

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station  
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 765

Note: Cowles Foundation Discussion Papers are preliminary materials circulated to stimulate discussion and critical comment. Requests for single copies of a Paper will be filled by the Cowles Foundation within the limits of the supply. References in publications to Discussion Papers (other than mere acknowledgment by the writer that he has access to such unpublished material) should be cleared with the author to protect the tentative character of these papers.

ASYMPTOTIC EXPANSIONS IN NONSTATIONARY VECTOR AUTOREGRESSIONS

Peter C. B. Phillips

August 1985

ASYMPTOTIC EXPANSIONS IN  
NONSTATIONARY VECTOR AUTOREGRESSIONS

P. C. B. Phillips\*

Cowles Foundation for Research in Economics  
Yale University

O. ABSTRACT

This paper studies the statistical properties of vector autoregressions (VAR's) for quite general multiple time series which are integrated 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 specified to the scalar case and are related to other recent work by the author in [17] and [19].

August, 1985

---

\*My thanks, as always, go to Glenna Ames for her skill and effort in typing the manuscript of this paper. The research reported here was supported by the NSF under Grant No. SES 8218792.

## 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 [21] and has been recently reviewed and discussed in [4]. Almost all of this research has been empirical. Much of the most recent work in the field has been motivated by issues of economic policy as in [4] and by problems of prediction as in [12].

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, there appears to be no general statistical theory of estimation and inference in models of nonstationary time series. Moreover, while empirical investigators recognize that many of the time series that are used in their regressions actually display nonstationary behavior, they have been forced out of necessity to rely on traditional methods of inference for stationary time series in interpreting the results of these regressions.

The present paper draws much of its motivation from the needs expressed in the last paragraph. Our concern will focus on the statistical properties of the estimated regression coefficients in VAR's. More specifically, we shall provide an extension to the multivariate case of the theory developed by the author in other recent work [17] for scalar autoregressions with a unit root. As in [17], the conditions we impose on the time series are quite weak and allow for a wide class of weakly dependent and heterogeneously distributed processes.

A major aim of the paper 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 [3], Prohorov [20] and others to assist in the study of Banach valued random variables. The first order asymptotics we develop here and in related work [17, 19] rely on central limit theory on Banach spaces (functional central limit theory). The first step in the development of higher order asymptotics, therefore, involves a refinement of this functional central limit theory. In the second step, the 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 a multivariate functional central limit theory for partial sums of  $\rho$ -mixing innovations. This theory is related to an alternative multivariate result for  $\varphi$ - and  $\alpha$ -mixing processes which is proved in [19]. 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 an 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  $(\Omega, \mathcal{B}, P)$  with

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

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

that the limit

$$(2) \quad \Sigma = \lim_{T \rightarrow \infty} T^{-1} M_T$$

exist and be positive definite. More generally, we may allow  $M_T$  to have the representation  $M_T = TH(T)$ , where  $H(T)$  is a positive definite and slowly varying matrix function of  $T$ , so that  $\lim_{T \rightarrow \infty} H(kT)H(T)^{-1} = I_n$  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, p. 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 stronger requirement (2) in what follows.

We introduce the random element

$$(3a) \quad X_T(t) = T^{-1/2} \Sigma^{-1/2} S_{[Tt]} = T^{-1/2} \Sigma^{-1/2} S_{j-1} ; ((j-1)/T \leq t < j/T, j=1, \dots, T)$$

$$(3b) \quad X_T(1) = T^{-1/2} \Sigma^{-1/2} S_T$$

where  $[a]$  denotes the integer part of  $a$ .  $X_T(t)$  lies in the product metric space  $D^n = D[0,1] \times \dots \times D[0,1]$  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

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

where  $d_0(,)$  is the modified Skorohod metric (see [2, p. 112]) under which  $D[0,1]$  is separable and complete.

As in [17] and [19] 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 the maximal correlation coefficient  $\rho_m$ , which we define by:

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

where  $L_r^S$  denotes the subspace generated by the variables  $\{u_t, r \leq t \leq s\}$ .

Sequences  $\{u_t\}_1^\infty$  for which  $\rho_m \rightarrow 0$  as  $m \rightarrow \infty$  are said to be  $\rho$ -mixing.

Such sequences were introduced for scalar processes by Kolmogorov and Rozanov [11]. The mixing condition  $\rho_m \rightarrow 0$  has the simple interpretation that members of the sequence  $\{u_t\}$  which are separated by at least  $m$  time periods have correlation which tends to zero as  $m \rightarrow \infty$ . The condition is easy to verify for many commonly occurring time series models.

Thus, stationary ARMA processes have correlation sequences which are known to decay exponentially and are therefore  $\rho$ -mixing. Moreover, under weak moment conditions such as  $\sup_t E|u_t|^{2+\delta} < \infty$  for some  $\delta > 0$ , sequences which are strong (or,  $\alpha$ -) mixing (and a fortiori  $\phi$ - or  $\psi$ -mixing) are also  $\rho$ -mixing [10, p. 307]. In addition, simple processes such as the stable AR(1) driven by Bernoulli innovations that are known to be not strong mixing [1] are, in fact,  $\rho$ -mixing. The class of processes included by this condition is, therefore, rather wide.

One reason for the popularity of other measures of weak dependence, such as strong and  $\phi$ -mixing, in recent econometric work (see, for example, [17], [22], [23]) is that these conditions and also the mixing decay rates continue to apply to measurable functions of the mixing processes [22, p. 47]. The same result does not apply to  $\rho$ -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  $\rho$ -mixing. In such situations, the  $\rho$ -mixing condition is usually sufficient to determine

the limiting distribution of the functional. Moreover, as we see below, no additional condition is required on the mixing decay rate, in contrast to the limit theory based on  $\alpha$ - or  $\varphi$ -mixing processes in [17] and [19].

The main result of this section is:

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

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

$$E|u_{iT}|^{2+b} = O(T^{b/2-\varepsilon}) , \quad (i = 1, \dots, n)$$

*then  $X_T(t) \Rightarrow W(t)$  as  $T \uparrow \infty$  where  $W(t)$  is a vector Wiener process.*

We use the symbol  $\Rightarrow$  to signify the weak convergence of the associated probability measures [2]. Each element of the limit process  $W(t)$  is a univariate Wiener process, the elements of  $W(t)$  are independent and the sample paths of  $W(t)$  lie almost surely in the function space  $C^n = C[0,1] \times \dots \times C[0,1]$ , the product space of  $n$  copies of  $C[0,1]$  (the space of all real valued continuous functions on the interval  $[0,1]$ ).

Theorem 2.1 is a multivariate generalization of a functional central limit theorem for  $\rho$ -mixing scalar sequences established recently by Herrndorf [8]. It provides an alternative to the closely related multivariate limit theorem presented recently in [19]. The latter theorem may also be relied upon in our theoretical development and is especially useful in cases where there are other reasons (such as those given above) for employing  $\alpha$ - or  $\varphi$ -mixing conditions. For convenience, we shall state the result here, using  $\alpha_m$  and  $\varphi_m$  to represent the ( $\alpha$ - and  $\varphi$ -, respectively) mixing coefficients which measure the dependence between events separated

by  $m$  time periods.

*THEOREM 2.2.* If  $\{u_t\}_1^\infty$  is a sequence of random  $n$ -vectors satisfying (1),

(2) and

- (a)  $\{u_{it}^2\}$  is uniformly integrable for all  $i = 1, \dots, n$  ;
- (b)  $\sup_t (E|u_{it}|^\beta) < \infty$  for some  $2 \leq \beta < \infty$  and all  $i = 1, \dots, n$  ;
- (c)  $E\{T^{-1}(S_{k+T} - S_k)(S_{k+T} - S_k)'\} \rightarrow \Sigma$  as  $\min(k, T) \uparrow \infty$  ;
- (d) either  $\varphi_m$  is of size  $-\beta/(2\beta-2)$  or  $\beta > 2$  and  $\alpha_m$  is of size  $-\beta(\beta-2)$

then  $X_T(t) \Rightarrow W(t)$  , a vector Wiener process, as  $T \uparrow \infty$  .

The conditions of Theorem 2.1 are generally weaker than those of Theorem 2.2. In particular, the requirement that  $\{u_t\}_1^\infty$  be  $\rho$ -mixing eliminates the need for the mixing decay rate condition (d) of Theorem 2.2. Moreover, the moment condition (b) of Theorem 2.2 ensures that the  $\varphi$ - or  $\alpha$ -mixing processes are actually  $\rho$ -mixing [10, p. 307 and p. 309]. Finally, in contrast to (b) of Theorem 2.2, the moment condition (c) of Theorem 2.1 allows for moderate growth in the higher moments of  $u_t$  as the process evolves.

We shall also have occasion to use the random element:

$$(6a) \quad Z_T(t) = T^{-1/2} \Sigma^{-1/2} S_{[Tt]} + T^{-1/2} (Tt - [Tt]) \Sigma^{-1/2} u_{[Tt]+1}$$

for  $(j-1)/T \leq t < j/T$  ( $j = 1, \dots, T$ )

$$(6b) \quad Z_T(1) = T^{-1/2} \Sigma^{-1/2} S_T .$$

Now  $Z_T(t) \in C^n$  , which we may endow with the uniform metric

$$d_u(f, g) = \max_i \sup_t |f_i(t) - g_i(t)| \quad \text{for } f, g \in C^n .$$

Note that since  $W(t) \in C^n$  almost surely (with respect to Wiener



measure) we may, according to [2, p. 151], deduce from Theorem 2.1 that  $X_T(t) \Rightarrow W(t)$  where the weak convergence to multivariate Wiener measure is interpreted in the sense of the uniform topology induced by the metric  $d_u$  on  $D^n$ . We also have:

*THEOREM 2.3.*  $X_T(t) \Rightarrow W(t)$  iff  $Z_T(t) \Rightarrow W(t)$ .

### 3. CHARACTERISTIC FUNCTIONALS AND REFINEMENTS OF FUNCTIONAL CENTRAL LIMIT THEORY

The multivariate limit process  $W(t)$  of Theorems 2.1 and 2.2 is a Banach valued random variable. Its distribution is determined by the multivariate Wiener measure on  $(C^n, C^n)$  where  $C^n$  (a Banach space) is the support of  $W(t)$  (i.e.  $W(C^n) = 1$  where  $W(\cdot)$  denotes Wiener measure) and  $C^n$  is the class of Borel sets on  $C^n$  (i.e. the  $\sigma$ -field generated by the open subsets of  $C^n$  with the uniform metric  $d_u$ ). This distribution is also uniquely determined by the characteristic functional of the generalized random process<sup>1</sup> corresponding to  $W(t)$ .

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, which would require methods outside the scope of the present paper. As pointed out recently in [17], to the author's

---

<sup>1</sup>The reader is referred to [3], [6], [9] and [20] for an introduction to the theory of characteristic functionals and to [6], [9] and [24] for the theory of generalized random processes.

knowledge, no work at all has yet been done on asymptotic expansions for central limit theorems on function spaces. What follows, therefore, is a preliminary 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 on  $K_n$  [6] and will be denoted by  $\Phi(\varphi)$ ,  $\varphi \in K_n$ . For the multivariate Wiener process  $W(t)$  we may define the corresponding generalized random process by the integral

$$(7) \quad \Phi(\varphi) = \int_0^1 \varphi(t)' W(t) dt$$

which is well defined for all  $\varphi \in K_n$ . The correlation functional of  $\Phi$  is given by ( $\varphi, \psi \in K_n$ ):

$$\begin{aligned} B(\varphi, \psi) &= E\{\Phi(\varphi)\Phi(\psi)\} \\ &= \int_0^1 \int_0^1 \varphi(t)' \psi(s) \min(t, s) dt ds \\ (8) \quad &= \int_0^1 \hat{\varphi}(t)' \hat{\psi}(t) dt \end{aligned}$$

where

$$\hat{\varphi}(t) = \int_t^1 \varphi(s) ds, \quad \hat{\psi}(t) = \int_t^1 \psi(s) ds.$$

Formula (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 [9, p. 125].

Given  $\varphi \in K$ , the distribution of the linear functional  $\Phi$  is

$N(0, B(\varphi, \varphi))$  . It follows that the characteristic functional of the generalized random process  $\Phi(\varphi)$  is given by:

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

As in the case of distributions on finite dimensional spaces, the characteristic functional uniquely determines the probability measure. Here, the measure is Wiener measure on  $(C^n, C^n)$  . 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 [9, p. 122]. We note that: (i)  $L(\varphi)$  is a continuous functional in the sense that  $L(\varphi_k) \rightarrow L(\varphi)$  whenever  $\varphi_k \rightarrow \varphi$  as  $k \rightarrow \infty$  for any sequence  $\{\varphi_k\}$  and limit function  $\varphi$  in  $K$  ; (ii)  $L(\varphi)$  is positive definite in the sense that for any functions  $\varphi_1, \dots, \varphi_m$  in  $K_n$  and any complex numbers  $\alpha_1, \dots, \alpha_m$  the inequality  $\sum_1^m \sum_1^m L(\varphi_j - \varphi_k) \alpha_j \bar{\alpha}_k \geq 0$  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  $\Phi_T(\varphi) = \int_0^1 \phi(t)' X_T(t) dt$  corresponding to  $X_T(t)$  . The correlation functional of  $\Phi_T$  is  $B_T(\varphi, \psi) = E\{\Phi_T(\varphi)\Phi_T(\psi)\}$  , with  $\varphi, \psi \in K_n$  , 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  $W(t)$  . There is a related expansion for  $L_T(\varphi)$  and a stochastic expansion of  $X_T(t)$  about  $W(t)$  . Our first result is the following:

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

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

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

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

for any  $\varphi \in K_n$  and

$$(12) \quad X_T(t) = W(t) + O_p(T^{-1}) .$$

The condition  $\sum_1^\infty m\rho_m < \infty$  on the mixing coefficients  $\rho_m$  is not very restrictive. It is, for example, satisfied by all stationary finite order ARMA processes, since  $\rho_m = O(\lambda^{-m})$  with  $\lambda > 1$ , for such processes and thus  $\sum_1^\infty m\rho_m = \lambda/(\lambda-1)^2 < \infty$ .

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:

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

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

$$c_{kT}(t_1, \dots, t_k) = \text{cumulant}\{X_T(t_1), \dots, X_T(t_k)\} ; k = 1, 2, \dots$$

as

$$(14) \quad C_T(\varphi) = \sum_{k=1}^\infty \frac{i^k}{k!} \int_0^1 \dots \int_0^1 \varphi(t_1) \dots \varphi(t_k) c_{kT}(t_1, \dots, t_k) dt_1 \dots dt_k .$$

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

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

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  $r$ 'th cumulants of  $X_T(t)$  are  $O(T^{-r/2+1})$  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 (10). The characteristic functional is now:

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

and we obtain:

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

and

$$X_T(t) = W(t) + O_p(T^{-1/2}).$$

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

we have:

*THEOREM 3.2.* Let  $\{u_t\}_1^\infty$  be a weakly stationary  $\rho$ -mixing sequence of random  $n$ -vectors satisfying (1), (2) and  $\sum_{m=1}^\infty m\phi_m < \infty$ . If third order cumulants of  $\{u_t\}_1^\infty$  are zero, if  $r$ 'th cumulants of  $X_T(t)$  are  $O(T^{-r/2+1})$  and if the cumulant functional  $C_T(\varphi)$  admits an expansion of the form given in (14) then  $X_T(t) = W(t) + O_p(T^{-1})$ .

In some instances it is useful to consider generalized processes such as  $\Phi(\varphi)$  and  $\Phi_T(\varphi)$  where  $\varphi$  may lie in a function space that is larger than  $K_n$ . In general, the larger the space of test functions  $\varphi$  the narrower is the class of generalized random processes. However, in the present case where attention centers solely on the random elements  $W(t)$  and  $X_T(t)$ , 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  $K'_n$ . Since  $K'_n$  includes functionals such as the delta function for which

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

the generalized processes  $\Phi(\varphi)$  and  $\Phi_T(\varphi)$  now include all of the finite dimensional distributions of  $X_T(t)$  and  $W(t)$ . Thus:

$$\Phi(\delta(t - t_0)) = \int_0^1 \delta(t - t_0) W(t) dt = W(t_0)$$

and

$$\Phi(\sum_{i=1}^m a_i \delta(t - t_i)) = \sum_{i=1}^m \int_0^1 \delta(t - t_i) W(t) dt = \sum_{i=1}^m a_i W(t_i)$$

for arbitrary constants  $a_i$  ( $i = 1, \dots, m$ ). This method of extracting

the finite dimensional distributions of random elements such as  $W(t)$  is also useful in generating asymptotic expansions.

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

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

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 (8) above).

We may write (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

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

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 ( $0 < t_0 \leq 1$ ). The corresponding generalized process is  $\Phi_T(\delta(t - t_0))$ . Moreover, setting  $\varphi(t) = y\delta(t - t_0)$ , we obtain:

$$\bar{\varphi}(r) = y \quad r \geq t_0$$

$$0 \quad r < t_0$$

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

$$E(\varphi) = \left[ \bar{\varphi}\left(\frac{[Tt_0] + 1}{T}\right) - \bar{\varphi}\left(\frac{[Tt_0]}{T}\right) \right] \left[ \bar{\varphi}\left(\frac{[Tt_0] + 1}{T}\right) - T \int_{t_0}^{([Tt_0] + 1)/T} \bar{\varphi}(r) dr \right]$$

$$= y^2 \left\{ 1 - T \left( \frac{[Tt_0] + 1}{T} - t_0 \right) \right\}$$

$$= y^2 \{ Tt_0 - [Tt_0] \} .$$

Hence, from (16) we find:

$$(17) \quad g(y) = L_T(y\delta(t - t_0))$$

$$= \exp\{-(y^2 t_0/2)\} \left[ 1 + \frac{y^2}{2T} (Tt_0 - [Tt_0]) \right] + O(T^{-2}) .$$

Inverting (17) we obtain the Edgeworth expansion of the density of  $X(t_0)$ , viz.

$$\text{pdf}(x) = \frac{1}{2\pi} \int e^{-ixy} g(y) dy$$

$$= \frac{1}{t_0^{1/2}} i\left(\frac{x}{t_0^{1/2}}\right) - \frac{1}{2T} (Tt_0 - [Tt_0]) \left(\frac{x^2}{t_0} - 1\right) \frac{1}{t_0^{3/2}} i\left(\frac{x}{t_0^{1/2}}\right) + O(T^{-2})$$

where  $i(z) = (2\pi)^{-1/2} e^{-z^2/2}$  is the density of standard  $N(0,1)$  distribution.



Upon integration we find the expansion of the distribution function of

$X_T(t_0)$  :

$$(19) \quad \text{cdf}(x) = I((x/t_0^{1/2})(1+a_T)) + O(T^{-2})$$

where  $I(z)$  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  $I(x/t_0^{1/2})$  we deduce from (18) and (19) the stochastic expansion:

$$(20) \quad \begin{aligned} X_T(t_0) &= W(t_0)[1 - a_T] + O(T^{-2}) \\ &= W(t_0) \left[ 1 - \frac{Tt_0 - [Tt_0]}{2Tt_0} \right] + O(T^{-2}) . \end{aligned}$$

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

$$\begin{aligned} \bar{\varphi}(r) &= ya & , \quad r \geq t_1 \\ & y(a-b) & , \quad t_0 \leq r < t_1 \\ & 0 & , \quad r < t_0 \end{aligned}$$

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

and

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

From (16) we deduce that:

$$\begin{aligned} g(y) &= L_T(\varphi(t)) \\ &= \exp \left\{ -\frac{y^2}{2} [a^2 t_0 + b^2 (t_1 - t_0)] \right\} \left[ 1 + \frac{y^2}{2T} \{ (a^2 - b^2) (Tt_0 - [Tt_0]) + b^2 (Tt_1 - [Tt_1]) \} \right] + O(T^{-2}) \end{aligned}$$

and upon inversion we obtain the stochastic expansion:

$$\begin{aligned} (25) \quad & aX_T(t_0) + b(X_T(t_1) - X_T(t_0)) \\ &= [aW(t_0) + b(W(t_1) - W(t_0))] \left[ 1 - \frac{1}{2T\omega^2} \{ (a^2 - b^2) (Tt_0 - [Tt_0]) + b^2 (Tt_1 - [Tt_1]) \} \right] + O(T^{-2}) \end{aligned}$$

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

Both (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 of order one that are generated in discrete time according to:

$$(26a) \quad y_t = Ay_{t-1} + u_t ; \quad t = 1, 2, \dots$$

$$(26b) \quad A = I_n .$$

Either of the commonly proposed initial conditions may be used:

$$(27a) \quad y_0 = c , \quad \text{a constant; or}$$

$$(27b) \quad y_0 = \text{random with a certain specified distribution.}$$

The innovation sequence  $\{u_t\}_1^\infty$  in (26) will be required to satisfy the conditions of Theorem 2.1 or 2.2. As discussed in [19] (26) includes quite general vector ARMA specifications because of the weak conditions imposed on  $\{u_t\}_1^\infty$ .

Define the matrices  $Y' = [y_1, \dots, y_T]$ ,  $Y'_{-1} = [y_0, \dots, y_{T-1}]$ , and  $U' = [u_1, \dots, u_T]$ .  $A^* = Y'Y_{-1}(Y'_{-1}Y_{-1})^{-1}$  is the matrix of regression coefficients from the vector autoregression of  $y_t$  on  $y_{t-1}$ . The asymptotic theory for such a regression has been developed recently in [19], where the following result is proved:

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

$$(28) \quad T(A^* - I) \Rightarrow (1/2)\{\Sigma^{1/2}W(1)W(1)'\Sigma^{1/2} - \Sigma_U\}\{\Sigma^{1/2}\int_0^1 W(t)W(t)'\,dt\Sigma^{1/2}\}^{-1}$$

where  $\Sigma_U = \lim_{T \rightarrow \infty} \Sigma_1^T E(u_t u_t')$  and  $W(t)$  is a multivariate Wiener process on  $C^n$ . (28) applies irrespective of the initial conditions, (7a) or (7b).

(28) implies that  $A^* = I + O_p(T^{-1})$  and, of course,  $A^* \xrightarrow{p} I$  as  $T \rightarrow \infty$ . These results and (28) are especially interesting because of the generality of the underlying conditions on the innovation process  $\{u_t\}_1^\infty$  under which they are proved. In view of this generality, it is more than usually intriguing to study the adequacy of the asymptotic theory delivered by Theorem 4.1. The aim of our next main result, therefore, is to effect a refinement of these first order asymptotics. First, we shall prove:

LEMMA 4.2. Define  $v_t = (u_{it}^2 - E(u_{it}^2))_{n \times 1}$  and let the sequence  $\{v_t\}_1^\infty$  satisfy the conditions of either Theorem 2.1 or 2.2. If, additionally,  $\{u_t\}_1^\infty$  satisfies the conditions of either Theorem 3.1 or 3.2 and if  $\{y_t\}_1^\infty$  is generated by (26) then

$$(a) \quad T^{-2} Y_{-1}' Y_{-1} = \Sigma^{1/2} \int_0^1 W(t) W(t)' dt \Sigma^{1/2} + T^{-1/2} \left\{ \Sigma^{1/2} \int_0^1 W(t) dt y_0' + y_0 \int_0^1 W(t)' dt \Sigma^{1/2} \right\} + O_p(T^{-1})$$

$$(b) \quad T^{-1} Y_{-1}' U = (1/2) \left\{ \Sigma^{1/2} W(1) W(1)' \Sigma^{1/2} - \Sigma_u \right\} + T^{-1/2} \left\{ y_0 W(1)' \Sigma^{1/2} - (1/2) \xi \right\} + O_p(T^{-1})$$

where  $y_0$  satisfies either of the initial conditions (7a) or (7b) and where  $\xi$  is a random symmetric  $n \times n$  matrix distributed as matrix  $N(0, V)$  with

$$(29) \quad V = P_D \left[ \Sigma_{k=0}^\infty \{ \Psi_k - \text{vec}(\Sigma_u) \text{vec}(\Sigma_u)' \} \right] P_D$$

where

$$(30) \quad \Psi_k = E(u_t u_{t+k}' \otimes u_t u_{t+k}')$$

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

and  $D$  is the duplication matrix of [13].  $W(t)$  and  $\xi$  are statistically independent.

THEOREM 4.3. If  $\{u_t\}_1^\infty$  satisfies the conditions of Lemma 4.2 then  $A^*$  has the following asymptotic expansion as  $T \uparrow \infty$ :

$$(32) \quad T(A^* - I) = (1/2) \left\{ \Sigma^{1/2} W(1) W(1)' \Sigma^{1/2} - \Sigma_u \right\} \left\{ \Sigma^{1/2} \int_0^1 W(t) W(t)' dt \Sigma^{1/2} \right\}^{-1} + T^{-1/2} \left[ \left\{ \Sigma^{1/2} W(1) y_0' - (1/2) \xi \right\} \left\{ \Sigma^{1/2} \int_0^1 W(t) W(t)' dt \Sigma^{1/2} \right\}^{-1} - (1/2) \left\{ \Sigma^{1/2} W(1) W(1)' \Sigma^{1/2} - \Sigma_u \right\} \left\{ \Sigma^{1/2} \int_0^1 W(t) W(t)' dt \Sigma^{1/2} \right\}^{-1} \cdot \left\{ \Sigma^{1/2} \int_0^1 W(t) dt y_0' + y_0 \int_0^1 W(t)' dt \Sigma^{1/2} \right\} \left\{ \Sigma^{1/2} \int_0^1 W(t) W(t)' dt \Sigma^{1/2} \right\}^{-1} \right] + O_p(T^{-1})$$

where  $W(t)$  is a multivariate Wiener process on  $C^n$  and  $\xi$  is an independent  $N(0, V)$  matrix with covariance matrix  $V$  given by (29).

*COROLLARY 4.4.* If the conditions of Theorem 4.3 hold and if the initial value  $y_0 = 0$  then

$$(33) \quad T(A^*-I) = (1/2) \left\{ \Sigma^{1/2} W(1) W(1)' \Sigma^{1/2} - \Sigma_u \right\} \left\{ \Sigma^{1/2} \int_0^1 W(t) W(t)' dt \Sigma^{1/2} \right\}^{-1} \\ - (1/2\sqrt{T}) \xi \left\{ \Sigma^{1/2} \int_0^1 W(t) W(t)' dt \Sigma^{1/2} \right\}^{-1} + O_p(T^{-1}) .$$

Theorem 4.3 provides an asymptotic expansion of the distribution of  $T(A^*-I)$  that holds under very general conditions. These conditions apply for a wide class of weakly stationary sequences  $\{u_t\}_1^\infty$ . They are certainly satisfied by stationary Gaussian sequences which satisfy the mixing condition  $\sum_{m=1}^\infty \rho_m < \infty$  and thereby include all finite order 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 (32) may therefore be expected to have rather wide applicability.

We observe that, since the matrix variate  $\xi$  is independent of the vector Wiener process  $W(t)$  and since  $E(\xi) = 0$ , the correction term of  $O(1/\sqrt{T})$  in the expansion (33) contributes no adjustment to the mean of the limiting distribution of  $T(A^*-I)$ . 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 (32) that the initial conditions may have an important influence on the sampling distribution of  $A^*$ . The 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 reductions:

$$\Sigma_u = \Sigma$$

$$\Psi_k = (\text{vec } \Sigma)(\text{vec } \Sigma)' ; \quad k = 1, 2, \dots$$

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

where  $K_n$  is the commutation matrix, so that

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

since  $I + K_n = 2P_D$  [13, pp. 427-428]. Thus, in this case  $\xi$  is matrix  $N(0, 2P_D(\Sigma \otimes \Sigma))$ . When  $n = 1$  (33) reduces to

$$\begin{aligned} & \frac{(\sigma^2/2)(W(1)^2 - 1) - (1/2\sqrt{T})\xi}{\sigma^2 \int_0^1 W(t)^2 dt} + O_p(T^{-1}) \\ &= \frac{(1/2)(W(1)^2 - 1) - (1/\sqrt{2T})\eta}{\int_0^1 W(t)^2 dt} + O_p(T^{-1}) \end{aligned}$$

where  $\eta = (1/\sqrt{2} \sigma^2)\xi \equiv N(0, 1)$ . This highly specialized case of (32) and (33) was first derived by the author in [17, formula (38)].

## 5. CONCLUSIONS

In earlier work [14, 15] the author developed analytic formulae for Edgeworth-type expansions in a stationary first order autoregression. Simple derivations of these formulae and extensions of them to stationary vector autoregressions are provided in other ongoing research [18]. 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 (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 [16].

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 (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 (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. Research on some of these topics is now underway.



MATHEMATICAL APPENDIX

Proof of Theorem 2.1. It is easy to see that individual elements of the vector random element  $X_T(t)$  satisfy the conditions of Herrndorf's theorem [8, p. 142]. Thus  $X_{iT}(t) \Rightarrow W_i(t)$  as  $T \uparrow \infty$ , where  $W_i(t)$  is a Wiener process on  $C[0,1]$  for each  $i = 1, \dots, n$ . Moreover, since  $D[0,1]$  is separable and complete under the metric  $d_0$  [2, p. 112], the weak convergence of  $X_{iT}(t) \Rightarrow W_i(t)$  implies that the family of marginal probability measures associated with the sequence  $\{X_{iT}(t) : T = 1, 2, \dots\}$  is tight by Prohorov's theorem [2, p. 37]. Furthermore, tightness of these marginal probability measures ensures tightness of the family of probability measures associated with  $X_T(t)$  on the product metric space  $D^n$  [2, p. 41, exercise 6]. Since the finite dimensional distributions of  $X_T(t)$  converge to those of the multivariate Wiener process  $W(t)$  (this may be proved as in the proof of Theorem 2.1 of [19]) it follows that  $X_T(t) \Rightarrow W(t)$ , as required.

Proof of Theorem 2.3. The proof is the same as that of Theorem 2.11 of [7, p. 100] after modification for the product space metric.

Proof of Theorem 3.1

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

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

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

and

$$\Sigma = E(u_1 u_1') + \Sigma_{k=2}^{\infty} [E(u_1 u_k') + E(u_k u_1')]$$

so that

$$\begin{aligned}
(A2) \quad T^{-1} E(S_{j-1} S_{j-1}') &= (j-1)/T \Sigma - (j-1)/T \Sigma_{k=j}^{\infty} [E(u_1 u_k') + E(u_k u_1')] \\
&\quad - (1/T) \Sigma_{r=1}^{j-2} r [E(u_1 u_{1+r}') + E(u_{r+1} u_1')] .
\end{aligned}$$

Since  $\Sigma_{r=1}^{\infty} r \rho_r < \infty$  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

$$\begin{aligned} T^{-1} E(S_{j-1} S'_{j-1}) &= (j-1)/T \Sigma - (1/T) \sum_{r=1}^{j-2} r [E(u_1 u'_{1+r}) + E(u_{r+1} u'_1)] + o(T^{-1}) \\ (A3) \qquad \qquad \qquad &= (j-1)/T \Sigma + O(T^{-1}) . \end{aligned}$$

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

$$\begin{aligned} T^{-1} E(S_{j-1} S'_{k-1}) &= T^{-1} \sum_{r=1}^{j-1} \sum_{s=1}^{k-1} E(u_r u'_s) \\ &= T^{-1} \sum_{r=1}^{k-1} \sum_{s=1}^{k-1} E(u_r u'_s) + T^{-1} \sum_{r=k}^{j-1} \sum_{s=1}^{k-1} E(u_r u'_s) \\ &= (k-1)/T [E(u_1 u'_1) + \sum_{r=2}^{k-1} \{E(u_1 u'_r) + E(u_r u'_1)\}] \\ &\quad - (1/T) \sum_{r=1}^{k-2} r E(u_1 u'_{1+r} + u_{r+1} u'_1) + (1/T) \sum_{r=k}^{j-1} \sum_{s=1}^{k-1} E(u_r u'_s) . \end{aligned}$$

Note that

$$T^{-1} \sum_{r=k}^{j-1} \sum_{s=1}^{k-1} E(u_r u'_s) = T^{-1} \sum_{s=1}^{k-1} \sum_{q=k-s}^{j-1-s} E(u_{1+q} u'_1)$$

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

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

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

$$\begin{aligned}
B_T(\varphi, \psi) &= T^{-1} \sum_{j=1}^T \int_{(j-1)/T}^{j/T} \int_{(j-1)/T}^{j/T} [Tt] \varphi(t) \psi(s) dt ds \\
&+ T^{-1} \sum_{j=2}^T \int_{(j-1)/T}^{j/T} \sum_{k=1}^{j-1} \left[ [Ts] \varphi(t) \int_{(k-1)/T}^{k/T} \psi(s) ds dt \right. \\
&\left. + [Ts] \psi(t) \int_{(k-1)/T}^{k/T} \varphi(s) ds dt \right] + O(T^{-1}) .
\end{aligned}$$

We observe that

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

and thus

$$\begin{aligned}
B_T(\varphi, \psi) &= \int_0^1 \int_0^1 \varphi(t) \psi(s) \min(t, s) ds dt + O(T^{-1}) \\
&= B(\varphi, \psi) + O(T^{-1})
\end{aligned}$$

as required.

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

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

We deduce that

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

and

$$X_T(t) = W(t) + O_p(T^{-1})$$

as required.

Proof of Theorem 3.2. Note that  $X_T(t)$  is linear in  $\{u_t\}$ . Hence, third order cumulants of  $X_T(t)$  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 [19].

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

$$\begin{aligned}
 T^{-2} Y_{-1}' Y_{-1} &= T^{-2} \Sigma_{j=1}^T (S_{j-1} + y_0) (S_{j-1} + y_0)' \\
 &= \Sigma_{j=1}^T \int_{(j-1)/T}^{j/T} (\Sigma^{1/2} X_T(t) + T^{-1/2} y_0) (\Sigma^{1/2} X_T(t) + T^{-1/2} y_0)' dt \\
 (A5) \quad &= \Sigma^{1/2} \int_0^1 X_T(t) X_T(t)' dt \Sigma^{1/2} + T^{-1/2} y_0 \int_0^1 X_T(t)' dt \Sigma^{1/2} \\
 &\quad + \Sigma^{1/2} \int_0^1 X_T(t) dt (T^{-1/2} y_0)' + T^{-1} y_0 y_0' .
 \end{aligned}$$

By Theorem 3.1 or 3.2  $X_T(t) = W(t) + O_p(T^{-1})$ . We deduce from (A5) and the continuous mapping theorem that:

$$\begin{aligned}
 (A6) \quad T^{-2} Y_{-1}' Y_{-1} &= \Sigma^{1/2} \int_0^1 W(t) W(t)' dt \Sigma^{1/2} \\
 &\quad + T^{-1/2} \left\{ \Sigma^{1/2} \int_0^1 W(t) dt y_0' + y_0 \int_0^1 W(t)' dt \Sigma^{1/2} \right\} + O_p(T^{-1})
 \end{aligned}$$

proving (a).

To prove (b) we first write

$$(A7) \quad T^{-1} Y_{-1}' U = T^{-1} \Sigma_{j=1}^T y_{t-1} u_t' = T^{-1} \Sigma_{j=1}^T (S_{j-1} + y_0) u_j' = T^{-1} \Sigma_{j=1}^T S_{j-1} u_j' + y_0 T^{-1} \Sigma_{j=1}^T u_j' .$$

Next we employ the random element  $Z_T(t) \in C^n$  defined in (6). We note that  $dZ_T(t) = T^{1/2} \Sigma^{-1/2} u_j dt$  for  $(j-1)/T \leq t < j/T$  and hence, by direct integration:

$$(A8) \quad \int_{(j-1)/T}^{j/T} Z_T(t) dZ_T(t)' = T^{-1} \Sigma^{-1/2} S_{j-1} u_j' \Sigma^{-1/2} + (1/2T) \Sigma^{-1/2} u_j u_j' \Sigma^{-1/2}.$$

Alternatively, by integration by parts we have:

$$(A9) \quad \int_{(j-1)/T}^{j/T} Z_T(t) dZ_T(t)' = \frac{1}{2} [Z_T(t) Z_T(t)']_{(j-1)/T}^{j/T}.$$

Summing (A8) and (A9) over  $j = 1, \dots, T$  and solving we obtain:

$$\begin{aligned} T^{-1} \Sigma^{-1/2} (\Sigma_{j=1}^T S_{j-1} u_j') \Sigma^{-1/2} &= (1/2) Z_T(1) Z_T(1)' - (1/2T) \Sigma^{-1/2} (\Sigma_{j=1}^T u_j u_j') \Sigma^{-1/2} \\ &= (1/2) X_T(1) X_T(1)' - (1/2T) \Sigma^{-1/2} (\Sigma_{j=1}^T u_j u_j') \Sigma^{-1/2}. \end{aligned}$$

By Theorem 3.1 or 3.2 we deduce that:

$$(A10) \quad T^{-1} \Sigma_{j=1}^T S_{j-1} u_j' = (1/2) \Sigma^{1/2} W(1) W(1)' \Sigma^{1/2} - 1/2 \Sigma_u \\ - (1/2\sqrt{T}) \{T^{-1/2} \Sigma_{j=1}^T (u_j u_j' - \Sigma_u)\} + O_p(T^{-1}).$$

Now  $\{(u_t u_t' - \Sigma_u)\}_1^\infty$  is a weakly stationary sequence of random matrices with zero mean that satisfies the conditions of Theorem 2.1 or 2.2. Thus, following the argument of the proof of Theorem 3.3 of [19] we find that:

$$(A11) \quad T^{-1/2} \Sigma_{j=1}^T \text{vec}(u_j u_j' - \Sigma_u) \Rightarrow N(0, V)$$

where

$$V = P_D \sum_{k=0}^{\infty} \{\Psi_k - \text{vec}(\Sigma_u) \text{vec}(\Sigma_u)'\} P_D',$$

$$\Psi_k = E(u_t u_{t+k}' \otimes u_t u_{t+k}') ; \quad k = 0, 1, 2, \dots$$

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

and  $D$  is the duplication matrix of [13].

Since the error on the asymptotic approximation (A11) is at least as small as  $O_p(T^{-1/2})$  we may write (A10) in the alternative stochastic expansion form

$$T^{-1} \sum_{j=1}^T S_{j-1} u_j' = (1/2) \{ \Sigma^{1/2} W(1) W(1)' \Sigma^{1/2} - \Sigma_u \} - (1/2\sqrt{T}) \xi + O_p(T^{-1})$$

where  $\xi$  is a random symmetric matrix for which  $\text{vec } \xi$  is  $N(0, V)$ .

Returning to (A7) we observe that

$$T^{-1/2} \sum_{t=1}^T u_t = T^{-1/2} S_T = \Sigma^{1/2} \chi_T(1) = \Sigma^{1/2} W(1) + O_p(T^{-1}).$$

Thus:

$$(A12) \quad T^{-1} Y_{-1}' U = (1/2) \{ \Sigma^{1/2} W(1) W(1)' \Sigma^{1/2} - \Sigma_u \} - (1/2\sqrt{T}) \xi + (1/\sqrt{T}) y_0' W(1)' \Sigma^{1/2} + O_p(T^{-1})$$

giving (b) as required. Note, finally, that  $\xi$  depends on a quadratic function of  $u_t$  whereas  $W(t)$  depends on partial sums which are linear in the  $u_t$ . Hence  $\xi$  and  $W(t)$  are uncorrelated and, being normal, are therefore independent as stated in the Lemma.

Proof of Theorem 4.3

$$(A13) \quad T(A^*-I) = (T^{-1}U'Y_{-1})(T^{-2}Y'_{-1}Y_{-1})^{-1}$$

and

$$(T^{-2}Y'_{-1}Y_{-1})^{-1} = \left\{ \Sigma^{1/2} \int_0^1 W(t)W(t)' dt \Sigma^{1/2} \right\}^{-1}$$

$$\cdot \left[ I + T^{-1/2} \left\{ \Sigma^{1/2} \int_0^1 W(t) dy'_0 + y_0 \int_0^1 W(t)' dt \Sigma^{1/2} \right\} \left\{ \Sigma^{1/2} \int_0^1 W(t)W(t)' dt \Sigma^{1/2} \right\}^{-1} \right]^{-1} + O_p(T^{-1})$$

$$(A14) = \left\{ \Sigma^{1/2} \int_0^1 W(t)W(t)' dt \Sigma^{1/2} \right\}^{-1} \left[ 1 - T^{-1/2} \left\{ \Sigma^{1/2} \int_0^1 W(t) dy'_0 \right. \right.$$

$$\left. \left. + y_0 \int_0^1 W(t)' dt \Sigma^{1/2} \right\} \left\{ \Sigma^{1/2} \int_0^1 W(t)W(t)' dt \Sigma^{1/2} \right\}^{-1} \right] + O_p(T^{-1}) .$$

Combining (A12)-(A14) we have:

$$T(A^*-I) = (1/2) \{ \Sigma^{1/2} W(1)W(1)' \Sigma^{1/2} - \Sigma_u \} \left\{ \Sigma^{1/2} \int_0^1 W(t)W(t)' dt \Sigma^{1/2} \right\}^{-1}$$

$$+ T^{-1/2} \left[ \left\{ \Sigma^{1/2} W(1)y'_0 - (1/2)\xi \right\} \left\{ \Sigma^{1/2} \int_0^1 W(t)W(t)' dt \Sigma^{1/2} \right\}^{-1} \right.$$

$$\left. - (1/2) \{ \Sigma^{1/2} W(1)W(1)' \Sigma^{1/2} - \Sigma_u \} \left\{ \Sigma^{1/2} \int_0^1 W(t)W(t)' dt \Sigma^{1/2} \right\}^{-1} \right.$$

$$\left. \cdot \left\{ \Sigma^{1/2} \int_0^1 W(t) dy'_0 + y_0 \int_0^1 W(t)' dt \Sigma^{1/2} \right\} \left\{ \Sigma^{1/2} \int_0^1 W(t)W(t)' dt \Sigma^{1/2} \right\}^{-1} \right] + O_p(T^{-1})$$

as required.



6. REFERENCES

- [1] Andrews, D. W. K. "Non-strong mixing autoregressive processes," Journal of Applied Probability, 21 (1984), 930-934.
- [2] Billingsley, P. Convergence of Probability Measures. John Wiley: New York, 1968.
- [3] Bochner, S. Harmonic Analysis and the Theory of Probability. Berkeley: University of California Press, 1960.
- [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  $\phi$ -mixing sequences," Z. Wahrscheinlichkeitstheorie verw. Gebiete, 63 (1983), 97-108.
- [8] \_\_\_\_\_. "A functional central limit theorem for  $\rho$ -mixing sequences," Journal of Multivariate Analysis, 15 (1984), 141-146.
- [9] Hida, T. Brownian Motion. Springer Verlag: New York, 1980.
- [10] Ibragimov, I. A. and Y. V. Linnik. Independent and Stationary Sequences of Random Variables. Groningen: Woller-Nordhoff, 1971.
- [11] Kolmogorov, A. N. and Y. A. Rozanov. "On strong mixing conditions for stationary Gaussian processes," Theory of Probability and Applications, 5 (1960), 204-208.
- [12] Litterman, R. B. "Forecasting with Bayesian vector autoregressions: Four years of experience," mimeographed (Federal Reserve Bank of Minneapolis), 1984.
- [13] Magnus, J. R. and H. Neudecker. "The elimination matrix: Some lemmas and applications," SIAM Journal on Algebraic and Discrete Methods, 1 (1980), 422-449.
- [14] Phillips, P. C. B. "Approximations to some finite sample distributions associated with a first-order stochastic difference equation," Econometrica, 45 (1977), 463-485.
- [15] \_\_\_\_\_. "Edgeworth and saddlepoint approximations in a first order noncircular autoregression," Biometrika, 65 (1978), 91-98.
- [16] \_\_\_\_\_. "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.

- [17] \_\_\_\_\_. "Time series regression with unit roots," Cowles Foundation Discussion Paper No. 740, Yale University, April 1985.
- [18] \_\_\_\_\_. "Asymptotic expansions in stationary vector autoregressions" (in preparation).
- [19] \_\_\_\_\_ and S. N. Durlauf. "Multiple time series regression with integrated processes," mimeographed, Yale University, August 1985.
- [20] Prohorov, Y. V. "The method of characteristic functionals," in J. Neyman (Ed.), Proceedings of 4'th Berkeley Symposium on Mathematical Statistics and Probability, Vol. 2, pp. 403-419. Berkeley: University of California Press, 1961.
- [21] Sims, C. A. "Macroeconomics and reality," Econometrica, 48 (1980), 1-48.
- [22] White, H. Asymptotic Theory for Econometricians. New York: Academic Press, 1984.
- [23] \_\_\_\_\_ and I. Domowitz. "Nonlinear regression with dependent observations," Econometrica, 52 (1984), 143-162.
- [24] Yaglom, A. M. An Introduction to the Theory of Stationary Random Functions. New York: Dover, 1973.