The Annals of Statistics
1983, Vol. 11, No. 3, 856–871

CONSISTENCY PROPERTIES OF LEAST SQUARES ESTIMATES OF AUTOREGRESSIVE PARAMETERS IN ARMA MODELS[1]

BY GEORGE C. TIAO AND RUEY S. TSAY

University of Chicago and Carnegie-Mellon University

A unified treatment of the consistency properties of the ordinary least squares estimates in an autoregressive fitting of time series from nonstationary or stationary autoregressive moving average models is given. For a given model, the orders of autoregressions which produce consistent estimates are obtained and the limiting values, hence the biases, of the estimates of other autoregressions are investigated.

1. Introduction. This paper concerns the consistency properties of least squares estimators of autoregressive (AR) parameters in mixed stationary or nonstationary autoregressive moving average (ARMA) time series models. Specifically, we consider the following ARMA(p, d, q) model for the time series $\{Z_t\}$,

$$\Phi(B)Z_t = \theta(B)a_t \tag{1.1}$$

with $\Phi(B) = \phi(B)U(B)$ where $\Phi(B) = 1 - \Phi_1 B - \cdots - \Phi_{p+d}B^{p+d}$, $\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$, $U(B) = 1 - U_1 B - \cdots - U_d B^d$ and $\theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$ are polynomials in B, B is the backshift operator such that $BZ_t = Z_{t-1}$, and a_t is the error term. It is assumed that $\{a_t\}$ is a white noise process of i.i.d. continuous random variables with zero mean, variance σ_a^2 and finite fourth moment $E(a_t^4) = \kappa_4 + 3\sigma_a^4$. We shall require that all the zeros of $U(B)$ lie on, those of $\phi(B)$ lie outside and those of $\theta(B)$ lie on or outside the unit circle, and also that $\Phi(B)$ and $\theta(B)$ have no common factors. Thus, $U(B)$ and $\phi(B)$ are, respectively, the nonstationary and stationary autoregressive parts of the model (1.1). In recent years models of this form have been widely used in practice, primarily due to the work of Box and Jenkins (1970).

In the literature properties of the ordinary least squares (OLS) estimates of the autoregressive parameters in $\Phi(B)$ of (1.1) when $q = 0$ have been considered by a number of authors. In particular, Mann and Wald (1943) considered the estimation of AR parameters in the stationary case ($d = 0$); Dickey (1976), Fuller (1976) and Dickey and Fuller (1979) studied the case $U(B) = 1 - B$; Hasza and Fuller (1979) considered the case $U(B) = (1 - B)^2$; Kawashima (1980) investigated the situation $p > 0$ and $U(B) = (1 - B)^d$; Graupe (1980) considered the general case of $\Phi(B)$ but much of his proofs are in error; and Anderson and Taylor (1979) went further to show the strong consistency results for the case $d = 0$.

The asymptotic behavior of the OLS estimates for a pure autoregressive process ($q = 0$) with an explosive AR polynomial, i.e. some zeros of $\Phi(B)$ are lying inside the unit circle, has also been studied by several authors. For example, Rubin (1950), Anderson (1959), Rao (1961) and Stigum (1974) considered the estimation problem when at least one explosive root exists. Fuller, Hasza and Goebel (1981) discussed the case where either one nonstationary or one explosive root exists with some fixed sequences in the independent variables.

Received January 1982; revised December 1982.

[1] This research is supported in part by the Army Research Office under Grant No. DAAG29-78-G-0166, Bureau of the Census under JSA 80-10 and the ALCOA Foundation.

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

Key words and phrases. ARMA time series; autoregression; consistency; least squares; nonstationarity.

Although the OLS estimates have attracted a great deal of attention from time series researchers, to our knowledge, so far there has not been a comprehensive study of the properties of such estimates for the general ARMA models. The principal purpose of this paper is, therefore, to provide a unified treatment of the convergence properties of the OLS estimates for the ARMA models.

We first show, for the ARMA(p, d, q) process in (1.1), (a) that the OLS estimates of an AR(d) regression are consistent for the nonstationary AR coefficients in $U(B)$ and (b) that the OLS estimates of an AR$(p + d)$ fitting are inconsistent for the parameters in $\Phi(B)$ unless the underlying process follows a pure autoregressive $(q = 0)$ model. Furthermore, in the latter case the inconsistencies of the estimates are shown to be due to the effects of the moving average (MA) polynomial $\theta(B)$ on the OLS estimates of the stationary AR parameters in $\phi(B)$. In a later paper, Tsay and Tiao (1982), we then propose an iterative regression procedure which yields consistent estimates of the AR parameters in $\Phi(B)$.

For simplicity, we shall assume throughout that the initial values $(Z_{-p-d+1}, \cdots, Z_0)$ are known.

2. Preliminaries. Let the symbol $\rightarrow_P$ denote convergence in probability, $\rightarrow_L$ denote convergence in distribution, $\doteq$ be asymptotic equivalence in probability, and $\mathbf{A}'$ and $\det(\mathbf{A})$ denote, respectively, the transpose and the determinant of the matrix $\mathbf{A}$. We begin with some fundamental lemmas.

LEMMA 2.1. *Suppose $\{X_n\}$ is a sequence of random variables and C is a constant. Then $X_n = C + O_p(n^{-\delta})$, $\delta > 0$, implies that $X_n \rightarrow_P C$.*

This lemma follows directly from the definition of $O_p(n^{-\delta})$ and that of convergence in probability.

Next, we denote the magnitude of a $u \times v$ matrix $\mathbf{A} = (a_{ij})$ by

$$\mathrm{Mag}(\mathbf{A}) = \max|a_{ij}|. \tag{2.1}$$

Some useful properties of $\mathrm{Mag}(\mathbf{A})$ are summarized in the following lemma.

LEMMA 2.2. *For any two matrices $\mathbf{W}_{k\times r}$ and $\mathbf{A}_{u\times v}$*

(a) $\mathrm{Mag}(\mathbf{W}) = \mathrm{Mag}(\mathbf{W}')$
(b) $\mathrm{Mag}(\mathbf{W} + \mathbf{A}) \le \mathrm{Mag}(\mathbf{W}) + \mathrm{Mag}(\mathbf{A})$ *if $k = u$ and $r = v$*
(c) $\mathrm{Mag}(\mathbf{WA}) \le r\mathrm{Mag}(\mathbf{W})\mathrm{Mag}(\mathbf{A})$ *if $r = u$*
(d) $|\det(\mathbf{W}) - \det(\mathbf{A})| \le (r!)[\sum_{j=1}^{r} \binom{r}{j} \mathrm{Mag}(\mathbf{W})^{r-j}\mathrm{Mag}(\mathbf{W} - \mathbf{A})^j]$ *if $k = r = u = v$.*

The proof is straightforward and is left to the reader.

LEMMA 2.3. *For an* ARMA(p, d, q) *process $\{Z_t\}$ in* (1.1) *and a positive integer ℓ, let $\mathbf{Y}_t = (Z_t, Z_{t-1}, \cdots, Z_{t-\ell+1})'$. If $\{Z_t\}$ is not a purely deterministic process, then, for $n \ge 2\ell$, $\mathbf{A}_n = \sum_{\ell+1}^{n} \mathbf{Y}_{t-1}\mathbf{Y}'_{t-1}$ is a symmetric and positive definite matrix with probability* 1.

PROOF. It is clear that $\mathbf{A}_n$ is symmetric and nonnegative definite. To show positive definiteness, let $\mathbf{c} = (c_1, \cdots, c_\ell)'$ be an arbitrary vector and consider

$$\mathbf{c}'\mathbf{A}_n\mathbf{c} = \textstyle\sum_{t=1}^{n-\ell} (\sum_{j=1}^{\ell} c_j Z_{t+\ell-j})^2.$$

If $\mathbf{c}'\mathbf{A}_n\mathbf{c} = 0$, then

$$\textstyle\sum_{j=1}^{\ell} c_j Z_{t+\ell-j} = 0, \quad t = 1, \cdots, n - \ell$$

which, since $n \ge 2\ell$, is a system of linear equations of ℓ unknowns in at least ℓ equations.

In particular, consider $\sum_{j=1}^{\ell} c_j Z_{t+\ell-j} = 0$ for $t = 1, \cdots, \ell$. Since Z_t is continuous and not deterministic, we have that $\det(\mathbf{Z}) \neq 0$ with probability 1 where $\mathbf{Z}$ is a $\ell \times \ell$ matrix whose ith row is $\mathbf{Y}'_{\ell+i-1}$. This implies that $\mathbf{c} = \mathbf{0}$ and the conclusion follows. □

The implication of the above lemma is as follows. For any ARMA process $\{Z_t\}$, suppose we have n observations and consider fitting an autoregression of order ℓ,

$$Z_t = \beta_1 Z_{t-1} + \cdots + \beta_\ell Z_{t-\ell} + e_t, \quad t = \ell + 1, \cdots, n$$

where e_t is the error term. Then, the OLS estimates of $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_\ell)'$ is

$$\hat{\boldsymbol{\beta}} = \mathbf{A}_n^{-1} \sum_{\ell+1}^{n} \mathbf{Y}_{t-1} Z_t$$

and Lemma 2.3 implies that $\hat{\boldsymbol{\beta}}$ exists for any positive integer ℓ provided n is large enough. Since a linear model is usually written in the form $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, in what follows for convenience we shall sometimes call $\mathbf{A}_n$ the $\mathbf{X}'\mathbf{X}$ matrix in fitting an AR(ℓ) regression to Z_t with n observations.

Next, we discuss some order properties of Z_t. For the model (1.1), assume, for convenience, that $Z_t = 0$ and $a_t = 0$ for $t \leq 0$. We can then write Z_t as

$$Z_t = \sum_0^{t-1} \psi_i a_{t-i} \tag{2.2}$$

where ψ_i denotes the coefficient of B^i in the expression $\psi(B) = \sum_0^\infty \psi_i B^i$ and $\psi(B)$ satisfies the relation $\psi(B)\Phi(B) = \theta(B)$. The ψ_i's are usually referred to as the ψ-weights of the process. It is well known that for a stationary process, i.e. $U(B) = 1$, $\sum_0^\infty \psi_i^2 < \infty$ which in turn implies that $\sum_1^n Z_t^2 = O_p(n)$. On the other hand, if Z_t is nonstationary then $\sum_0^\infty \psi_i^2$ is no longer bounded. However, in this case, we may factor $U(B)$ into

$$U(B) = U_\alpha^m(B) U_\beta(B), \tag{2.3}$$

where

$$U_\alpha(B) = \prod_{i=1}^{I} (1 - \alpha_i B) \quad \text{and} \quad U_\beta(B) = \prod_{j=1}^{J} (1 - \beta_j B)^{m_j}$$

with $m > m_j \geq 0$ and the α_i's and β_j's are all distinct. Thus, m is the highest multiplicity of the characteristic roots of $U(B)$. Note that we can denote $m = 0$ for the case $U(B) = 1$.

LEMMA 2.4. *If Z_t follows the model* (1.1) *with $U(B)$ in*(2.3) *and $m > 0$, then $\sum_1^n Z_t^2 = O_p(n^{2m})$.*

PROOF. Since $U(B)\phi(B)\psi_k = 0$ for $k > \max\{p + d - 1, q\}$, from (2.3) and solutions of homogeneous difference equations, we have that

$$\psi_k = \sum_{i=1}^{I} \alpha_i^k \sum_{\ell=0}^{m-1} b_{i\ell} k^\ell + \sum_{j=1}^{J} \beta_j^k \sum_{\ell=0}^{m_j-1} c_{j\ell} k^\ell + R_k \tag{2.4}$$

where R_k satisfies $\phi(B)R_k = 0$, and the b's and c's are constants. Since $|\alpha_i| = |\beta_j| = 1$, $m > m_j$ and all the zeros of $\phi(B)$ lie outside the unit circle, it follows that the order of ψ_k is determined by the first term of the right hand side of (2.4). Therefore, $\psi_k = O(k^{m-1})$ and $\sum_0^n \psi_k^2 = O(n^{2m-1})$.

By (2.2),

$$\sum_1^n Z_t^2 = \mathbf{a}'\psi'(\mathbf{Z})\psi(\mathbf{Z})\mathbf{a} \tag{2.5}$$

where $\mathbf{a} = (a_1, \cdots, a_n)'$ and

$$\psi(\mathbf{Z}) = \begin{bmatrix} 1 & & & & \\ \psi_1 & 1 & & & \\ \vdots & \psi_1 & \ddots & & \\ & & & \ddots & \\ \psi_{n-1} & \cdot & \cdot & \psi_1 & 1 \end{bmatrix}_{n \times n}$$

Since the ℓth diagonal element of $\psi'(\mathbf{Z})\psi(\mathbf{Z})$ is $\sum^{n-\ell} \psi_k^2$ and each off diagonal term of it is in the form $\sum \psi_k \psi_{k+h}$ where h is a fixed integer, (2.5) can be rewritten as

$$\textstyle\sum_1^n Z_t^2 = n^{2m-1}\mathbf{a}'\mathbf{G}_n\mathbf{a}$$

where the symmetric $n \times n$ matrix $\mathbf{G}_n = O(1)$. Thus the expected value and variance of $\sum^n Z_t^2$ are, respectively,

$$E(\textstyle\sum_1^n Z_t^2) = n^{2m-1}\mathrm{tr}(\mathbf{G}_n)\sigma_a^2 = O(n^{2m})$$

and

$$\mathrm{Var}(\textstyle\sum_1^n Z_t^2) = n^{4m-2}[\kappa_4 \sum_{i=1}^n g_{ii}^2 + 2\sigma_a^4 \mathrm{tr}(\mathbf{G}_n^2)] = O(n^{4m})$$

where $\mathrm{tr}(\mathbf{A})$ denotes the trace of the matrix $\mathbf{A}$ and g_{ii} is the ith diagonal element of $\mathbf{G}_n$. Therefore, $\sum^n Z_t^2 = O_p(n^{2m})$. □

Note that the mean and variance of $\sum^n Z_t^2$ in the above lemma are

$$E(\textstyle\sum_1^n Z_t^2) = \sigma_a^2 \sum_{t=1}^n (n - t + 1)\psi_{t-1}^2$$

$$\mathrm{Var}(\textstyle\sum_1^n Z_t^2) = (\kappa_4 + 2\sigma_a^2) \sum_{t=1}^n (\sum_{\ell=0}^{t-1} \psi_\ell^2)^2 + 4\sigma_a^4 \sum_{t=1}^{n-1} \sum_{j=t+1}^n (\sum_{\ell=0}^{t-1} \psi_\ell \psi_{\ell+j-t})^2.$$

We next show a generalization of the above lemma. For $m > 0$, suppose that $U_*(B)$ is a factor of $U(B)$ such that the highest multiplicity of the roots of $U_*(B)^{-1}U(B)$ is m'. Then, $m \geq m' \geq 0$ and $X_t = U_*(B)Z_t$ will follow the model

(2.6) $$U_*^{-1}(B)U(B)\phi(B)X_t = \theta(B)a_t.$$

Lemma 2.5. *If Z_t and X_t follow respectively the model* (1.1) *and* (2.6) *with $m > 0$, then*

$$\textstyle\sum_1^n Z_t X_{t+h} = O_p(n^{m+m'})$$

where h is a fixed integer.

PROOF. Let $\psi_k(X)$ be the ψ-weights of X_t. Then

(2.7) $$\textstyle\sum^n Z_t X_{t+h} = \mathbf{a}'\psi'(\mathbf{Z})\mathbf{H}\psi(\mathbf{X})\mathbf{a}$$

where $\mathbf{a}$ and $\psi(\mathbf{Z})$ are defined in (2.5), $\psi(\mathbf{X})$ is the matrix of ψ-weights of X_t, i.e. $\psi(\mathbf{X})$ is defined as $\psi(\mathbf{Z})$ with ψ_k replaced by $\psi_k(X)$, and $\mathbf{H}$ is a $n \times n$ h-step shift matrix whose elements are either 0 or 1 depending on h. For example, if $h = 0$ then $\mathbf{H}$ is the $n \times n$ identity matrix and if $h = -1$ then the first subdiagonal elements of $\mathbf{H}$ are 1 and all the other terms are zero. From (2.7), the order in probability of $\sum^n Z_t X_{t+h}$ is determined by those of $\sum^n \psi_k \psi_{k+h}(X)$. If $m' > 0$, then, from (2.4) $\psi_k = O(k^{m-1})$ and $\psi_{k+h}(X) = O(k^{m'-1})$. So $\sum^n \psi_k \psi_{k+h}(X) = O(n^{m+m'-1})$. On the other hand, if $m' = 0$ then X_t is stationary and

(2.8) $$\psi(X)(B) = \phi(B)^{-1}\theta(B).$$

Since $\phi(B)$ has no root on the unit circle, by letting $B = 1, -1$ and $e^{i\omega}$ in (2.8) respectively, we may obtain that $\sum_0^\infty \psi_k(X) < \infty$, $\sum_0^\infty (-1)^k \psi_k(X) < \infty$ and $\sum_0^\infty \sin(k\omega)\psi_k(X) < \infty$. Next, by Lemma A.1 in the Appendix, we have that $\sum^n \psi_k \psi_{k+h}(X) = O(n^{m-1})$. Consequently, (2.7) can be rewritten as

(2.9) $$\textstyle\sum^n Z_t \mathrm{X}_{t+h} = n^{m+m'-1}\mathbf{a}'\mathbf{G}_n\mathbf{a}$$

where $\mathbf{G}_n$ denotes a $n \times n$ matrix and $\mathbf{G}_n = O(1)$. Moreover, since (2.9) is a quadratic form of $\mathbf{a}$ we can rewrite it as

$$\textstyle\sum^n Z_t X_{t+h} = n^{m+m'-1}\mathbf{a}'\mathbf{G}_n^*\mathbf{a}$$

where $\mathbf{G}_n^*$ is a symmetric $n \times n$ matrix and $\mathbf{G}_n^* = O(1)$. Adopting an argument analogous to that in the proof of Lemma 2.4, we obtain

$$\textstyle\sum_1^n Z_t X_{t+h} = O_p(n^{m+m'}). \square$$

We remark here that if both m and m' are positive, then the above lemma can be readily obtained from Lemma 2.4 by the Cauchy-Schwarz inequality. However, the above proof gives a tighter bound for the case $m' = 0$.

We now turn to discuss the orders in probability of determinants of some $\mathbf{X}'\mathbf{X}$ matrices in fitting autoregressions to Z_t. We begin with the order of $(\sum Z_t^2)^{-1}$.

LEMMA 2.6. *If Z_t follows the model* (1.1) *with $U(B)$ in* (2.3) *and $m > 0$, then* $(\sum_1^n Z_t^2)^{-1} = O_p(n^{-2m})$.

PROOF. By Lemma 2.4, $n^{-2m} \sum_1^n Z_t^2 = O_p(1)$. So, given any $\varepsilon > 0$ there exist $M_1 > 0$ and an integer $N_1 > 0$ such that for $n > N_1$

$$P(n^{-2m} \textstyle\sum_1^n Z_t^2 < M_1) \geq 1 - \varepsilon/2. \tag{2.10}$$

On the other hand, note that

$$\textstyle\sum_1^n Z_t^2 \geq n^{-1} (\sum_1^n Z_t)^2, \tag{2.11}$$

$$\textstyle\sum_1^n Z_t^2 = \sum_1^n Z_t^{*2} \tag{2.12}$$

where Z_t^* equals to either Z_t or $-Z_t$, and from (2.2)

$$\operatorname{Var}(\textstyle\sum_1^n Z_t) = \sigma_a^2 \sum_{t=1}^n (\sum_{k=0}^{t-1} \psi_k)^2. \tag{2.13}$$

From the ψ_k in (2.4), we have the following two cases:

(i) If $K_1 < n^{-m} |\sum^n \psi_k| < K_2$ for some constants $K_i > 0$, then $\lim_{n\to\infty} n^{-(2m+1)} \operatorname{Var}(\sum^n Z_t) = c_1$ where c_1 is a positive constant. In this case, by central limit theorem, $n^{-(2m+1)} (\sum^n Z_t)^2 \to_L c_1\chi_1^2$ where χ_1^2 denotes a Chi squared random variable with one degree of freedom. By (2.11), for the same $\varepsilon > 0$, there exist $M_2 > 0$ and an integer $N_2 > 0$ such that for $n > N_2$

$$P(n^{-2m} \textstyle\sum_1^n Z_t^2 > M_2) \geq 1 - \varepsilon/2. \tag{2.14}$$

Consequently, from (2.10) and (2.14), for any given $\varepsilon > 0$ there exist $M_1 > 0$, $M_2 > 0$ and an integer $N = \max\{N_1, N_2\} > 0$ such that for $n > N$

$$P(M_2 < n^{-2m} \textstyle\sum_1^n Z_t^2 < M_1) \geq 1 - \varepsilon.$$

Thus, $(\sum_1^n Z_t^2)^{-1} = O_p(n^{-2m})$.

(ii) If $K_1 < n^{-\delta} |\sum^n \psi_k| < K_2$ and $\delta < m$, then some cancellations between the coefficients of k^{m-1} exist in the ψ-weights. In this situation we can define $Z_t^* = Z_t$ or $Z_t^* = -Z_t$, according to the pattern of the ψ-weights, to nullify these cancellations and obtain $\lim_{n\to\infty} n^{-(2m+1)} \operatorname{Var}(\sum^n Z_t^*) = c_2$ where c_2 is a positive constant. Thus, we have that, from (2.11) and (2.12), $\sum^n Z_t^2 \geq n^{-1} (\sum^n Z_t^*)^2$ and that $n^{-(2m+1)}(\sum^n Z_t^*)^2 \to_L c_2\chi_1^2$. We can then adopt an argument analogous to that of part (i) to show that $(\sum_1^n Z_t^2)^{-1} = O_p(n^{-2m})$ also holds in this case. $\square$

COROLLARY 2.6. *Let $r_k = \sum^n Z_t Z_{t+k} / \sum^n Z_t^2$ be the lag k sample autocorrelation of Z_t. If Z_t follows the model* (2.3) *with $d > 0$ then $U_\alpha(B) r_k = O_p(n^{-1})$.*

PROOF. Notice that $U_\alpha(B) r_k = (\sum^n Z_t^2)^{-1} (\sum^n Z_t X_{t+k})$ where $X_t = U_\alpha(B) Z_t$ which has the highest multiplicity of nonstationary characteristic roots $m - 1$. Hence, $U_\alpha(B) r_k = O_p(n^{-1})$ follows directly from the above lemma and Lemma 2.5. $\square$

Corollary 2.6 can be regarded as a general proof of the fact that for nonstationary ARMA(p, d, q) processes the sample autocorrelations r_k's satisfy asymptotically a homogeneous difference equation, and it shows that the homogeneous difference equation is in fact determined by those nonstationary characteristic roots of $\Phi(B)$ which possess the highest multiplicity. Some related results can be found in Findley (1980) and Quinn (1980).

Next, we consider a generalization of Lemma 2.6. From (2.3), if $d > 0$, the model (1.1) can be written as

$$U_\alpha^m(B)U_\beta(B)\phi(B)Z_t = \theta(B)a_t. \tag{2.15}$$

Since all its roots are distinct and lying on the unit circle, $U_\alpha(B)$ can be factored as

$$U_\alpha(B) = (1 - B)^{\delta_1}(1 + B)^{\delta_2} \prod_{i=1}^{h} (1 - 2\cos\omega_i B + B^2) \tag{2.16}$$

where δ_1 and δ_2 are either 0 or 1, $2h + \delta_1 + \delta_2 = d_1$, $0 < \omega_i < \pi$, $\omega_i \neq \omega_j$ and d_1 is the degree of $U_\alpha(B)$. Thus, each of the factor $(1 - 2\cos\omega_i B + B^2)$ corresponds to a pair of complex roots on the unit circle.

LEMMA 2.7. *If Z_t follows the model* (2.15) *with $d > 0$; then* $\det^{-1}(\mathbf{A}_n) = O_p(n^{-2md_1})$ *where d_1 is the degree of $U_\alpha(B)$ and $\mathbf{A}_n$ is the* $\mathbf{X}'\mathbf{X}$ *matrix of an* AR(d_1) *fitting on Z_t.*

PROOF. For simplicity in presentation, we shall only discuss the special case of $U_\alpha(B)$ in (2.16) for which $\delta_1 = \delta_2 = 1$ and $h = 2$, but it will be obvious that the techniques employed can be extended to the general case. Let

$$x_{i,t} = g_i^{-1}(B)\,U_\alpha(B)Z_t, \quad i = 1, 2, 3, 5 \tag{2.17}$$

where $g_1(B) = (1 - B)$, $g_2(B) = (1 + B)$, $g_3(B) = (1 - 2\cos\omega_1 B + B^2)$ and $g_5(B) = (1 - 2\cos\omega_2 B + B^2)$. Further let $x_{4,t} = x_{3,t-1}$, $x_{6,t} = x_{5,t-1}$ and $\mathbf{x}_t = (x_{1,t}, \cdots, x_{6,t})'$. It is then clear that

$$\sum^n \mathbf{x}_{t-1}\mathbf{x}_{t-1}' = \mathbf{T}\mathbf{A}_n\mathbf{T}'$$

where $\mathbf{T}$ is a 6×6 nonsingular matrix determined by the definition of the $x_{i,t}$'s. Since the determinant of $\mathbf{T}$ is a constant for given ω_1 and ω_2,

$$\det^{-1}(\mathbf{A}_n) = C_0 \det^{-1}\left(\sum^n \mathbf{x}_{t-1}\mathbf{x}_{t-1}'\right), \tag{2.18}$$

where C_0 is a positive constant.

Notice that the transformed variates $x_{1,t}$'s follow the following models

$$\begin{cases} (1 - B)x_{1,t} = e_t \\ (1 + B)x_{2,t} = e_t \\ (1 - 2\cos\omega_1 B + B^2)y_t = e_t \text{ with } y_t = x_{3,t} \text{ or } x_{4,t+1} \\ (1 - 2\cos\omega_2 B + B^2)y_t = e_t \text{ with } y_t = x_{5,t} \text{ or } x_{6,t+1} \end{cases} \tag{2.19}$$

where $e_t = U_\alpha(B)Z_t$ which satisfies the model

$$U_\alpha^{m-1}(B)U_\beta(B)\phi(B)e_t = \theta(B)a_t.$$

Hence the highest multiplicity of the nonstationary characteristic roots of autoregressive polynomial for each $x_{i,t}$ is m which in turn implies, by Lemma 2.6, that

$$\prod_{i=1}^{6} \left(\sum^n x_{i,t-1}^2\right)^{-1} = O_p(n^{-12m}). \tag{2.20}$$

Moreover, by Corollary 2.6 and (2.19), $(\sum^n x_{3,t}^2)^{-1}(\sum^n x_{3,t}x_{4,t}) \to_P \cos\omega_1$ and $(\sum^n x_{5,t}^2)^{-1}(\sum^n x_{5,t}x_{6,t}) \to_P \cos\omega_2$. Next, by (2.19) and the ψ-weights of the $x_{i,t}$'s, it can also be shown, see e.g. Lemma A.2 in the Appendix, that

$$\sum^n x_{i,t}x_{j,t} = O_p(n^{2m-1}) \quad \text{for} \quad i \neq j,\ i, j = 1, 2, 3, 5.$$

Therefore,

$$\mathbf{V}^{-1}\left(\sum^n \mathbf{x}_{t-1}\mathbf{x}_{t-1}'\right) \to_P \mathbf{G} \tag{2.21}$$

where $\mathbf{V} = \mathrm{diag}\{\sum^n x_{1,t-1}^2, \cdots, \sum^n x_{6,t-1}^2\}$ and $\mathbf{G} = \mathrm{diag}\{1, 1, \mathbf{G}_1, \mathbf{G}_2\}$ with

$$\mathbf{G}_j = \begin{bmatrix} 1 & \cos\omega_j \\ \cos\omega_j & 1 \end{bmatrix}, \quad j = 1, 2.$$

Since $\det(\sum^n \mathbf{x}_{t-1} \mathbf{x}_{t-1}') = \det(\mathbf{V})\det(\mathbf{V}^{-1}\sum^n \mathbf{x}_{t-1} \mathbf{x}_{t-1}')$ and $\det(\mathbf{G})$ is positive, it follows from (2.18), (2.20) and (2.21) that

$$\det{}^{-1}(\mathbf{A}_n) = O_p(n^{-12m}) = O_p(n^{-2md_1}). \qquad \square$$

3. Convergence properties of the OLS estimates. In this and the next sections we discuss some convergence properties of the OLS estimates in time series. The result depends on the underlying true model of the process and the order of the fitted autoregression. For the nonstationary ARMA(p, d, q) process in (1.1), (1) the OLS estimates of an AR(d) fitting are shown to be consistent for the nonstationary AR parameters in $U(B)$, and (2) those of an AR($d + p$) regression are shown to be biased for the coefficients in $\Phi(B)$ if both p and q are positive.

Before going into detailed derivations, it may be useful to point out at the outset that (i) from the preliminaries in Section 2 it is very straightforward to establish these convergence properties when $m = 1$; (ii) algebraic complexity arises mainly because of multiplicities ($m > 1$) in the roots of $U(B)$; (iii) the methods used rely heavily on first transforming some of the regressors to linear combinations of the Z_t's having lower multiplicities, and this is motivated by the method used in Hasza and Fuller (1979) for the special case $U(B) = (1 - B)^2$.

3.1 Convergence properties of OLS estimates of transformed regressions. Rearranging the factors of $U(B)$ in (2.3), if $d > 0$, one can write the model (1.1) as

$$[\textstyle\prod_{i=1}^m U_i(B)]\phi(B)Z_t = \theta(B)a_t \tag{3.1}$$

where $U_i(B) = 1 - U_{1(i)}B - \cdots - U_{d_i(i)}B^{d_i}$ are polynomials in B of degrees d_i such that U(a) $\sum_{i=1}^m d_i = d$; (b) $\prod_{i=1}^m U_i(B) = U(B)$; (c) $U_i(B)$ is a factor of $U_{i+1}(B)$ and (d) the multiplicity of any root of $U_i(B)$ is 1. For instance, if $U(B) = (1 - B)^3(1 - \sqrt{2}B + B^2)$, then $m = 3$, $U_1(B) = U_2(B) = (1 - B)$ and $U_3(B) = (1 - B)(1 - \sqrt{2}B + B^2)$. Notice that in (3.1), (a) m is again the highest multiplicity of roots of $U(B)$; (b) $U_1(B) = U_\alpha(B)$ in (2.3) so that each root of $U_1(B)$ is a root of $U(B)$ with multiplicity m; (c) each root, if any, of $U_i^{-1}(B)\,U_{i+1}(B)$ is a root of $U(B)$ with multiplicity $m - i$. Further let

$$\begin{cases} Z_{1,t} = Z_t \quad \text{and} \\ Z_{j,t} = U_{j-1}(B)Z_{j-1,t}, \quad j = 2, \cdots, m+1. \end{cases} \tag{3.2}$$

Then, (a) $Z_{j,t}$ follows the model

$$[\textstyle\prod_{i=j}^m U_i(B)]\phi(B)Z_{j,t} = \theta(B)a_t.$$

where $\prod_{i=m+1}^m U_i(B) = 1$; (b) the highest multiplicity of the nonstationary characteristic roots of the AR polynomial for $Z_{j,t}$ is $m + 1 - j$; and (c) $Z_{m+1,t}$ is a stationary process. Now, with the above definition of the $Z_{j,t}$'s, the model (3.1) or (1.1) can be written in m alternative forms as

$$Z_t = \textstyle\sum_{j=1}^i \sum_{k=1}^{d_j} U_{k(j)}Z_{j,t-k} + Z_{i+1,t} \quad i = 1, \cdots, m. \tag{3.3}$$

Note that the ith equation of (3.3) can be regarded as a linear regression model with $Z_{i+1,t}$ the error term. For $1 \le i \le m$, let

$$\mathbf{A}_{i,n} = \textstyle\sum^n \mathbf{Y}_{i,t-1}\mathbf{Y}_{i,t-1}' \quad \text{and} \quad \mathbf{D}_{i,n} = \sum^n \mathbf{Y}_{i,t-1}Z_{i+1,t} \tag{3.4}$$

where $\mathbf{Y}_{i,t} = (\mathbf{W}_{1,t}', \cdots, \mathbf{W}_{i,t}')'$ and $\mathbf{W}_{j,t} = (Z_{j,t}, \cdots, Z_{j,t+1-d_j})'$, $j = 1, \cdots, i$. Thus, $\mathbf{A}_{i,n}$ is the $\mathbf{X}'\mathbf{X}$ matrix of the ith regression equation of (3.3) with n observations while $\mathbf{D}_{i,n}$ denotes the corresponding transformed error vector in the least squares sense. Also, let

$$h_i = 2\sum_{j=1}^{i} (m + 1 - j) d_j \quad \text{and} \quad \ell_i = \sum_{j=1}^{i} d_j, \quad i = 1, \cdots, m. \tag{3.5}$$

We now prove two important lemmas.

LEMMA 3.1. *For the* ith *regression equation of* (3.3), *if* $\det^{-1}(\mathbf{A}_{i,n}) = O_p(n^{-h_i})$, *then the* OLS *estimates* $\hat{U}_{k(j)} = U_{k(j)} + O_p(n^{j-i-1})$, $k = 1, \cdots, d_j; j = 1, \cdots, i$.

PROOF. First, since $\mathbf{A}_{i,n}$ is the $\mathbf{X}'\mathbf{X}$ matrix of an AR(ℓ_i) regression on Z_t, by Lemma 2.3, the estimates $\hat{U}_{k(j)}$ exist. Let $\boldsymbol{\beta}_i = (\mathbf{U}_1', \cdots, \mathbf{U}_i')'$ where $\mathbf{U}_j = (U_{1(j)}, \cdots, U_{d_j(j)})'$ be the coefficient vector of $U_j(B)$. Further let $\hat{\boldsymbol{\beta}}_i$ denote the vector of the OLS estimates of the ith regression of (3.3). Then

$$\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i = \mathbf{A}_{i,n}^{-1}\mathbf{D}_{i,n} = \det^{-1}(\mathbf{A}_{i,n})\mathrm{Adj}(\mathbf{A}_{i,n})\mathbf{D}_{i,n} \tag{3.6}$$

where $\mathrm{Adj}(\mathbf{A}_{i,n})$ denotes the adjoint of the matrix $\mathbf{A}_{i,n}$. By Lemma 2.5,

$$\sum^n \mathbf{W}_{j,t-1} Z_{i+1,t} = O_p(n^{2m+1-i-j}) \tag{3.7}$$

because the highest multiplicities of the nonstationary characteristic roots of $Z_{j,t}$ and $Z_{i+1,t}$ are $m + 1 - j$ and $m - i$, respectively. Now, we partition $\mathbf{A}_{i,n}$ into

$$\mathbf{A}_{i,n} = \{\mathbf{G}_{uv}\}_{i \times i}$$

where $\mathbf{G}_{uv} = \sum^n \mathbf{W}_{u,t-1}\mathbf{W}_{v,t-1}'$ with $\mathbf{W}_{j,t}$ defined in (3.4). Then

$$\mathbf{G}_{uv} = O_p(n^{2m+2-u-v}). \tag{3.8}$$

Since each $\mathbf{G}_{uv}$ is a $d_u \times d_v$ matrix, from (3.8), we have that

$$\mathrm{Cof}(\mathbf{G}_{uv}) = O_p(n^{h_i - 2m - 2 + u + v}) \tag{3.9}$$

where $\mathrm{Cof}(\mathbf{G}_{uv})$ denotes the (u, v)th block of $\mathrm{Adj}(\mathbf{A}_{i,n})$ corresponding to $\mathbf{G}_{uv}$ in $\mathbf{A}_{i,n}$. Then, the result of this lemma follows directly from (3.6), (3.7) and (3.9). □

LEMMA 3.2. *For the regressions of* (3.3), *if* $\det^{-1}(\mathbf{A}_{i,n}) = O_p(n^{-h_i})$ *then* $\det^{-1}(\mathbf{A}_{i+1,n}) = O_p(n^{-h_{i+1}})$, $i = 1, \cdots, m - 1$.

PROOF. Note that $\det(\mathbf{A}_{i+1,n}) = \det(\mathbf{A}_{i,n})\det(\hat{\mathbf{R}}_{i+1,n})$ where $\hat{\mathbf{R}}_{i+1,n} = \sum^n \hat{\mathbf{f}}_{t-1}\hat{\mathbf{f}}_{t-1}'$, $\hat{\mathbf{f}}_{t-1}$ is the estimated residual from the multivariate regression

$$\mathbf{W}_{i+1,t-1}' = \mathbf{W}_{1,t-1}'\gamma_1' + \cdots + \mathbf{W}_{i,t-1}'\gamma_i' + \mathbf{f}_{t-1}' \tag{3.10}$$

for $t = d + 1, \cdots, n$, the $\mathbf{W}_{j,t}$'s are defined in (3.4) and γ_j is a $d_{i+1} \times d_j$ matrix. Since (i) from (3.5) $h_{i+1} = h_i + 2(m - i)d_{i+1}$ and (ii) $\det^{-1}(\mathbf{A}_{i,n}) = O_p(n^{-h_i})$ by assumption, it follows that to prove Lemma 3.2 we only need to show that

$$\det^{-1}(\hat{\mathbf{R}}_{i+1,n}) = O_p(n^{-2(m-i)d_{i+1}}). \tag{3.11}$$

The main difficulty in proving (3.11) is to show that after proper normalization $\det(\hat{\mathbf{R}}_{i+1,n})$ is bounded away from zero in probability. Before jumping into the details, we shall briefly sketch the basic ideas of our proof. Consider the multivariate regression (3.10). Since, for any fixed integer k, $\sum^n \mathbf{W}_{j,t-1}Z_{i+1,t-k}$ and $\sum^n \mathbf{W}_{j,t-1}Z_{i+1,t}$ share the same order in probability, we have, by Lemma 3.1, that the OLS estimates

$$\hat{\gamma}_j = O_p(n^{j-i-1}), \quad j = 1, \cdots, i. \tag{3.12}$$

Also, $\det^{-1}(\sum^n \mathbf{W}_{i+1,t-1}\mathbf{W}_{i+1,t-1}') = O_p(n^{-2(m-i)d_{i+1}})$ by Lemma 2.7. From (3.12), one might attempt to argue by applying Slutsky's Theorem that $\hat{\mathbf{R}}_{i+1,n}$ and $\sum^n \mathbf{W}_{i+1,t-1}\mathbf{W}_{i+1,t-1}'$ are asymptotically equivalent statistics and hence (3.11) would follow. Unfortunately, since $\hat{\gamma}_i = O_p(n^{-1})$ and $\hat{\mathbf{R}}_{i+1,n}$ is obtained by summing $\hat{\mathbf{f}}_{t-1}\hat{\mathbf{f}}_{t-1}'$ over t from $d + 1$ to n, this asymptotic equivalence argument is not valid. However, from (3.12), we see that if $t \ll n$ then

$$\hat{\mathbf{f}}_{t-1} \approx \mathbf{W}_{i+1,t-1}.$$

Hence,

$$\mathbf{S}_{i+1,n_1} \approx \hat{\mathbf{R}}_{i+1,n_1} \quad \text{if} \quad n_1 \ll n \tag{3.13}$$

where $\mathbf{S}_{i+1,n_1} = \sum^{n_1} \mathbf{W}_{i+1,t-1}\mathbf{W}'_{i+1,t-1}$ and $\hat{\mathbf{R}}_{i+1,n_1} = \sum^{n_1} \hat{\mathbf{f}}_{t-1}\hat{\mathbf{f}}'_{t-1}$. By Lemma 2.7,

$$\det{}^{-1}(\mathbf{S}_{i+1,n_1}) = O_p(n^{-2(m-i)d_{i+1}}). \tag{3.14}$$

It follows from (3.13) that after proper normalization the $\det(\hat{\mathbf{R}}_{i+1,n_1})$ is bounded away from zero. Next, note that the asymptotic behaviors of $\hat{\mathbf{R}}_{i+1,n_1}$ and $\hat{\mathbf{R}}_{i+1,n}$ can be related through the relationship between n_1 and n. It is then clear that in order to establish (3.11) n_1 must be a properly chosen "fraction" of n. The basic idea of our proof below is to show the procedure of choosing n_1.

PROOF OF (3.11). For simplicity in presentation, we only demonstrate the proof of (3.11) for the special case $i = 1$, but it will be clear (see Remarks below) that the techniques employed can be readily extended to the general case $1 \le i \le m-1$. Now, for $i = 1$, (3.11) becomes

$$\det{}^{-1}(\hat{\mathbf{R}}_{2,n}) = \det{}^{-1}(\textstyle\sum^{n} \hat{\mathbf{f}}_{t-1}\hat{\mathbf{f}}'_{t-1}) = O_p(n^{-2(m-1)d_2}) \tag{3.15}$$

where

$$\hat{\mathbf{f}}'_{t-1} = \mathbf{W}'_{2,t-1} - \mathbf{W}'_{1,t-1}\hat{\gamma}'_1(n), \tag{3.16}$$

$\hat{\gamma}'_1(n) = \mathbf{A}^{-1}_{1,n}\mathbf{C}_n$ is a $d_1 \times d_2$ matrix, and $\mathbf{C}_n = \sum^n \mathbf{W}_{1,t-1}\mathbf{W}'_{2,t-1}$. From (3.14), $\det^{-1}(\mathbf{S}_{2,n_1}) = O_p(n_1^{-2(m-1)d_2})$ where $\mathbf{S}_{2,n_1}$ is a $d_2 \times d_2$ matrix. Therefore, given any $\varepsilon > 0$ there exist $M_\varepsilon > 0$ and an integer $N_\varepsilon > 0$ such that for $n_1 > N_\varepsilon$

$$P(n_1^{-2(m-1)d_2}\det(\mathbf{S}_{2,n_1}) > M_\varepsilon) \ge 1 - \varepsilon/2. \tag{3.17}$$

Now, by (3.12) and Lemma 2.5, we get

$$\mathrm{Mag}(\hat{\gamma}_1(n_1)) = O_p(n_1^{-1}), \quad \mathrm{Mag}(\mathbf{A}_{1,n_1}) = O_p(n_1^{2m}),$$
$$\mathrm{Mag}(\mathbf{S}_{2,n_1}) = O_p(n_1^{2(m-1)}) \quad \text{and} \quad \mathrm{Mag}(\mathbf{C}_{n_1}) = O_p(n_1^{2m-1})$$

where the definition and some properties of $\mathrm{Mag}(\cdot)$ are given in (2.1) and Lemma 2.2. Hence, given the same $\varepsilon > 0$ there exist $M_0 > 0$, $M_1 > 0$ and an integer $N_0 > 0$ such that for $n_1 > N_0$

$$P(\mathrm{Mag}(n_1^{-2(m-1)}\mathbf{S}_{2,n_1}) < M_0) \ge 1 - \varepsilon/4(d_2 - 1) \tag{3.18}$$

and

$$\begin{cases} \text{(i)}\ P(\mathrm{Mag}(\hat{\gamma}_1(n_1)) < M_1 n_1^{-1}) \ge 1 - \varepsilon/16 \\ \text{(ii)}\ P(n_1^{-2(m-1)}\mathrm{Mag}(\mathbf{A}_{1,n_1}) < M_1 n_1^2) \ge 1 - \varepsilon/16 \\ \text{(iii)}\ P(n_1^{-2(m-1)}\mathrm{Mag}(\mathbf{C}_{n_1}) < M_1 n_1) \ge 1 - \varepsilon/16. \end{cases} \tag{3.19}$$

Since M_0, M_1 and M_ε are fixed (after ε is given), the quantity

$$\delta(k) = 2d_1M_1^2/(k-1) + d_1^2M_1^3/(k-1)^2 \tag{3.20}$$

is a positive function of the integer variable k, $k > 1$, and is strictly monotone decreasing as k increases. So, there exists an integer $k_\varepsilon > 1$ such that

$$M_2 = \textstyle\sum_{j=1}^{d_2} \binom{d_2}{j} M_0^{d_2-j}\delta(k_\varepsilon)^j < M_\varepsilon/2(d_2!). \tag{3.21}$$

Let $N_1 = N_0 k_\varepsilon$, then for $n > N_1$ we have that $n_1 = [n/k_\varepsilon] > N_0$, where $[x]$ denotes the least integer greater than or equal to x, and, therefore, (3.18), (3.19) and the following property hold,

$$P(\mathrm{Mag}(\hat{\gamma}_1(n)) < M_1 n^{-1}) \geq 1 - \varepsilon/16. \tag{3.22}$$

By (3.16),

$$\hat{\mathbf{R}}_{2,n_1} = \sum^{n_1} \hat{\mathbf{f}}_{t-1}\hat{\mathbf{f}}'_{t-1} = \mathbf{S}_{2,n_1} - \hat{\gamma}_1(n)\mathbf{C}_{n_1} - \mathbf{C}'_{n_1}\hat{\gamma}_1(n) + \hat{\gamma}_1(n)\mathbf{A}_{1,n_1}\hat{\gamma}'_1(n). \tag{3.23}$$

Hence, by (3.19), (3.22) and Lemma 2.2,

$$P(n_1^{-2(m-1)}\mathrm{Mag}(\mathbf{S}_{2,n_1} - \hat{\mathbf{R}}_{2,n_1}) < 2d_1M_1^2 n_1/n + d_1^2 M_1^3 n_1^2/n^2) \geq 1 - \varepsilon/4. \tag{3.24}$$

From (3.20) and since $n_1 = [n/k_\varepsilon]$, (3.24) implies that

$$P(n_1^{-2(m-1)}\mathrm{Mag}(\mathbf{S}_{2,n_1} - \hat{\mathbf{R}}_{2,n_1}) < \delta(k_\varepsilon)) \geq 1 - \varepsilon/4. \tag{3.25}$$

Next, by (3.25) and Lemma 2.2(d),

$$P(n_1^{-2(m-1)d_2}|\det(\mathbf{S}_{2,n_1}) - \det(\hat{\mathbf{R}}_{2,n_1})| \leq (d_2!)h(\mathbf{S}_{2,n_1})) \geq 1 - \varepsilon/4 \tag{3.26}$$

where $h(\mathbf{S}_{2,n_1}) = \sum_{j=1}^{d_2} \binom{d_2}{j}\mathrm{Mag}(n_1^{-2(m-1)}\mathbf{S}_{2,n_1})^{d_2-j}\delta(k_\varepsilon)^j$. By (3.18),

$$P(h(\mathbf{S}_{2,n_1}) \leq M_2) \geq 1 - \varepsilon/4 \tag{3.27}$$

where M_2 is defined in (3.21). Finally, by (3.26), (3.27) and (3.21),

$$P(n_1^{-2(m-1)d_2}|\det(\mathbf{S}_{2,n_1}) - \det(\hat{\mathbf{R}}_{2,n_1})| \leq M_\varepsilon/2) \geq 1 - \varepsilon/2$$

which in turn implies that

$$P(n_1^{-2(m-1)d_2}\det(\hat{\mathbf{R}}_{2,n_1}) \geq n_1^{-2(m-1)d_2}\det(\mathbf{S}_{2,n_1}) - M_\varepsilon/2) \geq 1 - \varepsilon/2. \tag{3.28}$$

Consequently, from (3.17) and (3.28), for any given $\varepsilon > 0$ there exist $M_\varepsilon > 0$, $k_\varepsilon > 1$ and an integer $N = \max\{N_0 k_\varepsilon, N_\varepsilon k_\varepsilon\} > 0$ such that for $n > N$ we have $n_1 = [n/k_\varepsilon] > \max\{N_0, N\}$ and

$$P(n_1^{-2(m-1)d_2}\det(\hat{\mathbf{R}}_{2,n_1}) \geq M_\varepsilon/2) \geq 1 - \varepsilon.$$

Thus,

$$P(\det(\hat{\mathbf{R}}_{2,n_1}) \geq n^{2(m-1)d_2}M^*) \geq 1 - \varepsilon \tag{3.29}$$

where $M^* = M_\varepsilon/2k_\varepsilon^{2(m-1)d_2}$. Since $\hat{\mathbf{R}}_{2,n} - \hat{\mathbf{R}}_{2,n_1}$ is nonnegative definite, $\det(\hat{\mathbf{R}}_{2,n}) \geq \det(\hat{\mathbf{R}}_{2,n_1})$ and (3.29) implies that

$$P(\det(\hat{\mathbf{R}}_{2,n}) \geq M^* n^{2(m-1)d_2}) \geq 1 - \varepsilon.$$

Therefore, for any given $\varepsilon > 0$ there exist $M^* \geq 0$ and an integer $N > 0$ such that for $n > N$

$$P(\det(\hat{\mathbf{R}}_{2,n}) \geq M^* n^{2(m-1)d_2}) \geq 1 - \varepsilon.$$

On the other hand, we also have that

$$\det(\hat{\mathbf{R}}_{2,n}) \leq \det(\mathbf{S}_{2,n}) = O_p(n^{2(m-1)d_2}).$$

It is then clear that $\det^{-1}(\hat{\mathbf{R}}_{2,n}) = O_p(n^{-2(m-1)d_2})$. □

REMARKS. Consider the function $\delta(k)$ in (3.20) and the equations in (3.19) and (3.22). From the above proof, it is clear (a) that all of them are motivated by equation (3.23), and (b) that, apart from changing the subscript 2 to $i + 1$, those are the only places that have to be modified to prove the result for the general case $1 \leq i \leq m - 1$. Further, equation

(3.23) can be derived from (3.10) for the general i. Therefore, the above proof can be readily extended to cover the general case.

We can now establish the following theorem.

THEOREM 3.1. *Suppose that Z_t follows the model* (3.1) *with $m > 0$. Then, for any given i, $1 \leq i \leq m$, the* OLS *estimates of the* ith *regression of* (3.3) *are consistent. Specifically, $\hat{U}_{k(j)} = U_{k(j)} + O_p(n^{j-i-1})$, $k = 1, \cdots, d_j; j = 1, \ldots, i$.*

PROOF. Since $Z_{1,t} = Z_t$, by Lemma 2.7, $\det^{-1}(\mathbf{A}_{1,n}) = O_p(n^{-h_1})$. Next, by repeated application of Lemma 3.2, we have that $\det^{-1}(\mathbf{A}_{i,n}) = O_p(n^{-h_i})$. This theorem then follows from Lemma 3.1. □

3.2 Convergence properties of the OLS estimates for some autoregressions. For the model (1.1) or (3.1) with $d > 0$, we have discussed in Section 3.1 the convergence properties of OLS estimates of the transformed regressions in (3.3). Now, for $i = 1, \cdots, m$, let

$$\ell_i = \textstyle\sum_1^i d_j \quad \text{and} \quad \alpha_i(B) = 1 - \alpha_{1(i)}B - \cdots - \alpha_{\ell_i(i)}B^{\ell_i} = \prod_{j=1}^i U_j(B).$$

Consider the autoregression

$$Z_t = \beta_1 Z_{t-1} + \cdots + \beta_\ell Z_{t-\ell} + e_t \tag{3.30}$$

where e_t is the error term and $\ell = \ell_i$ for some i, $1 \leq i \leq m$.

THEOREM 3.2. *For any nonstationary* ARMA(p, d, q) *process Z_t and any given i, $1 \leq i \leq m$, the* OLS *estimates of the* AR(ℓ_i) *regression of* (3.30) *are consistent for the coefficients of $\alpha_i(B)$. Specifically,*

$$\hat{\beta}_j = \alpha_{j(i)} + O_p(n^{-1}), \quad j = 1, \cdots, \ell_i.$$

PROOF. For any given i, using the $Z_{j,t}$'s of (3.2) one can linearly transform the regressors of (3.30) to obtain a new regression equation

$$Z_t = \textstyle\sum_{j=1}^i \sum_{k=1}^{d_j} \beta_{k(j)} Z_{j,t-k} + e_t \tag{3.31}$$

where the $\beta_{k(j)}$'s are the new regression coefficients. From (3.30) and (3.31), the OLS estimates of both regressions are linearly related by

$$\textstyle\sum_{j=1}^{\ell} \hat{\beta}_j B^j = \sum_{v=1}^i \sum_{k=1}^{d_v} [\prod_{h=1}^{v-1} U_h(B)] \hat{\beta}_{k(v)} B^k \tag{3.32}$$

where $\prod_{h=1}^0 U_h(B) = 1$. Next, from Theorem 3.1,

$$\hat{\beta}_{k(v)} = U_{k(v)} + O_p(n^{v-i-1}), \quad k = 1, \cdots, d_v; \; v = 1, \cdots, i.$$

In particular, $\hat{\beta}_{k(i)} = U_{k(i)} + O_p(n^{-1})$, $k = 1, \cdots, d_i$. Hence, by (3.31) and the definition of $\alpha_i(B)$,

$$\textstyle\sum_{j=1}^{\ell} \hat{\beta}_j B^j = \sum_{v=1}^i \sum_{k=1}^{d_v} [\prod_{h=1}^{v-1} U_h(B)] U_{k(v)} B^k + O_p(n^{-1}) = \sum_{j=1}^{\ell} \alpha_j B^j + O_p(n^{-1}).$$

Now, if $i = m$ then $\ell_m = d$ and $\alpha_m(B) = U(B)$. Thus, as one of its special cases, Theorem 3.2 shows that for any nonstationary ARMA(p, d, q) process the OLS estimates of the AR(d) regression are consistent for the nonstationary AR coefficients in $U(B)$.

4. Estimates of AR$(d + \ell)$ regressions, $\ell > 0$. In this section, we shall consider for the model (1.1) with $d \geq 0$ the autoregression

$$Z_t = \beta_1 Z_{t-1} + \cdots + \beta_{d+\ell} Z_{t-d-\ell} + e_t \tag{4.1}$$

where $\ell > 0$ and e_t is again the error term.

If Z_t is stationary, i.e. $d = 0$, then it is easily shown that the OLS estimates of an AR(p) fitting are inconsistent for the true AR coefficients in $\phi(B)$ unless $q = 0$, see e.g. Tiao and Box (1981). Hence, in what follows we shall concentrate on the nonstationary case. For $\ell > 0$, from (3.2) we may again linearly transform the regressors of (4.1) to obtain a new regression equation

$$Z_t = \textstyle\sum_{j=1}^{m} \sum_{k=1}^{d_j} \beta_{k(j)} Z_{j,t-k} + \sum_{i=1}^{\ell} \eta_i Z_{m+1,t-i} + e_t. \tag{4.2}$$

In (4.2), $Z_{m+1,t}$ is a stationary process following the ARMA(p, o, q) model

$$\phi(B) Z_{m+1,t} = \theta(B) a_t.$$

From (4.1) and (4.2), the OLS estimates from both equations are linearly related by

$$\textstyle\sum_{j=1}^{d+\ell} \hat{\beta}_j B^j = \sum_{v=1}^{m} \sum_{k=1}^{d_v} [\prod_{h=1}^{v-1} U_h(B)]\, \hat{\beta}_{k(v)} B^k + U(B) \sum_{i=1}^{\ell} \hat{\eta}_i B^i. \tag{4.3}$$

Consider the two regression equations

$$Z_t = \textstyle\sum_{j=1}^{m} \sum_{k=1}^{d_j} H_{k(j)} Z_{j,t-k} + f_t \tag{4.4}$$

and

$$Z_{m+1,t} = \textstyle\sum_{i=1}^{\ell} \gamma_i Z_{m+1,t-i} + \varepsilon_t \tag{4.5}$$

and f_t and ε_t are the error terms. We now establish the following lemma.

LEMMA 4.1. *In the regression* (4.2), *the* OLS *estimates of the* η_i's *and the* $\beta_{k(j)}$'s *are, respectively, asymptotically equivalent to the* OLS *estimates of the* γ_i's *in* (4.5) *and the* $H_{k(j)}$'s *in* (4.4). *That is*

$$\hat{\eta}_i = \hat{\gamma}_i + O_p(n^{-1}), \quad i = 1, \cdots, \ell \tag{4.6}$$

$$\hat{\beta}_{k(j)} = \hat{H}_{k(j)} + O_p(n^{-1}), \quad j = 1, \cdots, m;\ k = 1, \cdots, d_j. \tag{4.7}$$

PROOF. From (4.2) the OLS estimate $\hat{\boldsymbol{\eta}}$ of $\boldsymbol{\eta} = (\eta_1, \cdots, \eta_\ell)'$ can be written as

$$\hat{\boldsymbol{\eta}} = (\textstyle\sum^n \hat{\mathbf{g}}_{t-1} \hat{\mathbf{g}}'_{t-1})^{-1} (\sum^n \hat{\mathbf{g}}_{t-1} \hat{f}_t) \tag{4.8}$$

where $\hat{f}_t$ *and* $\hat{\mathbf{g}}_t$ are, respectively, the estimated residuals from (4.4) and the following multivariate regression,

$$\mathbf{W}'_{m+1,t-1} = \mathbf{Y}'_{m,t-1} \mathbf{G}' + \mathbf{g}'_{t-1} \tag{4.9}$$

where $\mathbf{Y}_{m,t}$ is defined in (3.4), $\mathbf{W}_{m+1,t} = (Z_{m+1,t}, \cdots, Z_{m+1,t+1-\ell})'$ and $\mathbf{G}'$ is a $d \times \ell$ matrix of coefficients. Let $\mathbf{G}_{j\cdot}$ denote the jth row of the matrix $\mathbf{G}$. Then, by Theorem 3.1 and Lemma 3.1 with $i = m$, the OLS estimate $\hat{\mathbf{G}}_{j\cdot}$ of (4.9) has the property that

$$\hat{\mathbf{G}}_{j\cdot} = O_p(\boldsymbol{\Omega}), \quad j = 1, \cdots, \ell, \tag{4.10}$$

where $\boldsymbol{\Omega} = (\boldsymbol{\Omega}'_1, \cdots, \boldsymbol{\Omega}'_m)'$ with $\boldsymbol{\Omega}_v = n^{v-m-1} \mathbf{1}_{d_v}$ and $\mathbf{1}'_{d_v} = (1, \cdots, 1)_{1 \times d_v}$. Note that since $Z_{m+1,t}$ is a stationary process it follows that $(\sum^n Z_{m+1,t})^{-1} = O_p(n^{-1})$. So, from (4.10), Lemma 2.5 and the definition of $Z_{j,t}$'s we have $\hat{\mathbf{G}}'_{i\cdot} \mathbf{A}_{m,n} \hat{\mathbf{G}}_{j\cdot} = O_p(1)$, $i, j = 1, \cdots, \ell$. Hence, from (4.9),

$$\textstyle\sum^n \hat{\mathbf{g}}_{t-1} \hat{\mathbf{g}}'_{t-1} = \sum^n \mathbf{W}_{m+1,t-1} \mathbf{W}'_{m+1,t-1} + O_p(1). \tag{4.11}$$

Next, from (4.4) and (4.9),

$$\textstyle\sum^n \hat{\mathbf{g}}_{t-1} \hat{f}_t = \sum^n [\mathbf{W}_{m-1,t-1} - \hat{\mathbf{G}} \mathbf{Y}_{m,t-1}] \hat{f}_t = \sum^n \mathbf{W}_{m+1,t-1} \hat{f}_t \tag{4.12}$$

because $\sum^n \mathbf{Y}_{m,t-1} \hat{f}_t = \mathbf{0}$. Moreover,

$$\begin{aligned} \textstyle\sum^n \mathbf{W}_{m+1,t-1} \hat{f}_t &= \textstyle\sum^n \mathbf{W}_{m+1,t-1} [Z_t - \sum_{j=1}^{m} \sum_{k=1}^{d_j} \hat{H}_{k(j)} Z_{j,t-k}] \\ &= \textstyle\sum^n \mathbf{W}_{m+1,t-1} Z_{m+1,t} + \sum_{j=1}^{m} \sum_{k=1}^{d_j} (U_{k(j)} - \hat{H}_{k(j)}) \sum^n \mathbf{W}_{m+1,t-1} Z_{j,t-k}. \end{aligned} \tag{4.13}$$

By Lemma 2.5 and Theorem 3.1, respectively, we have that

$$\sum^n \mathbf{W}_{m+1,t-1} Z_{j,t-k} = O_p(n^{m-j+1})$$

and

$$U_{k(j)} - \hat{H}_{k(j)} = O_p(n^{j-m-1}).$$

Therefore, by (4.12) and (4.13),

$$\sum^n \hat{\mathbf{g}}_{t-1} \hat{f}_t = \sum^n \mathbf{W}_{m+1,t-1} Z_{m+1,t} + O_p(1). \tag{4.14}$$

Finally, by (4.8), (4.11) and (4.14),

$$\hat{\boldsymbol{\eta}} = \left(\sum^n \mathbf{W}_{m+1,t-1} \mathbf{W}'_{m+1,t-1}\right)^{-1} \left(\sum^n \mathbf{W}_{m+1,t-1} Z_{m+1,t}\right) + O_p(n^{-1}) \tag{4.15}$$

which establishes (4.6). On the other hand, consider the first d normal equations of (4.2). It is easily seen that

$$\hat{\boldsymbol{\beta}} = \mathbf{A}_{m,n}^{-1} \sum^n \mathbf{Y}_{m,t-1} Z_t - \sum_{i=1}^{\ell} \hat{\eta}_i [\mathbf{A}_{m,n}^{-1} \sum^n \mathbf{Y}_{m,t-1} Z_{m+1,t-i}] = \hat{\mathbf{H}} + O_p(n^{-1}) \tag{4.16}$$

where $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{H}}$ are, respectively, the vectors of $\hat{\beta}_{k(j)}$'s of (4.2) and the $\hat{\mathbf{H}}_{k(j)}$'s of (4.4), and (4.7) is proved. □

Now, in terms of the ordinary AR$(d + \ell)$ regression in (4.1) we have, by (4.3), (4.6) and (4.7), that

$$\sum_{j=1}^{d+\ell} \hat{\beta}_j B^j \doteq \sum_{v=1}^{m} \sum_{k=1}^{d_v} \left[\prod_{h=1}^{v-1} U_h(B)\right] \hat{H}_{k(v)} B^k + U(B) \sum_{i=1}^{\ell} \hat{\gamma}_i B^i. \tag{4.17}$$

Furthermore, consider the regression (4.4). From the linear transformation we get that

$$\sum_{v=1}^{m} \sum_{k=1}^{d_v} \left[\prod_{h=1}^{v-1} U_k(B)\right] \hat{H}_{k(v)} B^k = \sum_{j=1}^{d} \hat{\lambda}_j B^j$$

where $\hat{\lambda}_j$ are the OLS estimates of the regression

$$Z_t = \sum_{j=1}^{d} \lambda_j Z_{t-j} + f_t. \tag{4.18}$$

Therefore, by letting $\hat{\beta}(B) = 1 - \sum_{j=1}^{d+\ell} \hat{\beta}_j B^j$, $\lambda(B) = 1 - \sum_{j=1}^{d} \hat{\lambda}_j B^j$ and $\hat{\gamma}(B) = 1 - \sum_{i=1}^{\ell} \hat{\gamma}_i B^i$, we have that

$$\hat{\beta}(B) \doteq \hat{\lambda}(B) - U(B) + U(B)\hat{\gamma}(B). \tag{4.19}$$

In other words, the convergence properties of the OLS estimates of (4.1) are asymptotically determined by the OLS estimates of the regressions (4.5) and (4.18). By Theorem 3.2,

$$\hat{\lambda}_j = U_j + O_p(n^{-1}). \tag{4.20}$$

But, from the results of the stationary case,

$$\begin{cases} \text{(i)} \ \hat{\gamma}_i = \phi_i + O_p(n^{-1/2}) \text{ only if } \ell \geq p \text{ and } q = 0 \\ \text{(ii)} \ \hat{\gamma}_i \text{ will be biased otherwise} \end{cases} \tag{4.21}$$

where it is understood that $\phi_i = 0$ for $p < i \leq \ell$. Thus, by (4.19)–(4.21), we obtain the following results:

(a) If $\ell \geq p$ and $q = 0$ then

$$\hat{\beta}(B) = \Phi(B) + O_p(n^{-1/2}) \tag{4.22}$$

so that the $\hat{\beta}_j$'s are consistent in this case.

(b) If $\ell < p$ or $q \neq 0$ then

$$\hat{\beta}(B) \doteq U(B)\hat{\gamma}(B) \tag{4.23}$$

so that $\hat{\beta}(B)$ is biased because $\hat{\gamma}(B)$ in inconsistent. It is also clear that the biases actually come from the stationary AR estimates $\hat{\gamma}(B)$ only.

(c) Finally, notice that from (4.19) and (4.20) $\hat{\beta}(B)$ can also be written as

$$\hat{\beta}(B) \doteq \hat{\lambda}(B)\hat{\gamma}(B). \tag{4.24}$$

The implications of (4.22)–(4.24) can be summarized in the following theorem.

THEOREM 4.1. *For a nonstationary* ARMA(p, d, q) *process* Z_t *in* (1.1) *where* $d > 0$, *the* OLS *estimates of an* AR$(d + \ell)$ *regression,* $\ell > 0$, *can be asymptotically obtained from the* OLS *estimates of an* AR(d) *fitting of* Z_t *and those of an* AR(ℓ) *fitting on* W_t *where* $W_t = U(B)Z_t$ *is a stationary* ARMA(p, o, q) *process. The* OLS *estimates of the* AR$(d + \ell)$ *regression on* Z_t *are inconsistent unless* $\ell \geq p$ *and* $q = 0$; *and the biases of the estimates arise from the biases of* OLS *estimates corresponding to the* AR(ℓ) *fitting on* W_t.

5. Conclusion. In conclusion, it seems worthwhile to point out some practical implication of the main results of this paper. Theorems 3.2 and 4.1 suggest that the nonstationary factor $U(B)$ of the autoregressive polynomial $\Phi(B)$ in (1.1) can be fairly precisely determined by successively fitting autoregressions of increasing orders and calculating the zeros of the fitted autoregressive polynomials. Once the order of the fitted polynomial exceeds d, all the zeros correspond to $U(B)$ should remain stable. Thus, the stability of the estimated zeros near the unit circle in successive fittings will serve to signify the order of $U(B)$. This provides a useful alternative to the current practice of differencing the series on the basis of a slowly decaying sample autocorrelation function which only covers the situation $U(B) = (1 - B^s)^r$ and also can often be misleading. Our results on the inconsistency in estimating $\phi(B)$ when $\theta(B)$ is present can in fact be resolved through a process of iterated regressions, leading to the development of what we call "extended sample autocorrelation function" which can be used to help specify the order of the general ARMA(p, d, q) model. Details are given in Tiao and Tsay (1981) and Tsay and Tiao (1982).

APPENDIX

LEMMA A.1. *Let* $\{c_j\}$ *be a sequence of real numbers such that the partial sum satisfies* $\sum_1^n c_j = O(1)$. *Then* $\sum_1^n j^m c_j = O(n^m)$ *for any finite nonnegative integer* m.

PROOF. Let $C_k = \sum_1^k c_j$ and, for given n, let $b_j = (j/n)^m$. Then

$$n^{-m} \sum_1^n j^m c_j = \sum_1^n c_j b_j = C_n b_{n+1} - \sum_{k=1}^n C_k(b_{k+1} - b_k).$$

Since $\sum_1^n c_j = O(1)$, $|C_k| < M$ for all k where M is a positive constant. Hence

$$|n^{-m} \sum_1^n j^m c_j| \leq M b_{n+1} + M(b_{n+1} - b_1) = M(2b_{n+1} - b_1) < 2Mb_{n+1} < 2^{m+1}M.$$

Thus, $\sum_1^n j^m c_j = O(n^m)$. □

COROLLARY A.1. (a) $\sum_1^n j^m \sin(j\omega_1 + y) = O(n^m)$,
(b) $\sum_1^n j^m \sin(j\omega_1 + y) \sin(j\omega_2 + x) = O(n^m)$ *and*
(c) $\sum_1^n j^m (-1)^j \sin(j\omega_1 + y) = O(n^m)$

where $m \geq 0$, $0 < \omega_i < \pi$, $\omega_1 \neq \omega_2$ *and* x *and* y *are constants.*

Now, suppose that Z_t follows the model (2.15) with $m > 0$ and $U_\alpha(B) = (1 - B)(1 + B)(1 - 2\cos\omega_1 B + B^2)(1 - 2\cos\omega_2 B + B^2)$ and that $x_{i,t}$ and e_t are respectively defined in (2.17) and (2.19).

LEMMA A.2. $\sum^n x_{i,t}x_{j,t} = O_p(n^{2m-1})$ *for* $i \neq j$, $i, j = 1, 2, 3, 5$.

PROOF. Denote the ψ-weights and the $n \times n$ ψ-weights matrix of $x_{i,t}$ by $\psi_k(x_i)$ and $\psi(\mathbf{x}_i)$ respectively. Then

$$\sum_1^n x_{i,t}x_{j,t} = \mathbf{a}'\psi'(\mathbf{x}_i)\psi(\mathbf{x}_j)\mathbf{a}. \tag{A.1}$$

From (A.1), it is clear that the order of $\sum^n x_{i,t}x_{j,t}$ is determined by those of $\sum_{k=1}^{n-h} \psi_k(x_i)\psi_{k+h}(x_j)$ where h is a fixed integer. By the expression of ψ-weights, e.g. see (2.4), the order of $\psi_k(x_i)$ is in turn determined by the highest multiplicity of nonstationary characteristic roots of $x_{i,t}$ and by its corresponding roots. On the other hand, from (2.19), the highest multiplicity of each $x_{i,t}$ is m and its corresponding roots are 1, $-1, e^{i\omega_1}$ and $e^{-i\omega_1}$, $e^{i\omega_2}$ and $e^{-i\omega_2}$ respectively for $i = 1, 2, 3, 5$.

Therefore, to check the order of $\sum_{k=1}^{n-h} \psi_k(x_i)\psi_{k+h}(x_j)$ we only need to consider the following equations.

$$\psi_k(x_1) = c_1k^{m-1} + R_k(x_1), \quad \psi_k(x_2) = c_2(-1)^k k^{m-1} + R_k(x_2),$$

$$\psi_k(x_3) = c_3k^{m-1}\sin k\omega_1 + R_k(x_3) \quad \text{and} \quad \psi_k(x_5) = c_5k^{m-1}\sin k\omega_2 + R_k(x_5)$$

where c_i are constants and $R_k(x_i)$ denote the remaining terms in each equation which, of course, have no effect on our order determination. Now, by Corollary A.1 and the result

$$\sum_{k=1}^n (-1)^k k^{m-1}(h+k)^{m-1} = O(n^{2m-2}),$$

we have that

$$\sum_{k=1}^{n-h} \psi_k(x_i)\psi_{k+h}(x_j) = O(n^{2m-2}).$$

So, (A.1) can be rewritten as

$$\sum_1^n x_{i,t}x_{j,t} = n^{2m-2}\mathbf{a}'\mathbf{G}_{ij}\mathbf{a}$$

where $\mathbf{G}_{ij}$ is a $n \times n$ matrix such that $\mathbf{G}_{ij} = O(1)$. Using the same techniques as those in the proof of Lemma 2.5, we get $\sum_1^n x_{i,t}x_{j,t} = O_p(n^{2m-1})$. □

REFERENCES

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

ANDERSON, T. W. and TAYLOR, J. B. (1979). Strong consistency of least squares estimates in dynamic models. *Ann. Statist.* **7** 484–489.

BOX, G. E. P. and JENKINS, G. M. (1970). *Time Series Analysis: Forecasting and Control.* Holden-day, San Francisco.

DICKEY, D. A. (1976). *Estimation and hypothesis testing in nonstationary time series.* Unpublished Ph.D. Thesis, Iowa State University, Ames, Iowa.

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.

FINDLEY, D. F. (1980). Large-sample behavior of the S-array of seasonally nonstationary ARMA model. *In Time Series Analysis,* (Anderson, O. D. and Perryman, M. R., ed.) North Holland, Amsterdam 163–170.

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

FULLER, W. A., HASZA, D. P. and GOEBEL, J. J. (1981). Estimation of the parameters of stochastic difference equations. *Ann. Statist.* **9** 531–543.

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

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

KAWASHIMA, H. (1980). Parameter estimation of autoregressive integrated processes by least squares. *Ann. Statist.* **8** 423–435.

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

QUINN, B. G. (1980). Limiting behavior of autocorrelation function of ARMA process as several roots of characteristic equation approach unit circle. *Commun. Statist.* **B9** 195–198.

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

RUBIN, H. (1950). Consistency of maximum-likelihood estimates in the explosive case. In *Statistical Inference in Dynamic Economic Models.* (Koopmans, T. C., ed.) Wiley, New York.

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

TIAO, G. C. and BOX, G. E. P. (1981). Modelling multiple time series with applications. *J. Amer. Statist. Assoc.* **76** 802–816.

TIAO, G. C. and TSAY, R. S. (1981). Identification of nonstationary and stationary ARMA models. *Proc. Bus.and Econ. Statist. Sec., Amer. Statist. Assoc.* 308–312.

TSAY, R. S. and TIAO, G. C. (1982). Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and nonstationary ARMA models. Technical Report #683, Department of Statistics, University of Wisconsin-Madison.

GRADUATE SCHOOL OF BUSINESS
UNIVERSITY OF CHICAGO
1101 EAST 58TH STREET
CHICAGO, ILLINOIS 60637

DEPARTMENT OF STATISTICS
CARNEGIE-MELLON UNIVERSITY
PITTSBURGH, PENNSYLVANIA 15213