The Annals of Statistics
1990, Vol. 18, No. 1, 220–250

ASYMPTOTIC PROPERTIES OF MULTIVARIATE NONSTATIONARY PROCESSES WITH APPLICATIONS TO AUTOREGRESSIONS[1]

BY RUEY S. TSAY AND GEORGE C. TIAO

Carnegie-Mellon University and University of Chicago

Asymptotic properties of multivariate time series with characteristic roots on the unit circle are considered. For a vector autoregressive moving average (ARMA) process, we derive the limiting distributions of certain statistics which are useful in understanding nonstationary processes. These distributions are derived in a unified manner for all types of characteristic roots and are expressed in terms of stochastic integrals of Brownian motions. For applications, we use the limiting distributions to establish the consistency properties of the ordinary least squares (LS) estimates of various autoregressions of a vector process, e.g., the ordinary, forward and shifted autoregressions. For a purely nonstationary vector ARMA(p, q) process, the LS estimates are consistent if the order of the fitted autoregression is p; for a general ARMA model, the limits of the LS estimates exist, but these estimates can only provide consistent estimates of the nonstationary characteristic roots.

1. Introduction. A k-dimensional linear process $\mathbf{z}_t = (z_{1t}, \dots, z_{kt})^{\mathrm{T}}$ follows a vector ARMA(p, q) model if it satisfies

$$\Phi(B)\mathbf{z}_t = \theta(B)\mathbf{a}_t, \tag{1.1}$$

where $\Phi(B) = \mathbf{I} - \Phi_1 B - \cdots - \Phi_p B^p$ and $\theta(B) = \mathbf{I} - \theta_1 B - \cdots - \theta_q B^q$ are matrix polynomials in B of degrees p and q, respectively, B is the backshift operator such that $B\mathbf{z}_t = \mathbf{z}_{t-1}$, Φ_i's and θ_j's are $k \times k$ real-valued matrices and $\{\mathbf{a}_t = (a_{1t}, \dots, a_{kt})^{\mathrm{T}}\}$ is a sequence of martingale differences satisfying

$$\begin{aligned} &E(\mathbf{a}_t|\Psi_{t-1}) = \mathbf{O}, \qquad \operatorname{cov}(\mathbf{a}_t|\Psi_{t-1}) = \Sigma \quad \text{and} \\ &\sup_{i,t} E\left(|a_{it}|^{2+\delta}|\Psi_{t-1}\right) < \infty, \quad \text{a.s. for some } \delta > 0, \end{aligned} \tag{1.2}$$

where Ψ_{t-1} is the σ-field generated by $\{\mathbf{a}_{t-j}|j = 1, 2, \dots\}$ and Σ is a positive definite matrix. For model (1.1), we assume that $\Phi(B)$ and $\theta(B)$ are left coprime and all of the zeros of the determinants $\det[\Phi(B)]$ and $\det[\theta(B)]$ as functions of B are on or outside the unit circle. For convenience, we shall refer to the inverses of the zeros of $\det[\Phi(B)]$ as the characteristic roots of $\mathbf{z}_t$. If all of the characteristic roots are inside the unit circle, $\mathbf{z}_t$ is stationary; otherwise, it is nonstationary. If all of the characteristic roots are on the unit circle, $\mathbf{z}_t$ is *purely* nonstationary. For a nonstationary process, we assume that it starts at a fixed

Received October 1986; revised March 1989.
[1]Research supported in part by ALCOA Research Foundation, National Science Foundation and NASA.
AMS 1980 *subject classifications*. Primary 62M10; secondary 62J05.
Key words and phrases. Autoregression, Brownian motion, consistency, functional central limit theorem, least squares, multivariate ARMA time series, nonstationarity.

time point with *known* starting values. As will be seen later, this last assumption has no effect on the asymptotic properties considered.

In recent years, the theory and applications of model (1.1) have been extensively investigated by many authors. For example, among others, Tunnicliffe-Wilson (1973), Hillmer and Tiao (1979) and Nicholls and Hall (1979) studied the maximum likelihood estimation; Chan and Wallis (1978), Tiao and Box (1981) and Tiao and Tsay (1983b) proposed model building procedures; Hosking (1981) and Li and McLeod (1981) considered model checking techniques. Most of the investigations, however, are confined to the stationary case, leaving behind the wider class of nonstationary processes commonly encountered in practice. In the univariate case, this may not be a serious problem because stationarity can often be achieved by using the technique of differencing; see Box and Jenkins (1976). However, how to transform a multivariate nonstationary process into a stationary one is unclear. Differencing of the vector process $\mathbf{z}_t$ is no longer a practical solution because it may introduce unnecessary complications by making some of the zeros of $\det[\boldsymbol{\theta}(B)]$ on the unit circle. Roughly speaking, when the nonstationarity of a particular component of $\mathbf{z}_t$ is attributed to its dependence on other nonstationary components, differencing each individually nonstationary component in this situation will result in a noninvertible model for the differenced vector process; see the discussions in Hillmer and Tiao (1979), Tiao and Tsay (1983b) and Lütkepohl (1982). Furthermore, there is no convenient method currently available to identify the "genuine nonstationary components" of a vector process. Therefore, a better way to understand multivariate processes is to investigate nonstationary processes directly.

Some results of nonstationary processes are indeed available in the literature. Stigum (1975) investigated asymptotic properties of ARIMA processes including the normality of a process and a law of iterated logarithm. Lai and Wei (1983, 1985) showed that the ordinary least squares (OLS) estimates of an AR(p) regression are strongly consistent if the true model is a purely autoregressive process of order p, i.e., AR(p), and $\mathbf{a}_t$ satisfies (1.2). Chan and Wei (1988) derived the limiting distribution of these OLS estimates when $\mathbf{z}_t$ is a univariate series. Earlier Dickey and Fuller (1979) and Hasza and Fuller (1979) considered the problem of testing for one or two unit roots in a univariate AR model, and Ahtola and Tiao (1987) considered the case of a pair of complex roots on the unit circle. Fewer results, however, are available for the nonstationary mixed ARMA model. Phillips and Durlauf (1985) investigated multiple time series regression with unit roots. In Tiao and Tsay (1983a), we established some consistency properties of the OLS estimates of various autoregressions of a univariate nonstationary ARMA model under the condition that (1.2) holds for $\delta = 2$. Of particular interest there is that we allowed for the order of the fitted autoregression to be different from the underlying AR order p. This flexibility is valuable in application because the AR order p is usually unknown. For the multivariate ARMA models, the corresponding asymptotic properties are yet to be derived.

The primary goal of this paper, therefore, is to investigate asymptotic properties in estimating vector ARMA models when there are characteristic roots on the unit circle and, in particular, to extend the consistency results of Tiao and

Tsay (1983a) to the multivariate models. More specifically, we shall consider limiting distributions of certain statistics of nonstationary vector processes and use this result to establish consistency properties of OLS estimates of autoregressions of a vector ARMA model. For the purely nonstationary ARMA(p, q) processes, we show that the OLS estimates of an AR(p) regression are consistent. For the general ARMA(p, q) models, the OLS estimates are inconsistent if $q > 0$, but the estimates provide consistent results for the nonstationary characteristic roots, i.e., those roots on the unit circle. The limiting distribution is derived in a unified manner for all types of characteristic roots on the unit circle. Using some univariate results of Chan and Wei (1988), we first establish the results for multivariate AR models. The results are then extended to the general vector ARMA models through a relation between nonstationary ARMA($1, q$) and AR(1) models. Finally, we apply the results to forward and shifted least squares autoregressions and derive the consistency properties of the associated estimates.

The results of this paper are useful in multivariate time series analysis. First, they give theoretical justifications for some existing procedures proposed for handling nonstationary multivariate time series data. For example, Tjøtheim and Paulsen (1982) proposed using the least squares estimate $\hat{\mathbf{\Phi}}_1$ in an AR(1) fitting as the coefficient of transformation to achieve stationarity. The reason that this approach works in some cases is because of the consistency in estimating the nonstationary roots. Second, the results here provide the basis on which unified procedures for modeling stationary and nonstationary ARMA processes can be constructed. For example, based on the consistency results of Tiao and Tsay (1983a), Tsay and Tiao (1984, 1985) have developed the extended sample autocorrelation function and a canonical correlation approach for model specification of univariate time series. Details of using the results of this paper to derive a unified modeling procedure for multivariate ARMA processes is given in Tiao and Tsay (1989). Third, the results specify the distributional behavior of nonstationary vector ARMA models which is important in understanding nonstationary processes.

2. Preliminaries. In this section, we derive a link between ARMA and AR models. This link enables us to generalize asymptotic results for pure AR models to the mixed ARMA processes. Letting $\mathbf{X}_t = (\mathbf{z}_t^{\mathrm{T}}, \dots, \mathbf{z}_{t-p+1}^{\mathrm{T}})^{\mathrm{T}}$, the ARMA model (1.1) can be rewritten as

$$\mathbf{X}_t = \mathbf{G}\mathbf{X}_{t-1} + \sum_{i=0}^{q} \mathbf{\Theta}_i \mathbf{a}_{t-i}, \tag{2.1}$$

where the coefficient matrices are given by

$$\mathbf{G} = \begin{bmatrix} \mathbf{\Phi}_1, \mathbf{\Phi}_2, \dots, \mathbf{\Phi}_p \\ \mathbf{I}_{k(p-1)} \quad \mathbf{O} \end{bmatrix}$$

and $\mathbf{\Theta}_i = -\mathbf{L}\boldsymbol{\theta}_i$, where $\mathbf{L}$ is the first k columns of $\mathbf{I}_{kp}$, $\boldsymbol{\theta}_0 = -\mathbf{I}_k$ and $\mathbf{I}_s$ denotes the $s \times s$ identity matrix. Since the characteristic roots of $\mathbf{z}_t$ are the eigenvalues of $\mathbf{G}$, these eigenvalues are on or inside the unit circle under model (1.1). Also, we have $\det[\mathbf{I} - \mathbf{G}B] = \det[\mathbf{\Phi}(B)]$.

Now there exists a nonsingular matrix $\mathbf{T}$ that transforms $\mathbf{G}$ into its Jordan form, i.e.,

$$\mathbf{TGT}^{-1} = \mathbf{J} = \operatorname{diag}\{\mathbf{J}_1, \mathbf{J}_2, \ldots, \mathbf{J}_s\}, \tag{2.2}$$

where

$$\mathbf{J}_j = \begin{bmatrix} \lambda_j & 1 & 0 & \cdot & \cdot & 0 \\ 0 & \lambda_j & 1 & \cdot & \cdot & 0 \\ \cdot & & & & & \cdot \\ \cdot & & & & & 1 \\ 0 & 0 & 0 & \cdot & 0 & \lambda_j \end{bmatrix}$$

is a $r_j \times r_j$ Jordan block, λ_j is an eigenvalue of $\mathbf{G}$ and r_j is the multiplicity of λ_j, $j = 1, \ldots, s$. Here the matrix $\mathbf{T}$ and the eigenvalues λ_j are over the complex field, and $\sum_{j=1}^{s} r_j = kp$.

By (2.1) and (2.2), we have

$$(\mathbf{I} - \mathbf{J}B)\mathbf{TX}_t = \left(\sum_{i=0}^{q} \Theta_i^* B^i \right) \mathbf{a}_t, \tag{2.3}$$

where the moving average coefficient matrices are given by $\Theta_i^* = \mathbf{T}\Theta_i = -\mathbf{T}^* \theta_i$ with $\mathbf{T}^*$ being the first k columns of $\mathbf{T}$. We now derive some basic results of the above transformation. For the general theory of matrix polynomials, one may consult MacDuffee (1956) and Kailath (1980) with the latter emphasizing applications to linear systems. From (2.3), the model for each individual Jordan block is of the form

$$(\mathbf{I} - \mathbf{D}B)\mathbf{u}_t = \left(\sum_{i=0}^{q} \mathbf{C}_i B^i \right) \mathbf{a}_t, \tag{2.4}$$

where $\mathbf{D}$ is a Jordan block corresponding to some eigenvalue λ of $\mathbf{G}$, $\mathbf{u}_t$ is a subvector of $\mathbf{TX}_t$ and $\mathbf{C}_i$ is a submatrix of Θ_i^* for $i = 1, \ldots, q$.

LEMMA 2.1. *Suppose that* $\mathbf{z}_t$ *follows the vector ARMA model of* (1.1). *Then the two matrix polynomials of each individual Jordan block in the form of* (2.4) *are left coprime.*

PROOF. Denote the AR and MA matrix polynomials of (2.4) by $\mathbf{D}(B)$ and $\mathbf{C}(B)$, respectively. Since $\det[\mathbf{D}(B)] = (1 - \lambda B)^h$, where h is the dimension of $\mathbf{u}_t$, we only need to consider the case $\lambda \neq 0$. If $\mathbf{D}(B)$ and $\mathbf{C}(B)$ are *not* coprime, then $(1 - \lambda B)$ is a factor of each and every element of $\operatorname{adj}[\mathbf{D}(B)]\mathbf{C}(B)$, where $\operatorname{adj}[\mathbf{D}(B)]$ is the adjoint matrix of $\mathbf{D}(B)$. Consider the model (2.3). Without loss of generality, we may treat $\mathbf{D}$ as the first Jordan block of (2.2). Then, by the

Jordan structure, we have

$$\text{(2.5)} \qquad \det[\mathbf{I} - \mathbf{J}B]\mathbf{T}\mathbf{X}_t = \operatorname{adj}\begin{bmatrix} \mathbf{D}(B) & \mathbf{O} \\ \mathbf{O} & \mathbf{D}^c(B) \end{bmatrix}\begin{bmatrix} \mathbf{C}(B) \\ \mathbf{C}^c(B) \end{bmatrix}\mathbf{a}_t,$$

where $\mathbf{D}^c(B)$ and $\mathbf{C}^c(B)$ are, respectively, the compliment parts of $\mathbf{D}(B)$ and $\mathbf{C}(B)$ in the AR and MA matrix polynomials of (2.3). Writing

$$\operatorname{adj}\begin{bmatrix} \mathbf{D}(B) & \mathbf{O} \\ \mathbf{O} & \mathbf{D}^c(B) \end{bmatrix} = \begin{bmatrix} \mathbf{H}_1(B) & \mathbf{O} \\ \mathbf{O} & \mathbf{H}_2(B) \end{bmatrix},$$

it is easy to see that every element of $\mathbf{H}_2(B)$ is divisible by $\det[\mathbf{D}(B)]$ and every element of $\mathbf{H}_1(B)$ is divisible by the corresponding element in $\operatorname{adj}[\mathbf{D}(B)]$. Therefore, if $(1 - \lambda B)$ is a factor of each element of $\operatorname{adj}[\mathbf{D}(B)]\mathbf{C}(B)$, then every element of the right-hand matrix of (2.5) is divisible by $(1 - \lambda B)$. Since $\det[\mathbf{I} - \mathbf{J}B] = \det[\mathbf{I} - \mathbf{G}B] = \det[\boldsymbol{\Phi}(B)]$, we have, by (2.1) and (2.2),

$$\text{(2.6)} \qquad \det[\boldsymbol{\Phi}(B)]\mathbf{X}_t = \mathbf{T}^{-1}\operatorname{adj}[\mathbf{I} - \mathbf{J}B]\boldsymbol{\Theta}^*(B)\mathbf{a}_t,$$

where $\boldsymbol{\Theta}^*(B)$ is the MA matrix polynomial of (2.3). The above result says that every element of the right-hand matrix of (2.6) is divisible by $(1 - \lambda B)$. However, by (1.1), $\det[\boldsymbol{\Phi}(B)]\mathbf{z}_t = \operatorname{adj}[\boldsymbol{\Phi}(B)]\boldsymbol{\theta}(B)\mathbf{a}_t$. In particular, by considering the first k rows of the right-hand matrix of (2.6), every element of $\operatorname{adj}[\boldsymbol{\Phi}(B)]\boldsymbol{\theta}(B)$ is divisible by $(1 - \lambda B)$. This contradicts the left coprime assumption of model (1.1). □

For model (2.4) with $\lambda \neq 0$, define the matrices ψ_i by

$$(\mathbf{I} - \mathbf{D}B)\sum_{i=0}^{\infty} \psi_i B^i = \mathbf{C}_0 + \mathbf{C}_1 B + \cdots + \mathbf{C}_q B^q.$$

Then $\psi_0 = \mathbf{C}_0$, $\psi_i = \sum_{v=0}^{i}\mathbf{D}^{i-v}\mathbf{C}_v$ for $i = 1, \dots, q$, and $\psi_i = \mathbf{D}^{i-q}\psi_q$ for $i > q$. Next, under the assumption of zero starting values we may rewrite the model (2.4) as

$$\text{(2.7)} \qquad \mathbf{u}_t = \mathbf{M}_t + \mathbf{Y}_t,$$

where

$$\text{(2.8)} \qquad \mathbf{M}_t = \sum_{i=0}^{q-1} \psi_i \mathbf{a}_{t-i}, \qquad \mathbf{Y}_t = \sum_{i=0}^{t-q-1} \mathbf{D}^i \mathbf{f}_{t-i} \quad \text{with } \mathbf{f}_t = \psi_q \mathbf{a}_{t-q}.$$

From (2.8), it is obvious that $\mathbf{Y}_t$ is a multivariate AR(1) process, i.e.,

$$\text{(2.9)} \qquad \mathbf{Y}_t = \mathbf{D}\mathbf{Y}_{t-1} + \mathbf{f}_t.$$

Equations (2.7) and (2.9) establish a link between a vector ARMA$(1, q)$ model and a vector AR(1) model that enables us to extend the results of a pure AR model to those of a mixed ARMA model. We now consider some properties of the AR(1) process $\mathbf{Y}_t$.

LEMMA 2.2. *Suppose that $\mathbf{z}_t$ follows the multivariate ARMA model of* (1.1). *Then the last row of ψ_q of* (2.8) *is nonzero provided that the eigenvalue λ of $\mathbf{D}$ is nonzero.*

PROOF. Denote the jth element of the last row of the matrix $\mathbf{A}$ by A_j. From $\psi_q = \sum_{v=0}^{q} \mathbf{D}^{q-v}\mathbf{C}_v$ and by the Jordan structure of $\mathbf{D}$, we have

$$\psi_{q,j} = \sum_{v=0}^{q} \lambda^{q-v} C_{v,j} \quad \text{for } j = 1, \dots, k.$$

If the last row of ψ_q is 0, then we have

$$C_{q,j} = -\sum_{v=0}^{q-1} \lambda^{q-v} C_{v,j}. \tag{2.10}$$

Now, since

$$\operatorname{adj}[\mathbf{D}(B)] = \begin{bmatrix} (1-\lambda B)^h & B(1-\lambda B)^{h-1} & \cdot & B^{h-2}(1-\lambda B) & B^{h-1} \\ 0 & (1-\lambda B)^h & \cdot & \cdot & B^{h-2}(1-\lambda B) \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & 0 & (1-\lambda B)^h \end{bmatrix},$$

where h is the number of rows of $\mathbf{D}$, the only element of adj$[\mathbf{D}(B)]$ that is not divisible by $(1 - \lambda B)$ is the $(1, h)$th element. Thus, we need only consider the last rows of the matrix polynomial $\mathbf{C}(B)$. The result of multiplying B^{h-1} by the jth element of the last row of $\mathbf{C}(B)$ is

$$g_j(B) = \sum_{v=0}^{q} C_{v,j} B^{h+v-1} \quad \text{for } j = 1, \dots, k. \tag{2.11}$$

Consequently, if the last row of ψ_q is 0, then by (2.10) and (2.11) we have $g_j(B) = \sum_{v=0}^{q-1} C_{v,j} B^{h+v-1}(1 - \lambda^{q-v} B^{q-v})$, which says that $g_j(B)$ is divisible by $(1 - \lambda B)$ for $j = 1, \dots, k$. This implies that each and every element of adj$[\mathbf{D}(B)]\mathbf{C}(B)$ is divisible by $(1 - \lambda B)$ which is in contradiction with Lemma 2.1. □

Next consider another Jordan block following the model

$$(\mathbf{I} - \mathbf{D}^* B)\mathbf{u}_t^* = \left(\sum_{i=0}^{q} \mathbf{C}_i^* B^i \right) \mathbf{a}_t,$$

where the eigenvalue of $\mathbf{D}^*$ is λ^*, $\mathbf{u}_t^*$ is a subvector of $\mathbf{TX}_t$ and $\mathbf{C}_i^*$'s are submatrices of Θ_i^*'s, respectively. Suppose that $\lambda^* \neq 0$. Then, similar to (2.7) we have

$$\mathbf{u}_t^* = \mathbf{M}_t^* + \mathbf{Y}_t^* \quad \text{with } \mathbf{Y}_t^* = \mathbf{D}^* \mathbf{Y}_{t-1}^* + \mathbf{f}_t^*. \tag{2.12}$$

Comparing the models of $\mathbf{Y}_t$ and $\mathbf{Y}_t^*$, we have the following result.

LEMMA 2.3. *Suppose that* $\mathbf{z}_t$ *follows the vector ARMA model of* (1.1) *and that the two Jordan blocks* $\mathbf{D}$ *and* $\mathbf{D}^*$ *of* (2.9) *and* (2.12) *are different in the sense that* $\mathbf{D} = \mathbf{J}_i$ *and* $\mathbf{D}^* = \mathbf{J}_j$ *for some* $i \neq j$. *Assume further that the two eigenvalues satisfy* $\lambda = \lambda^* \neq 0$. *Then,* f_t *and* $f_t{}^*$ *are not linearly dependent, where* f_t *and* $f_t{}^*$ *are the last elements of* $\mathbf{f}_t$ *and* $\mathbf{f}_t^*$, *respectively.*

PROOF. In this proof, denote the last element of the vector $\mathbf{w}_t$ by w_t. Since the eigenvalues are nonzero, f_t and $f_t{}^*$ are nonzero linear combinations of $\mathbf{a}_t$. If f_t and $f_t{}^*$ are linearly dependent, then $f_t = \beta f_t{}^*$ for some nonzero constant β. By (2.9) and (2.12), this implies $(1 - \lambda B)Y_t = \beta(1 - \lambda^* B)Y_t{}^*$. Since $\lambda = \lambda^*$, we have $Y_t = \beta Y_t{}^* + d$, where d is a constant. Therefore, from (2.7) and (2.12), we have

$$u_t - \beta u_t^* = M_t - \beta M_t^* + d.$$

Two cases are possible. First, if $M_t - \beta M_t{}^* = c$, a constant, then $u_t = \beta u_t^* + c + d$. However, since u_t and u_t^* come from two different rows of the transformation matrix $\mathbf{T}$, the result says that there is an exact (lagged) linear relationship in the original process $\mathbf{z}_t$. This, of course, contradicts the condition that Σ is positive definite. Second, if $M_t - \beta M_t{}^*$ is not a constant, then $u_t - \beta u_t^*$ is an MA(q) process, because both M_t and M_t^* are MA(q). This says that $u_t - \beta u_t^*$ belongs to a Jordan block of (2.2) which is associated with a zero eigenvalue. Thus, $u_t - \beta u_t^*$ can also be obtained by a third row of $\mathbf{T}$. This, however, contradicts the nonsingularity of $\mathbf{T}$. Consequently, f_t and $f_t{}^*$ are not linearly dependent. □

For convenience, we rearrange the Jordan blocks of (2.2) so that $\mathbf{TX}_t$ can be partitioned as

$$\mathbf{TX}_t = \left(\mathbf{N}_t^{\mathrm{T}}, \mathbf{S}_t^{\mathrm{T}}\right)^{\mathrm{T}}, \tag{2.13}$$

where $\mathbf{N}_t$ and $\mathbf{S}_t$ consist of all the nonstationary and stationary components of $\mathbf{TX}_t$, respectively. For each Jordan block in $\mathbf{N}_t$ the associated eigenvalue is on the unit circle. By Lemma 2.1, the two matrix polynomials of (2.4) are left coprime so that the multiplicity of the nonstationary characteristic root is just the multiplicity of the associated eigenvalue. Next, (2.7) and (2.9) show that in this case we may rewrite the process as a sum of an MA process and a purely nonstationary AR(1) process. Furthermore, by Lemma 2.2, the last element of the innovational series $\mathbf{f}_t$ of the nonstationary AR(1) process is nonzero. Thus, the corresponding process $\mathbf{Y}_t$ is not degenerated. This provides the needed background for the discussion in Section 3.

Finally, it is interesting to see the implications of Lemmas 2.2 and 2.3 in the univariate case. These lemmas show that for each nonstationary root there is at most a single Jordan block available, because any two nonzero functions of the scalar variable $\mathbf{a}_t$ are linearly dependent. Therefore, when $k = 1$ and $\boldsymbol{\theta}(B) = \mathbf{I}$, by combining complex conjugate pairs of eigenvalues to produce a real-valued system, one can further transform the process so that the resulting transformation matrix $\mathbf{T}$ becomes that of Chan and Wei (1988). The results of this paper thus reduce to those of Chan and Wei when $\mathbf{z}_t$ follows a univariate AR model.

3. Basic properties of purely nonstationary vector AR(1) processes. From the results of the preceding section, for each Jordan block associated with the nonstationary component $\mathbf{N}_t$ of (2.13) there is a corresponding purely nonstationary process $\mathbf{Y}_t$ in the form of (2.9). Furthermore, since $\mathbf{M}_t$ of (2.7) is stationary (because it is a finite sum of martingale differences), the asymptotic properties of any process in $\mathbf{N}_t$ are determined mainly by those of the corresponding purely nonstationary process in $\mathbf{Y}_t$. It is, therefore, important to investigate the asymptotic properties of purely nonstationary AR(1) processes in the form of (2.9) with the eigenvalue λ of $\mathbf{D}$ on the unit circle.

Since each component of $\mathbf{Y}_t$ is nonstationary, the asymptotic properties to be discussed, e.g., the limiting distribution of least squares estimates, are different from those obtained by the central limit or ergodic theorem. The results involve stochastic integrals of Brownian motions, and we use the following notation.

1. $D[0,1]$ is the space of functions $f(t)$ on the unit interval $[0,1]$ which are right continuous and have left-hand limits; see Billingsley (1968).
2. $D[0,1]$ is equipped with the Skorohod topology.
3. For the weak convergence $X_n \to_d X$ in D, we shall use $X_n(t) \to_d X(t)$ from time to time to indicate the time variable t of the random elements.
4. For $0 \le s \le 1$, $[ns]$ denotes the largest integer less than or equal to ns.

For a nonstationary process $\mathbf{Y}_t$ in the form of (2.9), write the eigenvalue of the Jordan block $\mathbf{D}$ as $\lambda = \exp(i\omega)$, where $0 \le \omega < 2\pi$. If the dimension of $\mathbf{Y}_t$ is h, then the component models are

$$(3.1)\qquad Y_{h,t} = \exp(i\omega) Y_{h,t-1} + f_{h,t},$$

$$(3.2)\qquad Y_{v,t} = \exp(i\omega) Y_{v,t-1} + Y_{v+1,t-1} + f_{v,t} \quad \text{for } v = h-1,\dots,1.$$

Assuming that the starting value $\mathbf{Y}_0 = \mathbf{O}$, (3.1) and (3.2) give

$$(3.3)\qquad Y_{h,t} = \sum_{j=1}^{t} [\exp(i\omega)]^{t-j} f_{h,j},$$

$$(3.4)\qquad Y_{v,t} = \sum_{j=1}^{t-1} [\exp(i\omega)]^{t-1-j} Y_{v+1,j} + \sum_{j=1}^{t} [\exp(i\omega)]^{t-j} f_{v,j} \quad \text{for } v = h-1,\dots,1.$$

Consequently,

$$(3.5)\qquad \exp(-i\omega t) Y_{h,t} = \sum_{j=1}^{t} \exp(-i\omega j) f_{h,j},$$

$$(3.6)\qquad \exp(-i\omega t) Y_{v,t} = \exp(-i\omega) \sum_{j=1}^{t-1} \exp(-i\omega j) Y_{v+1,j} + \sum_{j=1}^{t} \exp(-i\omega j) f_{v,j}.$$

These two equations give a nice recursion concerning $Y_{v,t}$. When $\omega = 0$ or $\omega = \pi$, i.e., the eigenvalue is either 1 or -1, the equations are real-valued; otherwise, they are in the complex field. Since $\mathbf{z}_t$ of (1.1) is a real-valued series, any complex process $Y_{v,t}$, innovation $f_{v,t}$ and root $\exp(i\omega)$ must exist in conjugate pairs. Therefore, $\{Y_{v,t}\}$ and $\{Y_{v,t}^*\}$ exist simultaneously, where $Y_{v,t}^*$ denotes the complex conjugate of $Y_{v,t}$, and we may assume $0 \le \omega \le \pi$. Notice that $Y_{v,t}^*$ satisfies (3.1) through (3.4) with $\exp(i\omega)$ and $f_{v,j}$ being replaced by $\exp(-i\omega)$ and $f_{v,j}^*$. Let $\mathbf{Y}_{v,t} = (Y_{v,t}, Y_{v,t}^*)^{\mathrm{T}}$, $\mathbf{f}_{v,t} = (f_{v,t}, f_{v,t}^*)^{\mathrm{T}}$ and $\mathbf{H} = \operatorname{diag}\{\exp(i\omega), \exp(-i\omega)\}$ for $0 < \omega < \pi$. Then we have

$$(3.7) \qquad \mathbf{Y}_{h,t} = \mathbf{H}\mathbf{Y}_{h,t-1} + \mathbf{f}_{h,t},$$

$$(3.8) \qquad \mathbf{Y}_{v,t} = \mathbf{H}\mathbf{Y}_{v,t-1} + \mathbf{Y}_{v+1,t-1} + \mathbf{f}_{v,t} \quad \text{for } v = h-1, \dots, 1.$$

These complex-valued equations can be transformed into real-valued ones by letting

$$(3.9) \qquad \mathbf{y}_{v,t} = \mathbf{r}\mathbf{Y}_{v,t} \quad \text{and} \quad \mathbf{e}_{v,t} = \mathbf{r}\mathbf{f}_{v,t},$$

where $\mathbf{r}$ is a 2×2 matrix with first row $(1,1)/\sqrt{2}$ and second $(-i, i)/\sqrt{2}$.The model of $\mathbf{y}_{v,t}$ then becomes

$$(3.10) \qquad \mathbf{y}_{h,t} = \mathbf{R}\mathbf{y}_{h,t-1} + \mathbf{e}_{h,t},$$

$$(3.11) \qquad \mathbf{y}_{v,t} = \mathbf{R}\mathbf{y}_{v,t-1} + \mathbf{y}_{v+1,t-1} + \mathbf{e}_{v,t} \quad \text{for } v = h-1, \dots, 1,$$

where $\mathbf{R}$ is a 2×2 orthogonal matrix given by

$$\mathbf{R} = \mathbf{R}(\omega) = \begin{bmatrix} \cos(\omega) & -\sin(\omega) \\ \sin(\omega) & \cos(\omega) \end{bmatrix}.$$

Note that the following properties of $\mathbf{R}(\omega)$ are useful

$$(3.12) \qquad [\mathbf{R}(\omega)]^j = \mathbf{R}(j\omega) \quad \text{and} \quad [\mathbf{R}(\omega)]^{\mathrm{T}} = \mathbf{R}(-\omega).$$

By (3.10) and (3.11), a bivariate recursion similar to (3.5) and (3.6) is obtained

$$(3.13) \quad \mathbf{R}^{-t}\mathbf{y}_{h,t} = \sum_{j=1}^{t} \mathbf{R}^{-j}\mathbf{e}_{h,j},$$

$$(3.14) \quad \mathbf{R}^{-t}\mathbf{y}_{v,t} = \mathbf{R}^{-1} \sum_{j=1}^{t-1} \mathbf{R}^{-j}\mathbf{y}_{v+1,j} + \sum_{j=1}^{t} \mathbf{R}^{-j}\mathbf{e}_{v,j} \quad \text{for } v = h-1, \dots, 1.$$

Notice that the recursion (3.13) and (3.14) actually applies to all ω satisfying $0 \le \omega \le \pi$ because $\mathbf{y}_{v,t} = (2^{-1/2}Y_{v,t}, 0)^{\mathrm{T}}$, $\mathbf{e}_{v,t} = (2^{-1/2}f_{v,t}, 0)^{\mathrm{T}}$ and $\mathbf{R}$ is diagonal if $\omega = 0$ or $\omega = \pi$. More precisely, the first component of each equation of (3.13) and (3.14) reduces to (3.5) and (3.6) when $\omega = 0$ or $\omega = \pi$. For this reason, we use (3.13) and (3.14) in the derivation below and assume $0 \le \omega \le \pi$.

To derive some basic properties of $\mathbf{y}_{v,t}$, we need a fundamental lemma. Let $\{a_t\}$ be a sequence of martingale differences satisfying (1.2) with variance σ_a^2. For $0 \leq \omega \leq \pi$, define

$$C_t(\omega) = \sum_{j=1}^{t} \cos(j\omega) a_j \quad \text{and} \quad S_t(\omega) = \sum_{j=1}^{t} \sin(j\omega) a_j. \tag{3.15}$$

Since

$$n^{-1} \sum_{t=1}^{n} \cos^2(t\omega) = 1 \quad \text{if } \omega = 0 \text{ or } \omega = \pi; \qquad \to \tfrac{1}{2}, \quad \text{o.w.},$$

$$n^{-1} \sum_{t=1}^{n} \sin^2(t\omega) = 0 \quad \text{if } \omega = 0 \text{ or } \omega = \pi; \qquad \to \tfrac{1}{2}, \quad \text{o.w.},$$

define the constants

$$C(\omega) = 1 \quad \text{if } \omega = 0 \text{ or } \omega = \pi; \qquad = 1/\sqrt{2} \quad \text{o.w.}, \tag{3.16}$$

$$S(\omega) = 0 \quad \text{if } \omega = 0 \text{ or } \omega = \pi; \qquad = 1/\sqrt{2} \quad \text{o.w.} \tag{3.17}$$

Next, define two functions on $[0, 1]$ by

$$\begin{aligned} Z_n(s) &= \left[C(\omega) n^{1/2} \sigma_a\right]^{-1} C_{[ns]}(\omega) \quad \text{and} \\ X_n(s) &= \left[S(\omega) n^{1/2} \sigma_a\right]^{-1} S_{[ns]}(\omega), \end{aligned} \tag{3.18}$$

where $Z_n(0) = X_n(0) = 0$. Using Theorem 2.2 of Chan and Wei (1988), we have the following lemma.

LEMMA 3.1. *Suppose that* $\{a_t\}$ *satisfies* (1.2) *and* $0 \leq \omega \leq \pi$. *Then*

$$(Z_n, X_n) \to_d (W_c, W_s),$$

where W_c *and* W_s *are two independent standard Brownian motions and the subscripts c and s are used to signify limits of cosines and sines, respectively.*

In Lemma 3.1, it is understood that $X_n = W_s = 0$ if $\omega = 0$ or π and that the independence of W_c and W_s follows directly from the orthogonal properties of trigonometric series.

Next we apply Lemma 3.1 to establish some basic limiting distributions for $\mathbf{y}_{v,t}$. Let

$$\mathbf{E}_{v,t} = \sum_{j=1}^{t} \mathbf{R}^{-j} \mathbf{e}_{v,j} \quad \text{for } v = 1, 2, \dots, h. \tag{3.19}$$

Since the covariance matrix of $\mathbf{f}_{h,t}$ of (3.7) is nonzero (see Lemma 2.2) and $\mathbf{r}$ of (3.9) is nonsingular, $\{\mathbf{e}_{h,t}\}$ are martingale differences and satisfy (1.2) because they are nonzero functions of $\{\mathbf{a}_t\}$. For any other v, if $\mathbf{e}_{v,t} = \mathbf{O}$, then it is understood that $\mathbf{E}_{v,t} = \mathbf{O}$ and the variance of any component of $\mathbf{e}_{v,t}$ is 0.

By (3.12), elements of $\mathbf{E}_{v,t}$ are linear combinations of terms similar to $C_t(\omega)$ and $S_t(\omega)$ of (3.15) with a_t replaced by components of $\mathbf{e}_{v,t}$. Define

$$\Upsilon_{v,n}(s) = n^{-1/2}\mathbf{E}_{v,[ns]} \quad \text{for } 0 \le s \le 1.$$

Then, by Lemma 3.1 and using the same argument as that of Theorem 2.2 in Chan and Wei (1988), we have the following result.

LEMMA 3.2. *Suppose that* $\mathbf{z}_t$ *follows the vector ARMA model* (1.1) *with* $\mathbf{a}_t$ *satisfying* (1.2) *and* $0 \le \omega \le \pi$. *Then*

$$[\Upsilon_{1,n}(s),\dots,\Upsilon_{h,n}(s)] \to_d [\Xi_1(s),\dots,\Xi_h(s)]$$

with

$$\Xi_v(s) = \begin{bmatrix} C(\omega)W_{1,c}^{(v)}(s) & -S(\omega)W_{2,s}^{(v)}(s) \\ S(\omega)W_{1,s}^{(v)}(s) & C(\omega)W_{2,c}^{(v)}(s) \end{bmatrix} \begin{bmatrix} \sigma_{v,1} \\ \sigma_{v,2} \end{bmatrix},$$

where $\sigma_{v,m}^2$ *is the variance of the mth element* $e_{v,m,t}$ *of* $\mathbf{e}_{v,t}$ *and* $W_{m,c}^{(v)}(s)$ *and* $W_{m,s}^{(v)}(s)$ *are independent standard Brownian motions associated with* $e_{v,m,t}$.

In Lemma 3.2, since the element $e_{v,m,t}$'s may not be linearly independent, $\Xi_i(s)$'s are in general linearly correlated. For simplicity, we do not give expressions for the covariances between the $\Xi_i(s)$'s variables. Some expressions, however, can be obtained by orthogonalizing the $e_{v,m,t}$'s [see (4.13) for an example].

To illustrate the application of Lemma 3.2 and show that the assumption of zero starting value, i.e., $\mathbf{y}_{v,0} = \mathbf{O}$, is immaterial in studying asymptotic results of $\mathbf{y}_{v,t}$, we consider the process $\mathbf{y}_{h,t}$ in detail. In this case, the multiplicity of the nonstationary characteristic root is unity, see Lemma 2.1.

LEMMA 3.3. *Suppose that* $\mathbf{y}_{h,t}$ *follows the model* (3.10) *which is obtained from* (1.1) *by the transformations of Section* 2. *Assume that* $\mathbf{a}_t$ *of* (1.1) *satisfies* (1.2) *and that the starting value* $\mathbf{y}_{h,0}$ *has a well-defined probability distribution function. Then*

(a) $$n^{-3/2}\sum_{t=1}^{n}\mathbf{R}^{-t}\mathbf{y}_{h,t} \to_d \int_0^1 \Xi_h(s)\,ds,$$

(b) $$n^{-2}\sum_{t=1}^{n}\mathbf{R}^{-t}\mathbf{y}_{h,t}\mathbf{y}_{h,t}^{\mathrm{T}}\mathbf{R}^{t} \to_d \int_0^1 \Xi_h(s)\Xi_h^{\mathrm{T}}(s)\,ds,$$

(c $$n^{-1}\mathbf{R}^{-n}\mathbf{y}_{h,n}\mathbf{y}_{h,n}^{\mathrm{T}}\mathbf{R}^{n} \to_d \Xi_h(1)\Xi_h^{\mathrm{T}}(1),$$

(d) $$n^{-1}\sum_{t=1}^{n}\left(\mathbf{R}^{1-t}\mathbf{y}_{h,t-1}\mathbf{e}_{h,t}^{\mathrm{T}}\mathbf{R}^{t} + \mathbf{R}^{-t}\mathbf{e}_{h,t}\mathbf{y}_{h,t-1}^{\mathrm{T}}\mathbf{R}^{t-1}\right) \to_d \Xi_h(1)\Xi_h^{\mathrm{T}}(1) - \mathbf{F},$$

where $\mathbf{F}$ *is the limiting matrix of* $n^{-1}\sum_{t=1}^{n}\mathbf{R}^{-t}\mathbf{e}_{h,t}\mathbf{e}_{h,t}^{\mathrm{T}}\mathbf{R}^{t}$.

PROOF. Given the starting value $\mathbf{y}_{h,0}$, we have $\mathbf{R}^{-t}\mathbf{y}_{h,t} = \sum_{j=1}^{t}\mathbf{R}^{-j}\mathbf{e}_{h,j} + \mathbf{y}_{h,0}$. To show part (a), consider

$$
\begin{aligned}
n^{-3/2}\sum_{t=1}^{n}\mathbf{R}^{-t}\mathbf{y}_{h,t} &= n^{-3/2}\sum_{t=1}^{n}\left(\mathbf{E}_{h,t-1} + \mathbf{R}^{-t}\mathbf{e}_{h,t} + \mathbf{y}_{h,0}\right) \\
&= n^{-1}\sum_{t=1}^{n} n^{-1/2}\mathbf{E}_{h,t-1} + n^{-1/2}\mathbf{y}_{h,0} + n^{-1/2}\left(n^{-1}\sum_{t=1}^{n}\mathbf{R}^{-t}\mathbf{e}_{h,t}\right) \\
&= \sum_{t=1}^{n} n^{-1/2}\mathbf{E}_{h,t-1}[t/n - (t-1)/n] + o_p(1) \\
&= \sum_{t=1}^{n} \Upsilon_{h,n}[(t-1)/n][t/n - (t-1)/n] + o_p(1) \\
&= \sum_{t=1}^{n}\int_{(t-1)/n}^{t/n}\Upsilon_{h,n}(s)\,ds + o_p(1) \\
&= \int_0^1 \Upsilon_{h,n}(s)\,ds + o_p(1).
\end{aligned}
\tag{3.20}
$$

In the above, the fact that $n^{-1}\Sigma\mathbf{R}^{-t}\mathbf{e}_{h,t}$ is bounded in probability can be obtained by applying a central limit theorem of martingale differences under the condition (1.2), e.g., Theorem 2.5 of Helland (1982). By continuous mapping theorem [Billingsley (1968), Theorem 5.2] and Lemma 3.2, (3.20) implies that (a) holds.

To show part (b), consider

$$
\begin{aligned}
n^{-2}\sum_{t=1}^{n}\mathbf{R}^{-t}\mathbf{y}_{h,t}\mathbf{y}_{h,t}^{\mathrm{T}}\mathbf{R}^{t} &= n^{-2}\sum_{t=1}^{n}(\mathbf{E}_{h,t} + \mathbf{y}_{h,0})(\mathbf{E}_{h,t} + \mathbf{y}_{h,0})^{\mathrm{T}} \\
&= n^{-2}\sum_{t=1}^{n}\mathbf{E}_{h,t}\mathbf{E}_{h,t}^{\mathrm{T}} + n^{-2}\left(\sum_{t=1}^{n}\mathbf{E}_{h,t}\right)\mathbf{y}_{h,0}^{\mathrm{T}} \\
&\quad + n^{-2}\mathbf{y}_{h,0}\left(\sum_{t=1}^{n}\mathbf{E}_{h,t}\right)^{\mathrm{T}} + n^{-1}\mathbf{y}_{h,0}\mathbf{y}_{h,0}^{\mathrm{T}}.
\end{aligned}
$$

Using the same techniques as those in the proof of part (a), the above equation can be rewritten as

$$
\begin{aligned}
n^{-2}\sum_{t=1}^{n}\mathbf{R}^{-t}\mathbf{y}_{h,t}\mathbf{y}_{h,t}^{\mathrm{T}}\mathbf{R}^{t} &= \int_0^1 \Upsilon_{h,n}(s)\Upsilon_{h,n}^{\mathrm{T}}(s)\,ds \\
&\quad + n^{-1/2}\left\{\int_0^1 \Upsilon_{h,n}(s)\,ds\right\}\mathbf{y}_{h,0}^{\mathrm{T}} \\
&\quad + n^{-1/2}\mathbf{y}_{h,0}\left\{\int_0^1 \Upsilon_{h,n}(s)\,ds\right\}^{\mathrm{T}} + n^{-1}\mathbf{y}_{h,0}\mathbf{y}_{h,0}^{\mathrm{T}}.
\end{aligned}
$$

Part (b), again, follows from the continuous mapping theorem and Lemma 3.2.

For part (c), we use

$$\begin{aligned} n^{-1}\mathbf{R}^{-n}\mathbf{y}_{h,n}\mathbf{y}_{h,n}^{\mathrm{T}}\mathbf{R}^{n} &= n^{-1}(\mathbf{E}_{h,n}+\mathbf{y}_{h,0})(\mathbf{E}_{h,n}+\mathbf{y}_{h,0})^{\mathrm{T}} \\ &= \left(n^{-1/2}\mathbf{E}_{h,n}\right)\left(n^{-1/2}\mathbf{E}_{h,n}\right)^{\mathrm{T}} + \left(n^{-1/2}\mathbf{E}_{h,n}\right)\left(n^{-1/2}\mathbf{y}_{h,0}\right)^{\mathrm{T}} \\ &\quad + \left(n^{-1/2}\mathbf{y}_{h,0}\right)\left(n^{-1/2}\mathbf{E}_{h,n}\right)^{\mathrm{T}} + n^{-1}\mathbf{y}_{h,0}\mathbf{y}_{h,0}^{\mathrm{T}} \\ &= \Upsilon_{h,n}(1)\left[\Upsilon_{h,n}(1)\right]^{\mathrm{T}} + \Upsilon_{h,n}(1)\left(n^{-1/2}\mathbf{y}_{h,0}^{\mathrm{T}}\right) \\ &\quad + n^{-1/2}\mathbf{y}_{h,0}\left[\Upsilon_{h,n}(1)\right]^{\mathrm{T}} + n^{-1}\mathbf{y}_{h,0}\mathbf{y}_{h,0}^{\mathrm{T}}. \end{aligned}$$

The result follows immediately. Finally, for part (d), we first premultiply (3.10) by $\mathbf{R}^{-t}$ to get $\mathbf{R}^{-t}\mathbf{y}_{h,t} = \mathbf{R}^{1-t}\mathbf{y}_{n,t-1} + \mathbf{R}^{-t}\mathbf{e}_{h,t}$, next postmultiply each side by its transpose and then sum over t, yielding

$$\begin{aligned} &n^{-1}\left(\sum_{t=1}^{n}\mathbf{R}^{1-t}\mathbf{y}_{h,t-1}\mathbf{e}_{h,t}^{\mathrm{T}}\mathbf{R}^{t} + \sum_{t=1}^{n}\mathbf{R}^{-t}\mathbf{e}_{h,t}\mathbf{y}_{h,t-1}^{\mathrm{T}}\mathbf{R}^{t-1}\right) \\ &\quad = n^{-1}\left(\mathbf{R}^{-n}\mathbf{y}_{h,n}\mathbf{y}_{h,n}^{\mathrm{T}}\mathbf{R}^{n} - \sum_{t=1}^{n}\mathbf{R}^{-t}\mathbf{e}_{h,t}\mathbf{e}_{h,t}^{\mathrm{T}}\mathbf{R}^{t} - \mathbf{y}_{h,0}\mathbf{y}_{h,0}^{\mathrm{T}}\right). \end{aligned}$$

Part (d) then follows from part (c). Here the existence of $\mathbf{F}$ is implied by $\mathbf{e}_{h,t}$ satisfying (1.2) and the properties of trigonometric series. □

Note that in Lemma 3.3 we may replace the upper limit of the summation n by $[ns]$ for $0 < s \leq 1$ and obtain general results of the limiting distribution. For instance, suppose that $\omega = 0$, i.e., the unit root case. Then the general results become

(a) $$n^{-3/2}\sum_{t=1}^{[ns]} Y_{ht} \to_d \sigma_h \int_0^s W(t)\,dt,$$

(b) $$n^{-2}\sum_{t=1}^{[ns]} Y_{ht}^2 \to_d \sigma_h^2 \int_0^s W^2(t)\,dt,$$

(c) $$n^{-1}Y_{h[ns]}^2 \to_d \sigma_h^2 W^2(s),$$

where $0 < s \leq 1$, σ_h is the standard deviation of $\mathbf{e}_{h,t}$ and $W(t)$ is a standard Brownian motion.

From the proof of Lemma 3.3, it is clear that the starting value $\mathbf{y}_{h,0}$ is immaterial in studying the limiting distribution as long as $\mathbf{y}_{h,0}$ has a well-defined distribution function. For this reason, we assume $\mathbf{y}_0 = \mathbf{O}$ for the rest of this paper. This observation has been made by Phillips (1987) in a study concerning asymptotic properties of univariate random walks.

Next consider the general process $\mathbf{y}_{v,t}$. Define

(3.21) $$\mathbf{K}_{v,t} = \sum_{j=1}^{t}\mathbf{R}^{-j}\mathbf{y}_{v,j} \quad \text{for } v = 1, 2, \ldots, h.$$

Then, from (3.13) and (3.14) and (3.19),

$$\mathbf{R}^{-t}\mathbf{y}_{h,t} = \mathbf{E}_{h,t}, \tag{3.22}$$

$$\mathbf{R}^{-t}\mathbf{y}_{v,t} = \mathbf{R}^{-1}\mathbf{K}_{v+1,t-1} + \mathbf{E}_{v,t} \quad \text{for } v = h-1,\dots,1. \tag{3.23}$$

Based on the recursion (3.22) and (3.23), we define a continuous recursion for $0 \le t \le 1$,

$$\Gamma_h(t) = \Xi_h(t), \tag{3.24}$$

$$\Gamma_v(t) = \mathbf{R}^{-1}\int_0^t \Gamma_{v+1}(s)\,ds \quad \text{for } v = h-1,\dots,1, \tag{3.25}$$

where the integration is a componentwise operation and $\Xi_h(t)$ is defined in Lemma 3.2. Note that $\Gamma_v(t)$ depends only on $\Xi_h(t)$ which comes from $\mathbf{e}_{h,t}$. This occurs because the multiplicity of a nonstationary characteristic root accumulates from $\mathbf{e}_{h,t}$.

To obtain a general result similar to Lemma 3.3(d), we make use of the following result of Chan and Wei (1988).

LEMMA 3.4. *Let* $\{X_v\}$ *and* $\{Y_v\}$ *be two sequences of random variables. Define*

$$U_n(s) = \sum_{v=1}^{[ns]} X_v,\ V_n(s) = \sum_{v=1}^{[ns]} Y_v \quad \textit{and} \quad T_n(s) = X_{[ns]} \quad \textit{for } 0 \le s \le 1.$$

(a) *Suppose that there exists a sequence* $\{k_n\}$ *with* $k_n \to \infty$ *such that*

$$\left(k_n^{-1}T_n, n^{-1/2}V_n\right) \to_d (T, W_1),$$

where $W_1(t)$ *is a standard Brownian motion with respect to an increasing sequence of σ-fields* G_t *and* T *is* G_t*-adaptive. Then*

$$\left(n^{-1/2}V_n, n^{-1}k_n^{-1}U_n, n^{-3/2}k_n^{-1}\sum_{v=1}^{n-1} U_n(v/n)Y_{v+1}\right) \to_d \left(W_1, H, \int_0^1 H\,dW_1\right),$$

where $H(t) = \int_0^t T(s)\,ds$.

(b) *Suppose that there exists an increasing sequence* F_v *of σ-fields such that* $(X_v, Y_v)^{\mathrm{T}}$ *is a sequence of martingale differences with respect to* F_v*. Moreover,* $E(X_v^2 + Y_v^2 | F_{v-1})$ *is uniformly bounded almost surely and* $n^{-1/2}(U_n, V_n) \to_d (W_2, W_1)$*, where* W_1 *and* W_2 *are two Brownian motions with respect to an increasing sequence of σ-fields* G_t*. Then*

$$n^{-1}\sum_{v=1}^{n-1} U_n(v/n)Y_{v+1} \to_d \int_0^1 W_2\,dW_1.$$

Using the above results and notation, we show the main result of this section.

THEOREM 3.1. *Suppose that* $\mathbf{y}_{v,t}$ *follows the purely nonstationary* $AR(1)$ *model of* (3.10) *and* (3.11) *obtained from the vector ARMA model* (1.1) *by the transformation of Section* 2 *and that* $\mathbf{a}_t$ *satisfies the martingale difference* *condition* (1.2). *Then, for* $v = h, h - 1, \dots, 1$,

(a) $$n^{-(h-v+1)-1/2} \sum_{t=1}^{n} \mathbf{R}^{-t} \mathbf{y}_{v,t} \to_d \int_0^1 \Gamma_v(s)\, ds,$$

(b) $$n^{-(2h-v-u+2)} \sum_{t=1}^{n} \mathbf{R}^{-t} \mathbf{y}_{v,t} \mathbf{y}_{u,t}^{\mathrm{T}} \mathbf{R}^{t} \to_d \int_0^1 \Gamma_v(s) [\Gamma_u(s)]^{\mathrm{T}}\, ds$$

for $v \le u \le h$,

(c) $$n^{-2(h-v+1)+1} \mathbf{R}^{-n} \mathbf{y}_{v,n} \mathbf{y}_{v,n}^{\mathrm{T}} \mathbf{R}^{n} \to_d \Gamma_v(1) [\Gamma_v(1)]^{\mathrm{T}},$$

(d) $$n^{-(h-v+1)} \sum_{t=1}^{n} \mathbf{R}^{1-t} \mathbf{y}_{v,t-1} \mathbf{e}_{u,t}^{\mathrm{T}} \mathbf{R}^{t} \to_d \int_0^1 \Gamma_v(s)\, d[\Xi_u(s)]^{\mathrm{T}}$$

for $v \le u \le h$.

Proof. We prove this theorem by backward induction with the help of Lemmas 3.3 and 3.4. For $v = h$, parts (a) through (c) are given by Lemma 3.3, while part (d) follows from Lemmas 3.2 and 3.4(b).

Next, consider $v = h - 1$. For part (a), by (3.22) and (3.21) we have

$$\begin{aligned} n^{-5/2} \sum_{t=1}^{n} \mathbf{R}^{-t} \mathbf{y}_{h-1,t} &= n^{-5/2} \left(\mathbf{R}^{-1} \sum_{t=1}^{n} \mathbf{K}_{h,t-1} + \sum_{t=1}^{n} \mathbf{E}_{h-1,t} \right) \\ &= n^{-1} \sum_{t=1}^{n} \mathbf{R}^{-1} \left[n^{-3/2} \left(\sum_{j=1}^{t-1} \mathbf{R}^{-j} \mathbf{y}_{h,j} \right) \right] + n^{-2} \left(n^{-1/2} \sum_{t=1}^{n} \mathbf{E}_{h-1,t} \right) \\ &= n^{-1} \sum_{t=1}^{n} \mathbf{R}^{-1} \left[n^{-3/2} \left(\sum_{j=1}^{[ns]} \mathbf{R}^{-j} \mathbf{y}_{h,j} \right) \right] + o_p(1) \\ &\qquad \text{with } s = (t-1)/n \\ &\to_d \int_0^1 \Gamma_{h-1}(s)\, ds. \end{aligned}$$

In the above, $o_p(1)$ is obtained by Lemma 3.2 and the last step is based on Lemma 3.3(a) and the continuous mapping theorem.

For part (b), there are two possible values of u: $u = h - 1$ or $u = h$. We shall only demonstrate the case of $u = v = h - 1$ because the same techniques apply to the other case. By (3.23),

$$\begin{aligned} &n^{-4} \sum_{t=1}^{n} \mathbf{R}^{-t} \mathbf{y}_{h-1,t} \mathbf{y}_{h-1,t}^{\mathrm{T}} \mathbf{R}^{t} \\ &= n^{-4} \sum_{t=1}^{n} \left(\mathbf{R}^{-1} \mathbf{K}_{h,t-1} + \mathbf{E}_{h-1,t} \right) \left(\mathbf{R}^{-1} \mathbf{K}_{h,t-1} + \mathbf{E}_{h-1,t} \right)^{\mathrm{T}} \\ &= n^{-1} \sum_{t=1}^{n} \mathbf{R}^{-1} \left(n^{-3/2} \sum_{j=1}^{t-1} \mathbf{R}^{-j} \mathbf{y}_{h,j} \right) \left(n^{-3/2} \sum_{j=1}^{t-1} \mathbf{R}^{-j} \mathbf{y}_{h,j} \right)^{\mathrm{T}} \mathbf{R} + o_p(1) \\ &\to_d \int_0^1 \Gamma_{h-1}(s) [\Gamma_{h-1}(s)]^{\mathrm{T}}\, ds. \end{aligned}$$

For part (c), notice that

$$n^{-3}\mathbf{R}^{-n}\mathbf{y}_{h-1,n}\mathbf{y}_{h-1,n}^{\mathrm{T}}\mathbf{R}^{n} = n^{-3}\big(\mathbf{R}^{-1}\mathbf{K}_{h,n-1} + \mathbf{E}_{h-1,n}\big)\big(\mathbf{R}^{-1}\mathbf{K}_{h,n-1} + \mathbf{E}_{h-1,n}\big)^{\mathrm{T}}$$
$$= \mathbf{R}^{-1}\left(n^{-3/2}\sum_{j=1}^{n-1}\mathbf{R}^{-j}\mathbf{y}_{h,j}\right)\left(n^{-3/2}\sum_{j=1}^{n-1}\mathbf{R}^{-j}\mathbf{y}_{h,j}\right)^{\mathrm{T}} + o_p(1)$$
$$\to_d \Gamma_{h-1}(1)\left[\Gamma_{h-1}(1)\right]^{\mathrm{T}}.$$

Finally, part (d) follows from part (a) and Lemmas 3.2 and 3.4. The results thus hold for $v = h - 1$. By the backward induction, the proof is complete. □

In Theorem 3.1, it is understood that $\mathbf{e}_{h,t}$ satisfies (1.2) and that part (d) changes to convergence in probability to 0 when $\mathbf{e}_{u,t}$ is 0. Note that parts (a) through (c) only depend on $\mathbf{e}_{h,t}$ and on the corresponding eigenvalue being on the unit circle.

4. Applications to vector AR(1) regressions. In this section we apply the results of Theorem 3.1 to derive asymptotic distributions of LS estimates of autoregressions of a purely nonstationary AR(1) process. The approach of Chan and Wei (1988) for obtaining the joint limiting distribution of the estimates is adopted. More specifically, using the result in Appendix III of Chan and Wei's paper, we may establish the joint results by investigating separately Jordan blocks with different nonstationary characteristic roots.

For a time series $\mathbf{y}_t$ and the AR(1) regression

$$\mathbf{y}_t = \boldsymbol{\beta}\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t \quad \text{for } t = 1, 2, \dots, n,$$

we denote the LS estimate of $\boldsymbol{\beta}^{\mathrm{T}}$ by $\hat{\boldsymbol{\beta}}_n^{\mathrm{T}} = (\sum_{t=1}^{n}\mathbf{y}_{t-1}\mathbf{y}_{t-1}^{\mathrm{T}})^{-1}(\sum_{t=1}^{n}\mathbf{y}_{t-1}\mathbf{y}_t^{\mathrm{T}})$. To gain insight into the result and obtain complete formulas, we first consider each individual Jordan block.

4.1. *Individual Jordan block*

Case 1: *Jordan block with eigenvalue* 1 *or* −1. Denote the eigenvalue by λ. In this case, $\omega = 0$ or π, (3.10) and (3.11) reduce to (3.1) and (3.2) and the model of $\mathbf{y}_t = (y_{1t}, \dots, y_{ht})^{\mathrm{T}}$ is

$$\mathbf{y}_t = \mathbf{D}\mathbf{y}_{t-1} + \mathbf{e}_t \quad \text{or}$$
$$y_{ht} = \lambda y_{h,t-1} + e_{ht} \quad \text{and} \quad y_{vt} = \lambda y_{v,t-1} + y_{v+1,t-1} + e_{vt}, \tag{4.1}$$

for $v = h - 1, \dots, 1$. To apply Theorem 3.1, we simply note that the elements of matrices in the right-hand side of

$$\hat{\boldsymbol{\beta}}_n^{\mathrm{T}} - \mathbf{D}^{\mathrm{T}} = \left(\sum_{t=1}^{n}\mathbf{y}_{t-1}\mathbf{y}_{t-1}^{\mathrm{T}}\right)^{-1}\left(\sum_{t=1}^{n}\mathbf{y}_{t-1}\mathbf{e}_t^{\mathrm{T}}\right)$$

are exactly the same as those in parts (b) and (d) of Theorem 3.1. Therefore, let $W_i(t)$ be the limiting standard Brownian motion of $(n^{1/2}\sigma_i)^{-1}\sum\lambda^t e_{it}$ and denote the recursion (3.24) and (3.25) by

(4.2) $$F_h(t) = W_h(t) \text{ and } F_v(t) = \lambda \int_0^t F_{v+1}(s)\, ds \quad \text{for } v = h-1, \dots, 1.$$

Also, define the notation:

$\mathbf{V} = (\mathbf{V}_1, \dots, \mathbf{V}_h)_{h\times h}$ where the ith column $\mathbf{V}_i$ is defined by

$$\mathbf{V}_i = \lambda \sigma_h \sigma_i \left(\int_0^1 F_1(t)\, dW_i(t), \int_0^1 F_2(t)\, dW_i(t), \dots, \int_0^1 F_h(t)\, dW_i(t) \right)^{\mathrm{T}},$$

$$\Omega = (\Omega_{ij})_{h\times h} \quad \text{where } \Omega_{ij} = \sigma_h^2 \int_0^1 F_i(t) F_j(t)\, dt,$$

$$\mathbf{L}_n = \operatorname{diag}\{n^h, n^{h-1}, \dots, n\},$$

where σ_v^2 is the variance of e_{vt}. Then, by Theorem 3.1, we have the following result.

THEOREM 4.1. *Suppose that* $\mathbf{y}_t$ *follows the model* (4.1) *which is obtained from* (1.1) *by the transformation of Section* 2 *and that the innovational series* $\mathbf{a}_t$ *of* (1.1) *satisfies* (1.2). *Then*

(a) $$\mathbf{L}_n^{-1}\left(\sum_{t=1}^{n} \mathbf{y}_{t-1}\mathbf{y}_{t-1}^{\mathrm{T}} \right)\mathbf{L}_n^{-1} \to_d \Omega,$$

(b) $$\mathbf{L}_n^{-1} \sum_{t=1}^{n} \mathbf{y}_{t-1}\mathbf{e}_t^{\mathrm{T}} \to_d \mathbf{V},$$

(c) $$\mathbf{L}_n(\hat{\beta}_n - \mathbf{D})^{\mathrm{T}} \to_d \Omega^{-1}\mathbf{V},$$

(d) $$\hat{\beta}_n \to_p \mathbf{D}.$$

Two remarks are in order. First, the identity

$$y_{v,t-1} e_{ut} = -(-1)^{t-1} y_{v,t-1} (-1)^t e_{ut}$$

is used in defining $\mathbf{V}_i$ when $\lambda = -1$. Second, for a fixed integer h, the random matrix Ω for $\lambda = 1$ and that of $\lambda = -1$ are equivalent in distribution. This can easily be seen from the definition. Similarly, $\mathbf{V}$ for $\lambda = 1$ and that of $\lambda = -1$ are equivalent in distribution. Consequently, there is only a sign change between the two limiting results. This property has been observed by Fuller (1976) and Chan and Wei (1988) for the univariate processes.

Case 2: *Jordan blocks with eigenvalues* $\exp(i\omega)$ *and* $\exp(-i\omega)$. Here Theorem 3.1 is not directly applicable because the matrix $\mathbf{R}$ is nontrivial. We need two lemmas.

LEMMA 4.1. *Let* $\mathbf{A} = \{a_{ij}\}$ *be a real-valued* 2×2 *matrix. Then*

$$\mathbf{R}^t \mathbf{A} \mathbf{R}^{-t} = \mathbf{A}_1 + \cos(2t\omega)\mathbf{A}_2 - \sin(2t\omega)\mathbf{A}_3,$$

where

$$2\mathbf{A}_1 = \begin{bmatrix} a_{11} + a_{22} & a_{12} - a_{21} \\ a_{21} - a_{12} & a_{11} + a_{22} \end{bmatrix},$$

$$2\mathbf{A}_2 = \begin{bmatrix} a_{11} - a_{22} & a_{12} + a_{21} \\ a_{12} + a_{21} & a_{22} - a_{11} \end{bmatrix},$$

$$2\mathbf{A}_3 = \begin{bmatrix} a_{21} + a_{12} & a_{22} - a_{11} \\ a_{22} - a_{11} & -a_{12} - a_{21} \end{bmatrix}.$$

LEMMA 4.2. *For $u, v = 1, \ldots, h$ and $0 < \omega < \pi$,*

(a) $$\sup_{\{1 \le t \le n\}} \left\| \sum_{j=1}^{t} \exp(ij\omega)\mathbf{E}_{v,j} \right\| = o_p(n^{3/2}),$$

(b) $$\sup_{\{1 \le t \le n\}} \left\| \sum_{j=1}^{t} \exp(ij\omega)\mathbf{E}_{v,j}\mathbf{E}_{u,j}^{\mathrm{T}} \right\| = o_p(n^{2}),$$

(c) $$\sup_{\{1 \le t \le n\}} \left\| \sum_{j=1}^{t} \exp(ij\omega)\mathbf{K}_{v,j} \right\| = o_p(n^{(h-v+1)+3/2}),$$

(d) $$\sup_{\{1 \le t \le n\}} \left\| \sum_{j=1}^{t} \exp(ij\omega)\mathbf{K}_{v,j}\mathbf{K}_{u,j}^{\mathrm{T}} \right\| = o_p(n^{2h-u-v+2}),$$

where the matrix norm is defined as $\|\mathbf{c}\| = \max_u\{\sum_v |c_{uv}|\}$.

PROOF. Lemma 4.1 can be proved by identities of trigonometric series and direct algebraic calculation, while Lemma 4.2 follows from Lemma 3.3.6 of Chan and Wei (1988) with the help of the triangular and Cauchy–Schwarz inequalities. □

Now consider the least squares AR(1) regression of $\mathbf{y}_t$. For $v, u = 1, \ldots, h$, define

(4.3) $$\mathbf{H}_n = \sum_{t=1}^{n} \mathbf{y}_{t-1}\mathbf{y}_{t-1}^{\mathrm{T}} = [\mathbf{H}_n(v,u)] \quad \text{where } \mathbf{H}_n(v,u) = \sum_{t=1}^{n} \mathbf{y}_{v,t-1}\mathbf{y}_{u,t-1}^{\mathrm{T}},$$

(4.4) $$\mathbf{Q}_n = \sum_{t=1}^{n} \mathbf{y}_{t-1}\mathbf{e}_t^{\mathrm{T}} = [\mathbf{Q}_n(v,u)] \quad \text{where } \mathbf{Q}_n(v,u) = \sum_{t=1}^{n} \mathbf{y}_{v,t-1}\mathbf{e}_{ut}^{\mathrm{T}},$$

and rewrite $\mathbf{H}_n(v,u)$ and $\mathbf{Q}_n(v,u)$ as

$$\mathbf{H}_n(v,u) = \sum_{t=1}^{n} \mathbf{R}^{t-1}(\mathbf{R}^{-t+1}\mathbf{y}_{v,t-1}\mathbf{y}_{u,t-1}^{\mathrm{T}}\mathbf{R}^{t-1})\mathbf{R}^{1-t},$$

$$\mathbf{Q}_n(v,u) = \sum_{t=1}^{n} \mathbf{R}^{t-1}(\mathbf{R}^{-t+1}\mathbf{y}_{v,t-1}\mathbf{e}_{u,t}^{\mathrm{T}}\mathbf{R}^{t})\mathbf{R}^{-t+1}\mathbf{R}^{-1}.$$

For $v, u = 1, \dots, h$, define the 2×2 matrices:

$$(4.5)\quad \mathbf{P}(v,u) = \int_0^1 \Gamma_v(s)[\Gamma_u(s)]^{\mathrm{T}}\,ds \quad\text{and}\quad \mathbf{O}(v,u) = \int_0^1 \Gamma_v(s)\,d[\Xi_u(s)]^{\mathrm{T}},$$

where $\Gamma_v(s)$ and $\Xi_u(s)$ are defined in (3.25) and Lemma 3.2, respectively. For a given 2×2 matrix $\mathbf{A} = (a_{ij})$, define an operation ∇ by

$$(4.6)\qquad \nabla\mathbf{A} = \mathbf{A}_1,$$

where $\mathbf{A}_1$ is defined in Lemma 4.1. Using this operation, we further define two $2h \times 2h$ matrices:

$$(4.7)\qquad \mathbf{H} = [\mathbf{H}(v,u)] \quad \text{with } \mathbf{H}(v,u) = \nabla\mathbf{P}(v,u),$$

$$(4.8)\qquad \mathbf{Q} = [\mathbf{Q}(v,u)] \quad \text{with } \mathbf{Q}(v,u) = [\nabla\mathbf{O}(v,u)]\mathbf{R}^{-1},$$

where $v, u = 1, \dots, h$. Finally, let $\mathbf{L}_n = \operatorname{diag}\{n^h, n^h, n^{h-1}, n^{h-1}, \dots, n, n\}$. We may apply Lemmas 4.1 and 4.2 and Theorem 3.1 to obtain the asymptotic properties of $\hat{\boldsymbol{\beta}}$.

THEOREM 4.2. *Suppose that* $\mathbf{y}_t$ *follows the model* (3.10) *and* (3.11) *with* $0 < \omega < \pi$ *that is transformed from model* (1.1) *by the transformation of Section* 2. *Also, assume that* $\mathbf{a}_t$ *of* (1.1) *satisfies* (1.2). *Then*

(a) $$\mathbf{L}_n^{-1}\left(\sum_{t=1}^{n} \mathbf{y}_{t-1}\mathbf{y}_{t-1}^{\mathrm{T}}\right)\mathbf{L}_n^{-1} \to_d \mathbf{H},$$

(b) $$\mathbf{L}_n^{-1} \sum_{t=1}^{n} \mathbf{y}_{t-1}\mathbf{e}_t^{\mathrm{T}} \to_d \mathbf{Q},$$

(c) $$\mathbf{L}_n(\hat{\boldsymbol{\beta}}_n - \mathbf{D})^{\mathrm{T}} \to_d \mathbf{H}^{-1}\mathbf{Q},$$

(d) $$\hat{\boldsymbol{\beta}}_n \to_p \mathbf{D},$$

where $\mathbf{H}$ *and* $\mathbf{Q}$ *are defined in* (4.7) *and* (4.8) *and* $\mathbf{D}$ *is the Jordan block of model* (3.10) *and* (3.11).

In Theorems 4.1 and 4.2, we have used the fact that $\boldsymbol{\Omega}$ and $\mathbf{H}$ are nonsingular almost surely. This is shown in Chan and Wei (1988) for the univariate case. For the multivariate case, if the characteristic root is 1 or -1, the univariate result is applicable because the limiting property is determined by the multiplicity of the root, and within an individual Jordan block the multiplicity is the same as that of the univariate situation. For complex roots, the univariate result is not directly applicable. However, the same idea can be used to show that $\mathbf{H}$ is nonsingular almost surely. Details can be found in an appendix of Tsay and Tiao (1986).

4.2. *The general Jordan matrix.* We now turn to the situation where eigenvalues of various types exist simultaneously. Clearly, it suffices to consider the case where the Jordan matrix of the nonstationary part $\mathbf{N}_t$ of (2.13) is in the form

$$\mathbf{D} = \operatorname{diag}\{\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_8\}, \tag{4.9}$$

where $\mathbf{D}_i$ is an $h_i \times h_i$ Jordan block such that the eigenvalues of $\mathbf{D}_1$ and $\mathbf{D}_2$ are 1, those of $\mathbf{D}_3$ and $\mathbf{D}_4$ are -1, those of $\mathbf{D}_5$ and $\mathbf{D}_6$ are $\exp(i\omega_1)$ and $\exp(-i\omega_1)$, respectively, and those of $\mathbf{D}_7$ and $\mathbf{D}_8$ are $\exp(i\omega_2)$ and $\exp(-i\omega_2)$, respectively, with $0 < \omega_1, \omega_2 < \pi$. Partition $\mathbf{Y}_t = (\mathbf{Y}_{1t}^{\mathrm{T}}, \ldots, \mathbf{Y}_{8t}^{\mathrm{T}})^{\mathrm{T}}$ and $\mathbf{f}_t = (\mathbf{f}_{1t}^{\mathrm{T}}, \ldots, \mathbf{f}_{8t}^{\mathrm{T}})^{\mathrm{T}}$ according to the dimensions of $\mathbf{D}_j$'s. Since $\mathbf{Y}_{5t}$ and $\mathbf{Y}_{6t}$ are conjugate pairs, one may rearrange their elements into the form of (3.7) and (3.8) based on the multiplicities of $\exp(i\omega_1)$ and $\exp(-i\omega_1)$ and transform the joint process $(\mathbf{Y}_{5t}^{\mathrm{T}}, \mathbf{Y}_{6t}^{\mathrm{T}})^{\mathrm{T}}$ into a real-valued process $\mathbf{y}_{5t}$ by using the $\mathbf{r}$ matrix of (3.9). Here the transformation matrix is of the form $\operatorname{diag}\{\mathbf{r}, \mathbf{r}, \ldots, \mathbf{r}\}$. The same technique applies to $\mathbf{Y}_{7t}$ and $\mathbf{Y}_{8t}$ and we denote the resulting real-valued process by $\mathbf{y}_{6t}$. Consequently, the complex-valued process $\mathbf{Y}_t$ can be transformed into a real-valued one $\mathbf{y}_t = (\mathbf{y}_{1t}^{\mathrm{T}}, \ldots, \mathbf{y}_{6t}^{\mathrm{T}})^{\mathrm{T}}$, where $\mathbf{y}_{vt} = \mathbf{Y}_{vt}$ for $v = 1, \ldots, 4$. Let $\mathbf{e}_t = (\mathbf{e}_{1t}^{\mathrm{T}}, \ldots, \mathbf{e}_{6t}^{\mathrm{T}})^{\mathrm{T}}$ be the corresponding transformation of the innovation $\mathbf{f}_t$.

From the results of the preceding subsection, asymptotic distributions for each individual Jordan block are available. Therefore, we shall concentrate here on the results *between* Jordan blocks, i.e., the cross-product terms. Moreover, since the limiting distributions concerning $\Sigma \mathbf{y}_{t-1}\mathbf{e}_t^{\mathrm{T}}$ are available, we need only consider the cross-product term of the $\Sigma \mathbf{y}_{t-1}\mathbf{y}_{t-1}^{\mathrm{T}}$ matrix. For convenience, we write the model of $\mathbf{y}_t$ as

$$\mathbf{y}_t = \mathbf{U}\mathbf{y}_{t-1} + \mathbf{e}_t \quad \text{with } \mathbf{U} = \operatorname{diag}\{\mathbf{U}_1, \ldots, \mathbf{U}_6\} \tag{4.10}$$

and define

$$\mathbf{L}_{v,n} = \operatorname{diag}\{n^h, n^{h-1}, \ldots, n^2, n\} \quad \text{with } h = \dim(\mathbf{y}_{vt}) \text{ for } v = 1,2,3,4, \tag{4.11}$$

$$\mathbf{L}_{v,n} = \operatorname{diag}\{n^h, n^h, n^{h-1}, n^{h-1}, \ldots, n, n\} \quad \text{with } h = \dim(\mathbf{y}_{vt})/2 \text{ for } v = 5,6. \tag{4.12}$$

Case 1: *Cross products of Jordan blocks with different eigenvalues.* The main result here is that cross products between different types of eigenvalues converge to 0 when they are normalized by proper orders. This result greatly simplifies the limiting distribution considered in this paper.

THEOREM 4.3. *Suppose that* $\mathbf{y}_t$ *follows the model* (4.10) *obtained from model* (1.1) *by the transformation of Section* 2 *and that* $\mathbf{a}_t$ *of* (1.1) *satisfies* (1.2). *Then*

(a) $\mathbf{L}_{i,n}^{-1} \sum_{t=1}^{n} \mathbf{y}_{i,t-1}\mathbf{y}_{j,t-1}^{\mathrm{T}}\mathbf{L}_{j,n}^{-1} \to_p \mathbf{O}$ *for* $i = 1,2$ *and* $j = 3,4,5,6$,

(b) $\mathbf{L}_{i,n}^{-1} \sum_{t=1}^{n} \mathbf{y}_{i,t-1}\mathbf{y}_{j,t-1}^{\mathrm{T}}\mathbf{L}_{j,n}^{-1} \to_p \mathbf{O}$ *for* $i = 3,4$ *and* $j = 5,6$,

(c) $\mathbf{L}_{5,n}^{-1} \sum_{t=1}^{n} \mathbf{y}_{5,t-1}\mathbf{y}_{6,t-1}^{\mathrm{T}}\mathbf{L}_{6,n}^{-1} \to_p \mathbf{O}$ *if* $\omega_1 \neq \omega_2$.

PROOF. This theorem can be proved by using the same techniques as those of Theorem 3.4.1 of Chan and Wei (1988) or as those of Lemmas A.2 of Tiao and Tsay (1983a). □

Case 2: *Cross products of Jordan blocks with the same eigenvalue.* The cross products in this case do not converge to 0. Instead, they possess certain limiting distributions which, again, can be derived by using the results of Section 3. Consequently, to obtain the properties of least squares estimates we must show that the cross products do not affect the nonsingularity of the limit of the $\Sigma \mathbf{y}_{t-1}\mathbf{y}_{t-1}^{\mathrm{T}}$ matrix. For simplicity in presentation, we again separate the problem into three cases according to the type of eigenvalues. Furthermore, since the basic idea is the same for all eigenvalues, we only provide details for the case where the eigenvalue is unity.

Case 2(a): *Eigenvalues are* 1. Here we consider the cross product between $\mathbf{y}_{1t}$ and $\mathbf{y}_{2t}$. That is, consider the matrix $\Sigma \mathbf{y}_{1,t-1}\mathbf{y}_{2,t-1}^{\mathrm{T}}$. To simplify the notation, we denote the dimensions of $\mathbf{y}_{1t}$ and $\mathbf{y}_{2t}$ by h and g, respectively, and the last components of the innovations $\mathbf{e}_{1t}$ and $\mathbf{e}_{2t}$ by $e_{1,h,t}$ and $e_{2,g,t}$. From (3.24) and (3.25) and (4.2), it is clear that the limiting distributions of the statistics of $\mathbf{y}_{1t}$ discussed depend only on the Brownian motion generated by $e_{1,h,t}$. Similarly, the Brownian motion corresponding to $e_{2,g,t}$ governs the limiting property of statistics of $\mathbf{y}_{2t}$. Therefore, it suffices in this section to concentrate on these two last components.

Since the eigenvalues are 1, the results of Section 4.1 are applicable. Therefore, similar to (4.2) with $\lambda = 1$, define for $0 \le t \le 1$,

$$F_{1,h}(t) = W_{1,h}(t) \text{ and } F_{1,v}(t) = \int_0^1 F_{1,v+1}(s)\,ds \quad \text{for } v = h-1,\dots,1,$$

$$F_{2,g}(t) = W_{2,g}(t) \text{ and } F_{2,u}(t) = \int_0^1 F_{2,u+1}(s)\,ds \quad \text{for } u = g-1,\dots,1,$$

where $W_{1,h}(t)$ and $W_{2,g}(t)$ are the limiting standard Brownian motions of $(n^{1/2}\sigma_{1,h})^{-1}\Sigma e_{1,h,t}$ and $(n^{1/2}\sigma_{2,g})^{-1}\Sigma e_{2,g,t}$, respectively. Here $\sigma_{1,h}$ and $\sigma_{2,g}$ are the standard deviations of $e_{1,h,t}$ and $e_{2,g,t}$.

LEMMA 4.3. *Suppose that $\mathbf{y}_{1t}$ and $\mathbf{y}_{2t}$ follow the models* (4.10) *with eigenvalue being unity. These processes are obtained from model* (1.1) *by the transformation of Section* 2. *Also, assume that $\mathbf{a}_t$ of* (1.1) *satisfies* (1.2). *Then*

$$n^{-(h+g-v-u+2)} \sum_{t=1}^{n} y_{1,v,t-1} y_{2,u,t-1} \to_d \sigma_{1,h}\sigma_{2,g} \int_0^1 F_{1,v}(s)F_{2,u}(s)\,ds,$$

for $1 \le v \le h$ and $1 \le u \le g$, where $y_{1,v,t}$ is the vth component of $\mathbf{y}_{1t}$ and $y_{2,u,t}$ is the uth component of $\mathbf{y}_{2t}$.

PROOF. This lemma can be proved by using Lemma 3.2 and the same techniques as those of Theorem 3.1. In fact, it is a generalization of Theorem 3.1(b). □

Next, to gain further insight into the result of cross products, we consider the relation between $e_{1,h,t}$ and $e_{2,g,t}$. By Lemma 2.3, we have

$$e_{2,g,t} = \beta e_{1,h,t} + \varepsilon_t, \tag{4.13}$$

where $\beta = \operatorname{cov}(e_{1,h,t}, e_{2,g,t})/\operatorname{var}(e_{1,h,t})$ and $\{\varepsilon_t\}$ is a sequence of martingale difference such that $\operatorname{var}(\varepsilon_t) \neq 0$ and $\{\varepsilon_t\}$ and $\{e_{1,h,t}\}$ are uncorrelated, i.e.,

$$E(e_{1,h,t}\varepsilon_j) = 0 \quad \text{for all } t \text{ and } j.$$

By (4.13) and Lemma 3.2, we may rewrite $W_{2,g}(t)$ as a linear combination of two independent standard Brownian motions, i.e.,

$$F_{2,g}(t) = W_{2,g}(t) = \frac{\sigma_{1,h}}{\sigma_{2,g}}\beta W_{1,h}(t) + \frac{\sigma_\varepsilon}{\sigma_{2,g}}W_\varepsilon(t), \tag{4.14}$$

where σ_ε is the standard deviation of ε_t and $W_\varepsilon(t)$ is the limiting standard Brownian motion of $(n^{1/2}\sigma_\varepsilon)^{-1}\Sigma\varepsilon_t$. Note that the independence of $W_{1,h}(t)$ and $W_\varepsilon(t)$ follows from the fact that $\{e_{1,h,t}\}$ and $\{\varepsilon_t\}$ are uncorrelated. Using (4.14), $F_{2,u}(t)$ can be rewritten as a linear combination of stochastic integrals of $W_{1,h}(t)$ and $W_\varepsilon(t)$.

Now, let $\mathbf{Z}_t = (\mathbf{y}_{1t}^{\mathrm{T}}, \mathbf{y}_{2t}^{\mathrm{T}})^{\mathrm{T}}$. By Theorem 4.1 and Lemma 4.3, the limit of $\Sigma\mathbf{Z}_t\mathbf{Z}_t^{\mathrm{T}}$ exists after proper normalization. Denote the limit by

$$\Omega = (\Omega_{ij}) \quad \text{where } \Omega_{ij} \text{ is the limit of } \mathbf{L}_{i,n}^{-1}\sum_{t=1}^{n}\mathbf{y}_{i,t-1}\mathbf{y}_{j,t-1}^{\mathrm{T}}\mathbf{L}_{j,n}^{-1},$$

where $L_{i,n}$ is defined in (4.11) and $i, j = 1, 2$.

LEMMA 4.4. *Suppose that* $\mathbf{y}_{1t}$ *and* $\mathbf{y}_{2t}$ *satisfy the conditions of Lemma* 4.3. *Then,* Ω *is nonsingular almost surely.*

PROOF. Let $S = \{\omega\colon W_{1,h}(\omega, s)$ and $W_\varepsilon(\omega, s)$ are continuous and nondifferentiable for $0 \le s \le 1\}$. It is well known that $\Pr(S) = 1$. We shall show that $\Omega(\omega)$ is nonsingular for any ω in S. Let $m = h + g$. If the contrary holds, then there exist an ω in S and a vector $\mathbf{c} = (c_1, \dots, c_m)^{\mathrm{T}} \neq \mathbf{O}$ such that $\mathbf{c}^{\mathrm{T}}\Omega\mathbf{c} = 0$, i.e.,

$$\int_0^1\left[\sum_{j=1}^{h} c_jF_{1,j}(\omega, s) + \sum_{j=h+1}^{m} c_jF_{2,j-h}(\omega, s)\right]^2 ds = 0.$$

By the choice of ω, $\Sigma_{j=1}^{h}c_jF_{1,j}(\omega, s) + \Sigma_{j=h+1}^{m}c_jF_{2,j-h}(\omega, s)$ is a continuous function in s. Hence, the above equation implies

$$\sum_{j=1}^{h} c_jF_{1,j}(\omega, s) + \sum_{j=h+1}^{m} c_jF_{2,j-h}(\omega, s) = 0 \quad \text{for } 0 \le s \le 1.$$

Next, since Ω_{11} is nonsingular, one element of $\{c_1, \dots, c_h\}$ must be nonzero. Similarly, one element of $\{c_{h+1}, \dots, c_m\}$ must be nonzero. Let $v_1 = \max\{j\colon c_j \neq 0 \text{ and } 1 \le j \le h\}$ and $v_2 = \max\{j\colon c_{j+h} \neq 0 \text{ and } 1 \le j \le g\}$. Now the technique used in the proof of Lemma 3.1.1 of Chan and Wei (1988) can be employed to prove the lemma. For instance, suppose that $v = \max\{v_1, v_2\} = v_1$.

Then, one may write $F_{1,v}(\omega, s)$ as a linear combination of $F_{1,j}(\omega, s)$ for $1 \le j < v$ and $F_{2,j+h}(\omega, s)$ for $1 \le j \le v_2$. By differentiating the linear combination $(h - v)$ times, a contradiction is realized. The same technique applies if $v = v_2$. □

COROLLARY 4.4. *For the* $\mathbf{y}_{1t}$ *and* $\mathbf{y}_{2t}$ *of Lemma* 4.4, *let* $\mathbf{Z}_t = (\mathbf{y}_{1t}^{\mathrm{T}}, \mathbf{y}_{2t}^{\mathrm{T}})^{\mathrm{T}}$. *Then*

$$\left[\det\left(\sum_{t=1}^{n} \mathbf{Z}_{t-1}\mathbf{Z}_{t-1}^{\mathrm{T}}\right)\right]^{-1} = O_p\left(n^{-h(h+1)-g(g+1)}\right),$$

where $h = \dim(\mathbf{y}_{1t})$ *and* $g = \dim(\mathbf{y}_{2t})$.

This corollary follows directly from the nonsingularity of Ω and the normalization matrices $\mathbf{L}_{i,n}$. It has several applications in time series analysis; see, for instance, Section 4.3.

Case 2(b): *Eigenvalues are* -1. Here we consider the cross product of $\mathbf{y}_{3t}$ and $\mathbf{y}_{4t}$. It is easy to see that this case is parallel to case 2(a) and similar results can be obtained. The only change that needs to be made is that a minus sign should be added in defining the stochastic integrals. [See the coefficient λ in (4.2).]

Case 2(c): *Eigenvalues are* $\exp(i\omega)$ *and* $\exp(-i\omega)$. Here we consider $\mathbf{y}_{5t}$ and $\mathbf{y}_{6t}$ with $\omega_1 = \omega_2 = \omega$ which is between 0 and π. Clearly, $\mathbf{y}_{5t}$ and $\mathbf{y}_{6t}$ satisfy the recursion of (3.10) and (3.11) and we need the result of Lemma 3.2. Let $h = \dim(\mathbf{y}_{5t})/2$ and $g = \dim(\mathbf{y}_{6t})/2$. Also, denote the hth bivariate subvector of $\mathbf{e}_{5t}$ by $\mathbf{e}_{5,h,t}$ and the gth bivariate subvector of $\mathbf{e}_{6t}$ by $\mathbf{e}_{6,g,t}$. Let $\Xi_{5,h}(t)$ and $\Xi_{6,g}(t)$, respectively, be the limiting vectors of Brownian motions corresponding to $\mathbf{e}_{5,h,t}$ and $\mathbf{e}_{6,g,t}$ (again, see Lemma 3.2). Then, similar to (3.24) and (3.25), we define

$$\Gamma_{5,h}(t) = \Xi_{5,h}(t) \text{ and } \Gamma_{5,v}(t) = \mathbf{R}^{-1}\int_0^t \Gamma_{5,v+1}(s)\,ds \quad \text{for } v = h-1,\ldots,1,$$

$$\Gamma_{6,g}(t) = \Xi_{6,g}(t) \text{ and } \Gamma_{6,u}(t) = \mathbf{R}^{-1}\int_0^t \Gamma_{6,u+1}(s)\,ds \quad \text{for } u = g-1,\ldots,1.$$

Next, let $\mathbf{Z}_t = (\mathbf{y}_{5t}^{\mathrm{T}}, \mathbf{y}_{6t}^{\mathrm{T}})$ and define

$$\Lambda = \text{the limit of } \mathbf{L}_n^{-1}\sum_{t=1}^{n}\mathbf{Z}_{t-1}\mathbf{Z}_{t-1}^{\mathrm{T}}\mathbf{L}_n^{-1} \quad \text{with } \mathbf{L}_n = \operatorname{diag}\{\mathbf{L}_{5,n}, \mathbf{L}_{6,n}\},$$

where $\mathbf{L}_{5,n}$ and $\mathbf{L}_{6,n}$ are defined in (4.12). The result in this case is as follows.

LEMMA 4.5. *Suppose that* $\mathbf{y}_{5t}$ *and* $\mathbf{y}_{6t}$ *are given by model* (4.10) *and obtained from* (1.1) *by the transformation of Section* 2. *Assume also that* $\mathbf{a}_t$ *of* (1.1) *satisfies* (1.2). *Then*

(a) $$n^{-(h+g-v-u+2)}\sum_{t=1}^{n}\mathbf{R}^{-t}\mathbf{y}_{5,v,t}\mathbf{y}_{6,u,t}^{\mathrm{T}}\mathbf{R}^{t}\to_d\int_0^1\Gamma_{5,v}(s)[\Gamma_{6,u}(s)]^{\mathrm{T}}\,ds$$

for $1\le v\le h$ *and* $1\le u\le g$,

(b) Λ *is nonsingular almost surely*,

(c) $$\left[\det\left(\sum_{t=1}^{n}\mathbf{Z}_{t-1}\mathbf{Z}_{t-1}^{\mathrm{T}}\right)\right]^{-1}=O_p\left(n^{-2h(h+1)-2g(g+1)}\right),$$

where $\mathbf{y}_{5,v,t}$ *is the vth bivariate subvector of* $\mathbf{y}_{5t}$ *and* $\mathbf{y}_{6,u,t}$ *is the uth bivariate subvector of* $\mathbf{e}_{6t}$.

Finally, combining the results of Theorems 4.1 through 4.3 and Lemmas 4.3 through 4.5, we summarize the asymptotic properties of $\mathbf{y}_t$ of (4.10) into a theorem. For simplicity, we treat $\omega_1=\omega_2$ so that the three types of eigenvalues all appear in different Jordan blocks.

THEOREM 4.4. *Suppose that* $\mathbf{y}_t$ *follows the model* (4.10) *and is a transformed process of model* (1.1) *by the method of Section* 2. *Assume also that* $\mathbf{a}_t$ *of* (1.1) *satisfies* (1.2) *and* $\omega_1=\omega_2$. *Then*

(a) $$\mathbf{L}_n^{-1}\sum_{t=1}^{n}\mathbf{y}_{t-1}\mathbf{y}_{t-1}^{\mathrm{T}}\mathbf{L}_n^{-1}\to_d\mathbf{A},$$

(b) $$\mathbf{L}_n(\hat{\boldsymbol\beta}_n-\mathbf{U})^{\mathrm{T}}\to_d\mathbf{A}^{-1}\mathbf{P},$$

(c) $$\hat{\boldsymbol\beta}_n\to_p\mathbf{U},$$

(d) $$\left[\det\left(\sum_{t=1}^{n}\mathbf{y}_{t-1}\mathbf{y}_{t-1}^{\mathrm{T}}\right)\right]^{-1}=O_p(n^{-m}),$$

where

$\mathbf{L}_n=\mathrm{diag}\{\mathbf{L}_{1,n},\mathbf{L}_{2,n},\dots,\mathbf{L}_{6,n}\}$ *with* $\mathbf{L}_{i,n}$ *given by* (4.11) *and* (4.12),

$\mathbf{A}=\mathrm{diag}\{\Omega,\Delta,\Lambda\}$ *with* Ω *and* Λ *given by Lemmas* 4.4 *and* 4.5, *respectively, whereas* Δ *is the counterpart of* Ω *for eigenvalue* $=-1$,

$\mathbf{P}=\mathrm{diag}\{\mathbf{V},\mathbf{V}^*,\mathbf{Q}\}$ *which are defined in a similar manner as those of Theorems* 4.1 *and* 4.2 *with* $\mathbf{V}^*$ *for eigenvalue* $=-1$,

$m=\sum_{i=1}^{4}h_i(h_i+1)+\sum_{v=5}^{6}2d_v(d_v+1)$ *with* $d_v=\dim(\mathbf{y}_{vt})/2$ *and* $h_i=\dim(\mathbf{y}_{it})$.

Since (4.10) covers all possible types of nonstationary characteristic roots, Theorem 4.4 in effect applies to the general purely nonstationary vector AR(1) processes.

4.3. *Some further applications.* In this section we consider two more applications, namely the shifted and forward autoregressions of $\mathbf{y}_t$ of (4.10). Again, we treat $\omega_1 = \omega_2$. These two applications are useful in the canonical analysis of vector time series; see Section 7 and Tiao and Tsay (1989). The results are also useful in studying seasonal behavior of a time series.

First, consider the "lag-s shifted" autoregression of $\mathbf{y}_t$,

$$\mathbf{y}_t = \boldsymbol{\beta}(s)\mathbf{y}_{t-s} + \boldsymbol{\varepsilon}_t, \qquad s > 0, \tag{4.15}$$

and the associated LS estimate $\hat{\boldsymbol{\beta}}_n^{\mathrm{T}}(s) = (\Sigma_{t=1}^n \mathbf{y}_{t-s}\mathbf{y}_{t-s}^{\mathrm{T}})^{-1}\Sigma_{t=1}^n \mathbf{y}_{t-s}\mathbf{y}_t^{\mathrm{T}}$.

THEOREM 4.5. *Suppose that* $\mathbf{y}_t$ *satisfies the conditions of Theorem* 4.4. *Then, for the lag-s shifted autoregression* (4.15), *we have*

(a) $$\hat{\boldsymbol{\beta}}_n(s) \to_p \mathbf{U}^s,$$

(b) $$\mathbf{L}_n\left[\hat{\boldsymbol{\beta}}_n(s) - \mathbf{U}^s\right]^{\mathrm{T}} \to_d \mathbf{A}^{-1}\mathbf{P}\left(\sum_{v=0}^{s-1}\mathbf{U}^v\right)^{\mathrm{T}},$$

where $\mathbf{L}_n$, $\mathbf{A}$ *and* $\mathbf{P}$ *are defined in Theorem* 4.4.

PROOF. Since

$$\mathbf{y}_t - \mathbf{U}^s\mathbf{y}_{t-s} = \sum_{v=0}^{s-1}\mathbf{U}^v\mathbf{e}_{t-v}, \tag{4.16}$$

the theorem can be proved by using the same techniques as those used in showing Theorem 4.4. The only difference is that here we apply the techniques s times. □

Next, consider the "forward" autoregression of $\mathbf{y}_t$,

$$\mathbf{y}_t = \boldsymbol{\delta}(s)\mathbf{y}_{t+s} + \boldsymbol{\varepsilon}_t, \qquad s > 0, \tag{4.17}$$

with LS estimate $\hat{\boldsymbol{\delta}}_n^{\mathrm{T}}(s) = (\Sigma_{t=1}^{n-s}\mathbf{y}_{t+s}\mathbf{y}_{t+s}^{\mathrm{T}})^{-1}\Sigma_{t=1}^{n-s}\mathbf{y}_{t+s}\mathbf{y}_t^{\mathrm{T}}$.

THEOREM 4.6. *Suppose that* $\mathbf{y}_t$ *satisfies the conditions of Theorem* 4.4. *Then for the forward autoregression* (4.17) *of* $\mathbf{y}_t$, $\hat{\boldsymbol{\delta}}_n(s) \to_p \mathbf{U}^{-s}$.

PROOF. By the model (4.16),

$$\mathbf{y}_t = \mathbf{U}^{-s}(\mathbf{y}_{t+s} - \boldsymbol{\eta}_{t+s}) \quad \text{with } \boldsymbol{\eta}_{t+s} = \sum_{v=0}^{s-1}\mathbf{U}^v\mathbf{e}_{t+s-v}.$$

Therefore,

$$\left[\hat{\boldsymbol{\delta}}_n(s) - \mathbf{U}^{-s}\right]^{\mathrm{T}} = -\left(\sum_{t=1}^{n-1}\mathbf{y}_{t+s}\mathbf{y}_{t+s}^{\mathrm{T}}\right)^{-1}\left(\sum_{t=1}^{n-s}\mathbf{y}_{t+s}\boldsymbol{\eta}_{t+s}^{\mathrm{T}}\right)(\mathbf{U}^{-s})^{\mathrm{T}}.$$

By (4.16) again, we may rewrite the second summation on the right-hand side as

$$\sum_{t=1}^{n-s} (\mathbf{U}^s \mathbf{y}_t + \boldsymbol{\eta}_{t+s})\boldsymbol{\eta}_{t+s}^{\mathrm{T}}.$$

The result then follows from Theorem 4.4 and the fact that $n^{-1}\Sigma_t \boldsymbol{\eta}_t \boldsymbol{\eta}_t^{\mathrm{T}}$ converges to a constant matrix. □

In summary, in this section we have established various asymptotic properties of purely nonstationary vector AR(1) processes. This in effect provides a framework for investigating the asymptotic results of general nonstationary vector ARMA models.

5. Purely nonstationary vector ARMA(1, q) processes. We now extend the results of the preceding two sections to the purely nonstationary ARMA(1, q) process $\mathbf{N}_t$ of (2.13). The key to the extension is the link established in Section 2. More specifically, by the result (2.7), we have

$$\mathbf{N}_t = \mathbf{M}_t + \mathbf{Y}_t, \tag{5.1}$$

where $\mathbf{M}_t$ is an MA(q) series and $\mathbf{Y}_t$ follows a purely nonstationary vector AR(1) model. Since an MA series is stationary, the asymptotic behavior of $\mathbf{N}_t$ is dominated by those of $\mathbf{Y}_t$ which are available in Sections 3 and 4. For simplicity, we treat $\mathbf{Y}_t$ as being the real-valued process $\mathbf{y}_t$, because, as shown before, the transformation is trivial.

LEMMA 5.1. *Suppose that* $\mathbf{N}_t$ *follows a purely nonstationary vector ARMA*(1, q) *model with each individual block in the form of* (2.4), *which is obtained from* (1.1) *by the transformation of Section* 2, *and that* $\mathbf{Y}_t$ *is related to* $\mathbf{N}_t$ *by* (5.1). *Also, assume that* $\mathbf{a}_t$ *of* (1.1) *satisfies* (1.2). *Then, for a fixed integer* s,

$$\left(\sum_{t=1}^{n} \mathbf{Y}_{t-s}\mathbf{Y}_{t-s}^{\mathrm{T}}\right)^{-1}\left(\sum_{t=1}^{n} \mathbf{N}_{t-s}\mathbf{N}_{t-s}^{\mathrm{T}}\right) = \mathbf{I}_h + o_p(1),$$

where $h = \dim(\mathbf{N}_t)$.

PROOF. From (5.1),

$$\sum_{t=1}^{n} \mathbf{N}_{t-s}\mathbf{N}_{t-s}^{\mathrm{T}} = \sum_{t=1}^{n} \mathbf{Y}_{t-s}\mathbf{Y}_{t-s}^{\mathrm{T}} + \sum_{t=1}^{n} \mathbf{M}_{t-s}\mathbf{M}_{t-s}^{\mathrm{T}} + \sum_{t=1}^{n} \mathbf{Y}_{t-s}\mathbf{M}_{t-s}^{\mathrm{T}} + \sum_{t=1}^{n} \mathbf{M}_{t-s}\mathbf{Y}_{t-s}^{\mathrm{T}}.$$

By Theorem 4.4, the stochastic order of $[\det(\Sigma \mathbf{Y}_{t-s}\mathbf{Y}_{t-s}^{\mathrm{T}})]^{-1}$ is available. Since $\mathbf{M}_t$ is stationary, $\Sigma \mathbf{M}_{t-s}\mathbf{M}_{t-s}^{\mathrm{T}} = O_p(n)$. The result then follows directly from the stochastic orders involved and an application of the Cauchy–Schwarz inequality. □

Next, consider the ordinary, the lag-s shifted and the forward AR(1) autoregressions of $\mathbf{N}_t$. The LS estimates are, for $s \geq 1$,

$$\hat{\mathbf{D}}_n^{\mathrm{T}}(s) = \left(\sum_{t=1}^{n} \mathbf{N}_{t-s}\mathbf{N}_{t-s}^{\mathrm{T}}\right)^{-1} \sum_{t=1}^{n} \mathbf{N}_{t-s}\mathbf{N}_t^{\mathrm{T}}$$

and

$$\hat{\boldsymbol{\delta}}_n^{\mathrm{T}}(s) = \left(\sum_{t=1}^{n-s} \mathbf{N}_{t+s}\mathbf{N}_{t+s}^{\mathrm{T}}\right)^{-1} \sum_{t=1}^{n-s} \mathbf{N}_{t+s}\mathbf{N}_t^{\mathrm{T}}.$$

THEOREM 5.1. *Suppose that $\mathbf{N}_t$ satisfies the conditions of Lemma* 5.1. *Then, for a fixed $s \geq 1$,* (a) $\hat{\mathbf{D}}_n(s) \to_p \mathbf{D}^s$, *and* (b) $\hat{\boldsymbol{\delta}}_n(s) \to_p \mathbf{D}^{-s}$.

PROOF. For part (a), from (5.1) we have

$$\left[\hat{\mathbf{D}}_n(s) - \mathbf{D}^s\right]^{\mathrm{T}} = \left(\sum_{t=1}^{n} \mathbf{N}_{t-s}\mathbf{N}_{t-s}^{\mathrm{T}}\right)^{-1} \sum_{t=1}^{n} \mathbf{N}_{t-s}\mathbf{F}_t^{\mathrm{T}}, \tag{5.2}$$

where $\mathbf{F}_t = \sum_{v=0}^{s-1}(\mathbf{D}^v\sum_{i=0}^{q}\mathbf{C}_i\mathbf{a}_{t-v-i})$ with $\mathbf{C}_i$ being a submatrix of Θ_i^* of (2.3). This can be written as

$$\left[\hat{\mathbf{D}}_n(s) - \mathbf{D}^s\right]^{\mathrm{T}} = \left[\left(\sum_{t=1}^{n} \mathbf{Y}_{t-s}\mathbf{Y}_{t-s}^{\mathrm{T}}\right)^{-1}\left(\sum_{t=1}^{n} \mathbf{N}_{t-s}\mathbf{N}_{t-s}^{\mathrm{T}}\right)\right]^{-1} \times\left[\left(\sum_{t=1}^{n} \mathbf{Y}_{t-s}\mathbf{Y}_{t-s}^{\mathrm{T}}\right)^{-1}\left(\sum_{t=1}^{n} \mathbf{N}_{t-s}\mathbf{F}_t^{\mathrm{T}}\right)\right].$$

Part (a) then follows from Lemma 5.1 and the fact that

$$\sum_{t=1}^{n} \mathbf{N}_{t-s}\mathbf{F}_t^{\mathrm{T}} = O_p\left(\sum_{t=1}^{n} \mathbf{Y}_{t-s}\mathbf{f}_t^{\mathrm{T}}\right),$$

where $\mathbf{f}_t$ is the innovational series of $\mathbf{Y}_t$. Similarly, part (b) can be shown by using Lemma 5.1, the techniques employed in Theorem 4.6 and the multivariate ergodic theorem of stationary processes. □

By Theorem 5.1 with $s = 1$ in (a), the LS estimate of AR(1) regression of a purely nonstationary vector ARMA(1, q) process is consistent. This extends the result of pure AR models to the mixed ARMA models. Next, from Lemma 5.1 and Theorem 4.4, for $s > 0$ we have

$$\mathbf{L}_n^{-1} \sum_{t=1}^{n} \mathbf{N}_{t-s}\mathbf{N}_{t-s}^{\mathrm{T}}\mathbf{L}_n^{-1} \to_d \mathbf{A},$$

where $\mathbf{L}_n$ and $\mathbf{A}$ are defined in Theorem 4.4. However, the limiting distribution of $\hat{\mathbf{D}}_n(s)$ is different from that of the LS estimate of AR(1) regression of $\mathbf{Y}_t$ shown in Theorem 4.5 if $q > 0$. To derive the limiting distribution of $\hat{\mathbf{D}}_n(s)$, it is necessary, from (5.2), to establish the asymptotic distribution of $\mathbf{L}_n^{-1}\Sigma_t \mathbf{N}_{t-s}\mathbf{a}_{t-i}^{\mathrm{T}}$ for $i = 0, 1, \dots, q + s - 1$. This can be achieved as follows.

1. Theorem 3.1(a) still holds when $\mathbf{y}_{v,t}$ is replaced by the corresponding subvector $\mathbf{N}_{v,t}$ of $\mathbf{N}_t$.
2. A result similar to Theorem 3.1(d) can be established for $\mathbf{N}_t$ and $\mathbf{a}_t$.
3. Using the results of parts 1 and 2 and the techniques used in Section 4, one can complete the proof.

Note that the limiting distribution of $\hat{\mathbf{D}}_n(s)$ depends on the MA coefficients $\mathbf{C}_i$, see the definition of $\mathbf{F}_t$ of (5.2). To use this distribution, one must replace $\mathbf{C}_i$'s with their consistent estimates. This procedure has been used in Tsay (1986) where the limiting distribution of the least squares estimate of AR(1) regression of univariate ARIMA$(p, 1, q)$ processes is derived.

6. The general vector ARMA processes. In this section, we investigate properties of the general vector process $\mathbf{TX}_t$ of (2.3). From (2.13), consider first the stationary part $\mathbf{S}_t$ of $\mathbf{TX}_t$. Let $\mathbf{L}_n^* = n^{1/2}\mathbf{I}_s$, where s is the dimension of $\mathbf{S}_t$, and let $\mathbf{F}_t = \Sigma_{i=0}^{q}\Theta_i^*\mathbf{a}_{t-i}$ be the moving average part of (2.3). Then, since $\mathbf{S}_t$ is stationary and by Lai and Wei (1985), we have

$$n^{-1}\sum_{t=1}^{n}\mathbf{S}_{t-1}\mathbf{S}_{t-1}^{\mathrm{T}} \to_p \Gamma(0) \quad \text{and} \quad n^{-1}\sum_{t=1}^{n}\mathbf{S}_{t-1}\mathbf{F}_t^{\mathrm{T}} \to_p \Gamma_q^*, \tag{6.1}$$

where $\Gamma(0)$ is the covariance matrix of $\mathbf{S}_t$, and $\Gamma_q^* = \mathbf{O}$ if $q = 0$ and $\Gamma_q^* \neq \mathbf{O}$ if $q > 0$. Note that for a given process $\mathbf{z}_t$ of (1.1) that admits alternative model representations the value of q may change. However, once the order p is chosen, q is fixed and the consistency of least squares estimates of the stationary part addressed below is with respect to this chosen model. Thus, the consistency through a unimodular matrix transformation is not discussed in this paper.

Since properties of the nonstationary part $\mathbf{N}_t$ are available in Lemma 5.1 and Theorem 5.1, it remains to consider the cross product between stationary and nonstationary parts. Let $\mathbf{K}_n = \operatorname{diag}\{\mathbf{L}_n, \mathbf{L}_n^*\}$ with $\mathbf{L}_n$ being given in Theorem 4.4.

LEMMA 6.1. $\mathbf{L}_n^{-1}\Sigma_{t=1}^{n}\mathbf{N}_{t-1}\mathbf{S}_{t-1}^{\mathrm{T}}(\mathbf{L}_n^*)^{-1} \to_p \mathbf{O}$.

PROOF. This lemma is a generalization of Theorem 3.4.2 of Chan and Wei (1988) and can be shown along the same line. The only change is that now $\mathbf{a}_t$ is a vector innovational process. This change does not affect the basic argument of the proof. It can also be shown by using the techniques of Lemma A.2 in Tiao and Tsay (1983a) if the fourth moment of each component of $\mathbf{a}_t$ is finite. □

By Lemma 6.1, we have

$$\mathbf{K}_n^{-1}\mathbf{T}\left(\sum_{t=1}^{n}\mathbf{X}_{t-1}\mathbf{X}_{t-1}^{\mathrm{T}}\right)\mathbf{T}^{\mathrm{T}}\mathbf{K}_n^{-1} \to_d \operatorname{diag}\{\mathbf{A}, \Gamma(0)\}, \tag{6.2}$$

where $\mathbf{A}$, again, is given in Theorem 4.4. This result says that to study properties of the LS estimate of AR(1) regression of $\mathbf{TX}_t$ we may consider $\mathbf{N}_t$ and $\mathbf{S}_t$ separately. Therefore, from (6.1) and Theorem 5.1 we may summarize the result in a theorem.

THEOREM 6.1. *Suppose that* $\mathbf{U}_t = \mathbf{TX}_t$ *follows the vector ARMA*(1, q) *model* (2.3) *which is obtained from* (1.1) *by the transformation of Section* 2. *Assume that* $\mathbf{a}_t$ *of* (1.1) *satisfies* (1.2). *Also, partition* $\mathbf{U}_t = (\mathbf{N}_t^{\mathrm{T}}, \mathbf{S}_t^{\mathrm{T}})^{\mathrm{T}}$ *according to the stationarity of each component and let* $\hat{\beta}_n^{\mathrm{T}} = (\Sigma_{t=1}^n \mathbf{U}_{t-1}\mathbf{U}_{t-1}^{\mathrm{T}})^{-1}(\Sigma_{t=1}^n \mathbf{U}_{t-1}\mathbf{U}_t^{\mathrm{T}})$ *be the least squares estimate of AR*(1) *regression of* $\mathbf{U}_t$. *Then* $\hat{\beta}_n \to_p \mathbf{J}$ *if either* $q = 0$ *or* $\mathbf{U}_t = \mathbf{N}_t$. *When* $q > 0$ *and* $\mathbf{U}_t \neq \mathbf{N}_t$, $\hat{\beta}_n$ *is inconsistent. However,* $\hat{\beta}_n$ *always provides consistent estimates for those eigenvalues of* $\mathbf{J}$ *that are on the unit circle.*

Finally, consider the vector ARMA(p, q) model $\mathbf{z}_t$ of (1.1). From (2.1), an AR(1) regression of $\mathbf{TX}_t$ is an AR(p) regression of $\mathbf{z}_t$. Therefore, Theorem 6.1 applies to $\mathbf{z}_t$ with the AR(1) regression being replaced by an AR(p) regression. Moreover, since the eigenvalues of the $\mathbf{G}$ matrix of (2.1) are the roots of the AR polynomial $\Phi(B)$, Theorem 6.1 says that the roots of the LS estimate $\hat{\Phi}(B)$ of an AR(p) regression of $\mathbf{z}_t$ provide consistent estimates of the nonstationary characteristic roots of $\mathbf{z}_t$. This generalizes the univariate result of Tiao and Tsay (1983a) to the vector case and provides a means by which the nonstationary structure of a vector ARMA model may be obtained. We summarize the result in a theorem.

THEOREM 6.2. *For the vector ARMA*(p, q) *process* $\mathbf{z}_t$ *of* (1.1), *stationary or nonstationary, with* $\mathbf{a}_t$ *satisfying* (1.2), *let* $\hat{\Phi}_p(B)$ *be the polynomial of the ordinary least squares estimate of an AR*(p) *regression of* $\mathbf{z}_t$. *Then*

$$\hat{\Phi}_p(B) \to_p \Phi(B)$$

if either $q = 0$ *or* $\mathbf{z}_t$ *is purely nonstationary. Furthermore, the roots of* $\hat{\Phi}_p(B)$ *provide consistent estimates of the nonstationary characteristic roots of* $\Phi(B)$ *as long as* q *is finite.*

7. Applications to model specification. The limiting distribution of Section 4 can be used to test the nonstationarity of a vector process, e.g., testing the existence of unit roots in a vector AR model. A second application of the results of this paper is that they can be used to develop a unified approach for modeling stationary and nonstationary vector ARMA processes. Since the dynamic structure of a vector process may be complex, it is useful to develop methods which linearly transform the series to uncover possibly simpler underlying structures. The consistency results of this paper form a basis on which such a method can be derived; see Tiao and Tsay (1989). The method uses a canonical correlation procedure to obtain simplifying structures which may be hidden in an observed vector process. Roughly speaking, canonical correlations of two constructed vector-valued variables $\mathbf{Y}_{m,t}$ and $\mathbf{Y}_{m,t-j-1}$, where $\mathbf{Y}_{m,t} = (\mathbf{z}_t^{\mathrm{T}}, \dots, \mathbf{z}_{t-m}^{\mathrm{T}})^{\mathrm{T}}$, are used to measure the dependence between lagged variables. This, however, is equivalent to considering the eigenvalues and eigenvectors of the matrix

$$\mathbf{A}(m, j) = \hat{\boldsymbol{\beta}}^*(m, j)\hat{\boldsymbol{\beta}}(m, j),$$

where

$$\hat{\boldsymbol{\beta}}^*(m, j) = \Big(\sum_t \mathbf{Y}_{m,t}\mathbf{Y}_{m,t}^{\mathrm{T}}\Big)^{-1}\Big(\sum_t \mathbf{Y}_{m,t}\mathbf{Y}_{m,t-j-1}^{\mathrm{T}}\Big),$$

$$\hat{\boldsymbol{\beta}}(m, j) = \Big(\sum_t \mathbf{Y}_{m,t-j-1}\mathbf{Y}_{m,t-j-1}^{\mathrm{T}}\Big)^{-1}\Big(\sum_t \mathbf{Y}_{m,t-j-1}\mathbf{Y}_{m,t}^{\mathrm{T}}\Big).$$

Consequently, it is essential to show the existence of the limit of $\mathbf{A}(m, j)$ and to understand the properties of such a limit for both nonstationary and stationary processes $\mathbf{z}_t$ when m satisfies certain conditions.

Since $\hat{\boldsymbol{\beta}}(m, j)$ consists entirely of the LS estimates of the ordinary and "lag-s" shifted AR(m) regressions of $\mathbf{z}_t$, Theorems 5.1 and 6.1 show that the limit of $\hat{\boldsymbol{\beta}}(m, j)$ exists as the sample size increases provided that $m \geq p$. Furthermore, using Theorems 4.6 and 5.1 and the results of forward autoregressions of stationary processes, one can show that the limit of $\hat{\boldsymbol{\beta}}^*(m, j)$ also exists. Consequently, the limit of $\mathbf{A}(m, j)$ exists as the sample size increases. Furthermore, the results of this paper can also be used to establish the consistency properties of the eigenvectors of $\mathbf{A}(m, j)$. This provides a method for obtaining consistent estimates of hidden simplifying structures and the autoregressive parameters of a multivariate ARMA model.

Acknowledgments. The authors wish to thank C. Z. Wei of the University of Maryland at College Park for useful suggestions and a referee for detailed comments.

REFERENCES

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

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

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

CHAN, N. H. and WEI, C. Z. (1988). Limiting distributions of least squares estimates of unstable autoregressive processes. *Ann. Statist.* **16** 367–401.

CHAN, W. Y. T. and WALLIS, K. F. (1978). Multiple time series modeling: Another look at the Mink–Muskrat interaction. *J. Roy. Statist. Soc. Ser. C* **27** 168–175.

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

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

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

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

HILLMER, S. E. and TIAO, G. C. (1979). Likelihood function of stationary multiple autoregressive moving average models. *J. Amer. Statist. Assoc.* **74** 652–660.

HOSKING, J. R. M. (1981). Lagrange-multiplier tests of multivariate time series. *J. Roy. Statist. Soc. Ser. B* **43** 219–230.

KAILATH, T. (1980). *Linear Systems*. Prentice-Hall, Englewood Cliffs, N.J.

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

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

LI, W. K. and MCLEOD, A. I. (1981). Distribution of the residual autocorrelations in multivariate ARMA time series models. *J. Roy. Statist. Soc. Ser. B* **43** 231–239.

LÜTKEPOHL, H. (1983). Differencing multiple time series: Another look at Canadian money and income data. *J. Time Ser. Anal.* **3** 235–243.

MACDUFFEE, C. C. (1956). *The Theory of Matrices*. Chelsea, New York.

MANN, H. B. and WALD, A. (1943). On the statistical treatment of linear stochastic difference equations. *Econometrica* **11** 173–200.

NICHOLLS, D. F. and HALL, A. D. (1979). The exact likelihood of multivariate autoregressive moving average models. *Biometrika* **66** 259–264.

PHILLIPS, P. C. B. (1987). Time series regression with a unit root. *Econometrica* **55** 277–301.

PHILLIPS, P. C. B. and DURLAUF, S. N. (1985). Multiple time series regression with integrated processes. *Rev. Econom. Stud.* **53** 473–496.

STIGUM, B. P. (1975). Asymptotic properties of autoregressive integrated moving average processes. *Stochastic Process. Appl.* **3** 315–344.

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

TIAO, G. C. and TSAY, R. S. (1983a). Consistency properties of least squares estimates of autoregressive parameters in ARMA models. *Ann. Statist.* **11** 856–871.

TIAO, G. C. and TSAY, R. S. (1983b). Multiple time series modeling and extended sample cross-correlations. *J. Bus. Econom. Statist.* **1** 43–56.

TIAO, G. C. and TSAY, R. S. (1989). Model specification in multivariate time series (with Discussion). *J. Roy. Statist. Soc. Ser. B* **51**, 157–213.

TJØSTHEIM, D. and PAULSEN, J. (1982). Empirical identification of multiple time series. *J. Time Ser. Anal.* **3** 265–282.

TSAY, R. S. (1986). Asymptotic inference of nearly nonstationary ARMA processes. Technical Report 360, Dept. Statistics, Carnegie-Mellon Univ.

TSAY, R. S. and TIAO, G. C. (1984). Consistent estimates of autoregressive parameters and the extended sample autocorrelation function of nonstationary and stationary ARMA models. *J. Amer. Statist. Assoc.* **79** 84–96.

TSAY, R. S. and TIAO, G. C. (1985). Use of canonical analysis in time series model identification. *Biometrika* **72** 299–315.

TSAY, R. S. and TIAO, G. C. (1986). Asymptotic properties of multivariate nonstationary processes with applications to autoregressions. Technical Report 293, Dept. Statistics, Carnegie-Mellon Univ.

TUNNICLIFFE-WILSON, G. (1973). The estimation of parameters in multivariate time series models. *J. Roy. Statist. Soc. Ser. B* **35** 76–85.

DEPARTMENT OF STATISTICS
CARNEGIE-MELLON UNIVERSITY
PITTSBURGH, PENNSYLVANIA 15213

GRADUATE SCHOOL OF BUSINESS
UNIVERSITY OF CHICAGO
CHICAGO, ILLINOIS 60637