Cowles Foundation Paper #679

ASYMPTOTIC EXPANSIONS IN NONSTATIONARY VECTOR AUTOREGRESSIONS

P. C. B. PHILLIPS

Cowles Foundation, Yale University

This paper studies the statistical properties of vector autoregressions (VAR's) for quite general multiple time series which are integrated processes of order one. Functional central limit theorems are given for multivariate partial sums of weakly dependent innovations and these are applied to yield first-order asymptotics in nonstationary VAR's Characteristic and cumulant functionals for generalized random processes are introduced as a means of developing a refinement of central limit theory on function spaces. The theory is used to find asymptotic expansions of the regression coefficients in nonstationary VAR's under very general conditions. The results are specialized to the scalar case and are related to other recent work by the author [21].

1. INTRODUCTION

Econometric modeling techniques based on VAR's have attracted a good deal of interest in recent years. The research in this field is largely inspired by the work of Sims [27], although it has its origins in earlier contributions by Mann and Wald [16] and the Cowles Commission researchers [13]. The field has recently been reviewed and discussed [4], compared with alternative methodologies [17] and subjected to a detailed critical evaluation [3]. But, whatever its shortcomings as an econometric modeling technique, this approach seems destined to continue to play an important role in empirical research. Indeed, almost all of the research in the field since [16] has been empirical. Much of the most recent work has been motivated by issues of economic policy as in [4] and by problems of prediction as in [14].

This paper was presented at the conference on Finite Sample Econometrics held at the University of Western Ontario, August 1985. I am grateful to Peter Hall and a referee for helpful comments on the first draft of the paper. My thanks also go to Glena Ames for her skill and effort in typing the manuscript. The research reported here was supported by the National Science Foundation under Grant No SES 8519595.

Econometric Theory, 3, 1987, 45–68 Printed in the United States of America
 0266–4666/87 $5.00

Against the background of this applied work, there has been a rather noticeable absence of analytical research on the statistical properties of regressions of this type. This is unfortunate. For the theory developed by Mann and Wald [16] applies to stationary time series, whereas many of the time series that are used in these regressions actually display nonstationary behavior. Out of necessity, therefore, applied researchers have often been forced to rely on traditional methods of inference for stationary time series in interpreting empirical results from VAR's with demonstrably nonstationary time series. As yet there is no general statistical theory of estimation and inference in models of nonstationary time series. However, a theory of regression for models that involve integrated time series of the ARIMA type seems to be well within reach.

The present paper draws much of its motivation from the needs expressed in the last paragraph. Our central concern is with the statistical properties of the estimated regression coefficients in VAR's. We first provide an extension to the multivariate case of the theory developed by the author in other recent work [21] for scalar autoregressions with a unit root. As in [21], the conditions we impose on the time series are quite weak and they allow for a wide class of weakly dependent and heterogeneously distributed innovation sequences.

Our main aim is to study higher-order asymptotic properties of regression coefficients in models of nonstationary time series. The methods we develop for this analysis involve characteristic and cumulant functionals for generalized random processes. Characteristic functionals were developed by Bochner [2], Prohorov [25] and others to assist in the study of Banach-valued random variables. The first-order asymptotics that we develop here and in related work [21, 24] rely on central limit theory in Banach spaces or functional central limit theory. The first step in our development of higher-order asymptotics involves a refinement of this functional central limit theory to accommodate asymptotic expansions. In the second step, these refinements are extended to apply to certain functionals of the Banach-valued random variables. The resulting asymptotic expansions help in analyzing the adequacy of asymptotic theory for nonstationary VAR's.

The plan of the paper is as follows. Section 2 develops some multivariate functional central limit theory for partial sums of ρ-mixing and strong mixing innovations. Characteristic and cumulant functionals are studied in Section 3 and used to establish a refinement of the functional central limit theory of Section 2. This theory is applied in Section 4 to develop asymptotic expansions of the distribution of the regression coefficients in a first-order nonstationary VAR. Some conclusions and extensions are given in Section 5. Proofs of results presented in the body of the paper are provided in the mathematical appendix.

2. MULTIVARIATE FUNCTIONAL CENTRAL LIMIT THEOREMS

Let $\{u_t\}_1^\infty$ be a sequence of random n-vectors on a probability space $(\Omega, \mathscr{B}, P)$ with

$$E(u_t) = 0,\ E(u_{it}^2) < \infty; \qquad (i = 1, \ldots, n), (t = 1, 2, \ldots). \tag{1}$$

We introduce the vector of partial sums $S_t = \Sigma_{j=1}^t u_j$ ($S_0 = 0$) and set $M_t = E(S_t S_t')$. It will be convenient although not essential to require that the limit

$$\Sigma = \lim_{T\to\infty} T^{-1} M_T \text{ exists and is positive definite.} \tag{2}$$

More generally, we could allow M_T to have the representation $M_T = TH(T)$, where $H(T)$ is a positive definite and (element-wise) slowly varying matrix function of T, so that $\lim_{T\to\infty} H(kT)H(T)^{-1} = I_n$ for any positive integer k. In the univariate case the latter representation is a necessary condition for the validity of the functional central limit theorem [7, pp. 98]. However, to work at this level of generality here would inhibit some of the applications intended for the present paper and we shall therefore employ the simpler and more convenient requirement of equation (2) in what follows.

We introduce the random elements

$$\begin{aligned} &X_T(r) = T^{-1/2} S_{[Tr]} = T^{-1/2} S_{j-1};\\ &((j-1)/T \leqslant r < j/T, j = 1, \ldots, T)\\ &X_T^*(r) = \Sigma^{-1/2} X_T(r), \end{aligned} \tag{3}$$

where $[Tr]$ denotes the integer part of Tr and $\Sigma^{1/2}$ is the positive definite square root of Σ. $X_T(r)$ and $X_T^*(r)$ lie in the product metric space $D^n = D[0,1] \times \cdots \times D[0,1]$ (n copies), where $D[0,1]$ is the space of all real valued functions on the interval $[0,1]$ that are right continuous and have finite left limits. We endow D^n with the metric

$$d_n(f,g) = \max_i \{d_0(f_i, g_i): i = 1, \ldots, n; f_i, g_i \in D[0,1]\} \tag{4}$$

where $d_0(\,,)$ is the modified Skorohod metric (see [1, pp. 112]) under which $D[0,1]$ is separable and complete.

As in [21] and [24] we want to allow for both temporal dependence and heterogeneity in the process $\{u_t\}_1^\infty$. As a measure of dependence we shall use primarily, but not exclusively, the maximal correlation coefficient ρ_m. We define ρ_m by:

$$\rho_m = \sup_t \sup_{\xi,\eta} \max_{i,j} \{|E(\xi_i \eta_j)|: \xi = (\xi_i) \in L_1^t, \eta = (\eta_j) \in L_{t+m}^{\infty};$$
$$E(\xi_i^2) = 1, E(\eta_j^2) = 1\}. \qquad \textbf{(5)}$$

Here $L_r^s = L_2(\mathscr{F}_r^s)$, the space of all square integrable and $\mathscr{F}_r^s$-measurable random variables, where $\mathscr{F}_r^s$ denotes the σ-field generated by $\{u_t, r \leqslant t \leqslant s\}$. Sequences $\{u_t\}_1^\infty$ for which $\rho_m \downarrow 0$ as $m \uparrow \infty$ are said to be ρ-mixing. Such sequences were introduced for scalar processes by Kolmogorov and Rozanov [12]. The mixing condition $\rho_m \downarrow 0$ has the simple interpretation that square integrable random variables generated by members of the sequence $\{u_t\}$ which are separated by at least m time periods have correlation which tends to zero as $m \uparrow \infty$. The condition is easy to verify for many commonly occurring time series models. Thus, stationary (ARMA) processes of finite order have correlation sequences which are known to decay exponentially. This implies that the maximal linear correlation coefficient [26, pp. 186] of $\{u_t\}$ also decays exponentially. For Gaussian sequences this coefficient is equal to the maximal correlation coefficient [26, pp. 181]. Thus, $\rho_m \downarrow 0$ at an exponential rate as $m \uparrow \infty$ for stationary Gaussian ARMA processes of finite order.

One reason for the popularity of α- and φ-mixing measures of weak dependence in recent econometric work (see, for example, [21], [28], [29]) is that conditions based upon these measures and the associated mixing decay rates continue to apply to measurable functions of the mixing processes provided the functions involve only a finite stretch of the original process [28, pp. 47]. The same result does not apply to ρ-mixing processes. However, in many of the applications of the theory that we develop here, the functions of interest turn out to be functionals of the partial sums of sequences of primitive innovations which we may quite reasonably require to be ρ-mixing. In such situations, the ρ-mixing condition is usually sufficient to determine the limiting distribution of the functional. Moreover, as we shall see, no additional condition is required on the ρ-mixing decay rate, in contrast to the limit theory based on α- or φ-mixing processes in [21], [24], and in Theorem 2.2 below.

THEOREM 2.1. *Let $\{u_t\}_1^\infty$ be a ρ-mixing sequence of random n-vectors satisfying conditions (1) and (2). If*

a. $\sup\{T^{-1}E(S_{k+T} - S_k)'(S_{k+T} - S_k): k \geqslant 0, T \geqslant 1\} < \infty$;
b *there exist $\delta > 0$ and $\varepsilon \in (0, \delta/2)$ such that*

$$E|u_{iT}|^{2+\delta} = O(T^{\delta/2-\varepsilon}), \qquad (i = 1, \ldots, n);$$

then $X_T(r) \Rightarrow B(r)$ *as* $T \uparrow \infty$ *where* $B(r)$ *is n-vector Brownian motion with covariance matrix* Σ.

We use the symbol "$\Rightarrow$" to signify the weak convergence of the associated probability measures [1]. Each element of the limit process $B(r)$ is a scalar Brownian motion and the sample paths of $B(r)$ lie almost surely in the function space $C^n = C[0, 1]^n$, the product space of n copies of $C[0, 1]$ (the space of all real valued continuous functions on the interval $[0, 1]$). Note that $W(r) = \Sigma^{-1/2}B(r)$ is n-vector standard Brownian motion on C^n and, by Theorem 2.1, $X_T^*(r) \Rightarrow W(r)$.

Theorem 2.1 is a multivariate generalization of a functional central limit theorem for ρ-mixing scalar sequences established recently by Herrndorf [8]. It provides an alternative to the closely related multivariate limit theorem proved recently in [24] (Theorem 2.1 of [24]). The following theorem is an improvement on Theorem 2.1 of [24] for strong mixing innovations. It is especially useful in situations where there are other reasons (such as those given above) for employing strong mixing rather than ρ-mixing conditions. In our statement of the result we shall use $\{\alpha_m\}_1^\infty$ to denote the sequence of strong mixing numbers.

THEOREM 2.2. *Let* $\{u_t\}_1^\infty$ *be a sequence of random n-vectors satisfying conditions (1) and (2). If*

a. $\sup_{i,t} E|u_{it}|^\beta < \infty$ *for some* $\beta > 2$;

b. $\sum_1^\infty \alpha_m^{1-2/\beta} < \infty$;

then $X_T(r) \Rightarrow B(r)$ *as* $T \uparrow \infty$.

The conditions of Theorem 2.1 are in some respects weaker than those of Theorem 2.2. Thus, the requirement that $\{u_t\}_1^\infty$ be ρ-mixing in Theorem 2.1 eliminates the need for the summability (and implied mixing decay rate) condition (b) of Theorem 2.2. Moreover, in contrast to (a) of Theorem 2.2, the moment condition (b) of Theorem 2.1 allows for moderate growth in the higher moments of u_t as the process evolves.

3. CHARACTERISTIC FUNCTIONALS AND REFINEMENTS OF FUNCTIONAL CENTRAL LIMIT THEORY

The multivariate limit process $W(r)$ is a Banach-valued random variable. Its distribution is determined by the multivariate Wiener measure on $(C^n, \mathscr{C}^n)$ where C^n (a Banach space) is the support of $W(r)$ (i.e., $\mathscr{W}(C^n) = 1$ where $\mathscr{W}(\ )$

denotes Wiener measure) and $\mathscr{C}^n$ is the class of Borel sets on C^n (i.e., the σ-field generated by the open subsets of C^n with the uniform metric). This distribution is also uniquely determined by the characteristic functional of the generalized random process[1] corresponding to $W(r)$.

We shall work with the characteristic functional rather than Wiener measure because the characteristic functional provides a natural tool for the refinement of central limit theorems on function spaces such as those in Section 2. Our approach will be rather formal and is inspired by the needs of the following sections of this paper. We shall not attempt a fully rigorous mathematical theory. According to the author's present information there has been no research published in the probability literature on asymptotic expansions for central limit theorems on function spaces. What follows, therefore, is a first step in this direction.

Let K_n denote the space of all real valued $n \times 1$ functions $\varphi(x)$ with continuous derivatives of all orders and with bounded support. A generalized random process is a continuous linear random functional [6] on K_n and will be denoted by $\Phi(\varphi)$, $\varphi \in K_n$. For the multivariate standard Wiener process $W(r)$ we may define the corresponding generalized random process by the integral

$$\Phi(\varphi) = \int_0^1 \phi(r)' W(r)\, dr \tag{6}$$

which is well defined for all $\varphi \in K_n$. Given φ, $\psi \in K_n$, the correlation functional of Φ is:

$$B(\varphi, \psi) = E\{\Phi(\varphi)\Phi(\psi)\} \tag{7}$$

$$= \int_0^1 \int_0^1 \varphi(r)' \psi(s) \min(r, s)\, dr ds$$

$$= \int_0^1 \hat{\varphi}(r)' \hat{\psi}(r)\, dr \tag{8}$$

where

$$\hat{\varphi}(r) = \int_r^1 \varphi(s)\, ds, \qquad \hat{\psi}(r) = \int_r^1 \psi(s)\, ds.$$

Equation (8) may be established quite easily by integration by parts. The scalar ($n = 1$) case of (8) is well known and may be found for example in [10, pp. 125].

Given $\varphi \in K$, the distribution of the linear functional Φ is $N(0, B(\varphi, \varphi))$. It follows that the characteristic functional of the generalized random process $\Phi(\varphi)$ is:

$$L(\varphi) = E\{e^{i\Phi(\varphi)}\} = \exp\{-(1/2)B(\varphi, \varphi)\}, \qquad \varphi \in K. \tag{9}$$

As in the case of distributions on finite dimensional spaces, the characteristic functional uniquely determines the probability measure. Here, the measure is standard Wiener measure on $(C^n, \mathscr{C}^n)$. The relevant extension of the continuity theorem for characteristic functions which achieves this unique correspondence is known as the Bochner–Minlos theorem and is given in [10, pp. 122]. We note that: (i) $L(\varphi)$ is a continuous functional in the sense that $L(\varphi_k) \to L(\varphi)$ whenever $\varphi_k \to \varphi$ as $k \uparrow \infty$ for any sequence $\{\varphi_k\}$ and limit function φ in K; (ii) $L(\varphi)$ is positive definite in the sense that for any functions $\varphi_1, \ldots, \varphi_m$ in K_n and any complex numbers $\alpha_1, \ldots, \alpha_m$ the inequality $\Sigma_1^m \Sigma_1^m L(\varphi_j - \varphi_k)\alpha_j\bar{\alpha}_k \geqslant 0$ holds; and (iii) $L(0) = 1$. Properties (i), (ii), and (iii) parallel the conventional properties of characteristic functions on finite dimensional spaces.

Define the generalized random process $\Phi_T(\varphi) = \int_0^1 \phi(r)' X_T^*(r)\, dr$ corresponding to $X_T^*(r)$. The correlation functional of Φ_T is $B_T(\varphi, \psi) = E\{\Phi_T(\varphi)\Phi_T(\varphi)\}$, with φ, $\psi \in K_n$, and its characteristic functional is $L_T(\varphi) = E[\exp\{i\Phi_T(\varphi)\}]$. Under quite general conditions on the process $\{u_t\}_1^\infty$ we may develop an asymptotic expansion of $B_T(\varphi, \psi)$ about the correlation functional, $B(\varphi, \psi)$, of the limit process $W(r)$. There is a related expansion for $L_T(\varphi)$ and an associated stochastic representation of $X_T^*(r)$ in terms of the limit process $W(r)$. Our first result is the following:

THEOREM 3.1. *Let $\{u_t\}_1^\infty$ be a weakly stationary ρ-mixing sequence of random n-vectors satisfying (1) and (2). If $\Sigma_{m=1}^\infty m\rho_m < \infty$ then*

$$B_T(\varphi, \psi) = B(\varphi, \psi) + O(T^{-1}) \tag{10}$$

for any φ, $\psi \in K_n$. Moreover, if $\{u_t\}_1^\infty$ is Gaussian, then

$$L_T(\varphi) = L(\varphi)[1 + O(T^{-1})] \tag{11}$$

for any $\varphi \in K_n$ and

$$X_T^*(r) \equiv W(r) + O_p(T^{-1}). \tag{12}$$

The symbol "$\equiv$" in equation (12) signifies equality in distribution. Thus, equation (12) tells us that the random element $X_T^*(r)$ is stochastically characterized by the expansion $W(r) + O_p(T^{-1})$. More precisely, in the present context, if A and B are two random elements on D^n we write $A \equiv B$ if and only if the characteristic functionals of the generalized random processes corresponding to A and B are the same. Since these characteristic functionals uniquely determine the probability measures associated with A and B, $A \equiv B$

if and only if the probability measures of A and B are the same. Thus, equation (12) is a stochastic representation (*not* a stochastic expansion) of the random element $X_T^*(r)$ in terms of the limit process $W(r)$.

The condition $\Sigma_1^\infty m\rho_m < \infty$ on the mixing sequence $\{\rho_m\}$ in Theorem 3.1 is not very restrictive. It is, for example, satisfied by all stationary Gaussian finite order ARMA processes, since $\rho_m = O(\lambda^{-m})$ with $\lambda > 1$ for such processes and thus $\Sigma_1^\infty m\rho_m = \lambda/(\lambda - 1)^2 < \infty$.

The requirement that $\{u_t\}_1^\infty$ is Gaussian in the second half of Theorem 3.1 is, of course, not necessary. To show how the condition may be relaxed, we first define the cumulant functional:

$$C_T(\varphi) = \ln L_T(\varphi) = \ln E\{e^{i\Phi_T(\varphi)}\} \tag{13}$$

and assume that it may be expanded in terms of the cumulant functions (which are assumed to exist)

$$c_{kT}(t_1, \ldots, t_k) = \text{cum}\{X_T^*(t_1), \ldots, X_T^*(t_k)\}; \qquad k = 1, 2, \ldots$$

as

$$C_T(\varphi) = \Sigma_{k=1}^\infty \frac{i^k}{k!} \int_0^1 \cdots \int_0^1 \varphi(t_1) \cdots \varphi(t_k) c_{kT}(t_1, \ldots, t_k)\, dt_1 \cdots dt_k. \tag{14}$$

Note that for a sequence of innovations $\{u_t\}_1^\infty$ satisfying (1) and (2) we have

$$c_{1T}(t) = 0, \qquad c_{2T}(t_1, t_2) = E\{X_T^*(t_1)X_T^*(t_2)\}$$

and thus

$$\int_0^1 \varphi(t_1) c_{1T}(t_1)\, dt_1 = 0,$$

$$\int_0^1 \int_0^1 \varphi(t_1)\varphi(t_2) c_{2T}(t_1, t_2)\, dt_1 dt_2 = B_T(\varphi, \varphi).$$

If we now assume that kth cumulants of $X_T^*(r)$ are $O(T^{-k/2+1})$ we obtain

$$\begin{aligned} C_T(\varphi) &= -(1/2)B_T(\varphi, \varphi) + O(T^{-1/2}) \\ &= -(1/2)B(\varphi, \varphi) + O(T^{-1/2}) \end{aligned}$$

under conditions which ensure the validity of equation (10). The characteristic functional is now:

$$L_T(\varphi) = \exp\{\ln C_T(\varphi)\} = \exp\{-(1/2)B(\varphi, \varphi)\}[1 + O(T^{-1/2})]$$
$$= L(\varphi)[1 + O(T^{-1/2})]$$

and we obtain:

$$\Phi_T(\varphi) \equiv \Phi(\varphi) + O_p(T^{-1/2})$$

and

$$X_T^*(r) \equiv W(r) + O_p(T^{-1/2}).$$

In cases where third-order cumulants are zero (as they are when $\{u_t\}_1^\infty$ is Gaussian) we obtain the improved result $X_T^*(r) \equiv W(r) + O_p(T^{-1})$. Thus, we have:

THEOREM 3.2. *Let $\{u_t\}_1^\infty$ be a strictly stationary ρ-mixing sequence of random n-vectors satisfying equations (1), (2) and $\Sigma_{m=1}^\infty m\rho_m < \infty$. If third-order cumulants of $\{u_t\}_1^\infty$ are zero, if kth cumulants of $X_T^*(r)$ are $O(T^{-k/2+1})$ and if the cumulant functional $C_T(\varphi)$ admits an expansion of the form given in equation (14) then $L_T(\varphi) = L(\varphi)[1 + O(T^{-1})]$ and $X_T^*(r) \equiv W(r) + O_p(T^{-1})$.*

In some instances it is useful to consider generalized processes such as $\Phi(\varphi)$ and $\Phi_T(\varphi)$ where φ may lie in a function space that is larger than K_n. In general, the larger the space of test functions φ the narrower is the class of generalized random processes. However, in the present case where attention centers solely on the random elements $W(r)$ and $X_T^*(r)$, it is very convenient to replace K_n with the set of all generalized functions K_n' (i.e., the set of all linear continuous functionals defined on K_n). The generalized random processes $\Phi(\varphi)$ and $\Phi_T(\varphi)$ are now continuous linear random functionals on the set of continuous linear functionals K_n'. K_n' includes functionals such as the delta function, which we interpret as the mapping:

$$(\delta(t - t_0), \varphi(t)) = \int_0^1 \delta(t - t_0)\varphi(t)\,dt = \varphi(t_0).$$

Correspondingly, the generalized processes $\Phi(\varphi)$ and $\Phi_T(\varphi)$ now include all of the finite dimensional distributions of $X_T^*(r)$ and $W(r)$. Thus:

$$\Phi(\delta(t - t_0)) = \int_0^1 \delta(r - t_0)W(r)\,dr = W(t_0)$$

and

$$\Phi\left(\sum_{i=1}^{m} a_i\delta(t-t_i)\right) = \Sigma_{i=1}^{m} \int_0^1 \delta(r-t_i)W(r)\,dr = \Sigma_{i=1}^{m} a_i W(t_i)$$

for arbitrary constants a_i $(i = 1, \ldots, m)$. This method of extracting the finite dimensional distributions of random elements such as $W(r)$ is also useful in generating asymptotic expansions.

As an example we shall take the simple case of a sequence $\{u_t\}_1^\infty$ of identically and independently distributed (i.i.d.) $N(0,1)$ variates. Here $n = 1, \Sigma = 1$ and we find from formula (A1) in the Appendix that:

$$\begin{aligned}
B_T(\varphi,\varphi) &= \sum_{j=1}^{T}\left(\frac{j-1}{T}\right)\int_{(j-1)/T}^{j/T}\int_{(j-1)/T}^{j/T}\varphi(t)\varphi(s)\,ds\,dt \\
&\quad +2\sum_{j=2}^{T}\int_{(j-1)/T}^{j/T}\sum_{k=1}^{j-1}\left(\frac{k-1}{T}\right)\varphi(t)\int_{(k-1)/T}^{k/T}\varphi(s)\,ds\,dt \\
&= B(\varphi,\varphi) - T^{-1}\sum_{j=1}^{T}\int_{(j-1)/T}^{j/T}\int_{(j-1)/T}^{j/T}\varphi(t)\varphi(s)(tT-[tT])\,dt\,ds \\
&\quad -(2/T)\sum_{j=2}^{T}\int_{(j-1)/T}^{j/T}\sum_{k=1}^{j-1}\varphi(t)\int_{(k-1)/T}^{k/T}\varphi(s)(sT-[sT])\,ds\,dt \\
&= B(\varphi,\varphi) - T^{-1}\sum_{j=1}^{T}\{\bar\varphi(j/T)-\bar\varphi((j-1)/T)\} \\
&\quad \times\left\{\bar\varphi(j/T) - T\int_{(j-1)/T}^{j/T}\bar\varphi(r)\,dr\right\} \\
&\quad -(2/T)\sum_{j=2}^{T}\{\bar\varphi(j/T)-\bar\varphi((j-1)/T)\} \\
&\quad \times\left\{\sum_{k=1}^{j-1}\bar\varphi(k/T) - T\int_0^{(j-1)/T}\bar\varphi(r)\,dr\right\},
\end{aligned} \tag{15}$$

where $\bar\varphi(r) = \int_0^r \varphi(t)\,dt$ and $B(\varphi,\varphi) = \int_0^1(\bar\varphi(1)-\bar\varphi(t))^2\,dt$ (compare equation (8) above). We may write equation (15) in the abbreviated form:

$$B_T(\varphi,\varphi) = B(\varphi,\varphi) - (1/T)E(\varphi)$$

and the characteristic functional $L_T(\varphi)$ now has the expansion

$$L_T(\varphi) = \exp\{-(1/2)B(\varphi,\varphi)\}[1 + (1/2T)E(\varphi) + O(T^{-2})]. \tag{16}$$

Let us suppose that we are concerned with developing Edgeworth expansions of the finite dimensional distributions of $X_T(t)$. We may start with the unidimensional case of $X_T(t_0)$ with t_0 fixed ($0 < t_0 \leqslant 1$). The corresponding generalized process is $\Phi_T(\delta(t - t_0))$. Moreover, setting $\varphi(t) = y\delta(t - t_0)$, we obtain:

$$\bar{\varphi}(r) = \begin{cases} y & r \geqslant t_0 \\ 0 & r < t_0 \end{cases}$$

$$B(\varphi, \varphi) = y^2 t_0$$

$$\begin{aligned} E(\varphi) &= \left[\bar{\varphi}\left(\frac{[Tt_0] + 1}{T}\right) - \bar{\varphi}\left(\frac{[Tt_0]}{T}\right) \right] \\ &\quad \times \left[\varphi\left(\frac{[Tt_0] + 1}{T}\right) - T \int_{t_0}^{([Tt_0]+1)/T} \bar{\varphi}(r)\, dr \right] \\ &= y^2 \left\{ 1 - T\left(\frac{[Tt_0] + 1}{T} - t_0\right) \right\} \\ &= y^2 \{ Tt_0 - [Tt_0] \}. \end{aligned}$$

Hence, from equation (16) we find:

$$\begin{aligned} g(y) &= L_T(y\delta(t - t_0)) \\ &= \exp\{-(y^2 t_0/2)\} \left[1 + \frac{y^2}{2T}(Tt_0 - [Tt_0]) \right] + O(T^{-2}). \end{aligned} \qquad \textbf{(17)}$$

Inverting equation (17) we obtain the Edgeworth expansion of the density of $X(t_0)$, namely

$$\begin{aligned} \mathrm{pdf}(x) &= \frac{1}{2\pi} \int e^{-\imath xy} g(y)\, dy \\ &= \frac{1}{t_0^{1/2}} f\left(\frac{x}{t_0^{1/2}}\right) - \frac{1}{2T}(Tt_0 - [Tt_0])\left(\frac{x^2}{t_0} - 1\right)\frac{1}{t_0^{3/2}} f\left(\frac{x}{t_0^{1/2}}\right) \\ &\quad + O(T^{-2}), \end{aligned} \qquad \textbf{(18)}$$

where $f(z) = (2\pi)^{-1/2} e^{-z^2/2}$ is the density of the standard $N(0, 1)$ distribution. Upon integration we find the expansion of the distribution function of $X_T(t_0)$:

$$\mathrm{cdf}(x) = F((x/t_0^{1/2})(1 + a_T)) + O(T^{-2}), \qquad \textbf{(19)}$$

where $F(z)$ is the cdf of $N(0, 1)$ and

$$a_T = \frac{1}{2Tt_0}(Tt_0 - [Tt_0]).$$

Noting that $W(t_0)$ is $N(0, t_0)$ with cdf $F(x/t_0^{1/2})$ we deduce from equations (18) and (19) the stochastic representation:

$$X_T(t_0) \equiv W(t_0)[1 - a_T] + O_p(T^{-2})$$
$$\equiv W(t_0)\left[1 - \frac{Tt_0 - [Tt_0]}{2Tt_0}\right] + O_p(T^{-2}). \quad \textbf{(20)}$$

Higher-order finite dimensional distributions of $X_T(t)$ may be expanded in a similar way. Let us consider $(X_T(t_0), X(t_1))$ with $0 < t_0 < t_1 < 1$, or equivalently $(X(t_0), X(t_1) - X(t_0))$. Take any linear combination such as $aX(t_0) + b(X(t_1) - X(t_0))$ and define the corresponding generalized process $\Phi_T(\varphi) = \int_0^1 \varphi(t)X_T(t)\,dt$ with $\varphi(t) = y\{(a-b)\delta(t - t_0) + b\delta(t - t_1)\}$. Then

$$\bar{\varphi}(r) = \begin{cases} ya & r \geqslant t_1 \\ y(a-b) & t_0 \leqslant r < t_1 \\ 0 & r < t_0 \end{cases} \quad \textbf{(21)}$$

$$B(\varphi, \varphi) = y^2\{a^2 t_0 + b^2(t_1 - t_0)\} \quad \textbf{(22)}$$

and

$$E(\varphi) = y^2\{b^2(Tt_1 - [Tt_1]) + (a^2 - b^2)(Tt_0 - [Tt_0])\}. \quad \textbf{(23)}$$

From equation (16) we deduce that:

$$g(y) = L_T(\varphi(t))$$
$$= \exp\left\{-\frac{y^2}{2}[a^2 t_0 + b^2(t_1 - t_0)]\right\}$$
$$\times\left[1 + \frac{y^2}{2T}\left\{(a^2 - b^2)(Tt_0 - [Tt_0]) + b^2(Tt_1 - [Tt_1])\right\}\right]$$
$$+ O(T^{-2}) \quad \textbf{(24)}$$

and upon inversion we obtain the stochastic representation:

$$aX_T(t_0) + b(X_T(t_1) - X_T(t_0))$$
$$\equiv [aW(t_0) + b(W(t_1) - W(t_0)]$$
$$\times\left[1 - \frac{1}{2T\omega^2}\{(a^2 - b^2)(Tt_0 - [Tt_0]) + b^2(Tt_1 - [Tt_1])\}\right]$$
$$+ O_p(T^{-2}) \quad \textbf{(25)}$$

where $\omega^2 = a^2 t_0 + b^2(t_1 - t_0)$.

Both equations (20) and (25) may be checked by conventional methods for the asymptotic expansion of finite dimensional distributions.

4. ASYMPTOTIC EXPANSIONS IN VECTOR AUTOREGRESSIONS WITH INTEGRATED PROCESSES

Our concern in this section is with multiple time series of integrated processes that are generated in discrete time according to:

$$y_t = Ay_{t-1} + u_t; \qquad t = 1, 2, \ldots \tag{26}$$

$$A = I_n. \tag{27}$$

We set initial conditions of equation (26) at $t = 0$ and allow y_0 to be any random variable (with a fixed distribution) including, of course, a constant. The innovation sequence $\{u_t\}_1^\infty$ in equation (26) will be required to satisfy the conditions of Theorem 2.1 or 2.2. As discussed in [24] equation (26) includes quite general vector ARMA specifications because of the weak conditions imposed on $\{u_t\}_1^\infty$.

Define the matrices $Y' = [y_1, \ldots, y_T]$, $Y'_{-1} = [y_0, \ldots, y_{T-1}]$, and $U' = [u_1, \ldots, u_T]$. $A^* = Y'Y_{-1}(Y'_{-1}Y_{-1})^{-1}$ is the matrix of least squares regression coefficients from the vector autoregression on y_t on y_{t-1}. The first-order asymptotic theory for such a regression has been developed recently in [24]. In view of equation (27) we shall work with the following closely related estimator of A:

$$\tilde{A} = (1/2)(Y'_{-1}Y_{-1})^{-1/2}(Y'Y_{-1} + Y'_{-1}Y)(Y'_{-1}Y_{-1})^{-1/2}.$$

The first-order asymptotic theory for $\tilde{A}$ is given by:

THEOREM 4.1. *If $\{u_t\}_1^\infty$ satisfies the condition of Theorem 2.1 or 2.2 and if $\{y_t\}_0^\infty$ is generated by equation (26) then as $T \uparrow \infty$:*

$$T(\tilde{A} - 1) \Rightarrow (1/2)\left\{\int_0^1 B(r)B(r)'\,dr\right\}^{-1/2} \{B(1)B(1)' - \Sigma_0\} \times \left\{\int_0^1 B(r)B(r')\,dr\right\}^{-1/2} \tag{28}$$

where $\Sigma_0 = \lim_{T\to\infty} T^{-1}\sum_1^T E(u_t u_t')$.

Equation (28) implies that $\tilde{A} = I + O_p(T^{-1})$ and, of course, $\tilde{A}\vec{p}I$ as $T \uparrow \infty$. These results and equation (28) are especially interesting because of the generality of the underlying conditions on the innovation process $\{u_t\}_1^\infty$ under which they hold. In view of this generality, it is more than usually intriguing to study the adequacy of the asymptotic theory delivered by Theorem 4.1. Our main aim is to effect a refinement of these first-order asymptotics. First, we define $v_t = \mathrm{vech}(u_t u_t' - Eu_t u_t')$, the $n(n+1)/2$ vector of nonredundant elements of $u_t u_t' - E(u_t u_t')$.

LEMMA 4.2. *If $\{v_t\}_1^\infty$ satisfies the conditions of either Theorem 2.1 or 2.2, if $\{u_t\}_1^\infty$ satisfies the conditions of Theorem 3.2 and if $\{y_t\}_0^\infty$ is generated by equation (26) then*

(a) $T^{-2}Y_{-1}'Y_{-1} \equiv \int_0^1 B(r)B(r)'\,dr + T^{-1/2}\{\int_0^1 B(r)\,dr y_0' + y_0 \int_0^1 B(r)'\,dr\}$
$+ O_p(T^{-1})$

(b) $T^{-1}(Y_{-1}'U + U'Y_{-1}) \equiv B(1)B(1)' - \Sigma_0 + T^{-1/2}\{y_0 B(1)'$
$+ B(1)y_0' - \xi\} + O_p(T^{-1})$.

In (b) *ξ is a random symmetric $n \times n$ matrix distributed as $N(0, V)$ with covariance matrix*

$$V = P_D\left(\sum_{k=0}^{\infty} \Psi_k\right)P_D' \tag{29}$$

where

$$\Psi_0 = E(u_t u_t' \otimes u_t u_t') - \mathrm{vec}(\Sigma_0)\,\mathrm{vec}(\Sigma_0)'$$

$$\Psi_k = \Phi_k + \Phi_k'; \qquad (k = 1, 2, \ldots) \tag{30}$$

$$\Phi_k = E(u_t u_{t+k}' \otimes u_t u_{t+k}') - \mathrm{vec}(\Sigma_0)\,\mathrm{vec}(\Sigma_0)'$$

$$P_D = D(D'D)^{-1}D' \tag{31}$$

and D is the duplication matrix of [13]. $B(r)$ and ξ are statistically independent.

THEOREM 4.3. *If $\{u_t\}_1^\infty$ and $\{v_t\}_1^\infty$ satisfy the conditions of Lemma 4.2 then $T(\tilde{A} - I)$ has the following asymptotic expansion as $T \uparrow \infty$:*

$$T(\tilde{A} - I) \equiv (1/2)F^{-1/2}EF^{-1/2} + (1/2\sqrt{T})\{F^{-1/2}JF^{-1/2} - HE - EH\} + O_p(T^{-1}) \tag{32}$$

where

$$F = \int_0^1 B(r)B(r)'\,dr$$

$$E = B(1)B(1)' - \Sigma_0$$

$$G = \int_0^1 B(r)\,dr y_0' + y_0 \int_0^1 B(r)'\,dr$$

$$J = B(1)y_0' + y_0 B(1)' - \xi$$

and H is the unique positive semi-definite solution of the matrix equation

$$F^{-1/2}H + HF^{-1/2} = F^{-1}GF^{-1}.$$

In these formulae (as before) $B(r)$ is n-vector Brownian motion with covariance matrix Σ and ξ is an independent $N(0, V)$ matrix whose covariance matrix V is given by equation (29).

COROLLARY 4.4. *If the conditions of Theorem 4.3 hold and if the initial value $y_0 = 0$ then as $T \uparrow \infty$:*

$$\begin{aligned} T(\tilde{A} - I) \equiv{} & \left\{\int_0^1 B(r)B(r)'\,dr\right\}^{-1/2} \{B(1)B(1)' - \Sigma_0\} \left\{\int_0^1 B(r)B(r)'\,dr\right\}^{-1/2} \\ & - \frac{1}{2\sqrt{T}} \left\{\int_0^1 B(r)B(r)'\,dr\right\}^{-1/2} \xi \left\{\int_0^1 B(r)B(r)'\,dr\right\}^{-1/2} \\ & + O_p(T^{-1}). \end{aligned} \tag{33}$$

Theorem 4.3 provides an asymptotic expansion of the distribution of $T(\tilde{A} - I)$ that holds under very general conditions. These conditions apply for a wide class of stationary sequences $\{u_t\}_1^\infty$. They are certainly satisfied by stationary Gaussian sequences which satisfy the mixing condition $\Sigma_{m=1}^\infty m\rho_m < \infty$ and thereby include all finite order vector ARMA processes that are stationary and Gaussian. Many non-Gaussian stationary sequences which obey the mixing condition and whose third-order cumulants are zero will also satisfy the conditions of Theorem 4.3. The asymptotic expansion given by equation (32) may therefore be expected to have rather wide applicability.

We observe that, since the matrix variate ξ is independent of the vector Brownian motion $B(r)$ and since $E(\xi) = 0$, the correction term of $O(1/\sqrt{T})$ in the expansion of equation (33) contributes no adjustment to the mean of the limiting distribution of $T(\tilde{A} - I)$. Thus, the location of the limiting distribution should be a fairly accurate approximation in moderately sized samples when $y_0 = 0$. For the special case $n = 1$ and $\{u_t\}_1^\infty$ i.i.d. $N(0, \sigma^2)$, this is confirmed by the experimental results of [5]. When $y_0 \neq 0$, it is also clear from equation (32) that the initial conditions may have an important influence on the sampling distribution of $\tilde{A}$. This conclusion too is corroborated by the specialized experimental results of [5].

Note that when $\{u_t\}_1^\infty$ is i.i.d. $N(0, \Sigma)$ we have the following reductions in the formula for the covariance matrix of ξ given by (29)–(31):

$$\Sigma_0 = \Sigma$$
$$\Psi_k = 0; \qquad k = 1, 2, \ldots$$

and

$$\Psi_0 = (I + K_n)(\Sigma \otimes \Sigma) + (\text{vec}\,\Sigma)(\text{vec}\,\Sigma)'$$

where K_n is the commutation matrix. Thus

$$\begin{aligned} V &= P_D(1 + K_n)(\Sigma \otimes \Sigma)P_D \\ &= 2P_D(\Sigma \otimes \Sigma) \end{aligned}$$

since $I + K_n = 2P_D$ [13, pp. 427–428]. Therefore, in this case the distribution of the matrix variate ξ is simply:

$$\xi \equiv N(0, 2P_D(\Sigma \otimes \Sigma)).$$

In the scalar case ($n = 1$, $\Sigma = \sigma^2$, $\Sigma_0 = \sigma_0^2$, $V = v$, $A = a$) we see that

$$\tilde{a} = \Sigma y_t y_{t-1} / \Sigma y_{t-1}^2$$

is the conventional least squares regression estimator. Here we find that

$$v = 2\pi f_2(0) = \sigma_2^2, \text{ say}$$

where $f_2(\lambda)$ is the spectral density of $\{u_t^2 - Eu_t^2\}$. Thus, equation (33) yields

$$\begin{aligned} T(\tilde{a} - 1) &\equiv \frac{(1/2)(B(1)^2 - \sigma_0^2)}{\int_0^1 B(r)^2\,dr} - \frac{1}{2\sqrt{T}} \frac{\xi}{\int_0^1 B(r)^2\,dr} + O_p(T^{-1}) \\ &\equiv \frac{(1/2)(W(1)^2 - \sigma_0^2/\sigma^2)}{\int_0^1 W(r)^2\,dr} - \frac{\sigma_2}{2\sigma^2\sqrt{T}} \frac{\eta}{\int_0^1 W(r)^2\,dr} + O_p(T^{-1}). \end{aligned} \tag{34}$$

This is a simple but rather general formula for the asymptotic expansion of the distribution of the least squares estimator in an autoregression with a unit root. Note that equation (34) depends only on the error variance ($Eu_t^2 = \sigma_0^2$), the value of spectral density of the errors at the origin ($\sigma^2 = 2\pi f_u(0)$) and the value of the spectral density of ($u_t^2 - E(u_t^2)$) at the origin ($\sigma_2^2 = 2\pi f_2(0)$).

When $\{u_t\}_1^\infty$ is i.i.d. $N(0, \sigma^2)$ we have $\sigma^2 = \sigma_0^2$ and $\sigma_2^2 = \mathrm{var}(u_t^2) = 2\sigma^4$. In this highly specialized case equation (34) reduces to the formula first derived by the author in [21, equation (38)].

5. CONCLUSIONS

In earlier work [18, 19] the author developed analytic formulae for Edgeworth-type expansions in a stationary first-order autoregression. Simple derivations of these formulae and their extensions to stationary vector autoregressions are provided in other ongoing research [23]. The present paper complements this research by providing higher-order asymptotics in nonstationary VAR's.

The asymptotic expansions derived in this paper are quite different in character from traditional Edgeworth expansions. In the first place, formulae such as equation (32) yield refinements of a limiting distribution theory that is nonnormal. The first-order asymptotics are obtained through weak convergence on function spaces rather than Euclidean spaces and the limiting distributions take the form of functionals of multivariate Wiener processes. Correction terms in the refinement of this limit theory take the form of new functionals of Wiener processes. The resulting asymptotic expansion is quite different from the prototypical form of an Edgeworth expansion: i.e., a limiting normal density scaled by a polynomial whose coefficients are functions of the sample size and the (pseudo-) moments of the statistic [20].

Secondly, and more significantly, the asymptotic expansions developed here have a much wider range of applicability than traditional Edgeworth expansions. This is because the new expansions have their genesis in invariance principles (such as those of Section 2) which apply in very general situations, allowing for a wide class of different models and processes. Thus, the validity and form of the asymptotic expansion given in Theorem 4.3 by equation (32) is unaffected by the misspecification of the VAR. The true model may be vector ARMA or even vector ARMAX with stationary exogenous inputs. All that is required is that the process $\{y_t\}_1^\infty$ be integrated of order one with stationary innovations that satisfy the quite general moment and mixing conditions of Theorem 4.3. In this way, the asymptotic expansion of equation (32) shares some of the invariance principle properties of the underlying limit theory that it refines.

Many extensions of the work reported here are now possible and merit further research. The most immediate are regressions with fitted means, higher-order regressions, general multivariate regressions, and problems of prediction. Attention also needs to be given to the conventional least squares regression coefficient matrix (in place of our symmetric matrix $\tilde{A}$). This appears to involve a refinement of the theory of weak convergence to the matrix stochastic integral $\int_0^1 B\,dB'$ as in [22] and it will be more difficult. Finally, it should be mentioned that in the general scalar case equation (34) can be used to obtain numerical computations of the asymptotic expansion of the distribution of $T(\tilde{a} - 1)$ up to $O(T^{-1/2})$. Since η and $B(r)$ are independent, the joint characteristic function of the numerator and denominator of equation (34) is easy to obtain. Numerical evaluations of the required density (or distribution function) then follow by direct numerical calculation of the usual inversion formulae for ratios of random variables. Research on some of these topics is now underway.

NOTES

1. The reader is referred to [2], [6], [10], and [25] for an introduction to the theory of characteristic functionals and to [6], [10], and [30] for the theory of generalized random processes.

REFERENCES

1. Billingsley, P *Convergence of Probability Measures.* John Wiley New York, 1968.
2 Bochner, S *Harmonic Analysis and the Theory of Probability.* Berkeley: University of California Press, 1960.
3. Cooley, T F. and S. F. LeRoy. Atheoretical macroeconometrics: A critique *Journal of Monetary Economics* 16 (1985): 283–308
4. Doan, T., R. B. Litterman, and C. Sims Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews* 3 (1984)· 1–100
5. Evans, G.B.A. and N E Savin. Testing for unit roots: 1. *Econometrica* 49 (1981): 753–779
6. Gelfand, I. M. and N. Y Vilenkin. *Generalized Functions,* Vol. 4. New York: Academic Press, 1964.
7. Herrndorf, N. The invariance principle for φ-mixing sequences *Z Wahrscheinlichkeitstheorie verw. Gebiete* 63 (1983)· 97–108.
8. Herrndorf, N. A functional central limit theorem for ρ-mixing sequences. *Journal of Multivariate Analysis* 15 (1984): 141–146.
9. Herrndorf, N A functional central limit theorem for weakly dependent sequences of random variables *Annals of Probability* 12 (1984). 141–153.
10. Hida, T. *Brownian Motion* Springer Verlag. New York, 1980
11. Ibragimov, I A. and Y. V. Linnik. *Independent and Stationary Sequences of Random Variables.* Groningen Woller-Nordhoff, 1971.
12. Kolmogorov, A. N. and Y. A. Rozanov. On strong mixing conditions for stationary Gaussian processes. *Theory of Probability and Applications* 5 (1960)· 204–208.
13. Koopmans, T. C. *Statistical Inference in Dynamic Economic Models* New York Wiley, 1950.
14 Litterman, R B. Forecasting with Bayesian vector autoregressions· Four years of experience. Mimeographed (Federal Reserve Bank of Minneapolis), 1984.
15. Magnus, J R. and H. Neudecker. The elimination matrix: Some lemmas and applications. *SIAM Journal on Algebraic and Discrete Methods* 1 (1980) 442–449
16 Mann, H B. and A Wald. On the statistical treatment of linear stochastic difference equations. *Econometrica* 11 (1943). 173–220.

17. Pagan, A. Three econometric methodologies A critical appraisal. Australian National University (mimeo), 1986.
18 Phillips, P.C.B Approximations to some finite sample distributions associated with a first-order stochastic difference equation. *Econometrica* 45 (1977): 463–485.
19 Phillips, P.C B Edgeworth and saddlepoint approximations in a first-order noncircular autoregression. *Biometrika* 65 (1978): 91–98
20. Phillips, P.C.B. Exact small sample theory in the simultaneous equations model Chapter 8 and pp 449–516 in M D Intriligator and Z. Griliches (Eds.), *Handbook of Econometrics* Amsterdam: North-Holland, 1983.
21 Phillips, P.C.B Time series regression with a unit root Cowles Foundation Discussion Paper No. 740R, Yale University (April 1985), forthcoming in *Econometrica*, 1986.
22 Phillips, P.C B. Weak convergence to the matrix stochastic integral $\int_0^1 B\,dB'$. Cowles Foundation Discussion Paper No. 796, Yale University, July 1986.
23 Phillips, P.C.B. Asymptotic expansions in stationary vector autoregressions (in preparation).
24 Phillips, P C.B. and S N. Durlauf. Multiple time series regression with integrated processes. *Review of Economic Studies* (July 1986), Vol. 53, pp. 473–496.
25 Prohorov, Y. V. The method of characteristic functionals, in J. Neyman (Ed), *Proceedings of 4th Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 2, pp 403–419. Berkeley: University of California Press, 1961
26 Rozanov, Y. A *Stationary Random Processes* San Francisco. Holden Day, 1967.
27 Sims, C. A. Macroeconomics and reality. *Econometrica* 48 (1980): 1–48.
28. White, H. *Asymptotic Theory for Econometricians* New York: Academic Press, 1984.
29. White, H. and I. Domowitz. Nonlinear regression with dependent observations. *Econometrica* 52 (1984): 143–162
30 Yaglom, A. M. *An Introduction to the Theory of Stationary Random Functions.* New York: Dover, 1973.

MATHEMATICAL APPENDIX

Proof of Theorem 2.1. Define $X_T^*(r) = \Sigma^{-1/2}X_T(r)$. It is easy to see that individual elements of the vector random element $X_T^*(r)$ satisfy the conditions of Herrndorf's theorem [8, pp 142]. Thus $X_{iT}^*(r) \Rightarrow W_i(r)$ as $T \uparrow \infty$, where $W_i(r)$ is a standard Wiener process on $C[0,1]$ for each $i = 1, \ldots, n$. Moreover, since $D[0,1]$ is separable and complete under the metric d_0 [1, pp 112], the weak convergence of $X_{iT}^*(r) \Rightarrow W_i(r)$ implies that the family of marginal probability measures associated with the sequence $\{X_{iT}^*(r)\colon T = 1, 2, \ldots\}$ is tight by Prohorov's theorem [1, pp 37] Furthermore, tightness of these marginal probability measures ensures tightness of the family of probability measures associated with $X_T^*(r)$ on the product metric space D^n [1, pp. 41, exercise 6] Since the finite dimensional distributions of $X_T(r)$ converge to those of the multivariate Wiener process $W(r)$ (this may be proved as in the proof of Theorem 2.1 of [24]), it follows that $X_T^*(r) \Rightarrow W(r)$. We deduce that $X_T(r) \Rightarrow \Sigma^{1/2}W(r) = B(r)$, as required.

Proof of Theorem 2.2. The proof follows the same line of argument as that of the proof of Theorem 2.1 given above except that the univariate result of Herrndorf [9, Corollary 1, pp. 142] is used to establish tightness of the marginal probability measures.

Proof of Theorem 3.1.

$$\begin{aligned}
B_T(\varphi, \psi) &= E\{\Phi_T(\varphi)\Phi_T(\psi)\} \\
&= \int_0^1 \int_0^1 \varphi(t)' E\{X_T^*(t) X_T^*(s)'\}\psi(s)\, ds dt \\
&= T^{-1} \sum_{j=1}^{T} \int_{(j-1)/T}^{j/T} \int_{(j-1)/T}^{j/T} \varphi(t)'\Sigma^{-1/2} E(S_{j-1}S_{j-1}')\Sigma^{-1/2}\psi(s)\, ds dt \\
&\quad + T^{-1} \sum_{j=2}^{T} \int_{(j-1)/T}^{j/T} \sum_{k=1}^{j-1} \left[\varphi(t)'\Sigma^{-1/2} E(S_{j-1}S_{k-1}')\Sigma^{-1/2} \int_{(k-1)/T}^{k/T} \psi(s)\, ds dt \right. \\
&\quad \left. + \psi(t)'\Sigma^{-1/2} E(S_{j-1}S_{k-1}')\Sigma^{-1/2} \int_{(k-1)/T}^{k/T} \varphi(s)\, ds dt \right] \qquad \textbf{(A1)}
\end{aligned}$$

Now since $\{u_t\}$ is weakly stationary

$$\begin{aligned}
T^{-1}E(S_{j-1}S_{j-1}') &= T^{-1} \sum_{r=1}^{j-1} \sum_{s=1}^{j-1} E(u_r u_s') \\
&= \frac{j-1}{T} \left[E(u_1 u_1') + \sum_{r=2}^{j-1} \{E(u_1 u_r') + E(u_r u_1')\} \right] \\
&\quad - \frac{1}{T} \sum_{r=1}^{j-2} r E(u_1 u_{1+r}' + u_{r+1} u_1')
\end{aligned}$$

and

$$\Sigma = E(u_1u_1') + \sum_{k=2}^{\infty} [E(u_1u_k') + E(u_ku_1')]$$

so that

$$T^{-1}E(S_{j-1}S_{j-1}') = (j-1)/T\Sigma - (j-1)/T \sum_{k=j}^{\infty} [E(u_1u_k') + E(u_ku_1')]$$
$$- (1/T)\sum_{r=1}^{j-2} r[E(u_1u_{1+r}') + E(u_{r+1}u_1')] \quad \textbf{(A2)}$$

Since $\sum_{r=1}^{\infty} r\rho_r < \infty$ the final term of (A2) is $O(T^{-1})$. Moreover, elements of the second term of (A2) are dominated by

$$(j-1)/T \sum_{k=j}^{\infty} \rho_k = O(T^{-1})$$

since $\sum_{j=1}^{\infty} \sum_{r=j}^{\infty} \rho_r = \sum_{r=1}^{\infty} r\rho_r < \infty$ and, hence, $\sum_{r=j}^{\infty} \rho_r = O(1/j)$ as $j \uparrow \infty$. It follows that

$$T^{-1}E(S_{j-1}S_{j-1}') = (j-1)/T\Sigma - (1/T)\sum_{r=1}^{j-2} r[E(u_1u_{1+r}') + E(u_{r+1}u_1')] + O(T^{-1})$$
$$= (j-1)/T\Sigma + O(T^{-1}). \quad \textbf{(A3)}$$

In a similar way we find that for $j > k$

$$T^{-1}E(S_{j-1}S_{k-1}') = T^{-1}\sum_{r=1}^{j-1}\sum_{s=1}^{k-1} E(u_ru_s')$$
$$= T^{-1}\sum_{r=1}^{k-1}\sum_{s=1}^{k-1} E(u_ru_s') + T^{-1}\sum_{r=k}^{j-1}\sum_{s=1}^{k-1} E(u_ru_s')$$
$$= (k-1)/T\left[E(u_1u_1') + \sum_{r=2}^{k-1}\{E(u_1u_r') + E(u_ru_1')\}\right]$$
$$- (1/T)\sum_{r=1}^{k-2} rE(u_1u_{1+r}' + u_{r+1}u_1') + (1/T)\sum_{r=k}^{j-1}\sum_{s=1}^{k-1} E(u_ru_s').$$

Note that

$$T^{-1}\sum_{r=k}^{j-1}\sum_{s=1}^{k-1} E(u_ru_s') = T^{-1}\sum_{s=1}^{k-1}\sum_{q=k-s}^{j-1-s} E(u_{1+q}u_1')$$

whose elements are dominated by

$$T^{-1}\sum_{r=1}^{j-1} r\rho_r < T^{-1}\sum_{r=1}^{\infty} r\rho_r.$$

Hence, we deduce as before that

$$T^{-1}E(S_{j-1}S'_{k-1}) = (k-1)/T\Sigma + O(T^{-1}). \tag{A4}$$

It now follows from (A1), (A3), and (A4) that

$$\begin{aligned} B_T(\varphi,\psi) = {} & T^{-1}\sum_{j=1}^{T}\int_{(j-1)/T}^{j/T}\int_{(j-1)/T}^{j/T}[Tt]\varphi(t)'\psi(s)\,dtds \\ & + T^{-1}\sum_{j=2}^{T}\int_{(j-1)/T}^{j/T}\sum_{k=1}^{j-1}\Bigg[[Ts]\varphi(t)'\int_{(k-1)/T}^{k/T}\psi(s)\,dsdt \\ & + [Ts]\psi(t)'\int_{(k-1)/T}^{k/T}\varphi(s)\,dsdt\Bigg] + O(T^{-1}). \end{aligned}$$

We observe that

$$\frac{[Tt]}{T} = t + O(T^{-1}), \qquad \frac{[Ts]}{T} = s + O(T^{-1})$$

and thus

$$\begin{aligned} B_T(\varphi,\psi) &= \int_0^1\int_0^1 \varphi(t)'\psi(s)\min(t,s)\,dsdt + O(T^{-1}) \\ &= B(\varphi,\psi) + O(T^{-1}) \end{aligned}$$

as required.

When $\{u_t\}$ is Gaussian, the characteristic functional of Φ_T is therefore

$$\begin{aligned} L_T(\varphi) &= \exp\{-(1/2)B_T(\varphi,\varphi)\} \\ &= \exp\{-(1/2)B(\varphi,\varphi)\}[1 + O(T^{-1})]. \\ &= L(\varphi)[1 + O(T^{-1})]. \end{aligned}$$

We deduce that

$$\Phi_T(\varphi) \equiv \Phi(\varphi) + O_p(T^{-1}), \qquad \text{all } \varphi \in K_n$$

and thus

$$X_T^*(r) \equiv W(r) + O_p(T^{-1})$$

as required.

Proof of Theorem 3.2. Note that $X_T^*(r)$ is linear in $\{u_t\}$ Hence, third-order cumulants of $X_T^*(r)$ are zero and the rest of the proof is given in the main text.

Proof of Theorem 4.1. This theorem is proved in the same way as Theorem 3.2 of [24].

Proof of Lemma 4.2. Note that $y_j = S_j + y_0$ and we may write:

$$
\begin{aligned}
T^{-2}Y_{-1}'Y_{-1} &= T^{-2}\sum_{j=1}^{T}(S_{j-1}+y_0)(S_{j-1}+y_0)' \\
&= \sum_{j=1}^{T}\int_{(j-1)/T}^{j/T}(X_T(r)+T^{-1/2}y_0)(X_T(r)+T^{-1/2}y_0)'\,dr \\
&= \int_0^1 X_T(r)X_T(r)'\,dr + T^{-1/2}y_0\int_0^1 X_T(r)'\,dr \\
&\quad + \int_0^1 X_T(t)\,dr(T^{-1/2}y_0') + T^{-1}y_0y_0'. \qquad \textbf{(A5)}
\end{aligned}
$$

By Theorem 3.2 $X_T(r) \equiv B(r) + O_p(T^{-1})$. We deduce from this representation of $X_T(r)$, from (A5) and from the continuous mapping theorem that.

$$
\begin{aligned}
T^{-2}Y_{-1}'Y_{-1} &\equiv \int_0^1 B(r)B(r)'\,dr \\
&\quad + T^{-1/2}\left\{\int_0^1 B(r)\,dry_0' + y_0\int_0^1 B(r)'\,dr\right\} + O_p(T^{-1}) \qquad \textbf{(A6)}
\end{aligned}
$$

proving (a).

To prove (b) we first note that

$$
Y_{-1}'U = \sum_1^T y_{t-1}u_t' = \sum_{j=1}^{T}(S_{j-1}+y_0)u_j' = \sum_{j=1}^{T}S_{j-1}u_j' + y_0\sum_{j=1}^{T}u_j'
$$

and

$$
S_TS_T' = \sum_1^T u_ju_j' + \sum_1^T(S_{j-1}u_j' + u_jS_{j-1}').
$$

Thus

$$
Y_{-1}'U + U'Y_{-1} = S_TS_T' - \sum_1^T u_ju_j' + y_0S_T' + S_Ty_0' \qquad \textbf{(A7)}
$$

Now from Theorem 3.2 we have

$$
T^{-1/2}S_T \equiv B(1) + O_p(T^{-1}). \qquad \textbf{(A8)}
$$

In addition, $\Sigma_0 = Eu_tu_t'$ and $\{(u_tu_t' - \Sigma_0)\}_1^\infty$ is a sequence of random matrices whose nonredundant elements $\{v_t\}_1^\infty$ satisfy the conditions of Theorem 2.1 or 2.2. Thus, following the argument of the proof of Theorem 3.3 of [24], we find that

$$
T^{-1/2}\sum_{t=1}^{T}(u_tu_t' - \Sigma_0) \Rightarrow N(0, V) \qquad \textbf{(A9)}
$$

as $T \uparrow \infty$, where the covariance matrix V of the limiting matrix normal distribution is given by equation (29). Since the error on the asymptotic approximation (A9) is at least as small as $O_p(T^{-1/2})$ we have

$$T^{-1/2}\sum_1^T (u_t u_t' - \Sigma_0) \equiv \xi + O_p(T^{-1/2}) \tag{A10}$$

where $\xi \equiv N(0, V)$ We now deduce from (A7), (A8), and (A10) that

$$\begin{aligned} T^{-1}(Y_{-1}'U + U'Y_{-1}) \equiv\ & B(1)B(1)' - \Sigma_0 \\ & + T^{-1/2}\{y_0 B(1)' + B(1)y_0' - \xi\} \\ & + O_p(T^{-1}) \end{aligned}$$

giving (b) as required. Note, finally, that ξ depends on a quadratic function of u_t whereas $B(r)$ depends on partial sums which are linear in the u_t. Hence, ξ and $B(r)$ are uncorrelated and, being normal, they are independent as stated in the Lemma.

Proof of Theorem 4.3. Note that

$$(F + T^{-1/2}G)^{-1} = F^{-1} - T^{-1/2}F^{-1}GF^{-1} + O_p(T^{-1})$$

and thus

$$(F + T^{-1/2}G)^{-1/2} = F^{-1/2} - T^{-1/2}H + O_p(T^{-1})$$

where H is the unique positive semi-definite solution of

$$F^{-1/2}H + HF^{-1/2} = F^{-1}GF^{-1}.$$

It now follows that

$$\begin{aligned} (T^{-2}Y_{-1}'Y_{-1})^{-1/2} &\equiv (F + T^{-1/2}G + O_p(T^{-1}))^{-1/2} \\ &= F^{-1/2} - T^{-1/2}H + O_p(T^{-1}). \end{aligned}$$

Recall from Lemma 4 2 part (b) that

$$T^{-1}(Y_{-1}'U + U'Y_{-1}) \equiv E + T^{-1/2}J + O_p(T^{-1}).$$

Now

$$\begin{aligned} T(\tilde{A} - I) &= (1/2)(T^{-2}Y_{-1}'Y_{-1})^{-1/2}\{T^{-1}(Y_{-1}'U + U'Y_{-1})\}(T^{-2}Y_{-1}'Y_{-1})^{-1/2} \\ &\equiv (1/2)(F^{-1/2} - T^{-1/2}H)(E + T^{-1/2}J)(F^{-1/2} - T^{-1/2}H) + O_p(T^{-1}) \\ &= (1/2)F^{-1/2}EF^{-1/2} + (1/2\sqrt{T})\{F^{-1/2}JF^{-1/2} - HE - EH\} + O_p(T^{-1}) \end{aligned}$$

as required for equation (32).