-\frac{1}{2}})\right]^{-1}\left[\mathbf{h}_{3n} + o_p(1)\right].$$

Because the limiting random vectors are continuous functions of $\mathbf{W}$ the result will follow if $\mathbf{H}_1$, $\mathbf{H}_2$, and $\mathbf{H}_3$ are nonsingular with probability one. That $\mathbf{H}_1$ is nonsingular with probability one follows from Theorem 2.1.

The determinant of $\mathbf{H}_2$ may be expressed as

$$\begin{aligned}\det(\mathbf{H}_2) &= \Sigma_{i,j}a_{ij}V_i^2V_j^2 + \Sigma_{i,j,k}b_{ijk}V_i^2V_jV_k \\ &\quad + \Sigma_{i,j,k,l}c_{ijkl}V_iV_jV_kV_l \\ &= \Sigma_i(a_{ii} + b_{iii} + c_{iiii})V_i^4 \\ &\quad + \text{cross product terms},\end{aligned}$$

where the coefficients a_{ij}, b_{ijk}, and c_{ijkl} depend only on the constants $\{\gamma_i\}$. Some tedious algebra will verify that $a_{11} + b_{111} + c_{1111} \neq 0$. The conditional distribution of $\det(\mathbf{H}_2)$ given $V_i = v_i$, $i = 2, 3, \cdots$ is the distribution of the random variable $\lambda V_1^4 + \mu V_1^3 + \nu V_1^2 + \omega V_1$ where λ and μ, ν, ω depend only on γ_i and on $\gamma_i, v_2, v_3, \cdots$, respectively. Since $\lambda \neq 0$ the above random variable is nonzero with probability one. Therefore, $P\{\det(\mathbf{H}_2) \neq 0 | V_2, V_3, \cdots\} = 1$ and by integration one obtains the result for $\mathbf{H}_2$. The result for $\mathbf{H}_3$ may be obtained in a similar manner. □

COROLLARY 3.1. *Let the assumptions of Theorem* 3.2 *hold. Then*

(i) $\Phi_1(2) \to_{\mathcal{L}} \frac{\sigma^{-2}}{2} \mathbf{h}_1'\mathbf{H}_1^{-1}\mathbf{h}_1$,

(ii) $\Phi_2(2) \to_{\mathcal{L}} \frac{\sigma^{-2}}{2} (\mathbf{h}_2'\mathbf{H}_2^{-1}\mathbf{h}_2 - W_1^2)$,

(iii) $\Phi_2(3) \to_{\mathcal{L}} \frac{\sigma^{-2}}{3} \mathbf{h}_2'\mathbf{H}_2^{-1}\mathbf{h}_2$,

(iv) $\Phi_3(2) \to_{\mathcal{L}} \frac{\sigma^{-2}}{2} [\mathbf{h}_3'\mathbf{H}_3^{-1}\mathbf{h}_3 - 12(\frac{1}{3} W_1^2 - W_1 W_2 + W_2^2)]$,

(v) $\Phi_3(4) \to_{\mathcal{L}} \frac{\sigma^{-2}}{4} \mathbf{h}_3'\mathbf{H}_3^{-1}\mathbf{h}_3$.

4. Simulation. Estimates of the percentiles of the Φ-statistics are presented in Table 4.1. For the finite sample sizes the Φ-statistics were computed for samples generated using the model (2.1), with $e_t \sim NID(0, 1)$ and $Y_0 = Y_{-1} = 0$.

To estimate the percentiles of the limiting distributions of the test statistics we use the method of simulation employed by Dickey (1976). Briefly the method consists of approximating the sequence $\{\gamma_i\}_{i=1}^{\infty}$ by the finite sequence $\{\tilde{\gamma}_i\}_{i=1}^{72}$, where $\tilde{\gamma}_i = \gamma_i$ for $i = 1, 2, \cdots, 14$ and the remaining $\tilde{\gamma}_i$ are chosen so that the first eight moments of the two sequences are in very close agreement. Using the Monte Carlo method sample distribution functions of the statistics are generated with the γ_i replaced by $\tilde{\gamma}_i$ in the definitions of the W_j. The percentiles were "smoothed" by fitting a regression function to the original set of estimates.

The estimated standard errors for the estimated percentiles are generally less than 0.3 percent of the table entry for the limiting distributions and are generally less than 1.2 percent of the table entry for the finite sample cases. Two hundred thousand independent sample statistics were used in constructing the percentiles for the asymptotic distributions. Details of the computations are given in Hasza (1977).

5. Higher order processes. In this section we consider the pth order process with two unit roots defined in (1.1). We shall outline the method for obtaining the asymptotic distributions of parameters and test statistics. The discussion is abbreviated and we refer the reader to the treatment of the pth order autoregressive process with one unit root given in Fuller (1976, pages 373–382).

TABLE 4.1
Empirical percentiles for test statistics.

	n	.50	.80	.90	.95	.975	.99
	25	0.96	2.05	2.89	3.78	4.66	6.01
	50	0.97	2.03	2.82	3.60	4.41	5.52
$\Phi_1(2)$	100	0.98	2.02	2.78	3.53	4.29	5.31
	250	0.98	2.01	2.76	3.49	4.22	5.20
	500	0.98	2.01	2.76	3.48	4.20	5.17
	∞	0.98	2.01	2.75	3.47	4.18	5.14
	25	2.56	4.44	5.78	7.17	8.61	10.55
	50	2.58	4.30	5.47	6.61	7.76	9.22
	100	2.58	4.24	5.33	6.36	7.38	8.65
$\Phi_2(2)$	250	2.59	4.20	5.25	6.23	7.18	8.36
	500	2.59	4.19	5.22	6.19	7.13	8.28
	∞	2.59	4.18	5.21	6.16	7.08	8.22
	25	2.07	3.34	4.28	5.22	6.23	7.59
	50	2.05	3.20	3.99	4.76	5.55	6.56
$\Phi_2(3)$	100	2.04	3.13	3.86	4.55	5.24	6.11
	250	2.03	3.09	3.79	4.44	5.08	5.88
	500	2.03	3.08	3.76	4.41	5.03	5.81
	∞	2.03	3.07	3.75	4.39	5.00	5.76
	25	4.97	7.70	9.54	11.41	13.34	15.88
	50	4.89	7.21	8.75	10.17	11.61	13.43
$\Phi_3(2)$	100	4.86	6.98	8.36	9.58	10.80	12.31
	250	4.83	6.86	8.13	9.25	10.34	11.70
	500	4.83	6.82	8.05	9.15	10.20	11.52
	∞	4.82	6.78	7.98	9.05	10.08	11.37
	25	3.13	4.59	5.63	6.66	7.74	9.25
	50	2.99	4.19	5.00	5.77	6.51	7.49
$\Phi_3(4)$	100	2.92	4.00	4.71	5.36	5.96	6.74
	250	2.89	3.90	4.55	5.14	5.68	6.38
	500	2.88	3.87	4.50	5.07	5.60	6.28
	∞	2.87	3.84	4.45	5.01	5.54	6.21

Let Δ denote the difference operator and consider the following models:

$$Y_t = \alpha_1 Y_{t-1} + \beta_1 \Delta Y_{t-1} + \Sigma_{j=1}^{p-2}\delta_{1j} \Delta^2 Y_{t-j} + e_t, \tag{5.1}$$

$$Y_t = \mu_2 + \alpha_2 Y_{t-1} + \beta_2 \Delta Y_{t-1} + \Sigma_{j=1}^{p-2}\delta_{2j} \Delta^2 Y_{t-j} + e_t, \tag{5.2}$$

$$Y_t = \mu_3 + \theta_3 t + \alpha_3 Y_{t-1} + \beta_3 \Delta Y_{t-1} + \Sigma_{j=1}^{p-2}\delta_{3j} \Delta^2 Y_{t-j} + e_t, \tag{5.3}$$

where the roots of

$$m^{p-2} - \Sigma_{j=1}^{p-2}\delta_{ij} m^{p-2-j} = 0$$

are less than one in absolute value for $i = 1, 2, 3$. Let $\xi_4' = (\alpha_1, \beta_1, \boldsymbol{\delta}_1')$, $\xi_5' = (\mu_2, \alpha_2, \beta_2, \boldsymbol{\delta}_2')$, and $\xi_6' = (\mu_3, \theta_3, \alpha_3, \beta_3, \boldsymbol{\delta}_3')$, where $\boldsymbol{\delta}_i' = (\delta_{i1}, \delta_{i2}, \cdots, \delta_{i,p-2})$. Let

$$\psi_{4t}' = \left(Y_{t-1}, \Delta Y_{t-1}, \Delta^2 Y_{t-1}, \cdots, \Delta^2 Y_{t-p+2}\right),$$

$$\psi_{5t}' = (1, \psi_{4t}') \quad \text{and} \quad \psi_{6t}' = (1, t, \psi_{4t}').$$

Then the least squares estimators are

$$\tilde{\xi}_i = (\Sigma_{t=1}^n \psi_{it}\psi_{it}')^{-1}\Sigma_{t=1}^n \psi_{it} Y_t. \tag{5.4}$$

It is noted that, for example, in model (5.1), the characteristic equation of Y_t will have two unit roots if and only if the true parameter values satisfy $\alpha_1 = \beta_1 = 1$. We define test statistics $\tilde{\Phi}_1(2)$, $\tilde{\Phi}_2(2)$, $\tilde{\Phi}_2(3)$, $\tilde{\Phi}_3(2)$, and $\tilde{\Phi}_3(4)$ (analogous to the likelihood ratio F-statistic under a fixed normal model) for testing $H_1 : (\alpha_1, \beta_1) = (1, 1)$, $H_2 : (\alpha_2, \beta_2) = (1, 1)$, $H_3 : (\mu_2, \alpha_2, \beta_2) = (0, 1, 1)$, $H_4 : (\alpha_3, \beta_3) = (1, 1)$, and $H_5 : (\mu_3, \theta_3, \alpha_3, \beta_3) = (0, 0, 1, 1)$ respectively. These will be likelihood ratio tests for the time series model when the e_t are normally distributed and $Y_{-p+1}, Y_{-p+2}, \cdots, Y_0$ are fixed.

Because Y_t has characteristic roots $m_1 = m_2 = 1$ and $|m_j| < 1$ for $j = 3, 4, \cdots, p$, $\Delta^2 Y_t$ is an autoregressive process of order $p - 2$ with characteristic roots less than one in modulus. We have

$$\begin{aligned} \Delta^2 Y_t &= \Sigma_{j=1}^{p-2}\delta_j \, \Delta^2 Y_{t-j} + e_t \qquad & t = 1, 2, \cdots \\ &= 0 & t = -p + 3, -p + 4, \cdots, 0, \end{aligned} \tag{5.5}$$

where the roots of

$$m^{p-2} - \Sigma_{j=1}^{p-2}\delta_j m^{p-2-j} = 0 \tag{5.6}$$

are $m_3, m_4, \cdots, m_p$. Let $c^{-1} = 1 - \Sigma_{j=1}^{p-2}\delta_j$ and note that $c^{-1} \neq 0$ because 1 is not a root of (5.6). We demonstrate that Y_t behaves much like a multiple c of the second order process (2.1).

LEMMA 5.1. *Define* $\{X_t\}_{t=-1}^{\infty}$ *by*

$$\begin{aligned} X_t &= X_{t-1} + \Delta X_{t-1} + e_t \qquad & t = 1, 2, \cdots \\ &= 0 & t = -1, 0. \end{aligned}$$

Then

$$\Sigma_{t=1}^n t^d Y_{t-1}^a (\Delta Y_{t-1})^b e_t^g = c^{a+b}\Sigma_{t=1}^n t^d X_{t-1}^a (\Delta X_{t-1})^b e_t^g + O_p(n^{1+3a/2+b/2+d-g/2}),$$

where a, b, d, g *are nonnegative integers with* $g \leqslant 1 \leqslant a + b + d + g \leqslant 2$. *Furthermore,*

$$\Sigma_{t=1}^n Y_{t-1}\Delta^2 Y_{t-j} = O_p(n^2) \qquad j = 1, 2, \cdots, p - 2,$$

$$\Sigma_{t=1}^n \Delta Y_{t-1}\Delta^2 Y_{t-j} = O_p(n) \qquad j = 1, 2, \cdots, p - 2.$$

PROOF. We have

$$X_t = Y_t - \Sigma_{j=1}^{p-2}\delta_j Y_{t-j} \qquad t = -1, 0, \cdots.$$

Therefore,

$$X_t = c^{-1}Y_t - \Sigma_{j=1}^{p-2}\delta_j(Y_t - Y_{t-j})$$
$$= c^{-1}Y_t - \Sigma_{j=1}^{p-2}\delta_j\Sigma_{k=1}^{j}\Delta Y_{t-k}.$$

Then, for example,

$$\Sigma_{t=1}^{n}X_{t-1}^2 = \Sigma_{t=1}^{n}\left[c^{-1}Y_{t-1} - \Sigma_{j=1}^{p-2}\delta_j\Sigma_{k=1}^{j}\Delta Y_{t-k-1}\right]^2$$
$$= c^{-2}\Sigma_{t=1}^{n}Y_{t-1}^2 + O_p(n^3)$$

because terms of the form $\Sigma_{t=1}^{n}Y_{t-1}\Delta Y_{t-k-1}$ and $\Sigma_{t=1}^{n}\Delta Y_{t-j-1}\Delta Y_{t-k-1}$ are $O_p(n^3)$ and $O_p(n^2)$ respectively. Similar techniques may be used to establish the other results. □

With Theorem 5.1 we establish the limiting behavior of the least squares estimators of the parameters of (5.1), (5.2) and (5.3).

THEOREM 5.1. *Let the assumptions and definitions of this section hold and let* $\mathbf{H}_i$ *and* $\mathbf{h}_i$ *be defined as in Section* 3. *Let* $\tilde{\boldsymbol{\delta}}_i$, $i = 1, 2, 3$, *be the* $p - 2$ *dimensional vector given by the last* $p - 2$ *elements of* $\tilde{\boldsymbol{\xi}}_i$ *of* (5.4). *Then*

(i) $[cn^2(\tilde{\alpha}_1 - 1), cn(\tilde{\beta}_1 - 1)]' \to_{\mathcal{L}} \mathbf{H}_1^{-1}\mathbf{h}_1$,

(ii) $[n^{\frac{1}{2}}\tilde{\mu}_2, cn^2(\tilde{\alpha}_2 - 1), cn(\tilde{\beta}_2 - 1)]' \to_{\mathcal{L}} \mathbf{H}_2^{-1}\mathbf{h}_2$,

(iii) $[n^{\frac{1}{2}}\tilde{\mu}_3, n^{\frac{3}{2}}\tilde{\theta}_3, cn^2(\tilde{\alpha}_3 - 1), cn(\tilde{\beta}_3 - 1)]' \to_{\mathcal{L}} \mathbf{H}_3^{-1}\mathbf{h}_3$,

(iv) $n^{\frac{1}{2}}(\tilde{\boldsymbol{\delta}}_j - \boldsymbol{\delta}) \to_{\mathcal{L}} N_{p-2}(\mathbf{0}, \sigma^2\boldsymbol{\Gamma}^{-1})$ *for* $j = 1, 2, 3$, *where the elements of* $\boldsymbol{\Gamma}$ *are given by*

$$\Gamma_{ij} = \lim_{t\to\infty} \operatorname{Cov}(\Delta^2 Y_t, \Delta^2 Y_{t+|i-j|}).$$

PROOF. We consider model (5.3). Let $\mathbf{G}_n = \Sigma_{t=1}^{n}\boldsymbol{\psi}_{6t}\boldsymbol{\psi}_{6t}'$ and let the $(p + 2) \times (p + 2)$ diagonal matrix

$$\mathbf{D}_n = \operatorname{diag}\{n^{\frac{1}{2}}, n^{\frac{3}{2}}, n^2, n, n^{\frac{1}{2}}, n^{\frac{1}{2}}, \cdots, n^{\frac{1}{2}}\}.$$

The normalized error in the least squares estimator of the parameters of model (5.3) is

$$\mathbf{D}_n(\tilde{\boldsymbol{\xi}}_6 - \boldsymbol{\xi}_6) = \mathbf{D}_n\mathbf{G}_n^{-1}\mathbf{g}_n$$
$$= \left[\mathbf{D}_n^{-1}\mathbf{G}_n\mathbf{D}_n^{-1}\right]^{-1}\mathbf{D}_n^{-1}\mathbf{g}_n$$

where $\mathbf{g}_n = \Sigma_{t=1}^{n}\boldsymbol{\psi}_{6t}e_t$.

Using the results of Lemma 5.1

$$\mathbf{D}_n^{-1}\mathbf{G}_n\mathbf{D}_n^{-1} \to_{\mathcal{L}} \begin{pmatrix} \mathbf{CH}_3\mathbf{C} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma} \end{pmatrix},$$

where $\mathbf{C} = \operatorname{diag}\{1, 1, c, c\}$. Also $\operatorname{diag}\{n^{-\frac{1}{2}}, n^{-\frac{3}{2}}, n^{-2}, n^{-1}\}(\Sigma_{t=1}^{n}e_t, \Sigma_{t=1}^{n}te_t, \Sigma_{t=1}^{n}Y_{t-1}e_t, \Sigma_{t=1}^{n}\Delta Y_{t-1}e_t)' \to_{\mathcal{L}} \mathbf{Ch}_3$. The result is then obtained by applying the results of Section 3 and the well-known limiting distributional theory for least squares estimators for processes with roots less than one in modulus. □

COROLLARY 5.1. *The limiting distributions of* $\tilde{\Phi}_1(2)$, $\tilde{\Phi}_2(2)$, $\tilde{\Phi}_2(3)$, $\tilde{\Phi}_3(2)$, *and* $\tilde{\Phi}_3(4)$ *are those given in Corollary* 3.1 *for* $\Phi_1(2)$, $\Phi_2(2)$, $\Phi_2(3)$, $\Phi_3(2)$, *and* $\Phi_3(4)$ *respectively.*

Because of Corollary 5.1, Table 4.1 may be used to test the respective hypotheses in the pth order autoregressive process.

6. Comments. In our discussion, $Y_{-p+1}, Y_{-p+2}, \cdots, Y_0$ are fixed, both under the null model and under the alternative model. The tests we have presented are likelihood ratio tests (for normal e_t) under such a model. The tests are not likelihood ratio and not the most powerful that can be constructed if, for example, the alternative model is that $(Y_{-p+1}, Y_{-p+2}, \cdots, Y_n)$ is a portion of a realization from a stationary autoregressive process. If it is desired to use these tests when the alternative model is that associated with (5.2) or (5.3) with all roots of the characteristic equation less than one in absolute value or $p - 1$ roots less than one in absolute value and one root equal to one in absolute value, tests $\Phi_2(2)$ and $\Phi_3(2)$ are recommended over $\Phi_2(3)$ and $\Phi_3(4)$ on the basis of power.

The models (3.2) and (3.3) permit testing against expanded classes of alternative models. In all cases the tabulated distributions are for the null model

$$Y_t = 2Y_{t-1} - Y_{t-2} + e_t$$

with $Y_{-1} = Y_0 = 0$. The limiting distribution does not depend upon the initial conditions (Y_{-1}, Y_0) but these values will influence the small sample distributions.

It can be shown that

$$\tilde{\Phi}_2(2) \to_{\mathcal{L}} \tfrac{1}{2}\chi_2^2$$

when $\mu_2 \neq 0$ and that

$$\tilde{\Phi}_3(2) \to_{\mathcal{L}} \tfrac{1}{2}\chi_2^2$$

when $\theta_3 \neq 0$. Therefore, if $\mu_2 \neq 0$ and $\alpha_2 = \beta_2 = 1$ the probability is greater than that tabled that the statistic $\Phi_2(2)$ will accept the hypothesis that $\alpha_2 = \beta_2 = 1$. The limiting distribution of $\Phi_3(2)$ has been obtained for the case $\mu_3 \neq 0$ and $\theta_3 = 0$. A Monte Carlo study demonstrated that if $\mu_3 \neq 0$, $\theta_3 = 0$, and $\Phi_3(2)$ is used to test the hypothesis that $\alpha_3 = \beta_3 = 1$ the hypothesis will be accepted with greater probability than that tabled.

7. Example. Box and Jenkins (1976, page 528) list 226 consecutive temperature readings on a chemical process. The readings were taken at one minute intervals. The following equations were fitted by least squares:

$$\begin{aligned} Y_t &= \underset{(0.1024)}{0.2747} + \underset{(0.0001)}{0.0001t} + \underset{(0.0044)}{0.9876Y_{t-1}} + \underset{(0.0383)}{0.8152(Y_{t-1} - Y_{t-2})} \qquad S^2 = 0.0175, \end{aligned} \tag{7.1}$$

$$Y_t = \underset{(0.1006)}{0.2766} + \underset{(0.0044)}{0.9876Y_{t-1}} + \underset{(0.0383)}{0.8151(Y_{t-1} - Y_{t-2})} \qquad S^2 = 0.0174, \tag{7.2}$$

$$Y_t - Y_{t-1} = \underset{(0.038)}{0.813(Y_{t-1} - Y_{t-2})} \qquad S^2 = 0.0180. \tag{7.3}$$

The numbers in parentheses are estimated standard errors calculated by the usual regression formulas used in most computer regression routines. From (7.1) we calculate $\Phi_3(2)$ to test the hypothesis that the time series was generated by an autoregressive process with two unit roots. The calculated value of $\Phi_3(2) = 16.68$ is compared with the tabulated value of 11.70 for $n = 250$ and the null hypothesis is rejected at the .01 level.

From (7.2) we calculate the "t-type" test statistic to test the hypothesis that $\alpha_2 = 1$, that is, that the second order autoregressive process has one unit root assuming the other root is less than one in absolute value. We have $\hat{\tau}_\mu = (0.9876 - 1)(0.0044)^{-1} = -2.82$. The percentiles of this test statistic when the process has one unit root are given in Table 8.5.2 of Fuller (1976). We have $P\{\hat{\tau}_\mu < -2.88\} \doteq 0.05$. To summarize, we conclude that the model with two unit roots is not compatible with the observed time series, while the model (7.3) with a single unit root is also suspect.

REFERENCES

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

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

CHUNG, K. L. (1974). *A Course in Probability Theory.* Academic Press, New York.

DIANANDA, P. H. (1953). Some probability limit theorems with statistical applications. *Proc. Cambridge Philos. Soc.* **49** 239–246.

DICKEY, D. A. (1976). Estimation and hypothesis testing in nonstationary time series. Ph.D. thesis, Dept. Statist. Iowa State Univ.

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

HANNAN, E. J. and HEYDE, C. C. (1972). On limit theorems for quadratic functions of discrete time series. *Ann. Math. Statist.* **43** 2058–2066.

HASZA, D. P. (1977). Estimation in nonstationary time series. Ph.D, thesis, Dept. Statist. Iowa State Univ.

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

RAO, M. M. (1961). Consistency and limit distributions of estimators of parameters in explosive stochastic difference equations. *Ann. Math. Statist.* **32** 195–218.

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

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

DEPARTMENT OF STATISTICS
KANSAS STATE UNIVERSITY
CALVIN HALL
MANHATTAN, KANSAS 66506

DEPARTMENT OF STATISTICS
IOWA STATE UNIVERSITY
AMES, IOWA 50010