What is dependence?

In the study of random processes, dependence is the rule rather than the exception. To facilitate the related statistical analysis, it is necessary to quantify the dependence between observations. In the talk I will briefly review the history of this fundamental problem. By interpreting random processes as physical systems, I will introduce physical and predictive dependence coefficients that quantify the degree of dependence of outputs on inputs.

Relations with nonlinear system theory and riskmetrics will be discussed.
Such dependence measures provide a new framework for the study of random processes and shed new light on a variety of problems including robust estimation of linear models with dependent errors, nonparametric inference of time series, representations of sample quantiles, bootstrap for time series, spectral estimation among others.

TWO papers provide background:

1. Asymptotic theory for stationary processes
2. Nonlinear system theory: Another look at dependence (starting on p22)

Statistics and Its Interface Volume 0 (2011) 1–20

Asymptotic theory for stationary processes

Wei Biao Wu

We present a systematic asymptotic theory for statistics of stationary time series. In particular, we consider properties of sample means, sample covariance functions, covariance matrix estimates, periodograms, spectral density estimates, U-statistics, kernel density and regression estimates of linear and nonlinear processes. The asymptotic theory is built upon physical and predictive dependence measures, a new measure of dependence which is based on nonlinear system theory. Our dependence measures are particularly useful for dealing with complicated statistics of time series such as eigenvalues of sample covariance matrices and maximum deviations of nonparametric curve estimates.

Keywords and phrases: Dependence, Covariance function, Covariance matrix estimation, Periodogram, Spectral density estimation, U-statistics, Kernel estimation, Invariance principle, Nonlinear time series.

1. INTRODUCTION

The exact probability distributions of statistics of time series can be too complicated to be useful and they are known only in very special situations. It can be impossible to derive close forms for exact finite-sample distributions of statistics of time series. Therefore it is necessary to resort to large sample theory. Asymptotics of linear time series have been discussed in many classical time series books; see for example Anderson (1971), Hannan (1970), Brillinger (1981), Brockwell and Davis (1991) and Hannan and Deistler (1988) among others. Since the pioneering work of Howell Tong on threshold processes, various nonlinear time series models have been proposed. It is more challenging to develop an asymptotic theory for such processes since one no longer assumes linearity.

This paper presents a systematic asymptotic theory for stationary processes of the form

$$X_i = H(\ldots, \varepsilon_{i-1}, \varepsilon_i), \tag{1}$$

where ε_i, $i \in \mathbb{Z}$, are independent and identically distributed (iid) random variables and H is a measurable function such that X_i is well-defined. In (1), (X_i) is causal in the sense that X_i does not depend on the future innovations $\varepsilon_j, j > i$. The causality is a reasonable assumption in the study of time series. As argued in Section 2, (1) provides a very general framework for stationary ergodic processes. Sections 3 and 4 present examples of linear and nonlinear processes that are of form (1).

In the past half century, following the influential work of Rosenblatt (1956b), there have been a substantial amount of results on limit theory for processes which are a strong mixing of various types, such as $\alpha-$, $\beta-$, $\rho-$, $\phi-$mixing and related concepts. See Ibragimov and Linnik (1971), the monograph edited by Eberlein and Taqqu (1986), Doukhan (1994) and Bradley (2007). Recently Doukhan and Louhichi (1999) and Dedecker and Prieur (2005) have proposed some new types of dependence measures which in a certain degree overcome some drawbacks of strong mixing conditions. In many cases it is not easy to compute strong mixing coefficients and verify strong mixing conditions.

In this paper we shall present a large-sample theory for statistics of stationary time series of form (1). In particular we shall discuss asymptotic properties of sample means, sample auto-covariances, covariance matrix estimates, periodograms, spectral density estimates, U-statistics and kernel density and regression estimates. Instead of using strong mixing conditions and their variants, we adopt physical and predictive dependence measure (Wu, 2005b) for our asymptotic theory. The framework, tools and results presented here can be useful for other time series asymptotic problems.

The rest of the paper is organized as follows. In Section 2 we shall review two types of representation theory for stationary processes: the Wold representation and (1), functionals of iid random variables. We argue that the latter representation is actually quite general. It can be viewed as a nonlinear analogue of the Wold representation. Based on (1), Section 3 defines physical and predictive dependence measures which in many situations are easy to work with. Examples of linear and nonlinear processes are given in Sections 3 and 4, respectively. Based on the physical and predictive dependence measures, we survey in Sections 5–12 asymptotic results for various statistics. Section 13 concludes the paper. Our dependence measures are particularly useful for dealing with complicated statistics of time series such as eigenvalues of sample covariance matrices, maxima of periodograms and maximum deviations of nonparametric curve estimates. In such problems it is difficult to apply the traditional strong mixing type of conditions. It would not be possible to include in this paper proofs of all surveyed results. We only present a few proofs so that readers can get a feeling of the techniques used. Nonetheless we shall provide detailed background information and references where proofs can be found.

2. REPRESENTATION THEORY OF STATIONARY PROCESSES

In 1938 Herman Wold proved a fundamental result which asserts that any weakly stationary process can be decomposed into a regular process (a moving average sum of white noises) and a singular process (a linearly deterministic component). The latter result, called Wold representation or decomposition theorem, reveals deep insights into structures of weakly stationary processes. On the other hand, however, one cannot apply the Wold representation theorem to obtain asymptotic distributions of statistics of time series since the white noises in the moving average process do not have properties other than being uncorrelated. The joint distributions of the white noises can be too complicated to be useful. Recently Volný, Woodroofe and Zhao (2011) proved that stationary processes can be represented as super-linear processes of martingale differences. Their useful and interesting decomposition reveals a finer structure than the one in Wold decomposition.

Here we shall adopt a different framework. It is based on quantile transformation. For a random vector $(X_1, \dots, X_n)$, let $\mathbf{X}_m = (X_1, \dots, X_m)$ and define $G_n(\mathbf{x}, u) = \inf\{y \in \mathbb{R} : F_{X_n|\mathbf{X}_{n-1}}(y|\mathbf{x}) \geq u\}$, $\mathbf{x} \in \mathbb{R}^{n-1}$, $u \in (0,1)$. Here $F_{X_n|\mathbf{X}_{n-1}}(\cdot|\cdot)$ is the conditional distribution function of X_n given $\mathbf{X}_{n-1}$. So G_n is the conditional quantile function of X_n given $\mathbf{X}_{n-1}$. In the theory of risk management, $G_n(\mathbf{X}_{n-1}, u)$ is the value-at-risk (VaR) at level u [cf. J. P. Morgan (1996)]. Then we have the distributional equality

$$\mathbf{X}_n =_{\mathcal{D}} (\mathbf{X}_{n-1}, G_n(\mathbf{X}_{n-1}, U_n)), \tag{2}$$

where U_n ~uniform$(0,1)$ and U_n is independent of $\mathbf{X}_{n-1}$. Let $\mathbf{U}_j = (U_1, \dots, U_j)$. Iterating (2), we can find measurable functions $H_1, \dots, H_n$ such that

$$\begin{pmatrix} X_1 \\ X_2 \\ \cdots \\ X_n \end{pmatrix} =_{\mathcal{D}} \begin{pmatrix} X_1 \\ G_2(\mathbf{X}_1, U_2) \\ \cdots \\ G_n(\mathbf{X}_{n-1}, U_n) \end{pmatrix} =_{\mathcal{D}} \begin{pmatrix} H_1(\mathbf{U}_1) \\ H_2(\mathbf{U}_2) \\ \cdots \\ H_n(\mathbf{U}_n) \end{pmatrix}. \tag{3}$$

In other words, we have the important and useful fact that *any finite dimensional random vector can be expressed in distribution as functions of iid uniforms.* The above construction was known for a long time; see for example Rosenblatt (1952), Wiener (1958) and Arjas and Lehtonen (1978). It can be used to simulate multivariate distributions (see e.g. Deák (1990), chapter 5) and Arjas and Lehtonen (1978). For more background see Wu and Mielniczuk (2010). They also discussed connections of their dependence concept with experimental design, reliability theory and risk measures. If $(X_i)_{i\in\mathbb{Z}}$ is a stationary ergodic process, one may expect that there exists a function H and iid standard uniform random variables U_i such that (1) holds. In Wiener (1958) it is called coding problem. The latter claim, however, is generally not true; see Rosenblatt (1959, 2009), Ornstein (1973) and Kalikow (1982). Nonetheless the above construction suggests that the class of processes that (1) represents can be very wide. For a more comprehensive account for representing stationary processes as functions of iid random variables see Wiener (1958), Kallianpur (1981), Priestley (1988), Tong (1990, p. 204), Borkar (1993) and Wu (2005b).

With the representation (1), together with the dependence measures that will be introduced in Section 3, we can establish a systematic asymptotic distributional theory for statistics of stationary time series. Such a theory would not be possible if one just applies the Wold representation theorem. On the other hand we note that in Wold decomposition one only needs weak stationarity while here we require strict stationarity.

3. DEPENDENCE MEASURES

To facilitate an asymptotic theory for processes of form (1), we need to introduce appropriate dependence measures. Here, based on the nonlinear system theory, we shall adopt dependence measures which quantify the degree of dependence of outputs on inputs in physical systems. Let the shift process

$$\mathcal{F}_i = (\dots, \varepsilon_{i-1}, \varepsilon_i). \tag{4}$$

Let $(\varepsilon'_i)_{i\in\mathbb{Z}}$ be an iid copy of $(\varepsilon_i)_{i\in\mathbb{Z}}$. Hence $\varepsilon'_i, \varepsilon_j, i, j \in \mathbb{Z}$, are iid. For a random variable X, we say $X \in \mathcal{L}^p$ $(p > 0)$ if $\|X\|_p := (\mathbb{E}|X|^p)^{1/p} < \infty$. Write the $\mathcal{L}^2$ norm $\|X\| = \|X\|_2$.

Definition 1 (Functional or physical dependence measure)**.** Let $X_i \in \mathcal{L}^p$, $p > 0$. For $j \geq 0$ define the physical dependence measure

$$\delta_p(j) = \|X_j - X_j^*\|_p, \tag{5}$$

where X_j^* is a coupled version of X_j with ε_0 in the latter being replaced by ε'_0:

$$X_j^* = H(\mathcal{F}_j^*), \quad \mathcal{F}_j^* = (\dots, \varepsilon_{-1}, \varepsilon'_0, \varepsilon_1, \dots, \varepsilon_{j-1}, \varepsilon_j).$$

Definition 2 (Predictive dependence measure)**.** For $j \in \mathbb{Z}$, define the projection operator

$$\mathcal{P}_j \cdot = \mathbb{E}(\cdot|\mathcal{F}_j) - \mathbb{E}(\cdot|\mathcal{F}_{j-1}). \tag{6}$$

Let $X_i \in \mathcal{L}^p$, $p \geq 1$. Define the predictive dependence measure $\theta_p(i) = \|\mathcal{P}_0 X_i\|_p$.

Lemma 1 (Wu, 2005)**.** *For $(X_i)_{i\in\mathbb{Z}}$ given in (1), assume $X_i \in \mathcal{L}^p$, $p \geq 1$. For $j \geq 0$ let g_j be a Borel function on $\mathbb{R} \times \mathbb{R} \times \cdots \mapsto \mathbb{R}$ such that $g_j(\mathcal{F}_0) = \mathbb{E}(X_j|\mathcal{F}_0)$. Let*

$$\omega_p(j) = \|g_j(\mathcal{F}_0) - g_j(\mathcal{F}_0^*)\|_p. \tag{7}$$

Then $\theta_p(i) \leq \omega_p(i) \leq 2\theta_p(i)$.

Definition 3 (Stability and weak stability)**.** We say that the process (X_i) is p-stable if

$$\Delta_p := \sum_{j=0}^{\infty} \delta_p(j) < \infty. \tag{8}$$

We say that it is weakly p-stable if $\Omega_p := \sum_{j=0}^{\infty} \theta_p(i) < \infty$.

In Definition 1 the pair (X_j, X_j^*) is exchangeable. Namely (X_j, X_j^*) and (X_j^*, X_j) have the same distribution. This interesting property is useful in applying our dependence measures. In Definition 2, the projection operators $\mathcal{P}_j$, $j \in \mathbb{Z}$, naturally lead to martingale differences. The function $g_j(\mathcal{F}_0)$ in Lemma 1 can be viewed as a nonlinear analogue of Kolmogorov's (1941) linear predictor which results from tail terms in the Wold decomposition. When $p = 2$, we write $\delta(j) = \delta_2(j)$, $\omega(j) = \omega_2(j)$ and $\theta(i) = \theta_2(i)$. The weak stability with $p = 2$ guarantees an invariance principle for the partial sum process $S_n = \sum_{i=1}^{n} X_i$; see Theorem 3 in Section 5.

Remark 1. The above dependence measures are defined for the one-sided process X_i given in (1). Clearly similar definitions can be given for the two-sided process

$$X_i = H(\ldots, \varepsilon_{i-1}, \varepsilon_i, \varepsilon_{i+1}, \ldots) \tag{9}$$

as well. We can show that with non-essential modifications, the majority of the results in the following sections remain valid. Since many processes encountered in practice are causal, we decide to use the one-sided representation.

Note that (9) can be naturally generalized to the spatial process $X_{\mathbf{i}} = H(\varepsilon_{\mathbf{i}-\mathbf{j}}, \mathbf{j} \in \mathbb{Z}^d)$, $\mathbf{i} \in \mathbb{Z}^d$, $d \geq 2$. Hallin, Lu and Tran (2001, 2004) considered kernel density estimation of such linear and non-linear random fields. Surgailis (1982) dealt with long-memory linear fields. El Machkouri, Volný and Wu (2010) established a very general central limit theorem for random fields of this type.

Remark 2. In Ibragimov (1962), Billingsley (1968), Bierens (1983), Andrews (1995) and Lu (2001), the following type of stationary processes has been considered: $X_i = H(V_{i-j}, j \in \mathbb{Z})$ or $X_i = H(\ldots, V_{i-1}, V_i)$, where V_i is another stationary process which can be $\alpha-$ or $\phi-$ mixing, and near-epoch dependence conditions are imposed. This framework and ours have different ranges of application. On one hand, our (1) does not seem to lose too much generality in view of (3) and Wiener's (1958) construction. On the other hand, the property that ε_i are independent greatly facilitates asymptotic studies of time series. For example, in Section 11, we review Liu and Wu's (2010a) asymptotic distributional theory for maximum deviations of nonparametric curve estimates for time series which can be possibly long-memory. It can be very difficult to establish results of such type by using the framework of functions of strong mixing processes under near-epoch dependence. In nonparametric inference it is important to have such an asymptotic distributional theory since one can use that to construct simultaneous, instead of point-wise, confidence bands. The simultaneous confidence bands are useful for assessing the overall variability of the estimated curves. Recently Lu and Linton (2007) and Li, Lu and Linton (2010) obtained asymptotic normality and uniform bounds for local linear estimates under near-epoch dependence. It seems not easy to apply their framework to establish the Gumbel type of convergence for maximum deviations of local linear estimates.

We interpret (1) as a physical system with $\mathcal{F}_i$ and X_i being the input and output, respectively, and H being a transform. With this interpretation, $\delta_p(j)$ quantifies the dependence of $X_j = H(\mathcal{F}_j)$ on ε_0 by measuring the distance between X_j and its coupled process $X_j^* = H(\mathcal{F}_j^*)$. The stability condition $\sum_{j=0}^{\infty} \delta_p(j) < \infty$ indicates that Δ_p, the cumulative impact of ε_0 on the future values $(X_i)_{i \geq 0}$, is finite. Hence it can be interpreted as a short-range dependence condition. For the predictive dependence measure $\omega_p(j)$, since $g_j(\mathcal{F}_0) = \mathbb{E}(X_j|\mathcal{F}_0)$ is the jth step ahead predicted mean, $\omega_p(j)$ measures the contribution of ε_0 in predicting X_j. Recently Escanciano and Hualde (2009) established a link between the persistence measure proposed by Granger (1995), the nonlinear impulse response (Koop et al. (1996)), and our predictive dependence measures.

Physical and predictive dependence measures provide a convenient way for a large-sample theory for stationary processes and they are directly related to the underlying data-generating mechanism H. The obtained results based on those dependence measures are often optimal or nearly optimal. The results in this paper extend to many previous theorems in classical textbooks which are mostly for the special case of linear processes.

In the rest of this section we present examples of linear processes and Volterra processes, a polynomial-type nonlinear process. We shall compute their physical and predictive dependence measures. Section 4 deals with nonlinear time series.

Example 1 (Linear Processes)**.** Let ε_i be iid random variables with $\varepsilon_i \in \mathcal{L}^p$, $p > 0$; let (a_i) be real coefficients such that

$$\sum_{i=0}^{\infty} |a_i|^{\min(2,p)} < \infty. \tag{10}$$

By Kolmogorov's Three Series Theorem (Chow and Teicher, 1988), the linear process

$$X_t = \sum_{i=0}^{\infty} a_i \varepsilon_{t-i} \tag{11}$$

exists and is well-defined. Then (11) is of form (1) with a linear functional H. We can view the linear process (X_t) in (11) as the output from a linear filter and the input $(\ldots, \varepsilon_{t-1}, \varepsilon_t)$ is a series of shocks that drive the system (Box, Jenkins and

Reinsel (1994), p. 8–9). Clearly $\omega_p(n) = \delta_p(n) = |a_n|c_0$, where $c_0 = \|\varepsilon_0 - \varepsilon'_0\|_p < \infty$. Let $p = 2$. If

$$(12) \qquad \sum_{i=0}^{\infty} |a_i| < \infty,$$

then the filter is said to be stable (Box, Jenkins and Reinsel, 1994) and the preceding inequality implies short-range dependence since the covariances are absolutely summable. In this sense Definition 3 extends the notion of stability to nonlinear processes.

Example 2 (Autoregressive Moving Average Process, ARMA)**.** An important special class of linear process (11) is the ARMA model which is of the form

$$(13) \qquad X_t - \sum_{j=1}^{p} \varphi_j X_{t-j} = \varepsilon_t + \sum_{l=1}^{q} \theta_l \varepsilon_{t-l},$$

where $(\varphi_j)_{j=1}^p$ (resp. $(\theta_l)_{l=1}^q$) are autoregressive (resp. moving average) parameters. Note that a_i is the coefficient of z^i of the infinite series $(1+\sum_{l=1}^q \theta_l z^l)/(1-\sum_{j=1}^p \varphi_j z^j)$. In the special case in which $q = 0$, we call (13) an AR (autoregressive) process. Let $\lambda_1, \dots, \lambda_p$ be the roots of the equation $\lambda^p - \sum_{j=1}^p \varphi_j \lambda^{p-j} = 0$. Assume $\lambda^* = \max_{m \le p} |\lambda_m| < 1$. Then $|a_i| = O(r^i)$ for all $r \in (\lambda^*, 1)$ and (10) holds.

Example 3 (Volterra Series)**.** Intuitively, if we perform first-order Taylor expansion of H in (1), then the corresponding linear process can viewed as a first-order approximation of X_i. To model nonlinearity, we can apply higher-order Taylor expansions. Suppose that H is sufficiently well-behaved so that it has the stationary and causal representation

$$(14) \qquad H(\dots, \varepsilon_{n-1}, \varepsilon_n) = \sum_{k=1}^{\infty} \sum_{u_1,\dots,u_k=0}^{\infty} g_k(u_1, \dots, u_k) \varepsilon_{n-u_1} \dots \varepsilon_{n-u_k},$$

where functions g_k are called the Volterra kernel. The right-hand side of (14) is called the Volterra expansion and it plays an important role in the nonlinear system theory (Schetzen 1980, Rugh 1981, Casti 1985, Priestley 1988, Bendat 1990, Mathews 2000). Assume that ε_t are iid with mean 0, variance 1 and $g_k(u_1, \dots, u_k)$ is symmetric in $u_1, \dots, u_k$ and it equals zero if $u_i = u_j$ for some $1 \le i < j \le k$, and

$$\sum_{k=1}^{\infty} \sum_{u_1,\dots,u_k=0}^{\infty} g_k^2(u_1, \dots, u_k) < \infty.$$

Then X_n exists and $X_n \in \mathcal{L}^2$. Wu (2005) computed the predictive dependence measure

$$\theta^2(n) = \sum_{k=1}^{\infty} \sum_{\min(u_1,\dots u_k)=n} g_k^2(u_1, \dots, u_k) = \sum_{k=1}^{\infty} k \sum_{u_2,\dots u_k=n+1}^{\infty} g_k^2(n, u_2, \dots, u_k)$$

and the physical dependence measure

$$\frac{\delta^2(n)}{2} = \sum_{k=1}^{\infty} k \sum_{u_2,\dots u_k=0}^{\infty} g_k^2(n, u_2, \dots, u_k).$$

4. NONLINEAR TIME SERIES

A wide class of nonlinear time series can be expressed as

$$(15) \qquad X_i = G(X_{i-1}, \xi_i) = G_{\xi_i}(X_{i-1}),$$

where ξ, ξ_i, $i \in \mathbb{Z}$, are iid random variables taking values in Ξ with distribution μ and $G : \mathcal{X} \times \Xi \mapsto \mathcal{X}$ is a measurable function. Here $(\mathcal{X}, \rho)$ is a complete and separable metric space. We can view (15) as an iterated random function. The problem of existence of stationary distributions of iterated random functions and the related convergence issues has been extensively studied (Barnsley and Elton (1988), Elton (1990), Duflo (1997), Arnold (1998), Diaconis and Freedman (1999), Steinsaltz (1999), Alsmeyer and Fuh (2001), Jarner and Tweedie (2001), Wu and Shao (2004)). Here we shall present a sufficient condition for (15) so that the representation (1) holds.

Define the forward iteration function

$$(16) \qquad X_n(x) = G_{\xi_n} \circ G_{\xi_{n-1}} \circ \dots \circ G_{\xi_1}(x),$$

where $n \in \mathbb{N}$, and the backward iteration function

$$(17) \qquad Z_n(x) = G_{\xi_1} \circ G_{\xi_2} \circ \dots \circ G_{\xi_n}(x).$$

Observe that, for all $x \in \mathcal{X}$, by independence of ξ_i, $X_n(x) \stackrel{\mathcal{D}}{=} Z_n(x)$. Note that the joint distributions $(X_n(x))_{n \ge 1}$ and $(Z_n(x))_{n \ge 1}$ are not the same. If $Z_n(x)$ converges almost surely to a random variable Z_∞ (say), then $X_n(x)$ converges in distribution to Z_∞.

Condition 1. There exist $y_0 \in \mathcal{X}$ and $\alpha > 0$ such that

$$(18) \qquad I(\alpha, y_0) := \mathbb{E}\{\rho^\alpha[y_0, G_\xi(y_0)]\} = \int_\Xi \rho^\alpha[y_0, G_\theta(y_0)] \mu(d\theta) < \infty.$$

Condition 2. There exist $x_0 \in \mathcal{X}, \alpha > 0$ and $r(\alpha) \in (0, 1)$ such that, for all $x \in \mathcal{X}$,

$$(19) \qquad \mathbb{E}\{\rho^\alpha[X_1(x), X_1(x_0)]\} \le r(\alpha) \rho^\alpha(x, x_0).$$

Theorem 1 (Wu and Shao, 2004)**.** *Suppose that Conditions 1 and 2 hold. Then there exists a random variable Z_∞ such that for all $x \in \mathcal{X}$, $Z_n(x) \to Z_\infty$ almost surely. The limit*

Z_∞ is $\sigma(\xi_1,\xi_2,\ldots)$-measurable and does not depend on x. Moreover, for every $n\in\mathbb{N}$,

$$(20)\qquad \mathbb{E}\{\rho^\alpha[Z_n(x),Z_\infty]\}\le Cr^n(\alpha),$$

where $C>0$ depends only on x,x_0,y_0, α and $r(\alpha)\in(0,1)$. In addition, we have the geometric-moment contracting (GMC) property:

$$(21)\qquad \mathbb{E}\{\rho^\alpha[Z_n(X_0'),Z_\infty]\}\le Cr^n(\alpha),$$

where $X_0'\sim\pi$ is independent of $\xi_1,\xi_2,\ldots$.

Remark 3. In applying Theorem 1, a useful sufficient condition for (19) is

$$(22)\qquad \mathbb{E}(K_\theta^\alpha)=\int_\Xi K_\theta^\alpha\mu(d\theta)<1,$$

$$\text{where } K_\theta=\sup_{x'\ne x}\frac{\rho[G_\theta(x'),G_\theta(x)]}{\rho(x',x)}.$$

To see this, by Fatou's lemma, we have (19) with $r(\alpha)=\mathbb{E}(K_\theta^\alpha)$ in view of

$$1>\mathbb{E}(K_\theta^\alpha)=\int_\Theta\sup_{x'\ne x}\frac{\rho^\alpha[G_\theta(x'),G_\theta(x)]}{\rho^\alpha(x',x)}\mu\{d\theta\}$$
$$\ge\sup_{x'\ne x}\int_\Theta\frac{\rho^\alpha[G_\theta(x'),G_\theta(x)]}{\rho^\alpha(x',x)}\mu\{d\theta\}.$$

Remark 4. Assume that K_θ has an algebraic tail. If there exists an α such that (19) holds, then $\mathbb{E}(\log K_\theta)<0$. The converse is also true. The latter is a key condition in Diaconis and Freedman (1999). Our Theorem 1 is an improved version of Theorem 1 in Diaconis and Freedman (1999) in that it states stronger results under weaker conditions.

The GMC property (21) asserts that $X_i, i\ge 0$, forgets the history $\mathcal{F}_0=(\ldots,\varepsilon_{-1},\varepsilon_0)$ geometrically quickly. It is equivalent to the following: the physical dependence measure $\delta_\alpha(n)=O(r^n(\alpha))$.

Theorem 1 can be generalized to nonlinear AR(p) models (Shao and Wu, 2007). Let $\varepsilon,\varepsilon_n$ be iid, $p,d\ge 1$; let $X_n\in\mathbb{R}^d$ be recursively defined by

$$(23)\qquad X_{n+1}=R(X_n,\ldots,X_{n-p+1};\varepsilon_{n+1}),$$

where R is a measurable function. Suitable conditions on R implies GMC.

Theorem 2 (Shao and Wu, 2007)**.** *Let $\alpha>0$ and $\alpha'=\min(1,\alpha)$. Assume that $R(y_0;\varepsilon)\in\mathcal{L}^\alpha$ for some y_0 and that there exist constants $a_1,\ldots,a_p\ge 0$ such that $\sum_{j=1}^p a_j<1$ and*

$$(24)\qquad \|R(y;\varepsilon)-R(y';\varepsilon)\|_\alpha^{\alpha'}\le\sum_{j=1}^p a_j|x_j-x_j'|^{\alpha'}$$

holds for all $y=(x_1,\ldots,x_p)$ and $y'=(x_1',\ldots,x_p')$. Then [i] (23) admits a stationary solution of the form (1) and [ii] X_n satisfies GMC(α). In particular, if there exist functions H_j such that $|R(y;\varepsilon)-R(y';\varepsilon)|\le\sum_{j=1}^p H_j(\varepsilon)|x_j-x_j'|$ for all y and y' and $\sum_{j=1}^p\|H_j(\varepsilon)\|_\alpha^{\alpha'}<1$, then we can let $a_j=\|H_j(\varepsilon)\|_\alpha^{\alpha'}$.

Duflo (1997) assumed $\alpha\ge 1$ and called (24) Lipschitz mixing condition. Here we allow $\alpha<1$. Similar conditions are given in Götze and Hipp (1994).

Doukhan and Wintenberger (2008) considered the AR(∞) or chain with infinite memory model

$$(25)\qquad X_{k+1}=R(X_k,X_{k-1},\ldots;\varepsilon_{k+1}),$$

where ε_k are iid innovations. Assume that there exists a non-negative sequence $(w_j)_{j\ge1}$ such that, for some $\alpha\ge 1$,

$$(26)\qquad \|R(x_{-1},x_{-2},\ldots;\varepsilon_0)-R(x_{-1}',x_{-2}',\ldots;\varepsilon_0)\|_\alpha\le\sum_{j=1}^\infty w_j|x_{-j}-x_{-j}'|.$$

Under suitable conditions on $(\omega_j)_{j\ge1}$, iterations of (25) lead to a stationary solution X_k of form (1). We now compute its physical dependence measure. Let $\delta_\alpha(k)=\|X_k-H(\mathcal{F}_k^*)\|_\alpha$. For $k\ge 0$, by (25) and (26), we have

$$(27)\qquad \delta_\alpha(k+1)\le\sum_{i=1}^{k+1}w_i\delta_\alpha(k+1-i).$$

Define recursively the sequence $(a_k)_{k\ge0}$ by $a_0=\delta_\alpha(0)$ and

$$(28)\qquad a_{k+1}=\sum_{i=1}^{k+1}w_ia_{k+1-i}.$$

Let $A(s)=\sum_{k=0}^\infty a_ks^k$ and $W(s)=\sum_{i=1}^\infty w_is^i$, $|s|\le 1$. By (28), we have $A(s)=a_0+A(s)W(s)$. Hence $A(s)=a_0(1-W(s))^{-1}$. Assume that, as $s\uparrow 1$, $1-W(s)\sim(1-s)^d$ with $d\in(0,1/2)$. Then $\delta_\alpha(k)\le a_k\sim a_0k^{d-1}/\Gamma(d)$, where $\Gamma(\cdot)$ is the Gamma function. The latter is the fractional integration model $(1-B)^dX_{k+1}=\varepsilon_{k+1}$. For a nonlinear functional R, (25) generates a nonlinear long-memory process.

Note that in our setting $W(1)=\sum_{j=1}^\infty w_j=1$, while $W(1)<1$ is required in Doukhan and Wintenberger (2008). Hence we can allow stronger dependence. If, as in Doukhan and Wintenberger (2008), $W(1)<1$, then $a_k=O(r^k)$ for some $r\in(0,1)$. This is analogous to Theorem 2 which ensures the GMC property.

Example 4 (Amplitude-dependent Exponential Autoregressive (EXPAR) Model)**.** Jones (1976) studied the following EXPAR model: let $\varepsilon_j\in\mathcal{L}^\alpha$ be iid and recursively define

$$X_n=[\alpha+\beta\exp(-aX_{n-1}^2)]X_{n-1}+\varepsilon_n,$$

where $\alpha, \beta, a > 0$ are real parameters. Then $H_1(\varepsilon) = |\alpha| + |\beta|$. By Theorem 1 (cf Remark 3), X_n is GMC(α) if $|\alpha| + |\beta| < 1$.

Example 5 (Nonlinear AR Process Based on the Clayton Copula)**.** Let $\alpha > 0$ and $U_i, i \in \mathbb{Z}$, be iid uniform$(0,1)$. Consider the model

$$Y_i = (U_i^{-\alpha/(1+\alpha)} - 1)Y_{i-1} + 1.$$

Then Y_i has the stationary distribution with $Y_i^{-1/\alpha} \sim$ uniform$(0,1)$. The above Markov process is generated by the Clayton copula (Chen and Fan, 2006) which is used to model tail dependence behavior of time series.

Example 6 (Bilinear time series)**.** Let $\varepsilon, \varepsilon_i, i \in \mathbb{Z}$, be iid and consider the recursion

$$X_i = (a + b\varepsilon_i)X_{i-1} + c\varepsilon_i, \tag{29}$$

where a, b and c are real parameters. When $b = 0$, then (29) reduces to an AR(1) process. The bilinear time series was first proposed by Tong (1981) to model sudden jumps in time series. Quinn (1982) derived the moment stability. By Theorem 1, if $\varepsilon \in \mathcal{L}^\alpha$, $\alpha > 0$, and $\mathbb{E}(|a + b\varepsilon|^\alpha) < 1$, then (29) admits a stationary solution. Consider the subdiagonal bilinear model [Granger and Anderson (1978), Subba Rao and Gabr (1984)]:

$$X_t = \sum_{j=1}^{p} a_j X_{t-j} + \sum_{j=0}^{q} c_j \varepsilon_{t-j} + \sum_{j=0}^{P}\sum_{k=1}^{Q} b_{jk} X_{t-j-k}\varepsilon_{t-k}. \tag{30}$$

Let $s = \max(p, P+q, P+Q)$, $r = s - \max(q, Q)$ and $a_{p+j} = 0 = c_{q+j} = b_{P+k,Q+j} = 0$, $k, j \geq 1$; let H be a $1 \times s$ vector with the $(r+1)$-th element 1 and all others 0, c be an $s \times 1$ vector with the first $r-1$ elements 0 followed by $1, a_1 + c_1, \ldots, a_{s-r} + c_{s-r}$, and d be an $s \times 1$ vector with the first r elements 0 followed by $b_{01}, \ldots, b_{0,s-r}$. Define the $s \times s$ matrices

$$A = \begin{pmatrix} 0 & 1 & & 0 & & 0 \\ & & \ddots & & 0 & \\ 0 & & & 1 & & 0 \\ 0 & 0 & & a_1 & \ddots & 0 \\ & & & \vdots & & 1 \\ a_s & \cdots & \cdots & a_{s-r} & & 0 \end{pmatrix},$$

$$B = \begin{pmatrix} 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 \\ b_{r1} & \cdots & b_{01} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ b_{r,s-r} & \cdots & b_{0,s-r} & 0 & \cdots & 0 \end{pmatrix}.$$

Let Z_t be an $s \times 1$ vector with the j-th entry X_{t-r+j} if $1 \leq j \leq r$ and $\sum_{k=j}^{r} a_k X_{t+j-k} + \sum_{k=j}^{s-r}(c_k + \sum_{l=0}^{P} b_{lk} X_{t+j-k-l})\varepsilon_{t+j-k}$ if $1 + r \leq j \leq s$. Pham (1985, 1993) discovered the representation

$$X_t = HZ_{t-1} + \varepsilon_t, \quad Z_t = (A + B\varepsilon_t)Z_{t-1} + c\varepsilon_t + d\varepsilon_t^2. \tag{31}$$

By (31), X_t is GMC(α), $\alpha \geq 1$ if $\varepsilon_1 \in \mathcal{L}^{2\alpha}$ and $\mathbb{E}(|A + B\varepsilon_1|^\alpha) < 1$. By (39), Z_t admits a causal representation and so does X_t.

Example 7 (Threshold AR model, TAR (Tong, 1990))**.** For $x \in \mathbb{R}$ let $x^+ = \max(x, 0)$ and $x^- = \min(x, 0)$. Tong (1990) considered the threshold autoregressive model (TAR)

$$X_i = \theta_1 X_{i-1}^+ + \theta_2 X_{i-1}^- + \varepsilon_i, \tag{32}$$

where θ_1, θ_2 are real parameters and $\varepsilon, \varepsilon_i, i \in \mathbb{Z}$, are iid. The above model suggests the regime switching phenomenon: if $X_{i-1} > 0$, then (32) becomes $X_i = \theta_1 X_{i-1} + \varepsilon_i$, while if $X_{i-1} < 0$, then X_i follows a different AR(1) process $X_i = \theta_2 X_{i-1} + \varepsilon_i$. By Theorem 1, if $\max(|\theta_1|, |\theta_2|) < 1$ and $\varepsilon \in \mathcal{L}^\alpha$, $\alpha > 0$, then (32) admits a stationary solution.

Example 8 (Autoregressive Conditional Heteroscedastic Models, ARCH (Engle, 1982))**.** Let $\varepsilon, \varepsilon_i, i \in \mathbb{Z}$, be iid. The ARCH with order 1 is defined by the recursion

$$X_i = \varepsilon_i \sqrt{a^2 + b^2 X_{i-1}^2}, \tag{33}$$

where a and b are real parameters. If $\mathbb{E}\varepsilon_i = 0$ and $\mathbb{E}\varepsilon_i^2 = 1$, then the conditional variance of X_i given X_{i-1} is $a^2 + b^2 X_{i-1}^2$, which depends on X_{i-1} and hence suggesting heteroscedasticity. The latter property is useful for modeling financial time series that exhibit time-varying volatility clustering. A sufficient condition for stationarity is $\mathbb{E}\log|b\varepsilon| < 0$. If there exists $\alpha > 0$ such that $\mathbb{E}(|b\varepsilon|^\alpha) < 1$, then X_i has a stationary solution with αth moment.

Example 9 (Generalized Autoregressive Conditional Heteroskedastic models, GARCH (Bollerslev, 1986))**.** Let $\varepsilon_t, t \in \mathbb{Z}$, be iid random variables with mean 0 and variance 1; let

$$X_t = \sqrt{h_t}\varepsilon_t, \tag{34}$$

where the conditional variance function follows the ARMA model

$$h_t = \alpha_0 + \alpha_1 X_{t-1}^2 + \cdots + \alpha_q X_{t-q}^2 + \beta_1 h_{t-1} + \cdots + \beta_p h_{t-p}, \tag{35}$$

where $\alpha_0 > 0$, $\alpha_j \geq 0$ for $1 \leq j \leq q$ and $\beta_i \geq 0$ for $1 \leq i \leq p$. Here (X_t) is called the generalized autoregressive conditional heteroscedastic model $GARCH(p, q)$. A sufficient condition for (X_t) being stationary is (Bollerslev, 1986):

$$\sum_{j=1}^{q} \alpha_j + \sum_{i=1}^{p} \beta_i < 1. \tag{36}$$

The existence of moments for GARCH models has been widely studied; see Chen and An (1998), He and Teräsvirta (1999), Ling (1999), and Ling and McAleer (2002) among others. Let $Y_t = (X_t^2, \dots, X_{t-q+1}^2, h_t, \dots, h_{t-p+1})^T$, $b_t = (\alpha_0\epsilon_t^2, 0, \dots, 0, \alpha_0, 0, \dots, 0)^T$ and $\theta = (\alpha_1, \dots, \alpha_q, \beta_1, \dots, \beta_p)^T$; let $e_i = (0, \dots, 0, 1, 0, \dots, 0)^T$ be the unit column vector with ith element being 1, $1 \le i \le p+q$. Then (34) admits the following autoregressive representation (Bougerol and Picard, 1992):

(37)

$$Y_t = M_t Y_{t-1} + b_t,$$
$$\text{where } M_t = (\theta\epsilon_t^2, e_1, \dots, e_{q-1}, \theta, e_{q+1}, \dots, e_{p+q-1})^\top.$$

For a square matrix M let $\rho(M)$ be its largest eigenvalue of $(M^T M)^{1/2}$. Let $\otimes$ be the usual Kronecker product; let $|Y|$ be the Euclidean length of a vector Y. Assume $\mathbb{E}(\varepsilon_t^4) < \infty$. Ling (1999) shows that if $\rho[\mathbb{E}(M_t^{\otimes 2})] < 1$, then (X_t) has a stationary distribution and $\mathbb{E}(X_t^4) < \infty$. Ling and McAleer (2002) argue that the condition $\rho[\mathbb{E}(M_t^{\otimes 2})] < 1$ is also necessary for the finiteness of the fourth moment. Our Proposition 1 asserts that the same condition actually implies (21) as well.

Proposition 1 (Wu and Min, 2005)**.** *For the GARCH model (34), assume that ε_t are iid with mean 0 and variance 1, $\mathbb{E}(\varepsilon_t^4) < \infty$ and $\rho[\mathbb{E}(M_t^{\otimes 2})] < 1$. Then $\mathbb{E}(|X_n - X_n'|^4) \le Cr^n$ for some $C < \infty$ and $r \in (0,1)$. Therefore (21) holds.*

Shao and Wu (2007) showed that (21) holds for the asymmetric GARCH processes of Ding, Granger and Engle (1993) and Ling and McAleer (2002).

Example 10 (Random Coefficients Model)**.** Let A_k be $p \times p$ random matrices and B_k be $p \times 1$ random vectors, $p \in \mathbb{N}$. Let (A_k, B_k), $k \in \mathbb{Z}$, be iid. The generalized random coefficient autoregressive process (X_i) is defined by

$$X_i = A_i X_{i-1} + B_i, \ i \in \mathbb{Z}. \tag{38}$$

Bilinear and GARCH models fall within the framework of (38). The stationarity, geometric ergodicity and β-mixing properties have been studied by Pham (1986), Mokkadem (1990) and Carrasco and Chen (2002). Their results require that innovations have a density, which is not needed in our setting.

For a $p \times p$ matrix A, let $|A|_\alpha = \sup_{z \ne 0} |Az|_\alpha / |z|_\alpha, \alpha \ge 1$ be the matrix norm induced by the vector norm $|z|_\alpha = (\sum_{j=1}^p |z_j|^\alpha)^{1/\alpha}$. Then X_i is GMC(α), $\alpha \ge 1$ if $\mathbb{E}(|A_0|_\alpha) < 1$ and $\mathbb{E}(|B_0|_\alpha) < \infty$. By Jensen's inequality, we have $\mathbb{E}(\log |A_0|_\alpha) < 0$. By Theorem 1.1 of Bougerol and Picard (1992),

$$X_n = \sum_{k=0}^{\infty} A_n A_{n-1} \dots A_{n-k+1} B_{n-k} \tag{39}$$

converges almost surely.

Example 11 (Nonlinear Heteroskedastic AR Models)**.** Let $\mu(\cdot)$ and $\sigma(\cdot) \ge 0$ be real valued functions; let $\varepsilon, \varepsilon_i, i \in \mathbb{Z}$, be iid random variables with $\varepsilon_i \in \mathcal{L}^\alpha$, $\alpha > 0$. Consider

$$X_i = \mu(X_{i-1}) + \sigma(X_{i-1})\varepsilon_i \tag{40}$$

If $\sigma(\cdot)$ is not a constant function, then (40) defines a heteroskedastic process. If ε_i is Gaussian, then we can view (40) as a discretized version of the stochastic diffusion model

$$dY_t = \mu(Y_t)dt + \sigma(Y_t)d\mathbb{B}(t) \tag{41}$$

where $\mathbb{B}$ is the standard Brownian motion. Many well-known financial models are special cases of (41); see Fan (2005) and references therein. For (40), assume that

$$\sup_x \|\mu'(x) + \sigma'(x)\varepsilon\|_\alpha < 1, \tag{42}$$

then by Theorem 1 it has a stationary solution.

5. CENTRAL LIMIT THEORY

This section presents a central limit theorem for the process (1). Let the mean $\mathbb{E}(X_i) = 0$ and $\gamma_k = \text{cov}(X_0, X_k)$ the covariance function. Let $S_n = \sum_{i=1}^n X_i$ and define the process

$$S_t = S_{\lfloor t \rfloor} + (t - \lfloor t \rfloor) X_{\lfloor t \rfloor + 1}, \quad t \ge 0, \tag{43}$$

where the floor function $\lfloor t \rfloor = \max\{k \in \mathbb{Z} : k \le t\}$. Note that S_t is continuous in t. We shall show that, under suitable weak dependence conditions, the central limit theorem

$$\frac{S_n}{\sqrt{n}} \Rightarrow N(0, \sigma^2) \tag{44}$$

holds for some $\sigma^2 < \infty$. Here $\Rightarrow$ denotes weak convergence (Billingsley, 1968). Central limit theorems of type (44) has a substantial history. The classical Lindeberg-Feller (cf Section 9.1 in Chow and Teicher (1988)) concerns independent random variables. Hoeffding and Robbins (1948) proved a central limit theorem under m-dependence. Rosenblatt (1956) introduced strong mixing processes, while Gänssler and Häeusler (1979) and Hall and Heyde (1980) considered martingales. For central limit theorems for stationary processes see Ibragimov (1962), Gordin (1969), Ibragimov and Linnik (1971), Gordin and Lifsic (1978), Peligrad (1996), Doukhan (1999), Maxwell and Woodroofe (2000), Rio (2000), Peligrad and Utev (2005), Dedecker et al (2007) and Bradley (2007).

Here we shall use the predictive dependence measure. It turns out that under a weak stability condition, one can actually have an invariance principle concerning the weak convergence of the re-scaled process of $\{S_{nu}, 0 \le u \le 1\}$ to a Brownian motion $\{\mathbb{B}(u), 0 \le u \le 1\}$. The latter automatically entails (44). Recall (6) for the projection operator $\mathcal{P}_i$.

Theorem 3. *Let $\theta_p(i) = \|\mathcal{P}_0 X_i\|_p$, $p > 1$. Assume $\mathbb{E}X_i = 0$ and*

$$\Theta_p := \sum_{i=0}^{\infty} \theta_p(i) < \infty. \tag{45}$$

Then (i) we have the moment inequality

$$\|S_n\|_p \le \begin{cases} (p-1)^{1/2} n^{1/2} \Theta_p, & p > 2, \\ (p-1)^{-1} n^{1/p} \Theta_p, & 1 < p \le 2. \end{cases} \tag{46}$$

(ii) Assume (45) holds with $p = 2$. Then the invariance principle holds:

$$\{S_{nu}/\sqrt{n},\, 0 \le u \le 1\} \Rightarrow \{\sigma \mathbb{B}(u),\, 0 \le u \le 1\}, \tag{47}$$

where the long-run variance σ^2 is given by

$$\sigma^2 = \left\| \sum_{i=0}^{\infty} \mathcal{P}_0 X_i \right\|^2 = \sum_{k \in \mathbb{Z}} \gamma_k. \tag{48}$$

Theorem 3(ii) follows from Hannan (1979) and Dedecker and Merlevéde (2003). See also Woodroofe (1992) and Volný (1993). A useful feature of Theorem 3 is that it provides an explicit probabilistic representation for the long-run variance $\sigma^2 = \| \sum_{i=0}^{\infty} \mathcal{P}_0 X_i\|^2$. The latter is also called a time-average variance constant or asymptotic variance. The inequality (46) is quite sharp if $p = 2$. Suppose we have a linear process $X_i = \sum_{j=0}^{\infty} a_j \varepsilon_{i-j}$, where ε_j are iid with mean 0 and variance 1, and $a_j \ge 0$ for all j. Then both σ and Θ_2 equal to $\sum_{j=0}^{\infty} a_j$ and $\lim_{n\to\infty} \|S_n\|/\sqrt{n} = \Theta_2$. In Theorem 3, (45) asserts that the cumulative contribution of ε_0 in predicting $(X_i)_{i\ge 0}$ is finite by noting that (45) is equivalent to $\sum_{i=0}^{\infty} \omega(i) < \infty$ in view of Lemma 1. If the latter condition is violated, then one may have long-range dependence and there is no $\sqrt{n}$-central limit theorem.

A basic problem in the inference of stationary processes is to estimate their means. Let $(X_i)_{i\in\mathbb{Z}}$ be a stationary process with unknown mean $\mu = \mathbb{E}(X_i)$. With observations $X_1, \ldots, X_n$, one can estimate μ by the sample average $\bar{X}_n = \sum_{i=1}^{n} X_i/n$. Let $\hat{\sigma}_n$ be a weak consistent estimate of σ. Namely $\hat{\sigma}_n \to \sigma$ in probability. By Theorem 3(ii), we can construct the $(1-\alpha)$th confidence interval for μ as

$$\bar{X}_n \pm \frac{\hat{\sigma}_n}{\sqrt{n}} z_{1-\alpha/2},$$

where $z_{1-\alpha/2}$ is the up $(\alpha/2)$th quantile of the standard Gaussian distribution. The estimation of σ^2 will be discussed in Section 10.

5.1 Proof of Theorem 3

By the triangle inequality, since $X_i = \sum_{l\in\mathbb{Z}} \mathcal{P}_{i-l} X_i$, we have

$$\|S_n\|_p = \left\| \sum_{i=1}^{n} \sum_{l\in\mathbb{Z}} \mathcal{P}_{i-l} X_i \right\|_p \le \sum_{l\in\mathbb{Z}} \left\| \sum_{i=1}^{n} \mathcal{P}_{i-l} X_i \right\|_p. \tag{49}$$

Note that $\mathcal{P}_{i-l}X_i$, $i = 1, \ldots, n$, are stationary martingale differences. If $p > 2$, by Theorem 2.1 in Rio's (2009), we have

$$\left\| \sum_{i=1}^{n} \mathcal{P}_{i-l} X_i \right\|_p^2 \le (p-1) n \|\mathcal{P}_0 X_l\|_p^2. \tag{50}$$

If $1 < p \le 2$, by Burkholder's (1988) moment inequality for martingale differences,

$$\left\| \sum_{i=1}^{n} \mathcal{P}_{i-l} X_i \right\|_p^p \le \frac{\mathbb{E}\{[\sum_{i=1}^{n} (\mathcal{P}_{i-l} X_i)^2]^{p/2}\}}{(p-1)^p} \le \frac{n\|\mathcal{P}_0 X_l\|_p^p}{(p-1)^p}, \tag{51}$$

where we applied the elementary inequality $(|a_1| + \cdots + |a_n|)^{p/2} \le |a_1|^{p/2} + \cdots + |a_n|^{p/2}$. Combining these two cases, we have (46).

Now we prove (ii). For $m \in \mathbb{N}$ let $\tilde{S}_n = \sum_{i=1}^{n} [X_i - \mathbb{E}(X_i|\mathcal{F}_{i-m})]$. Let l.i.m. denote the double limit $\limsup_{m\to\infty} \limsup_{n\to\infty}$. By Doob's inequality,

$$\begin{aligned} \text{l.i.m.} \frac{\| \max_{i\le n} |S_i - \tilde{S}_i| \|}{\sqrt{n}} &\le \text{l.i.m.} \frac{\sum_{k=m}^{\infty} \| \max_{i\le n} |\sum_{j=1}^{n} \mathcal{P}_{j-k} X_j| \|}{\sqrt{n}} \\ &\le \limsup_{m\to\infty} 2 \sum_{k=m}^{\infty} \|\mathcal{P}_0 X_k\| = 0. \end{aligned} \tag{52}$$

For fixed m, write $X_i - \mathbb{E}(X_i|\mathcal{F}_{i-m}) = \sum_{k=0}^{m-1} \mathcal{P}_{i-k} X_i$, since $(\mathcal{P}_{i-k} X_i)_{i=1}^{n}$ is a stationary martingale difference sequence, it is easily seen that the finite dimensional convergence and the tightness for the process $\{\tilde{S}_{nu}/\sqrt{n},\, 0 \le u \le 1\}$ hold. Hence it satisfies the invariance principle. By (52), (ii) follows. □

6. GAUSSIAN APPROXIMATIONS WITH RATES

The invariance principle Theorem 3(ii) does not have a convergence rate. With stronger moment conditions and faster decay rates of physical or predictive dependence measures, we can approximate the partial sum process S_n by a Brownian motion with nearly optimal rates. Such approximations are very useful in statistical inference of time series since Brownian motions have many attractive properties. In Wu and Zhao (2007) we applied Wu's (2007) Gaussian approximation (see Theorem 5 below) to perform statistical inference of trends in time series.

The celebrated strong invariance principle by Komlós, Major and Tusnady (1975, 1976) gives an optimal rate; see (53). The rate in (55) is optimal up to a multiplicative logarithmic factor. Theorem 2.1 in Liu and Lin's (2009a) leads to Theorem 6 which provides a strong invariance principle for vector-valued processes.

Theorem 4 (Komlós, Major and Tusnady, 1975, 1976)**.** *Assume that $X_i, i \in \mathbb{Z}$, are iid with mean 0 and $X_i \in \mathcal{L}^p$, $p > 2$. Let $\sigma = \|X_i\|$. Then on a richer probability space there exists a Brownian motion $\{\mathbb{B}(u), u \geq 0\}$ and a process $(X_i^\diamond)_{i\in\mathbb{Z}}$ such that $(X_i)_{i\in\mathbb{Z}} \stackrel{\mathcal{D}}{=} (X_i^\diamond)_{i\in\mathbb{Z}}$ and, for $S_n^\diamond = \sum_{i=1}^n X_i^\diamond$, we have*

$$\max_{0\leq u\leq n} |S_u^\diamond - \sigma\mathbb{B}(u)| = o_{a.s.}(n^{1/p}). \tag{53}$$

Theorem 5 (Wu, 2007)**.** *Let $(X_i)_{i\in\mathbb{Z}}$ be of the form (1) with mean 0 and $X_i \in \mathcal{L}^p$, $2 < p \leq 4$. Assume that*

$$\sum_{i=1}^{\infty}[\delta_p(i) + i\omega_p(i)] < \infty. \tag{54}$$

Then on a richer probability space there exists a Brownian motion $\{\mathbb{B}(u), u \geq 0\}$ and a process $(X_i^\diamond)_{i\in\mathbb{Z}}$ such that $(X_i)_{i\geq 0} \stackrel{\mathcal{D}}{=} (X_i^\diamond)_{i\geq 0}$ and

$$\max_{0\leq u\leq n} |S_u^\diamond - \sigma\mathbb{B}(u)| = o_{a.s.}(n^{1/p}(\log n)^{1/2+1/p}(\log\log n)^{2/p}), \tag{55}$$

where $\sigma = \|\sum_{i=0}^\infty \mathcal{P}_0 X_i\|$ is given in Theorem 3. A sufficient condition for (54) is

$$\sum_{i=1}^{\infty} i\delta_p(i) < \infty. \tag{56}$$

In the literature strong invariance principles obtained for dependent random variables typically have rates of the form $o_{a.s.}(n^{1/2-\delta})$, where $\delta > 0$ can be very small. See for example Philipp and Stout (1975) and Eberlein (1986). As pointed out in Wu and Zhao (2007), in nonparametric simultaneous inference of trends of time series, such error bounds are too crude to be useful.

Theorem 6 (Liu and Lin, 2009a)**.** *Let $(X_i)_{i\in\mathbb{Z}}$ be a d-dimensional stationary vector process of the form (1) with H taking values in $\mathbb{R}^d$, $d \geq 2$. Let $2 < p < 4$ and assume that, for some $\tau > 0$,*

$$\Delta_p(m) = \sum_{j=m}^{\infty} \delta_p(j) = O(m^{-(p-2)/(8-2p)-\tau}) \tag{57}$$

as $m \to \infty$. Let $D_k = \sum_{i=k}^\infty \mathcal{P}_k X_i$. Further assume that the covariance matrix $\Gamma = \mathbb{E}(D_k D_k^T)$ is positive definite. Then on a richer probability space, there exists an $\mathbb{R}^d$ valued Brownian motion $\mathbb{B}_d(t)$ such that

$$\max_{0\leq u\leq n} |S_u - \Gamma^{1/2}\mathbb{B}_d(u)| = o_{a.s.}(n^{1/p}). \tag{58}$$

7. SAMPLE COVARIANCE FUNCTIONS

Covariance functions characterize second order properties of stochastic processes and they play a fundamental role in the theory of time series. They are critical quantities that are needed in various inference problems including parameter estimation and hypothesis testing. Asymptotic properties of sample covariances have been studied in many classical time series textbooks; see for example Priestley (1981), Brockwell and Davis (1991), Hannan (1970) and Anderson (1971). For other contributions see Hall and Heyde (1980), Hannan (1976), Hosking (1996), Phillips and Solo (1992) and Wu and Min (2005). However, many of those results require that the underlying processes are linear.

Here we present an asymptotic theory for sample covariances for processes which can be nonlinear. Given observations $X_1, \ldots, X_n$, we estimate γ_k by the sample covariance

$$\hat\gamma_k = \frac{1}{n}\sum_{i=k+1}^{n}(X_i - \bar X_n)(X_{i-k} - \bar X_n), \quad 0 \leq k < n \tag{59}$$

and $\hat\gamma_{-k} = \hat\gamma_k$. If we know $\mu = 0$, then we use the estimate $\check\gamma_k = n^{-1}\sum_{i=k+1}^n X_i X_{i-k}$.

Theorem 7. *Let $k \in \mathbb{N}$ be fixed and $\mathbb{E}(X_i) = 0$; let $Y_i = (X_i, X_{i-1}, \ldots, X_{i-k})^T$ and $\Gamma_k = (\gamma_0, \gamma_1, \ldots, \gamma_k)^T$. (i) Assume $X_i \in \mathcal{L}^p$, $2 < p \leq 4$, and*

$$\Delta_p := \sum_{i=0}^{\infty} \delta_p(i) < \infty. \tag{60}$$

Then for all $0 \leq k \leq n-1$, we have

$$\|\hat\gamma_k - (1-k/n)\gamma_k\|_{p/2} \leq \frac{3p-3}{n}\Theta_p^2 + \frac{4n^{2/p-1}\|X_1\|_p\Delta_p}{p-2}. \tag{61}$$

(ii) Assume $X_i \in \mathcal{L}^4$ and (60) holds with $p = 4$. Then as $n \to \infty$,

$$\sqrt{n}(\hat\gamma_0 - \gamma_0, \hat\gamma_1 - \gamma_1, \ldots, \hat\gamma_k - \gamma_k) \Rightarrow N[0, \mathbb{E}(D_0 D_0^T)] \tag{62}$$

where $D_0 = \sum_{i=0}^\infty \mathcal{P}_0(X_i Y_i) \in \mathcal{L}^2$ and $\mathcal{P}_0$ is the projection operator defined by (6).

Proof of Theorem 7. Write $T_n = \sum_{i=1}^n X_i X_{i+j} - n\gamma_j$. First we show that for all $j \in \mathbb{Z}$,

$$\|T_n\|_{p/2} \leq \frac{4n^{2/p}\|X_1\|_p\Delta_p}{p-2}. \tag{63}$$

Let $q = p/2$ and assume $j \geq 0$. Recall that $X_i' = g(\xi_i')$ and, for $i < 0$, we have $X_i' = X_i$ and $\mathbb{E}(X_i X_{i+j}|\xi_{-1}) = \mathbb{E}(X_i' X_{i+j}'|\xi_{-1}) = \mathbb{E}(X_i' X_{i+j}'|\xi_0)$. By Jensen's and Schwarz's inequalities,

$$(64)\quad \begin{aligned}&\|\mathcal{P}_0(X_iX_{i+j})\|_q\\ &= \|\mathbb{E}(X_iX_{i+j} - X_i'X_{i+j}'|\xi_0)\|_q\\ &\le \|X_iX_{i+j} - X_i'X_{i+j}'\|_q\\ &\le \|X_i(X_{i+j} - X_{i+j}')\|_q + \|(X_i - X_i')X_{i+j}'\|_q\\ &\le \|X_i\|_p\delta_p(i+j) + \delta_p(i)\|X_{i+j}'\|_p.\end{aligned}$$

By the triangle inequality,

$$(65)\quad \|T_n\|_q = \left\|\sum_{i=1}^n\sum_{l\in\mathbb{Z}}\mathcal{P}_{i-l}X_iX_{i+j}\right\|_q \le \sum_{l\in\mathbb{Z}}\left\|\sum_{i=1}^n\mathcal{P}_{i-l}X_iX_{i+j}\right\|_q.$$

Note that $\mathcal{P}_{i-l}X_iX_{i+j}$, $i = 1, \dots, n$, form stationary martingale differences. By Burkholder's (1988) moment inequality for martingale differences, we have

$$(66)\quad \begin{aligned}&\left\|\sum_{i=1}^n \mathcal{P}_{i-l}X_iX_{i+j}\right\|_q^q\\ &\le \frac{\mathbb{E}\{[\sum_{i=1}^n(\mathcal{P}_{i-l}X_iX_{i+j})^2]^{q/2}\}}{(q-1)^q} \le \frac{n\|\mathcal{P}_0X_lX_{l+j}\|_q^q}{(q-1)^q}\end{aligned}$$

since $q/2 \le 1$. By (64) and (65), since $\delta_p(i) = 0$ if $i < 0$, we have (63). Write

$$\begin{aligned}\hat\gamma_n - \frac{n-k}{n}\gamma_k &= \frac{1}{n}\sum_{i=k+1}^n(X_iX_{i-k} - \gamma_k)\\ &\quad - \frac{\bar X_n}{n}\sum_{i=k+1}^n(X_{i-k} + X_{i+k}) + \frac{n-k}{n}\bar X_n^2\end{aligned}$$

By Theorem 3(i), the inequality $\|\bar X_n\sum_{i=k+1}^n X_{i-k}\|_q \le \|\bar X_n\|_p\|\sum_{i=k+1}^n X_{i-k}\|_p$ and (63), (61) follows via elementary manipulations.

By Theorem 3, (ii) follows from the Crámer-Wold device and (64) with $p = 4$. □

Theorem 7 provides a CLT for $\sqrt{n}(\hat\gamma_k - \gamma_k)$ with bounded k. It turns out that, for unbounded k, the asymptotic behavior is quite different in that the asymptotic distribution does not depend on the speed of $k_n \to \infty$; see (67). By Theorem 3.1 in Keenan (1997), one can have a CLT for strong mixing processes with $k_n = o(\log n)$. An open problem was posed in the latter paper on whether the severe restriction $k_n = o(\log n)$ can be relaxed. The latter restriction excludes many important applications. Harris, McCabe and Leybourne (2003) considered linear processes with larger ranges of k_n. Theorem 8(ii) gives a CLT for short-range dependent nonlinear processes under a natural and mild condition on k_n: $k_n \to \infty$ and $k_n/n \to 0$.

Theorem 8 (Wu, 2008)**.** *Let $Z_i = (X_i, X_{i-1}, \dots, X_{i-h+1})^T$, where $h \in \mathbb{N}$ is fixed. Let $k_n \to \infty$, $\mathbb{E}(X_i) = 0$ and assume (60). Then we have (i)*

$$(67)\quad \frac{1}{\sqrt{n}}\sum_{i=1}^n[X_iZ_{i-k_n} - \mathbb{E}(X_{k_n}Z_0)] \Rightarrow N(0, \Sigma_h),$$

where Σ_h is an $h \times h$ matrix with entries

$$(68)\quad \sigma_{ab} = \sum_{j\in\mathbb{Z}}\gamma_{j+a}\gamma_{j+b} = \sum_{j\in\mathbb{Z}}\gamma_j\gamma_{j+b-a} =: \sigma_{0,a-b}, \quad 1 \le a, b \le h,$$

and (ii) if additionally $k_n/n \to 0$, then

$$(69)\quad \sqrt{n}[(\hat\gamma_{k_n}, \dots, \hat\gamma_{k_n-h+1})^T - (\gamma_{k_n}, \dots, \gamma_{k_n-h+1})^T] \Rightarrow N(0, \Sigma_h).$$

Theorem 8 can be extended to long-memory linear processes. Wu, Huang and Zheng (2010) proved central and noncentral limit theorems for sample covariances of long-memory heavy-tailed linear processes with bounded as well as unbounded lags. They showed that the limiting distribution depends in an interesting way on the strength of dependence, the heavy-tailedness of the innovations, and the magnitude of the lags.

Remark 5. Bartlett (1946) derived approximate expressions of covariances of estimated covariances: for fixed $k, l \ge 0$,

$$(70)\quad n\mathrm{cov}(\hat\gamma_k, \hat\gamma_{k+l}) \sim \sum_{m=-\infty}^{\infty}(\gamma_m\gamma_{m+l} + \gamma_{m+k+l}\gamma_{m-k}).$$

If $k \to \infty$, then the above quantity converges to $\sum_{m=-\infty}^{\infty}\gamma_m\gamma_{m+l} = \sigma_{0,l}$. Theorem 8 provides an asymptotic distributional result. For large k, $\sqrt{n}(\hat\gamma_k - \mathbb{E}\hat\gamma_k)$ behaves as $\sum_{j\in\mathbb{Z}}\gamma_j\eta_{k-j}$, where η_j are iid standard normal random variables.

Remark 6. Theorem 8 suggests that the sample covariance $\hat\gamma_k$ is not a good estimate of γ_k if k is large, a folklore result in time series analysis. For example, if $k = k_n \to \infty$ with $k_n/n \to 0$ satisfies $\sqrt{n}\gamma_{k_n} \to 0$. The mean squared error (MSE) $\mathbb{E}(\hat\gamma_{k_n} - \gamma_{k_n})^2 \sim \sigma_{00}/n$. However for such k_n the estimate $\hat\gamma_{k_n}^o \equiv 0$ has a smaller MSE $\gamma_{k_n}^2 = o(n^{-1})$. The estimate $\hat\gamma_{k_n}$ is too noisy. The shrinkage estimate $\hat\gamma_k\mathbf{1}_{|\hat\gamma_k|\ge c_n}$ with a carefully chosen threshold $c_n \to 0$ can have a better performance in the sense that it can reduce the asymptotic MSE.

8. ESTIMATION OF COVARIANCE MATRICES

Theorems 7 and 8 provide asymptotic normality for sample covariances. This section deals with the estimation of the covariance matrix

$$(71)\quad \Sigma_n = (\gamma_{i-j})_{1\le i,j\le n}$$

based on the observations $X_1, \ldots, X_n$. Estimation of covariance matrices or their inverses is important in the study of prediction and various inference problems in time series. The entry-wise convergence results of Theorems 7 and 8 do not automatically lead to matrix convergence properties of estimates of Σ_n.

For an $n \times n$ matrix A with real entries the operator norm $\rho(A)$ is defined by

$$\rho(A) = \max_{x \in \mathbb{R}^n : |x|=1} |Ax|, \tag{72}$$

where, for an n-dimensional real vector $x = (x_1, \ldots, x_n)'$, $|x| = (\sum_{i=1}^n x_i^2)^{1/2}$. Hence $\rho^2(A)$ is the largest eigenvalue of $A^\top A$, where $^\top$ denotes matrix transpose.

Wu and Pourahmadi (2009) studied convergence of covariance matrix estimates. Theorem 9 shows that, under the operator norm $\rho(\cdot)$, the sample covariance matrix estimate

$$\hat{\Sigma}_n = (\hat{\gamma}_{i-j})_{1 \le i,j \le n} \tag{73}$$

is not a consistent estimate of Σ_n; see Theorem 9(i). Case (ii) asserts that $\rho(\hat{\Sigma}_n - \Sigma_n)$ has order $\log n$. We conjecture that, with proper centering and scaling, $\rho(\hat{\Sigma}_n - \Sigma_n)$ converges to Gumbel distribution. Geman (1980) and Yin, Bai and Krishnaiah (1988) considered the convergence problem of largest eigenvalues of sample covariance matrices of iid random vectors which has independent entries; see also Johnstone (2001), El Karoui (2007) and Bai and Silverstein (2010). Their techniques are not applicable here since, in time series analysis, we have only *one* realization with *dependent* observations, while they require multiple iid copies of vectors with independent entries.

The inconsistency of $\hat{\Sigma}_n$ is due to the fact that $\hat{\gamma}_k$ is not a good estimate of γ_k if k is large; see Remark 6. Hence, to obtain a consistent covariance matrix estimate, we shall use the truncation technique by shrinking the unreliable estimates $\hat{\gamma}_k$ to 0. Namely we can use *the banded covariance matrix estimate*

$$\hat{\Sigma}_{n,l_n} = (\hat{\gamma}_{i-j} \mathbf{1}_{|i-j| \le l_n})_{1 \le i,j \le n}, \tag{74}$$

where l_n is called the banding parameter. Under suitable conditions on l_n, $\hat{\Sigma}_{n,l_n}$ is consistent. Theorem 10 provides an explicit upper bound for $\rho(\hat{\Sigma}_{n,l_n} - \Sigma_n)$.

The estimate $\hat{\Sigma}_{n,l_n}$ in (74) is not guaranteed to be non-negative definite. This can be a serious shortcoming in applications. To rectify the latter issue, we propose to use the tapered estimate:

$$\tilde{\Sigma}_{n,l_n} = (\hat{\gamma}_{i-j} w(|i-j|/l_n))_{1 \le i,j \le n} = \hat{\Sigma}_n \star W_n, \tag{75}$$

where $\star$ is the Hadamard (or Schur) product, which is formed by element-wise multiplication of elements of matrices, and $w(\cdot)$ is a lag window function satisfying (i) $w(\cdot)$ is even and piecewise continuous; (ii) $w(0) = 1$, $\sup_u |w(u)| \le 1$ and (iii) $w(u) = 0$ if $|u| > 1$. Note that $\hat{\Sigma}_n$ is non-negative definite. If W_n is also non-negative definite, then by the Schur Product Theorem in matrix theory (Horn and Johnson, 1990), their Schur product $\tilde{\Sigma}_{n,l_n}$ is also non-negative definite. The truncated or rectangular window with $w(u) = \mathbf{1}_{|u| \le 1}$ is, unfortunately, not non-negative definite. The Bartlett or triangular window $w_B(u) = \max(0, 1 - |u|)$ leads to a positive definite weight matrix W_n in view of

$$w_B(u) = \int_{\mathbb{R}} w(x) w(x+u) dx, \tag{76}$$

where w is the rectangular window. To see this, let $c_i, u_i \in \mathbb{R}$, $i = 1, \ldots, n$. By (76),

$$\sum_{1 \le i,j \le n} c_i w_B(u_i - u_j) c_j = \int_{\mathbb{R}} \left[\sum_{i=1}^n c_i w(v - u_i) \right]^2 dv \ge 0.$$

Replacing $w(\cdot)$ in (76) by $\sqrt{3} w_B(\cdot)$, we obtain the Parzen window:

$$\begin{aligned} w_P(u) &= \int_{\mathbb{R}} w_B(x) w_B(x+u) dx \\ &= \begin{cases} 1 - 6u^2 + 6|u|^3, & |u| < 1/2, \\ \max[0,\, 2(1-|u|)^3], & |u| \ge 1/2. \end{cases} \end{aligned} \tag{77}$$

Theorem 9. *(i) (Wu and Pourahmadi (2009)) Assume that the process (X_t) in (1) is weakly stable, namely (45) holds with $p = 2$. If $\| \sum_{i=0}^\infty \mathcal{P}_0 X_i \| > 0$, then, $\rho(\hat{\Sigma}_n - \Sigma_n) \not\to 0$ in probability. (ii) (Xiao and Wu (2010b)) Let conditions in Theorem 13 be satisfied. Then there exists a constant $c > 0$ such that*

$$\lim_{n \to \infty} \mathbb{P}[c^{-1} \log n \le \rho(\hat{\Sigma}_n - \Sigma_n) \le c \log n] = 1. \tag{78}$$

Theorem 10. *Assume that (X_t) in (1) satisfies $\mathbb{E} X_i = 0$. Let $\hat{\gamma}_k = n^{-1} \sum_{i=|k|+1}^n X_i X_{i-|k|}$, $|k| < n$, $w_k = w(k/l)$, and $b_n = \sum_{k=1}^l |1 - w_k + k w_k / n| |\gamma_k| + \sum_{j=l+1}^n |\gamma_j|$. (i) If (8) holds with $2 < p \le 4$, then for $\tilde{\Sigma}_{n,l} = (\hat{\gamma}_{i-j} w(|i-j|/l))_{1 \le i,j \le n}$, we have*

$$\| \rho(\tilde{\Sigma}_{n,l} - \Sigma_n) \|_q \le 2 b_n + (l+1) \frac{4 \|X_1\|_p \Delta_p}{n^{1-1/q}(p-2)}, \ 0 \le l < n, \tag{79}$$

where $q = p/2$. Hence if $l = l_n \to \infty$ and $l_n n^{1/q-1} \to 0$, then

$$\| \rho(\tilde{\Sigma}_{n,l} - \Sigma_n) \|_q \to 0. \tag{80}$$

(ii) (Xiao and Wu (2010b)) Assume $X_i \in \mathcal{L}^p$, $p > 4$, and $\Theta_p(m) = O(m^{-\alpha})$, $\alpha > 0$. Let $l_n \asymp n^\lambda$, where $\lambda \in (0,1)$ satisfies $\lambda < p\alpha/2$ and $(1 - 2\alpha)\lambda < 1 - 4/p$. Then

$$\rho(\tilde{\Sigma}_{n,l} - \Sigma_n) = O(b_n) + O_{\mathbb{P}}[(n^{-1} l_n \log l_n)^{1/2}]. \tag{81}$$

Additionally assume that $X_0 \in \mathcal{L}^p$, $p > \max(4, 2/(1-\lambda))$, $\sum_{t=0}^{\infty} \min(\delta_{t,p}, \Psi_{n+1,p}) = O(n^{-T_1})$ *with* $T_1 > \max[1/2-(p-4)/(2p\lambda),\, 2\lambda/p]$ *and* $\Theta_{n,p} = O(n^{-T_2})$, $T_2 > \max[0,\, 1-(p-4)/(2p\lambda)]$. *Then there exists a constant* $c > 0$ *such that*

$$\lim_{n\to\infty} \mathbb{P}[c^{-1}(n^{-1}l_n \log l_n)^{1/2} - 2b_n \le \rho(\tilde{\Sigma}_{n,l} - \Sigma_n)] = 1. \tag{82}$$

Proof. (i) We shall use the argument in Wu and Pourahmadi (2009). Since $\tilde{\Sigma}_{n,l} - \Sigma_n$ is a symmetric Toeplitz matrix, from Golub and Van Loan (1989), we have

$$\begin{aligned}
&\rho(\tilde{\Sigma}_{n,l} - \Sigma_n) \\
&\quad \le \max_{1\le j\le n} \sum_{i=1}^{n} |\hat{\gamma}_{i-j} w_{|i-j|} - \gamma_{i-j}| \\
&\quad \le \sum_{i=1-n}^{n-1} |\hat{\gamma}_i w_i - \gamma_i| \le 2\sum_{i=0}^{l} |\hat{\gamma}_i w_i - \gamma_i| + 2\sum_{i=1+l}^{n} |\gamma_i|.
\end{aligned}$$

By Theorem 7(i), we have (79) since the bias $|\mathbb{E}\hat{\gamma}_i - \gamma_i| \le i|\gamma_i|/n$. (ii) Here we shall apply Theorem 3 in Liu and Wu (2010b). For details see Xiao and Wu (2010b). □

The bound in (79) is non-asymptotic in that it holds for all $l < n$. If $\mathbb{E}X_i$ is unknown, then we should estimate γ_k by $\hat{\gamma}_k$ defined in (59). By Theorem 7(i), the bound in (79) still holds with $4\|X_1\|_p \Delta_p / n^{1-1/q}(p-2)$ therein replaced by the slightly bigger one in (61). Relations (81) and (82) imply the sharp and elegant result: if $b_n = o[(n^{-1}l_n \log l_n)^{1/2}]$, then the *exact* order of magnitude of the operator norm $\rho(\tilde{\Sigma}_{n,l} - \Sigma_n)$ is $(n^{-1}l_n \log l_n)^{1/2}$.

Note that our setting is different from the one in Bickel and Levina (2008) and Wu and Pourahmadi (2003), where it is assumed that there exist multiple iid copies of $(X_i)_{i=1}^n$. In time series applications, however, oftentimes one has only one realization.

We now discuss some interesting special cases. Assume $p = 4$ and $\gamma_k = O(\rho^k)$ for some $0 < \rho < 1$. Choose $l = l_n = \lfloor (\log n)/\log \rho^{-2} \rfloor$. Then for the rectangle window with $w_k = 1$, $|k| \le l$, by (79), we have $\|\rho(\tilde{\Sigma}_{n,l} - \Sigma_n)\| = O(n^{-1/2}\log n)$, an optimal bound up to a multiplicative logarithmic factor. The drawback is that the estimated covariance matrix $\tilde{\Sigma}_{n,l}$ may not be non-negative definite. For the Bartlett window, choosing $l \asymp n^{1/4}$, we have

$$\begin{aligned}
\|\rho(\tilde{\Sigma}_{n,l} - \Sigma_n)\| &= O(1)\sum_{k=1}^{l}(1-w_k)|\gamma_k| + O(ln^{-1/2} + \rho^l) \\
&= O(l^{-1} + ln^{-1/2} + \rho^l) = O(n^{-1/4})
\end{aligned} \tag{83}$$

Using the Parzen window, since $1 - w_P(u) = O(u^2)$, letting $l \asymp n^{1/6}$, we have

$$\|\rho(\tilde{\Sigma}_{n,l} - \Sigma_n)\| = O(l^{-2} + ln^{-1/2} + \rho^l) = O(n^{-1/3}). \tag{84}$$

Example 12. In (76) if we let $w(x) = \sqrt{30}x(1-x)\mathbf{1}_{|x|\le 1}$, then the window

$$\int_{\mathbb{R}} w(x)w(x+u)dx = (1-|u|)^3(1+3|u|+u^2), \quad |u| \le 1, \tag{85}$$

also leads to a positive-definite weight matrix.

As an application of our covariance matrix estimates, we can apply the bound (79) to the celebrated problem of prediction and filtering of stationary time series. Kolmogorov (1939) and Wiener (1949) considered the fundamental problem of predicting unknown future values of a time series based on past observations. Their theory is one of the great achievements in time series analysis. For a detailed account see Doob (1953), Whittle (1963), Priestley (1981) and Pourahmadi (2001) among others. In many of such works, it is assumed that the covariances γ_k are known. For example, to predict X_n based on past observations, Kolmogorov and Wiener assumed that the whole past $(X_i)_{i=-\infty}^{n-1}$ is known and in this case by the ergodic theorem γ_k can be accurately estimated. In practice, however, one has only finitely many past observations, and thus γ_k should be replaced by its estimates. Then the question naturally appears as to whether a prediction theory can be obtained for finite samples. Jones (1964) and Bhansali (1974, 1977) investigated this problem by factorizing estimated spectral densities. The bound (79) enables us to establish a finite sample version of the Wiener-Kolmogorov prediction theory by using the asymptotic theory for sample covariances and covariance matrix estimates. Also, an asymptotic theory for estimates of coefficients in the Wold decomposition theorem and in the discrete Wiener-Hopf equations can be established.

9. PERIODOGRAMS

In spectral or frequency domain analysis of time series, the primary subjects of interest are periodograms and spectral density functions. Periodograms can be used to test the existence of hidden periodicities or seasonal components. Spectral density, power spectral density, or spectrum describes how the energy of a time series varies with frequency.

Definition 4 (Periodogram)**.** Let $\imath = \sqrt{-1}$ be the imaginary unit. Let $x_1, \ldots, x_n$ be a sequence of real numbers. Its periodogram is define as

$$I_n(\phi) = \frac{|S_n(\phi)|^2}{n}, \quad \phi \in \mathbb{R}, \tag{86}$$

where $S_n(\phi)$ is the Fourier transform of $\{x_1, \ldots, x_n\}$:

$$S_n(\phi) = \sum_{t=1}^{n} x_t e^{\imath t\phi}. \tag{87}$$

Definition 5 (Spectral density function)**.** Let (X_k) be a stationary process with covariance function $\gamma_k = \mathrm{cov}(X_0, X_k)$. We say that F is a spectral distribution function if it is right-continuous, non-decreasing and bounded on $[0, 2\pi]$ such that $\gamma_k = \int_0^{2\pi} e^{\imath k\phi} dF(\phi)$. If F is absolutely continuous, then its derivative $f = F'$ is called the spectral density.

Note that the process (1) is regular in the sense that $\mathbb{E}(X_j|\mathcal{F}_{-\infty}) = \mathbb{E}(X_j)$ since the sigma algebra $\sigma(\mathcal{F}_{-\infty}) = \cap_{i\in\mathbb{Z}}\sigma(\mathcal{F}_i) = \{\emptyset, \Omega\}$ is trivial. Theorem 1 in Peligrad and Wu (2010) asserts that, for a regular process, its spectral density function exists almost surely over $\phi \in [0, 2\pi]$ with respect to the Lebesgue measure. If

$$\sum_{k\in\mathbb{Z}} |\gamma_k| < \infty, \tag{88}$$

then spectral density function has the form

$$f(\phi) = \frac{1}{2\pi}\sum_{k\in\mathbb{Z}} \gamma_k e^{\imath k\phi} = \frac{1}{2\pi}\sum_{k\in\mathbb{Z}} \gamma_k \cos(k\phi), \tag{89}$$

which exists at all $\phi \in \mathbb{R}$ and is continuous. The spectral density function is even and has period 2π. Its continuity property is related to the decay rate of the covariances γ_k. If $\sum_{k=1}^\infty k^p|\gamma_k| < \infty$, $p > 0$, then $f \in \mathcal{C}^p(\mathbb{R})$. If the former holds for all $p > 0$, for example if $\gamma_k \to 0$ geometrically quickly, then f is an analytic function.

Let (X_k) be a stationary second order process with mean 0; let $I_n(\phi)$ be the periodogram of $X_1, \ldots, X_n$. Assume (88). Then as $n \to \infty$, elementary manipulations show that

$$\mathbb{E}I_n(\phi) = \sum_{k=1-n}^{n-1} (1 - |k|/n)\gamma_k \cos(k\phi) \to 2\pi f(\phi).$$

Hence $I_n(\phi)$ is an asymptotically unbiased estimate of $2\pi f(\phi)$. However, by Theorem 11 or Proposition 2, $I_n(\phi)$ is not consistent.

The central limit problem of $S_n(\phi)$ has been studied by Rosenblatt (Theorem 5.3, p 131, 1985) for mixing processes, Brockwell and Davis (Theorem 10.3.2., p 347, 1991), Walker (1965) and Terrin and Hurvich (1994) for linear processes. For other contributions see Olshen (1967), Rootzén (1976), Yajima (1989) and Walker (2000). Theorem 11 is very general and it allows nonlinear, non-strong mixing and/or even long-memory processes. It follows from Theorem 1 in Peligrad and Wu (2010). Proposition 2 concerns a fixed frequency $\vartheta \in (0, 2\pi)$ and it is established in Wu (2005). Note that the case in which $\vartheta = 0$ is covered by Theorem 3 since $S_n(0) = S_n$. Theorem 12 is for Fourier transforms at Fourier frequencies $\vartheta_k = 2\pi k/n$, $k = 1, \ldots, n$, where $\vartheta_1 = 2\pi/n$ is called the fundamental frequency. Central limit theorem of this type is a key ingredient in the Whittle likelihood method. For a complex number z, let $\Re z$ (resp. $\Im z$) denote the real (resp. imaginary) part of z.

Theorem 11. *Assume $\mathbb{E}X_k^2 < \infty$. (i) For almost all $\vartheta \in \mathbb{R}$ (Lebesgue), we have*

$$\begin{pmatrix} \Re \\ \Im \end{pmatrix} \frac{S_n(\vartheta)}{\sqrt{n}} \Rightarrow N[0, \pi f(\vartheta)\mathrm{Id}_2] \tag{90}$$

and consequently $I_n(\vartheta)/(2\pi f(\vartheta)) \Rightarrow Exp(1)$, the standard exponential distribution with scale parameter 1. (ii) Moreover, for almost all pairs (ϑ, φ) (Lebesgue), $S_n(\vartheta)/\sqrt{n}$ and $S_n(\varphi)/\sqrt{n}$ are asymptotically independent.

Proposition 2. *Assume that*

$$\sum_{i=0}^{\infty} \|\mathcal{P}_0 X_i - \mathcal{P}_0 X_{i+1}\| < \infty. \tag{91}$$

Then (90) holds for all $0 < \vartheta < 2\pi$. A sufficient condition for (91) is (45).

By the celebrated Fast Fourier Transform algorithm, one can compute $S_n(\vartheta_j)$, $j = 0, \ldots, n-1$, in a very efficient way with computational complexity $O(n \log n)$ and memory complexity $O(n)$. Historically this computational advantage fuels the development of spectral analysis. Theorem 12 concerns asymptotic distribution of $S_n(\vartheta_j)$. In the special case in which X_i are iid standard Gaussian random variables, $I(\vartheta_j)/2$, $j = 1, \ldots, \lfloor (n-1)/2 \rfloor$, are iid standard exponentials.

Theorem 12. *Assume that (X_i) defined in (1) satisfies (45) and $\min_\vartheta f(\vartheta) > 0$. Let $q \in \mathbb{N}$, $m = \lfloor (n-1)/2 \rfloor$ and let Y_k, $1 \le k \le 2q$, be iid standard normals. Then*

$$\left\{ \frac{S_n(\vartheta_{l_j})}{\sqrt{n\pi f(\vartheta_{l_j})}},\ 1 \le j \le q \right\} \Rightarrow \{Y_{2j-1} + \imath Y_{2j},\ 1 \le j \le q\} \tag{92}$$

for integers $1 \le l_1 < l_2 < \cdots < l_q \le m$, where the indices l_j may depend on n. Consequently, for $\tilde{I}_n(\vartheta) := I_n(\vartheta)/f(\vartheta)$,

$$\{\tilde{I}_n(\vartheta_{l_j}),\ 1 \le j \le q\} \Rightarrow \{E_j,\ 1 \le j \le q\}, \tag{93}$$

where E_j are iid standard exponential random variables (exp(1)).

By (93) of Theorem 12 and the continuous mapping theorem, if q is fixed, we have $\max_{j\le q} \tilde{I}_n(\theta_{l_j}) \Rightarrow \max_{j\le q} E_j$. Lin and Liu (2009b) proved a deep result that the latter convergence still holds by letting $q = m = \lfloor (n-1)/2 \rfloor$ in the sense of (95). Note that $\max_{j\le m} E_j - \log m$ converges to the standard Gumbel distribution since, for fixed $u \in \mathbb{R}$, as $m \to \infty$,

$$\begin{aligned} \mathbb{P}\left(\max_{1\le l\le m} E_j - \log m \le u \right) &= \mathbb{P}^m(E_j \le u + \log m) \\ &= (1 - e^{-u}/m)^m \to e^{-e^{-u}}. \end{aligned}$$

Theorem 13 (Lin and Liu, 2009b). *Assume that (X_i) defined in (1) satisfies $\min_\vartheta f(\vartheta) > 0$, $\mathbb{E}(X_i) = 0$, $X_i \in \mathcal{L}^p$, $p > 2$ and, as $j \to \infty$,*

$$(94) \qquad \sum_{i=j}^{\infty} \delta_p(i) = o(1/\log j).$$

Recall Theorem 12 for $\tilde{I}_n(\theta)$ and $m = \lfloor (n-1)/2 \rfloor$. Then

$$(95) \qquad \mathbb{P}\left(\max_{1\le l\le m} \tilde{I}_n(\theta_l) - \log m \le u\right) = e^{-e^{-u}}, \quad u \in \mathbb{R}.$$

10. ESTIMATION OF SPECTRAL DENSITIES

A fundamental problem in spectral analysis of time series is the estimation of spectral density functions. Section 9 demonstrates that $I_n(\vartheta)$ is an asymptotically unbiased, but inconsistent estimate of $f(\theta)$. To obtain a consistent estimate, one can introduce a *taper, data window* or *convergence factor* K and propose

$$(96) \qquad f_n(\theta) = \frac{1}{2\pi} \sum_{k=1-n}^{n-1} K(k/B_n)\hat{\gamma}_k e^{\imath k\theta},$$

where B_n satisfies $B_n \to \infty$ and $B_n/n \to 0$, and the function K is symmetric, bounded, $K(0) = 1$ and K is continuous at 0. If K has bounded support, since $B_n/n \to 0$, the summands for large k in (96) are zero. Here f_n is called the *lag window estimate.*

Properties of spectral density estimates have been discussed in many classical textbooks on time series; see Anderson (1971), Brillinger (1975), Brockwell and Davis (1991), Grenander and Rosenblatt (1984), Priestley (1981) and Rosenblatt (1985) among others. A classical problem in spectral analysis is to develop an asymptotic distributional theory for the spectral density estimate $f_n(\theta)$. With the latter results one can perform statistical inference such as hypothesis testing and construction of confidence intervals. However, it turns out that the central limit problem for $f_n(\theta)$ is highly nontrivial. Many of the previous results require that the underlying processes are linear or strong mixing, or satisfy stringent cumulant summability conditions that are not easily verifiable.

Here we shall present a central limit theorem for $f_n(\lambda)$ under very mild and natural conditions, thus substantially extending the applicability of spectral analysis to nonlinear and/or non-strong mixing processes. Let $\varpi(u) = 2$ if $u/\pi \in \mathbb{Z}$ and $\varpi(u) = 1$ if $u/\pi \notin \mathbb{Z}$.

Theorem 14 (Liu and Wu, 2010b). *Assume $\mathbb{E}(X_k) = 0$, $\mathbb{E}(X_k^4) < \infty$ and the 4-stability condition $\Delta_4 < \infty$. Let $B_n \to \infty$ and $B_n = o(n)$ as $n \to \infty$. Further assume that K is symmetric, bounded, $\lim_{u\to 0} K(u) = K(0) = 1$, $\kappa := \int_{-\infty}^{\infty} K^2(x)dx < \infty$, K is continuous at all but a finite number of points and $\sup_{0<w\le 1} \sum_{j\ge c/w} K^2(jw) \to 0$ as $c \to \infty$. Then for any fixed $0 \le \theta < 2\pi$,*

$$(97) \qquad \sqrt{\frac{n}{B_n}}\{f_n(\theta) - \mathbb{E}[f_n(\theta)]\} \Rightarrow N[0, s^2(\theta)], \quad \text{where } s^2(\theta) = \varpi(\theta) f^2(\theta)\kappa.$$

In Theorem 14, the short-range dependence condition $\Delta_4 < \infty$ is natural, since otherwise the process (X_j) may be long-range dependent and the spectral density function may not be well-defined. The bandwidth condition $B_n \to \infty$ and $B_n = o(n)$ is also natural.

A particularly interesting special case of Theorem 14 is $\theta = 0$. In this case $2\pi f(0) = \sigma^2$ is the long-run variance. Estimation of long-run variance is needed in the inference of means of stationary processes; see Theorems 3 and 5. By (97), we have

$$(98) \qquad \sqrt{\frac{n}{B_n}}\{f_n(0) - f(0)\} \Rightarrow N(0, s^2), \text{ where } s^2 = 2f^2(0)\kappa,$$

if the bandwidth $b_n = 1/B_n$ satisfies

$$2\pi\{\mathbb{E}[f_n(0)] - f(0)\} = \sum_{k=1-n}^{n-1} K(kb_n)(1-|k|/n)\gamma_k - \sum_{k=-\infty}^{\infty} \gamma_k = O((nb_n)^{-1/2}).$$

If K is the rectangle kernel $K(u) = \mathbf{1}_{|u|\le 1}$, then the above condition is reduced to

$$\frac{1}{n}\sum_{k=1}^{B_n} k\gamma_k + \sum_{k=1+B_n}^{\infty} \gamma_k = O((nb_n)^{-1/2}).$$

Hence, taking a logarithmic transformation of (98), we can stabilize the variance via

$$(99) \qquad \sqrt{\frac{n}{B_n}}\{\log f_n(0) - \log f(0)\} \Rightarrow N(0, 4).$$

Therefore the $(1-\alpha)$th, $0 < \alpha < 1$, confidence interval for $\log f(0)$ can be constructed by

$$\log f_n(0) \pm \frac{2z_{1-\alpha/2}}{\sqrt{nb_n}},$$

where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$th quantile of the standard normal distribution.

The spectral density estimate (96) is non-recursive in the sense that it cannot be updated within $O(1)$ computation once a new observation arrives. Xiao and Wu (2010a) proposed a recursive or single-pass algorithm which is computationally fast in that the update can be performed within $O(1)$ computation, and the required memory complexity is also only $O(1)$. The computational advantage becomes highly attractive for efficient and fast processing for extra long time series. Xiao and Wu (2010a) proved a central limit theorem for their recursive estimates by using physical dependence measures.

11. KERNEL ESTIMATION OF TIME SERIES

Kernel method is an important nonparametric approach in the inference of the data-generating mechanisms of time series. It is useful in situations in which the functional or parametric forms are unknown. Asymptotic properties for kernel estimates of iid observations have been studied in Silverman (1986), Devroye and Györfi (1985), Wand and Jones (1995), Prakasa Rao (1983), Nadaraya (1989) and Eubank (1999) among others, and for strong mixing processes in Robinson (1983), Singh and Ullah (1985), Castellana and Leadbetter (1986), Györfi et al (1989) and Bosq (1996), Yu (1993), Neumann (1998), Kreiss and Neumann (1998), Härdle et al (1997), Tjostheim (1994) and Fan and Yao (2003). Wu and Mielniczuk (2002) and Ho and Hsing (1996) considered long-memory processes.

Here we shall present an asymptotic theory for kernel estimates with predictive dependence measures. Consider the model

$$Y_i = G(X_i, \eta_i),\ X_i = H(\ldots, \varepsilon_{i-1}, \varepsilon_i), \tag{100}$$

where η_i, $i \in \mathbb{Z}$, are also iid and η_i is independent of $\mathcal{F}_{i-1} = (\ldots, \varepsilon_{i-2}, \varepsilon_{i-1})$. An important special example of (100) is the autoregressive model

$$X_{i+1} = R(X_i, \varepsilon_{i+1}) \tag{101}$$

by letting $\eta_i = \varepsilon_{i+1}$ and $Y_i = X_{i+1}$. Given the data (X_i, Y_i), $0 \le i \le n$, let

$$T_n(x) = \frac{1}{n}\sum_{t=1}^{n} Y_t K_{b_n}(x - X_t), \tag{102}$$

where $K_b(x) = K(x/b)/b$, the kernel K is symmetric and bounded on $\mathbb{R}$: $\sup_{u\in\mathbb{R}} |K(u)| \le K_0$, $\int_{\mathbb{R}} K(u)du = 1$ and K has bounded support; namely, $K(x) = 0$ if $|x| \ge c$ for some $c > 0$, and $b = b_n$ is a sequence of bandwidths satisfying the natural condition

$$b_n \to 0 \text{ and } nb_n \to \infty. \tag{103}$$

The Nadaraya-Watson estimator of the regression function

$$g(x_0) = \mathbb{E}(Y_n | X_n = x_0) = \mathbb{E}[G(x_0, \eta_0)] \tag{104}$$

has the form

$$g_n(x_0) = \frac{T_n(x_0)}{f_n(x_0)}, \tag{105}$$

where f_n is Rosenblatt's (1956) kernel density estimate

$$f_n(x_0) = \frac{1}{nb_n}\sum_{t=1}^{n} K(\frac{x_0 - X_t}{b_n}) = \frac{1}{n}\sum_{t=1}^{n} K_{b_n}(x_0 - X_t). \tag{106}$$

For $i \in \mathbb{Z}$, $l \in \mathbb{N}$, let $F_l(x|\mathcal{F}_i) = \mathbb{P}(X_{i+l} \le x|\mathcal{F}_i)$ be the l-step ahead conditional distribution function of X_{i+l} given $\mathcal{F}_i$ and $f_l(x|\mathcal{F}_i) = \frac{d}{dx} F_l(x|\mathcal{F}_i)$ be the conditional density.

Theorem 15 (Wu (2005), Wu, Huang and Huang (2010))**.** *Assume that exists a constant $c_0 < \infty$ such that $\sup_{x\in\mathbb{R}} f_1(x|\mathcal{F}_0) \le c_0$ almost surely, and*

$$\sum_{i=1}^{\infty} \sup_x \|\mathcal{P}_0 f_1(x|\mathcal{F}_i)\| < \infty. \tag{107}$$

Let $\kappa = \int_{\mathbb{R}} K^2(u)du$. Assume (103). (i) The central limit theorem $\sqrt{nb_n}[f_n(x_0) - \mathbb{E}f_n(x_0)] \Rightarrow N(0, f(x_0)\kappa)$ holds. (ii) Let $V_p(x) = \mathbb{E}[|G(x, \eta_n)|^p]$ and $\sigma^2(x) = V_2(x) - g^2(x)$. If $f(x_0) > 0$, V_2, $g \in \mathcal{C}(\mathbb{R})$ and that $V_p(x)$ is bounded on a neighborhood of x_0, then

$$\sqrt{nb_n}\left\{g_n(x_0) - \frac{\mathbb{E}T_n(x_0)}{\mathbb{E}f_n(x_0)}\right\} \Rightarrow N[0, \sigma^2(x_0)\kappa/f(x_0)]. \tag{108}$$

Using the Crámer-Wold device, we can have a multivariate version of (108). Liu and Wu (2010a) developed an asymptotic distributional theory for the maximum deviation

$$\Delta_n := \sup_{l\le x\le u} \frac{\sqrt{nb}}{\sqrt{\kappa f(x)}} |f_n(x) - \mathbb{E}f_n(x)|, \tag{109}$$

where l and u are fixed bounds. Similar asymptotic distributions hold for maximum deviations of the regression estimates as well. Such results can be used to construct uniform or simultaneous confidence bands for unknown density and regression functions. Liu and Wu's theorem substantially generalize earlier results which were obtained under independence (Bickel and Rosenblatt, 1973) or restrictive beta mixing assumptions (Neumann, 1998). The problem of generalizing Bickel and Rosenblatt's theorem to stationary processes is very challenging and it has been open for a long time. Fan and Yao (2003, p. 208) conjectured that similar results hold for stationary processes under certain mixing conditions. Using physical dependence measure, Liu and Wu solved this open problem and established an asymptotic theory for both short- and long-range dependent processes.

Theorem 16 (Liu and Wu (2010a))**.** *Assume $X_n = a_0\varepsilon_n + g(\ldots, \varepsilon_{n-2}, \varepsilon_{n-1}) \in \mathcal{L}^p$ for some $p > 0$, where g is a measurable function, $a_0 \ne 0$, and the density function f_ε of ε_1 is positive and $\sup_{x\in\mathbb{R}}[f_\varepsilon(x) + |f'_\varepsilon(x)| + |f''_\varepsilon(x)|] < \infty$. For the bandwidth b_n, assume that there exists $0 < \delta_2 \le \delta_1 < 1$ such that $n^{-\delta_1} = O(b_n)$ and $b_n = O(n^{-\delta_2})$. Let $p' = \min(p, 2)$ and*

$\Theta_n = \sum_{i=0}^{n} \delta_{p'}(i)^{p'/2}$. *Assume* $\Psi_{n,p'} = O(n^{-\gamma})$ *for some* $\gamma > \delta_1/(1-\delta_1)$ *and*

$$(110) \qquad \sum_{k=-n}^{\infty} (\Theta_{n+k} - \Theta_k)^2 = o(b_n^{-1} n \log n).$$

Let the kernel $K \in \mathcal{C}^1[-1,1]$ *with* $K(\pm 1) = 0$*; let* $l = 0$ *and* $u = 1$*. Then*

$$(111) \qquad \mathbb{P}\left((2\log b^{-1})^{1/2}\Delta_n - 2\log b^{-1} - \log K_3^{1/2} \le z\right) \to e^{-2e^{-z}}$$

holds for every $z \in \mathbb{R}$*, where* $K_3 = \int_{-1}^{1}(K'(t))^2 dt / (4\pi^2 \int_{-1}^{1} K^2(t)dt)$*.*

For the short-range dependent linear process $X_n = \sum_{j=0}^{\infty} a_j \varepsilon_{n-j}$ with $\mathbb{E}\varepsilon_1 = 0$ and $\mathbb{E}\varepsilon_1^2 = 1$, (110) is satisfied if $\sum_{j=0}^{\infty} |a_j| < \infty$ and $\sum_{j=n}^{\infty} a_j^2 = O(n^{-\gamma})$ for some $\gamma > 2\delta_1/(1-\delta_1)$. The latter condition can be weaker than $\sum_{j=0}^{\infty} |a_j| < \infty$ if $\delta_1 < 1/3$. Interestingly, (110) also holds for some long-range dependent processes. Let $a_j = j^{-\beta}\ell(j)$, $1/2 < \beta < 1$, where $\ell(\cdot)$ is a slowly varying function. If $\delta_1/(1-\delta_1) < \beta - 1/2$ and $b_n^{1/2} n^{1-\beta}\ell(n) = o(\log^{-1/2} n)$. then (111) holds. If $\log^{1/2} n = o(b_n^{1/2} n^{1-\beta}\ell(n))$, Liu and Wu showed that the limiting distribution of Δ_n is no longer Gumbel.

12. U-STATISTICS

Given a sample $X_1, \ldots, X_n$, consider the weighted U-statistic

$$(112) \qquad U_n = \sum_{1 \le i,j \le n} w_{i-j} K(X_i, X_j),$$

where w_i are weights with $w_i = w_{-i}$ and K is a symmetric measurable function. Many statistics can be expressed in the form of U_n. Hoeffding (1961), O'Neil and Redner (1993), Major (1994) and Rifi and Utzet (2000) considered properties of U_n for iid observations. Yoshihara (1976), Denker and Keller (1983, 1986), Borovkova, Burton and Dehling (1999, 2001, 2002) and Dehling, Wendler (2010) dealt with strong mixing processes. Hsing and Wu (2004) developed general results for processes satisfying (1) for both summable and non-summable weights. In the context of U-statistics, it is natural to define the predictive dependence measure

$$(113) \qquad \theta_{i,j} = \|\mathcal{P}_0 K(X_i, X_j)\|.$$

Theorem 17 (Hsing and Wu, 2004)**.** *(i) (Summable weights) Assume that*

$$(114) \qquad \sum_{k=0}^{\infty}\sum_{i=0}^{\infty} |w_k| \theta_{i,i-k} < \infty.$$

Then there exists $\sigma^2 < \infty$ *such that* $(U_n - \mathbb{E}U_n)/\sqrt{n} \Rightarrow N(0, \sigma^2)$*. (ii) (Non-summable weights) Let* $W_n(i) = \sum_{j=1}^{n} w_{i-j}$ *and* $W_n = [\sum_{i=1}^{n} W_n^2(i)/n]^{1/2}$*. Assume* $\sum_{i=1}^{\infty} |w_i| = \infty$, $\sum_{k=0}^{n}(n-k)w_k^2 = o(nW_n^2)$, $\liminf_{n\to\infty} W_n/(\sum_{i=0}^{n}|w_i|) > 0$ *and*

$$(115) \qquad \sum_{\ell=0}^{\infty} \sup_{j\in\mathbb{Z}} \|K(X_0, X_j) - K(\tilde{X}_0, \tilde{X}_j)\| < \infty,$$

$$\text{where } \tilde{X}_j = \mathbb{E}(X_j | \varepsilon_{j-\ell}, \ldots, \varepsilon_j).$$

Then there exists $\sigma_U^2 < \infty$ *such that* $(U_n - \mathbb{E}U_n)/(W_n\sqrt{n}) \Rightarrow N(0, \sigma_U^2)$*.*

Hsing and Wu (2004) applied Theorem 17(ii) with $w_i \equiv 1$ and derived a central limit theorem for the correlation integral $U = \sum_{i,j=1}^{n} \mathbf{1}_{|X_i - X_j| \le b}$, which measures the number of pairs (X_i, X_j) such that their distance is less than $b > 0$. Correlation integral is of critical importance in the study of dynamical systems (Grassberger and Procaccia (1983a, 1983b), Wolff (1990), Serinko (1994), Denker and Keller (1986), Borovkova et al (1999)). The central limit theorem is useful for the related statistical inference. A non-central limit theorem is also developed in Hsing and Wu (2004) for long memory linear processes.

13. CONCLUSION

Physical and predictive dependence measures shed new light on the asymptotic theory of time series. They are directly related to the underlying physical mechanisms of the processes and have the attractive input-output interpretation. In many cases they are easy to compute and results built upon them are often optimal and nearly optimal. They are particularly useful for dealing with complicated statistics of time series such as eigenvalues of sample covariance matrices and maxima of periodograms, where it is difficult to apply the traditional strong mixing type of conditions. We expect that our framework, tools and results can be useful for other asymptotic problems in the study of stationary time series.

ACKNOWLEDGEMENTS

This work was supported in part from DMS-0906073 and DMS-0448704. I am grateful to Mohamed El Machkouri, Martin Wendler, Jan Mielniczuk and a referee for their many helpful comments.

Received 10 August 2010

REFERENCES

ALSMEYER, G. and FUH, C. D. (2001). Limit theorems for iterated random functions by regenerative methods. *Stochastic Processes and their Applications* **96** 123–142. MR1856683

ANDERSON, T. W. (1971). *The Statistical Analysis of Time Series.* Wiley, New York. MR0283939

Anderson, T. W. and Walker, A. M. (1964). On the asymptotic distribution of the autocorrelations of a sample from a linear stochastic process. *The Annals of Mathematical Statistics* **35** 1296–1303. MR0165602

Andrews, D. W. K. (1995). Nonparametric kernel estimation for semiparametric models. *Econometric Theory* **11** 560–596. MR1349935

Arjas, E. and Lehtonen, T. (1978). Approximating many server queues by means of single server queues. *Math. Operation Research* **3** 205–223. MR0506659

Arnold, L. (1998). *Random Dynamical Systems*. Springer, Berlin. MR1723992

Bai, Z. and Silverstein, J. W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. Springer, New York. MR2567175

Barnsley, M. F. and Elton, J. H. (1988). A new class of Markov processes for image encoding. *Adv. Appl. Probab.* **20** 14–32. MR0932532

Bartlett, M. S. (1946). On the theoretical specification and sampling properties of autocorrelated time-series. *Suppl. J. Roy. Statist. Soc.* **8** 27–41. MR0018393

Bhansali, R. J. (1974). Asymptotic properties of the Wiener-Kolmogorov predictor. I. *Journal of the Royal Statistical Society. Series B* **36** 61–73. MR0368365

Bhansali, R. J. (1977). Asymptotic properties of the Wiener-Kolmogorov predictor. II. *Journal of the Royal Statistical Society. Series B* **39** 66–72. MR0445748

Bickel, P. J. and Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates. *Ann. Statist.* **1** 1071–1095. MR0348906

Bickel, P. J. and Levina, E. (2008). Regularized estimation of large covariance matrices. *Ann. Statist.* **36** 199–227. MR2387969

Bierens, H. (1983). Uniform consistency of kernel estimators of a regression function under generalized conditions. *J. Amer. Statist. Assoc.* **78** 699–707. MR0721221

Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley, New York. MR0233396

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31** 307–327. MR0853051

Borovkova, S., Burton, R. M. and Dehling, H. (1999). Consistency of the Takens estimator for the correlation dimension. *Annals of Applied Probability* **9** 376–390. MR1687339

Borovkova, S., Burton, R. M. and Dehling, H. (2001). Limit theorems for functionals of mixing processes with applications to U -statistics and dimension estimation. *Transactions of the American Mathematical Society* **353** 4261–4318. MR1851171

Borovkova, S., Burton, R. M. and Dehling, H. (2002). From dimension estimation to asymptotics of dependent U-processes. In: *Limit Theorems in Probability and Statistics I* (I. Berkes, E. Csaki and M. Csorgo, eds.) Budapest 2002, 201–234. MR1979966

Borkar, V. S. (1993). White-noise representations in stochastic realization theory. *SIAM J. Control Optim.* **31** 1093–1102. MR1233993

Bosq, D. (1996). *Nonparametric Statistics for Stochastic Processes. Estimation and Prediction. Lecture Notes in Statist.* **110**. Springer, New York. MR1441072

Bougerol, P. and Picard, N. (1992). Stationarity of GARCH processes and of some nonnegative time series. *J. Econometrics* **52** 115–127. MR1165646

Box, G. E. P., Jenkins, G. M. and Reinsel, G. C. (1994). *Time Series Analysis: Forecasting and Control*. Prentice-Hall, NJ. MR1312604

Bradley, R. C. (2007). *Introduction to Strong Mixing Conditions*. Kendrick Press, Utah.

Brillinger, D. R. (1981). *Time Series: Data Analysis and Theory*, 2nd ed. Holden-Day, San Francisco. MR0595684

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*, 2nd ed., Springer, New York. MR1093459

Burkholder, D. L. (1988). Sharp inequalities for martingales and stochastic integrals. In *Colloque Paul Lévy sur les Processus Stochastiques* (Palaiseau, 1987). Astérisque No. 157–158, pp. 75–94. MR0976214

Carrasco, M. and Chen, X. (2002). Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory* **18** 17–39. MR1885348

Castellana, J. V. and Leadbetter, M. R. (1986). On smoothed probability density estimation. *Stochastic Process. Appl.* **21** 179–193. MR0833950

Chan, K. S. and Tong, H. (2001). *Chaos: A Statistical Perspective*. Springer, New York. MR1851668

Chen, M. and An. H. (1998). A note on the stationarity and existence of moments of the GARCH models. *Statistica Sinica* **8** 505–510. MR1624371

Chen, X. and Fan, Y. (2006). Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification. *J. Econometrics* **135**, no. 1–2, 125–154. MR2328398

Chow, Y. S. and Teicher, H. (1988). *Probability Theory*, 2nd ed. Springer, New York. MR0953964

Deák, I. (1990). *Random Numbers Generators and Simulation*. Akadémiai Kiadó, Budapest. MR1080965

Dedecker, J. P., Doukhan, G., Lang, J. R., Leon R., Louhichi, S. and Prieur, C. (2007). *Weak Dependence: With Examples and Applications*. Springer, New York. MR2338725

Dedecker, J. and Merlevède, F. (2002). Necessary and sufficient conditions for the conditional central limit theorem. *Ann. Probab.* **30** 1044–1081. MR1920101

Dedecker, J. and Merlevède, F., (2003). The conditional central limit theorem in Hilbert spaces. *Stochastic Process. Appl.* **108** 229–262. MR2019054

Dedecker, J., Merlevède, F. and Volný, D. (2007). On the weak invariance principle for non-adapted sequences under projective criteria. *J. Theoret. Probab.* **20** 971–1004. MR2359065

Dedecker, J. and Prieur, C. (2005). New dependence coefficients. Examples and applications to statistics. *Probability Theory and Related Fields* **132** 203–236. MR2199291

Dehling, H. and Wendler, M. (2010). Central limit theorem and the bootstrap for U-statistics of strongly mixing data. *Journal Multivariate Anal.* **101** 126–137. MR2557623

Denker, M. and Keller, G. (1983). On U-statistics and von Mises' statistic for weakly dependent processes. *Z. Wahrsch. verw. Gebiete* **64** 505–552. MR0717756

Denker, M. and Keller, G. (1986). Rigorous statistical procedures for data from dynamical systems. *Journal of Statistical Physics* **44** 67–93. MR0854400

Devroye, L. and Györfi, L. (1984). *Nonparametric Density Estimation: The L^1 View*. Wiley, New York. MR0780746

Diaconis, P. and Freedman, D. (1999). Iterated random functions. *SIAM Rev.* **41** 45–76. MR1669737

Ding, Z., Granger, C. and Engle, R. (1993). A long memory property of stock market returns and a new model. *J. Empirical Finance* **1** 83–106.

Doob, J. L. (1953). *Stochastic Processes*. Wiley, New York. MR0058896

Doukhan, P. and Louhichi, S. (1999). A new weak dependence condition and applications to moment inequalities. *Stochastic Process. Appl.* **84** 313–342. MR1719345

Doukhan, P. (1994). *Mixing: Properties and Examples*. Springer, New York. MR1312160

Doukhan, P. and Wintenberger, O. (2008). Weakly dependent chains with infinite memory. *Stochastic Process. Appl.* **118** 1997–2013. MR2462284

Duflo, M. (1997). *Random Iterative Models*. Springer-Verlag, Heidelberg Germany. MR1485774

Eberlein, E. (1986). On strong invariance principles under dependence assumptions. *Ann. Probab.* **14** 260–270. MR0815969

Eberlein, E. and Taqqu, M. (ed.) (1986). *Dependence in Probability and Statistics: A Survey of Recent Results*. Birkhauser, Boston. MR0899982

El Karoui, N. (2007). Tracy-Widom limit for the largest eigenvalue of a large class of complex sample covariance matrices. *Ann. Probab.* **35** 663–714. MR2308592

El Machkouri, M., Volný, D. and Wu, W. B. (2010). Central limit theorems for random fields. *Preprint.*

Elton, J. H. (1990). A multiplicative ergodic theorem for Lipschitz maps. *Stoc. Proc. Appl.* **34** 39–47. MR1039561

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50** 987–1007. MR0666121

Escanciano, J. C. and Javier, H. (2009). Persistence in Nonlinear Time Series: A Nonparametric Approach. *CAEPR Working Paper No. 2009-003.*

Eubank, R. (1999). *Nonparametric Regression and Spline Smoothing*, 2nd ed. Marcel Dekker, New York. MR1680784

Fan, J. (2005). A selective overview of nonparametric methods in financial econometrics (with discussion). *Statistical Science* **20** 317–357. MR2210224

Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer, New York. MR1964455

Gänssler, P. and Häusler, E. (1979). Remarks on the functional central limit theorem for martingales. *Z. Wahrsch. Verw. Gebiete* **50** 237–243. MR0554543

Geman, S. (1980). A limit theorem for the norm of random matrices. *Ann. Probab.* **8** 252–261. MR0566592

Golub, G. H. and Van Loan, C. F. (1989). *Matrix Computations*, 2nd ed. The Johns Hopkins University Press, Baltimore, Maryland. MR1002570

Gordin, M. I. (1969). The central limit theorem for stationary processes. (Russian) *Dokl. Akad. Nauk SSSR* **188** 739–741. MR0251785

Gordin, M. I. and Lifšic, B. A. (1978). Central limit theorem for stationary Markov processes. (Russian) *Dokl. Akad. Nauk SSSR* **239** 766–767. MR0501277

Götze, F. and Hipp, C. (1994). Asymptotic distribution of statistics in time series. *Ann. Statist.* **22** 2062–2088. MR1329183

Granger, C. W. J. (1995). Modelling nonlinear relationships between extended-memory variables. *Econometrica* **63** 265–279. MR1323523

Granger, C. W. J. and Anderson, A. P. (1978). *An Introduction to Bilinear Time Series Models*. Vandenhoek and Ruprecht, Gottinger. MR0483231

Grassberger, P. and Procaccia, I. (1983a). Measuring the Strangeness of Strange Attractors. *Physica D* **9** 189–208. MR0732572

Grassberger, P. and Procaccia, I. (1983b). Characterization of Strange Attractors. *Phys. Rev. Let.* **50** 346–349. MR0689681

Grenander, U. and Rosenblatt, M. (1984). *Statistical Analysis of Stationary Time Series*, 2nd ed. Chelsea Publishing Co., New York. MR0890514

Grenander, U. and Szegö, G. (1958). *Toeplitz Forms and Their Applications*. Cambridge University Press, London. MR0094840

Györfi, L., Härdle, W., Sarda, P. and Vieu, P. (1989). *Nonparametric Curve Estimation From Time Series*. Lecture Notes in Statist. 60, Springer, Berlin. MR1027837

Hall, P and C.C. Heyde (1980). *Martingale Limit Theorem and its Application*. Academic Press, New York. MR0624435

Hallin, M., Lu, Z. and Tran, L. T. (2001). Density estimation for spatial linear processes. *Bernoulli* **7** 657–668. MR1849373

Hallin, M., Lu, Z. and Tran, L. T. (2004). Kernel Density Estimation for Spatial Processes: The L_1 Theory. *Journal of Multivariate Analysis* **88** 61–75. MR2021860

Hannan, E. J. (1970). *Multiple Time Series*. Wiley, New York. MR0279952

Hannan, E. J. (1973). Central limit theorems for time series regression. *Z. Wahrsch. Verw. Gebiete* **26** 157–170. MR0331683

Hannan, E. J. (1976). The asymptotic distribution of serial covariances. *Ann. Statist.* **4** 396–399. MR0398029

Hannan, E. J. (1979). The central limit theorem for time series regression. *Stochastic Process. Appl.* **9** 281–289. MR0562049

Hannan, E. J. and Deistler, M. (1988). *The Statistical Theory of Linear Systems*. Wiley, New York. MR0940698

Härdle, W., Lutkepohl, H. and Chen, R. (1997). A review of nonparametric time series analysis. *Int. Stat. Rev.* **65** 49–72.

Harris, D., McCabe, B. and Leybourne, S. (2003). Some limit theory for autocovariances whose order depends on sample size. *Econometric Theory* **19** 829–864. MR2002575

He, C. and Teräsvirta, T. (1999). Fourth moment structure of the GARCH(p,q) process. *Econometric Theory* **15** 824–846.

Ho, H. C. and Hsing, T. (1996). On the asymptotic expansion of the empirical process of long-memory moving averages. *Ann. Statist.* **24** 992–1024. MR1401834

Hoeffding, W. (1961). The strong law of large numbers for U-statistics. *Mimeograph Series No. 302*, Department of Statistics, University of North Carolina at Chapel Hill.

Hoeffding, W. and Robbins, H. (1948). The central limit theorem for dependent random variables. *Duke Math. J.* **15** 773–780. MR0026771

Horn, R. A. and Johnson, C. R. (1990). *Matrix Analysis*. Corrected reprint of the 1985 original. Cambridge University Press, Cambridge, UK. MR1084815

Hosking, J. R. M. (1996). Asymptotic distributions of the sample mean, autocovariances, and autocorrelations of long-memory time series. *J. Econom.* **73** 261–284. MR1410007

Hsing, T. and Wu, W. B. (2004). On weighted U-statistics for stationary processes. *Annals of Probability* **32** 1600–1631. MR2060311

Ibragimov, I. A. (1962). Some limit theorems for stationary processes. *Theory of Probability and its Applications* **7** 349–382. MR0148125

Ibragimov, I. A. and Linnik, Yu. V. (1971). *Independent and stationary sequences of random variables*. Groningen, Wolters-Noordhoff. MR0322926

Jones, D. A. (1976). *Non-linear autoregressive processes*. Unpublished Ph.D. Thesis, University of London.

Jarner, S. and Tweedie, R. (2001). Locally contracting iterated random functions and stability of Markov chains. *J. Appl. Probab.* **38** 494–507. MR1834756

Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327. MR1863961

Jones, R. H. (1964). Spectral analysis and linear prediction of meteorological time series. *J. Appl. Meteor.* **3** 45–52.

Kalikow, S. A. (1982). T, T^{-1} transformation is not loosely Bernoulli. *Ann. Math.* **115** 393–409. MR0647812

Kallianpur, G. (1981). Some ramifications of Wiener's ideas on nonlinear prediction. In: *Norbert Wiener, Collected Works with Commentaries.* MIT Press, Mass., 402–424.

Keenan, D. M. (1997). A central limit theorem for m(n) autocovariances. *J. Time Ser. Anal.* **18** 61–78. MR1437742

Kolmogorov, A. (1941). Interpolation und Extrapolation von stationären zufälligen Folgen. *Bull. Acad. Sci. URSS Sér. Math. [Izvestia Akad. Nauk. SSSR]* **5** 3–14. MR0004416

Komlós, J., Major, P. and Tusnády, G. (1975). An approximation of partial sums of independent RV's and the sample DF. I. *Z. Wahrsch. Verw. Gebiete* **32** 111–131; II. (1976) **34** 33–58.

Koop, G., Pesaran, M. H. and Potter, S. M. (1996). Impulse response analysis in nonlinear multivariate models. *Journal of Econometrics* **74** 119–147. MR1409037

Li, D., Lu, Z. and Linton, O. (2010). Local Linear Fitting under Near Epoch Dependence: Uniform Consistency with Convergence Rates. Discussion paper, London School of Economics. Available at: http://personal.lse.ac.uk/lintono/downloads/Li-Lu-Linton-4.pdf.

Ling, S. (1999). On the stationarity and the existence of moments of conditional heteroskedastic ARMA models. *Statistica Sinica* **9** 1119–1130. MR1744828

Ling, S. and McAleer, M. (2002). Necessary and sufficient moment conditions for the GARCH (r, s) and asymmetric power GARCH(r, s) models. *Econometric Theory* **18** 722–729. MR1906332

Liu, W. and Lin, Z. (2009a). Strong approximation for a class of stationary processes. *Stochastic Processes and their Applications* **119** 249–280. doi:10.1016/j.spa.2008.01.012. MR2485027

Liu, W. and Lin, Z. (2009b). On maxima of periodograms of stationary processes. *Ann. Statist.* **37** 5B, 2676–2695. MR2541443

Liu, W. and Wu, W. B. (2010a). Simultaneous nonparametric inference of time series. *Ann. Statist.* **38** 2388–2421. MR2676893

Liu, W. and Wu, W. B. (2010b). Asymptotics of spectral density estimates. *Econometric Theory* **26** 1218–1245.

Lu, Z. (2001). Asymptotic normality of kernel density estimators under dependence. *Annals of the Institute of Statistical Mathematics* **53** 447–468. MR1868884

Lu, Z. and Linton, O. (2007). Local linear fitting under near epoch dependence. *Econometric Theory* **23** 37–70. MR2338951

Major, P. (1994). Asymptotic distributions for weighted U-statistics. *Ann. Probab. 22*, 1514–1535. MR1303652

Maxwell, M. and Woodroofe, M. (2000). Central limit theorems for additive functionals of Markov chains. *Ann. Probab.* **28** 713–724. MR1782272

Morgan, J. P. (1996). *RiskMetrics*. Technical Document. New York.

Nadaraya, E. A. (1989). *Nonparametric Estimation of Probability Densities and Regression Curves*. Kluwer Academic Pub. MR1093466

Neumann, M.H. (1998). Strong approximation of density estimators from weakly dependent observations by density estimators from independent observations. *Ann. Statist.* **26** 2014–2048. MR1673288

Olshen, R. A. (1967). Asymptotic properties of the periodogram of a discrete stationary process. *J. Appl. Probab.* **4** 508–528. MR0228059

O'Neil, K. A. and Redner, R. A. (1993). Asymptotic Distributions of Weighted U-Statistics of Degree 2. *Annals of Probability* **21** 1159–1169. MR1217584

Ornstein, D. S. (1973). An example of a Kolmogorov automorphism that is not a Bernoulli shift. *Advances in Math.* **10** 49–62. MR0316682

Peligrad, M. (1996). On the asymptotic normality of sequences of weak dependent random variables. *J. Theoret. Probab.* **9** 703–715. MR1400595

Peligrad, M. and Utev, S. (2005). A new maximal inequality and invariance principle for stationary sequences. *Ann. Probab.* **33** 798–815. MR2123210

Peligrad, M. and Wu, W. B. (2010). Central limit theorem for Fourier transform of stationary processes. *Annals of Probability* **38** 2009–2022. MR2722793

Pham, D. T. (1985). Bilinear Markovian representation and bilinear models. *Stochastic Process. Appl.* **20** 295–306. MR0808163

Pham, D. T. (1986). The mixing property of bilinear and generalised random coefficient autoregressive models. *Stochastic Process. Appl.* **23** 291–300. MR0876051

Pham, D. T. (1993). Bilinear time series models. In *Dimension Estimation and Models* (H. Tong, ed.). World Scientific, Singapore. MR1307660

Philipp, W. and Stout, W. (1975). Almost sure invariance principles for partial sums of weakly dependent random variables. *Mem. Am. Math. Soc.* **161** 1–140. MR0433597

Phillips, P. C. B. and Solo, V. (1992). Asymptotics for linear processes. *Ann. Statist.* **20** 971–1001. MR1165602

Pourahmadi, M. (2001). *Foundations of Time Series Analysis and Prediction Theory*. Wiley, New York. MR1849562

Prakasa Rao, B. L. S. (1983). *Nonparametric Functional Estimation*. Academic Press, New York. MR0740865

Priestley, M. B. (1981). *Spectral Analysis and Time Series 1*. Academic Press. MR0628735

Priestley, M. B. (1988). *Nonlinear and Nonstationary Time Series Analysis*. Academic Press. MR0991969

Quinn, B. G. (1982). A note on the existence of strictly stationary solutions to bilinear equations. *J. Time Ser. Anal.* **3** 249–252. MR0703088

Rifi, M. and Utzet, F. (2000). On the asymptotic behavior of weighted U-statistics. *J. Theor. Probab.* **13** 141–167. MR1744988

Rio, E. (2000). Théorie asymptotique des processus aléatoires faiblement dépendants. *Mathématiques & Applications* **31**. Springer-Verlag, Berlin. MR2117923

Rio, E. (2009). Moment inequalities for sums of dependent random variables under projective conditions. *Journal of Theoretical Probability* **22** 146–163. MR2472010

Robinson, P. M. (1983). Nonparametric estimators for time series. *J. Time Ser. Anal.* **4** 185–207. MR0732897

Rootzén, H. (1976). Gordin's theorem and the periodogram. *J. Appl. Probab.* **13** 365–370. MR0410876

Rosenblatt, M. (1961). Independence and dependence, In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob. vol II*, Univ. California Press, Berkeley, CA, pp. 431–443. MR0133863

Rosenblatt, M. (1952). Remarks on a multivariate transformation. *Ann. Math. Statist.* **23** 470–472. MR0049525

Rosenblatt, M. (1956a). A central limit theorem and a strong mixing condition. *Proc. Nat. Acad. Sci. U. S. A.* **42** 43–47. MR0074711

Rosenblatt, M. (1956b). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* **27** 832–837. MR0079873

Rosenblatt, M. (1985). *Stationary Sequences and Random Fields*. Birkhäuser, Boston. MR0885090

Rosenblatt, M. (2009). A comment on a conjecture of N. Wiener. *Statist. Probab. Letters* **79** 347–348. MR2493017

Serinko, R. J. (1994). A consistent approach to least squares estimation of correlation dimension in weak Bernoulli dynamical systems. *Annals of Appl. Probab.* **4** 1234–1254. MR1304784

Shao, X. and Wu, W. B. (2007). Asymptotic spectral theory for nonlinear time series. *Ann. Statist.* **35** 1773–1801. MR2351105

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London. MR0848134

Singh, R. S. and Ullah, A. (1985). Nonparametric time-series estimation of joint DGP, conditional DGP, and vector autoregression. *Econometric Theory* **1** 27–52.

Steinsaltz, D. (1999). Locally contractive iterated function systems. *Ann. Probab.* **27** 1952–1979. MR1742896

Subba Rao, T. and Gabr, M. M. (1984). *An Introduction to Bispectral Analysis and Bilinear Time Series Models*. Lecture Notes in Statistics, 24. Springer-Verlag, New York. MR0757536

Surgailis, D. (1982). Zones of attraction of self-similar multiple integrals. *Lithuanian Mathematical Journal* **22** 327–340. MR0684472

Terrin, N. and Hurvich, C. M. (1994). An asymptotic Wiener-Ito representation for the low frequency ordinates of the periodogram of a long memory time series. *Stochastic Process. Appl.* **54** 297–307. MR1307342

Tjøstheim, D. (1994). Nonlinear time series, a selective review. *Scand. J. Statist.* **21** 97–130.

Tong, H. (1981). A note on a Markov bilinear stochastic process in discrete time. *J. Time Ser. Anal.* **2** 279–284. MR0648732

Tong, H. (1983). *Threshold Models in Non-linear Time Series Analysis*. Springer-Verlag. MR0717388

Tong, H. (1990). *Non-linear Time Series: A Dynamic System Approach*. Oxford University Press, Oxford. MR1079320

Volný, D., Woodroofe, M. and Zhao, O. (2011). Central limit theorems for superlinear processes. *Stochastics and Dynamics* **11** 71–80

Volný, D. (1993). Approximating martingales and the central limit theorem for strictly stationary processes. *Stochastic Process. Appl.* **44** 41–74. MR1198662

Walker, A. M. (1965). Some asymptotic results for the periodogram of a stationary time series. *J. Austral. Math. Soc.* **5** 107–128. MR0177457

Walker, A. M. (2000). Some results concerning the asymptotic distribution of sample Fourier transforms and periodograms for a discrete-time stationary process with a continuous spectrum. *J. Time Ser. Anal.* **21** 95–109. MR1766176

Wand, M. P. and Jones, M. C. (1995). *Kernel Smoothing*. Chapman and Hall, London. MR1319818

Whittle, P. (1963). *Prediction and Regulation by Linear Least Squares Methods*. Van Nostrand, Princeton. MR0157416

WIENER, N. (1949). *Extrapolation, Interpolation and Smoothing of Stationary Time Series*. Wiley, New York.

WIENER, N. (1958). *Nonlinear Problems in Random Theory*. MIT Press, Cambridge, MA. MR0100912

WOLFF, R. C. L. (1990). A note on the behaviour of the correlation integral in the presence of a time series. *Biometrika* **77** 689–697. MR1086682

WOODROOFE, M. (1992). A central limit theorem for functions of a Markov chain with applications to shifts. *Stochastic Processes and Their Applications* **41** 33–44. MR1162717

WU, W. B. (2005a). Fourier transforms of stationary processes. *Proc. Amer. Math. Soc.* **133** 285–293. MR2086221

WU, W. B. (2005b). Nonlinear system theory: Another look at dependence. *Proc. Nat. Journal Acad. Sci. USA* **102** 14150–14154. MR2172215

WU, W. B. (2007). Strong invariance principles for dependent random variables. *Ann. Probab.* **35** 2294–2320. MR2353389

WU, W. B. (2008). An asymptotic theory for sample covariances of Bernoulli shifts. *Stochastic Processes and their Applications* **119** 453–467. MR2493999

WU, W. B., HUANG, Y. and HUANG, Y. (2010). Kernel estimation for time series: An asymptotic theory. *Stochastic Processes and their Applications* **120** 2412–2431.

WU, W. B., HUANG, Y. and ZHENG, W. (2010). Covariances estimation for long-memory processes. *Adv. in Appl. Probab.* **42** 137–157. MR2666922

WU, W. B. and MIELNICZUK, J. (2002). Kernel density estimation for linear processes. *Annals of Statistics* **30** 1441–1459. MR1936325

WU, W. B. and MIELNICZUK, J. (2010) A new look at measuring dependence. In *Dependence in Probability and Statistics*. pp. 123–142 (Ed. P. Doukhan, G. Lang, G. Teyssière, and D. Surgailis) (Lecture Notes in Statistics) Springer, New York.

WU, W. B. and MIN, W. (2005). On linear processes with dependent innovations. *Stochastic Process. Appl.* **115** 939–958. MR2138809

WU, W. B. and POURAHMADI, M. (2003). Nonparametric estimation of large covariance matrices of longitudinal data. *Biometrika* **90** 831–844. MR2024760 MR2024760

WU, W. B. and POURAHMADI, M. (2009). Banding sample covariance matrices of stationary processes. *Statist. Sinica* **19** 1755–1768. MR2589209

WU, W. B. and SHAO, X. (2004). Limit theorems for iterated random functions. *J. Appl. Probab.* **41** 425–436. MR2052582

WU, W. B. and ZHAO, Z. (2007). Inference of trends in time series. *Journal of the Royal Statistical Society, Series B* **69** 391–410. MR2323759

XIAO, H. and WU, W. B. (2010a). A single-pass algorithm for spectrum estimation with fast convergence. *IEEE Transactions on Information Theory, To Appear.*

XIAO, H. and WU, W. B. (2010b). Covariance matrix estimation for stationary time series. *Preprint.*

YAJIMA, Y. (1989). A central limit theorem of Fourier transforms of strongly dependent stationary processes. *J. Time Ser. Anal.* **10** 375–383. MR1038470

YIN, Y. Q., BAI, Z. D. and KRISHNAIAH, P. R. (1988). On the limit of the largest eigenvalue of the large-dimensional sample covariance matrix. *Probab. Theory Related Fields* **78** 509–521. MR0950344

YOSHIHARA, K. (1976). Limiting behavior of U-statistics for stationary, absolutely regular processes. *Z. Wahrsch. verw. Gebiete* **35** 237–252. MR0418179

YU, B. (1993). Density estimation in the L^∞ norm for dependent data with applications to the Gibbs sampler. *Ann. Statist.* **21** 711–735. MR1232514

Wei Biao Wu
Department of Statistics
The University of Chicago
USA
E-mail address: wbwu@galton.uchicago.edu

Nonlinear system theory: Another look at dependence

Wei Biao Wu†

Department of Statistics, University of Chicago, 5734 South University Avenue, Chicago, IL 60637

Communicated by Murray Rosenblatt, University of California at San Diego, La Jolla, CA, August 4, 2005 (received for review April 29, 2005)

Based on the nonlinear system theory, we introduce previously undescribed dependence measures for stationary causal processes. Our physical and predictive dependence measures quantify the degree of dependence of outputs on inputs in physical systems. The proposed dependence measures provide a natural framework for a limit theory for stationary processes. In particular, under conditions with quite simple forms, we present limit theorems for partial sums, empirical processes, and kernel density estimates. The conditions are mild and easily verifiable because they are directly related to the data-generating mechanisms.

nonlinear time series | limit theory | kernel estimation | weak convergence

Let ε_i, $i \in \mathbb{Z}$, be independent and identically distributed (iid) random variables and g be a measurable function such that

$$X_i = g(\ldots, \varepsilon_{i-1}, \varepsilon_i), \qquad [1]$$

is a properly defined random variable. Then (X_i) is a stationary process, and it is causal or nonanticipative in the sense that X_i does not depend on the future innovations ε_j, $j > i$. The causality assumption is quite reasonable in the study of time series. Wiener (1) considered the fundamental coding and decoding problem of representing stationary and ergodic processes in terms of the form Eq. **1**. In particular, Wiener studied the construction of ε_i based on X_k, $k \le i$. The class of processes that Eq. **1** represents is huge and it includes linear processes, Volterra processes, and many time series models. In certain situations, Eq. **1** is also called the nonlinear Wold representation. See refs. 2–4 for other deep contributions of representing stationary and ergodic processes by Eq. **1**. To conduct statistical inference of such processes, it is necessary to consider the asymptotic properties of the partial sum $S_n = \sum_{i=1}^n X_i$ and the empirical distribution function $F_n(x) = n^{-1} \sum_{i=1}^n \mathbf{1}_{X_i \le x}$.

In probability theory, many limit theorems have been established for independent random variables. Those limit theorems play an important role in the related statistical inference. In the study of stochastic processes, however, independence usually does not hold, and the dependence is an intrinsic feature. In an influential paper, Rosenblatt (5) introduced the strong mixing condition. For a stationary process (X_i), let the sigma algebra $\mathcal{A}_m^n = \sigma(X_m, \ldots, X_n)$, $m \le n$, and define the strong mixing coefficients

$$\alpha_n = \sup\{|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| : A \in \mathcal{A}_{-\infty}^0, B \in \mathcal{A}_n^\infty\}. \qquad [2]$$

If $\alpha_n \to 0$, then we say that (X_i) is strong mixing. Variants of the strong mixing condition include ρ, ψ, and β-mixing conditions among others (6). A central limit theorem (CLT) based on the strong mixing condition is proved in ref. 5. Since then, as basic assumptions on the dependence structures, the strong mixing condition and its variants have been widely used and various limit theorems have been obtained; see the extensive treatment in ref. 6.

Since the quantity $|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|$ in Eq. **2** measures the dependence between events A and B and it is zero if A and B are independent, it is sensible to call α_n and its variants "probabilistic dependence measures." For stationary causal processes, the calculation of probabilistic dependence measures is generally not easy because it involves the complicated manipulation of taking the supremum over two sigma algebras (7–9). Additionally, many well-known processes are not strong mixing. A prominent example is the Bernoulli shift process. Consider the simple AR(1) process $X_n = (X_{n-1} + \varepsilon_n)/2$, where ε_i are iid Bernoulli random variables with success probability 1/2 (see refs. 10 and 11). Then X_n is a causal process with the representation $X_n = \sum_{i=0}^\infty 2^{-i}\varepsilon_{n-i}$ and the innovations ε_n, ε_{n-1}, ..., correspond to the dyadic expansion of X_n. The process X_n is not strong mixing since $\alpha_n \equiv 1/4$ for all n (12). Some alternative ways have been proposed to overcome the disadvantages of strong mixing conditions (8, 9).

Dependence Measures

In this work, we shall provide another look at the fundamental issue of dependence. Our primary goal is to introduce "physical or functional" and "predictive dependence measures" a previously undescribed type of dependence measures that are quite different from strong mixing conditions. In particular, following refs. 1 and 13, we shall interpret Eq. **1** as an input/output system and then introduce dependence coefficients by measuring the degree of dependence of outputs on inputs. Specifically, we view Eq. **1** as a physical system

$$x_i = g(\ldots, e_{i-1}, e_i), \qquad [3]$$

where $e_i, e_{i-1}, \ldots$ are inputs, g is a filter or a transform, and x_i is the output. Then, the process X_i is the output of the physical system **3** with random inputs. It is clearly not a good way to assess the dependence just by taking the partial derivatives $\partial g/\partial e_j$, which may not exist if g is not well-behaved. Nonetheless, because the inputs are random and iid, the dependence of the output on the inputs can be simply measured by applying the idea of coupling. Let (ε_i') by an iid copy of (ε_i); let the shift process $\xi_i = (\ldots, \varepsilon_{i-1}, \varepsilon_i)$, $\xi_i' = (\ldots, \varepsilon_{i-1}', \varepsilon_i')$. For a set $I \subset \mathbb{Z}$, let $\varepsilon_{j,I} = \varepsilon_j'$ if $j \in I$ and $\varepsilon_{j,I} = \varepsilon_j$ if $j \notin I$; let $\xi_{i,I} = (\ldots, \varepsilon_{i-1,I}, \varepsilon_{i,I})$ and $\xi_i^* = \xi_{i,\{0\}}$. Then $\xi_{i,I}$ is a coupled version of ξ_i with ε_j replaced by ε_j' if $j \in I$. For $p > 0$ write $X \in \mathcal{L}^p$ if $\|X\|_p := [\mathbb{E}(|X|^p)]^{1/p} < \infty$ and $\|X\| = \|X\|_2$.

Definition 1 (Functional or physical dependence measure): For $p > 0$ and $I \subset \mathbb{Z}$ let $\delta_p(I, n) = \|g(\xi_n) - g(\xi_{n,I})\|_p$ and $\delta_p(n) = \|g(\xi_n) - g(\xi_n^*)\|_p$. Write $\delta(n) = \delta_2(n)$.

Definition 2 (Predictive dependence measure): Let $p \ge 1$ and g_n be a Borel function on $\mathbb{R} \times \mathbb{R} \times \ldots \mapsto \mathbb{R}$ such that $g_n(\xi_0) = \mathbb{E}(X_n|\xi_0)$, $n \ge 0$. Let $\omega_p(I, n) = \|g_n(\xi_0) - g_n(\xi_{0,I})\|_p$ and $\omega_p(n) = \|g_n(\xi_0) - g_n(\xi_0^*)\|_p$. Write $\omega(n) = \omega_2(n)$.

Definition 3 (p-stability): Let $p \ge 1$. The process (X_n) is said to be p-stable if $\Omega_p := \sum_{n=0}^\infty \omega_p(n) < \infty$, and p-strong stable if $\Delta_p := \sum_{n=0}^\infty \delta_p(n) < \infty$. If $\Omega = \Omega_2 < \infty$, we say that (X_n) is stable.

By the causal representation in Eq. **1**, if $\min\{i : i \in I\} > n$, then $\delta_p(I, n) = 0$. Apparently, $\delta_p(I, n)$ quantifies the dependence of $X_n = g(\xi_n)$ on $\{\varepsilon_i, i \in I\}$ by measuring the distance between $g(\xi_n)$ and its coupled version $g(\xi_{n,I})$. In *Definition 2*, $\mathbb{E}(X_n|\xi_0)$ is the n-step ahead predicated mean, and $\omega_p(n)$ measures the contribution of ε_0 in predicting future expected values. In the

Abbreviation: iid, independent and identically distributed.

†E-mail: wbwu@galton.uchicago.edu.

classical prediction theory (14), the conditional expectation of the form $\mathbb{E}(X_n|X_0, X_{-1}, \ldots)$ is studied. The one $\mathbb{E}(X_n|\xi_0)$ used in *Definition 2* has a different form. It turns out that, in studying asymptotic properties and moment inequalities of S_n, it is convenient to use $\mathbb{E}(X_n|\xi_0)$ and predictive dependence measure (*cf. Theorems 2* and *3*), while the other version $\mathbb{E}(X_n|X_0, X_{-1}, \ldots)$ is generally difficult to work with. In the special case in which X_n are martingale differences with respect to the filter $\sigma(\xi_n)$, $g_n = 0$ almost surely and consequently $\omega(n) = 0$, $n \geq 1$.

Roughly speaking, since $g_n(\xi_0) = \mathbb{E}(X_n|\xi_0)$, the p-stability in *Definition 3* indicates that the cumulative contribution of ε_0 in predicting future expected values $\{\mathbb{E}(X_n|\xi_0)\}_{n\geq 0}$ is finite. Interestingly, the stability condition $\Omega_2 < \infty$ implies invariance principles with $\sqrt{n}$-norming in a natural way (*Theorem 3*). By (*i*) of *Theorem 1*, p-strong stability implies p-stability since $\delta_p(n) \geq \omega_p(n)$.

Our dependence measures provide a very convenient and simple way for a large-sample theory for stationary causal processes (see *Theorems 2–5* below). In many applications, functional and predictive dependence measures are easy to use because they are directly related to data-generating mechanisms and because the construction of the coupled process $g(\xi_{n,I})$ is simple and explicit. Additionally, limit theorems with those dependence measures have easily verifiable conditions and are often optimal or nearly optimal. On the other hand, however, our dependence measures rely on the representation **1**, whereas the strong mixing coefficients can be defined in more general situations (6).

Theorem 1. *(i) Let $p \geq 1$ and $n \geq 0$. Then $\delta_p(n) \geq \omega_p(n)$. (ii) Let $p \geq 1$ and the projection operator $\mathcal{P}_k Z = \mathbb{E}(Z|\xi_k) - \mathbb{E}(Z|\xi_{k-1})$, $Z \in \mathcal{L}^p$. Then for $n \geq 0$,*

$$\|\mathcal{P}_0 X_n\|_p \leq \omega_p(n) \leq 2\|\mathcal{P}_0 X_n\|_p . \quad [4]$$

(iii) Let $p > 1$, $C_p = 18p^{3/2}(p-1)^{-1/2}$ if $1 < p < 2$, $C_p = \sqrt{2p}$ if $p \geq 2$; let $I \subset \mathbb{Z}$. Then

$$\delta_p^{p'}(I, n) \leq 2^{p'} C_p^{p'} \sum_{i \in I} \delta_p^{p'}(n-i), \quad \text{where } p' = \min(p, 2). \quad [5]$$

Proof: (*i*) Since $\xi_n^* = (\xi_{-1}, \varepsilon_0', \varepsilon_1, \ldots, \varepsilon_n)$,

$$\begin{aligned}&\mathbb{E}[g(\xi_n) - g(\xi_n^*)|\xi_{-1}, \varepsilon_0', \varepsilon_0]\\ &\quad = \mathbb{E}[g(\xi_n)|\xi_{-1}, \varepsilon_0] - \mathbb{E}[g(\xi_n^*)|\xi_{-1}, \varepsilon_0']\\ &\quad = g_n(\xi_0) - g_n(\xi_0^*),\end{aligned}$$

which by Jensen's inequality implies $\delta_p(n) \geq \omega_p(n)$. (*ii*) Since $\mathbb{E}[g(\xi_n)|\xi_{-1}] = \mathbb{E}[g_n(\xi_0)|\xi_{-1}]$ and ε_0' and (ε_i) are independent, we have $\mathbb{E}[g_n(\xi_0)|\xi_{-1}] = \mathbb{E}[g_n(\xi_0^*)|\xi_0]$ and inequality **4** follows from

$$\begin{aligned}\|\mathcal{P}_0 X_n\|_p &= \|\mathbb{E}[g_n(\xi_0) - g_n(\xi_0^*)|\xi_0]\|_p\\ &\leq \|g_n(\xi_0) - g_n(\xi_0^*)\|_p\\ &\leq \|g_n(\xi_0) - \mathbb{E}[g_n(\xi_0)|\xi_{-1}]\|_p\\ &\quad + \|\mathbb{E}[g_n(\xi_0)|\xi_{-1}] - g_n(\xi_0^*)\|_p\\ &= 2\|\mathcal{P}_0 X_n\|_p.\end{aligned}$$

(*iii*) For presentational clarity, let $I = \{\ldots, -1, 0\}$. For $i \leq 0$ let

$$\begin{aligned}D_i = D_{i,n} &= \mathbb{E}(X_n|\varepsilon_{i+1}, \varepsilon_{i+2}, \ldots, \varepsilon_n) - \mathbb{E}(X_n|\varepsilon_i, \ldots, \varepsilon_n)\\ &= \mathbb{E}[g(\xi_{n,\{i\}}) - g(\xi_n)|\varepsilon_i, \ldots, \varepsilon_n].\end{aligned}$$

Then $D_0, D_{-1}, \ldots$ are martingale differences with respect to the sigma algebras $\sigma(\varepsilon_i, \ldots, \varepsilon_n)$, $i = 0, -1, \ldots$. By Jensen's inequality, $\|D_i\|_p \leq \delta_p(n-i)$. Let $V = \sum_{i=-\infty}^{0} D_i^2$, $M = \sum_{i=-\infty}^{0} D_i$ and $\tilde{X}_n = \mathbb{E}(X_n|\varepsilon_1, \ldots, \varepsilon_n)$. Then $X_n - \tilde{X}_n = -M$ and

$$\delta_p(I, n) \leq \|X_n - \tilde{X}_n\|_p + \|\tilde{X}_n - g(\xi_{n,I})\|_p = 2\|M\|_p .$$

To show Eq. **5**, we shall deal with the two cases $1 < p < 2$ and $p \geq 2$ separately. If $1 < p < 2$, then $V^{p/2} \leq \sum_{i=-\infty}^{0} |D_i|^p$. By Burkholder's inequality (15)

$$\|M\|_p^p \leq C_p^p \|V^{1/2}\|_p^p \leq C_p^p \sum_{i=-\infty}^{0} \delta_p^p(n-i).$$

If $p \geq 2$, by proposition 4 in ref. 16, $\|M\|_p^2 \leq 2p \sum_{i=-\infty}^{0} \|D_i\|_p^2$. So Eq. **5** follows.

Inequality **5** suggests the interesting reduction property: the degree of dependence of X_n on $\{\varepsilon_i, i \in I\}$ can be bounded in an element-wise manner, and it suffices to consider the dependence of X_n on individual ε_i. Indeed, our limit theorems and moment inequalities in *Theorems 2–5* involve conditions only on $\delta_p(n)$ and $\omega_p(n)$.

Linear Processes. Let ε_i be iid random variables with $\varepsilon_i \in \mathcal{L}^p$, $p \geq 1$; let (a_i) be real coefficients such that

$$X_t = \sum_{i=0}^{\infty} a_i \varepsilon_{t-i}, \quad [6]$$

is a proper random variable. The existence of X_t can be checked by Kolmogorov's three series theorem. The linear process (X_t) can be viewed as the output from a linear filter and the input $(\ldots, \varepsilon_{t-1}, \varepsilon_t)$ is a series of shocks that drive the system (ref. 17, pp. 8–9). Clearly, $\omega_p(n) = \delta_p(n) = |a_n| c_0$, where $c_0 = \|\varepsilon_0 - \varepsilon_0'\|_p < \infty$. Let $p = 2$. If

$$\sum_{i=0}^{\infty} |a_i| < \infty, \quad [7]$$

then the filter is said to be stable (17) and the preceding inequality implies short-range dependence since the covariances are absolutely summable. *Definition 3* extends the notion of stability to nonlinear processes.

Volterra Series. Analysis of nonlinear systems is a notoriously difficult problem, and the available tools are very limited (18). Oftentimes it would be unsatisfactory to linearize or approximate nonlinear systems by linear ones. The Volterra representation provides a reasonably simple and general way. The idea is to represent Eq. **3** as a power series of inputs. In particular, suppose that g in Eq. **3** is sufficiently well-behaved so that it has the stationary and causal representation

$$\begin{aligned}&g(\ldots, e_{n-1}, e_n)\\ &\quad = \sum_{k=1}^{\infty} \sum_{u_1, \ldots, u_k = 0}^{\infty} g_k(u_1, \ldots, u_k) e_{n-u_1} \ldots e_{n-u_k}, \quad [8]\end{aligned}$$

where functions g_k are called the Volterra kernel. The right-hand side of Eq. **8** is generically called the Volterra expansion, and it plays an important role in the nonlinear system theory (13, 18–22). There is a continuous-time version of Eq. **8** with summations replaced by integrals. Because the series involved has infinitely many terms, to guarantee the meaningfulness of the representation, there is a convergence issue that is often difficult to deal with, and the imposed conditions can be quite restrictive (18). Fortunately, in our setting, the difficulty can be circumvented because we are dealing with iid random inputs. Indeed, assume that e_t are iid with mean 0, variance 1 and $g_k(u_1, \ldots, u_k)$ is symmetric in $u_1, \ldots, u_k$ and it equals zero if $u_i = u_j$ for some $1 \leq i < j \leq k$, and

$$\sum_{k=1}^{\infty} \sum_{u_1,\ldots,u_k=0}^{\infty} g_k^2(u_1, \ldots, u_k) < \infty.$$

Then X_n exists and is in $\mathcal{L}^2$. Simple calculations show that

$$\frac{\omega^2(n)}{2} = \sum_{k=1}^{\infty} \sum_{\min(u_1,\ldots u_k)=n} g_k^2(u_1, \ldots, u_k)$$

$$= \sum_{k=1}^{\infty} k \sum_{u_2,\ldots u_k=n+1}^{\infty} g_k^2(n, u_2, \ldots, u_k),$$

and

$$\frac{\delta^2(n)}{2} = \sum_{k=1}^{\infty} k \sum_{u_2,\ldots u_k=0}^{\infty} g_k^2(n, u_2, \ldots, u_k).$$

The Volterra process is stable if $\sum_{i=1}^{\infty} \omega(i) < \infty$.

Nonlinear Transforms of Linear Processes. Let (X_t) be the linear process defined in Eq. **6** and consider the transformed process $Y_t = K(X_t)$, where K is a possibly nonlinear filter. Let $\omega(n, Y)$ be the predictive dependence measure of (Y_t). Assume that ε_i have mean 0 and finite variance. Under mild conditions on K, we have $\|\mathcal{P}_0 Y_n\| = O(|a_n|)$ (*cf.* theorem 2 in ref. 23). By *Theorem 1*, $\omega(n, Y) = O(|a_n|)$. In this case, if (X_t) is stable, namely Eq. **7** holds, then (Y_t) is also stable.

Quite interesting phenomena happen if (X_n) is unstable. Under appropriate conditions on K, (Y_n) could possibly be stable. With a nonlinear transform, the dependence structure of (Y_t) can be quite different from that of (X_n) (24–27). The asymptotic problem of $S_n(K) = \sum_{t=1}^{n} K(X_t)$ has a long history (see refs. 23 and 27 and references therein). Let $K_\infty(w) = \mathbb{E}[K(w + X_t)]$ and assume $K_\infty \in \mathcal{C}^\tau(\mathbb{R})$ for some $\tau \in \mathbb{N}$. Consider the remainder of the τ-th order Volterra expansion of Y_n

$$L^{(\tau)}(\xi_n) = Y_n - \sum_{r=0}^{\tau} \kappa_r U_{n,r}, \qquad [9]$$

where $\kappa_r = K_\infty^{(r)}(0)$, $r = 0, \ldots, \tau$, and

$$U_{n,r} = \sum_{0 \leq j_1 < \ldots < j_r < \infty} \prod_{s=1}^{r} a_{j_s} \varepsilon_{n-j_s}.$$

Let $\theta_n = |a_{n-1}|[|a_{n-1}| + A_n^{1/2}(4) + A_n^{\tau/2}(2)]$ and $A_n(j) = \sum_{t=n}^{\infty} |a_t|^j$. Under mild regularity conditions on K and ε_n, by theorem 5 in ref. 23, $\|\mathcal{P}_0 L^{(\tau)}(\xi_n)\| = O(\theta_{n+1})$. By *Theorem 1*, the predictive dependence measure $\omega^{(\tau)}(n)$ of the remainder $L^{(\tau)}(\xi_n)$ satisfies

$$\omega^{(\tau)}(n) = O(\theta_{n+1}). \qquad [10]$$

It is possible that $\sum_{n=1}^{\infty} \theta_n < \infty$ while $\sum_{n=1}^{\infty} |a_n| = \infty$. Consider the special case $a_n = n^{-\beta} l(n)$, where $1/2 < \beta < 1$ and l is a slowly varying function, namely, for any $c > 0$. $l(cn)/l(n) \to 1$ as $n \to \infty$. By Karamata's Theorem (28) for $j \geq 2$, $A_n(j) = O[n^{1-\beta j} l^j(n)]$. If $\tau > (2\beta - 1)^{-1} - 1$, then $\theta_n = O[n^{\tau(1/2-\beta)} l^\tau(n)]$ is summable. Therefore, if the function K satisfies $\kappa_r = 0$ for $r = 0, \ldots, \tau$ and $(\tau + 1)(2\beta - 1) > 1$, then $Y_t = K(X_t)$ is stable even though X_t is not. Appell polynomials (29) satisfy such conditions. For example, let $K(x) = x^2 - \mathbb{E}(X_n^2)$, then $K_\infty(w) = w^2$ and $\kappa_1 = 0$, $\kappa_2 = 2$. If $\beta \in (3/4, 1)$, then the process $X_t^2 - \mathbb{E}(X_t^2)$ is stable. If $1/2 < \beta < 3/4$, then $S_n(K)/\|S_n(K)\|$ converges to the Rosenblatt distribution.

Uniform Volterra expansions for $F_n(x)$ over $x \in \mathbb{R}$ are established in refs. 30 and 31. Wu (32) considered nonlinear transforms of linear processes with infinite variance innovations.

Nonlinear Time Series. Let ε_t be iid random variables and consider the recursion

$$X_t = R(X_{t-1}, \varepsilon_t), \qquad [11]$$

where R is a measurable function. The framework **11** is quite general, and it includes many popular nonlinear time series models, such as threshold autoregressive models (33), exponential autoregressive models (34), bilinear autoregressive models, autoregressive models with conditional heteroscedasticity (35), among others. If there exists $\alpha > 0$ and x_0 such that

$$\mathbb{E}(\log L_\varepsilon) < 0 \quad \text{and} \quad L_{\varepsilon_0} + |R(x_0, \varepsilon_0)| \in \mathcal{L}^\alpha, \qquad [12]$$

where

$$L_\varepsilon = \sup_{x \neq x'} \frac{|R(x, \varepsilon) - R(x', \varepsilon)|}{|x - x'|},$$

then Eq. **11** admits a unique stationary distribution (36), and iterations of Eq. **11** give rise to Eq. **1**. By theorem 2 in ref. 37, Eq. **12** implies that there exists $p > 0$ and $r \in (0, 1)$ such that

$$\|X_n - g(\xi_{n,I})\|_p = O(r^n), \qquad [13]$$

where $I = \{\ldots, -1, 0\}$. Recall $\xi_n^* = \xi_{n,\{0\}}$. By stationarity, $\|g(\xi_n^*) - g(\xi_{n,I})\|_p = \|g(\xi_{n+1}) - g(\xi_{n+1,I})\|_p$. So Eq. **13** implies $\delta_p(n) = \|g(\xi_n^*) - X_n\|_p = O(r^n)$. On the other hand, by *Theorem 1 (iii)*, if $\delta_p(n) = O(r^n)$ holds for some $p > 1$ and for some $r \in (0,1)$, then Eq. **13** also holds. So they are equivalent if $p > 1$. In refs. 37 and 38, the property **13** is called geometric-moment contraction, and it is very useful in studying asymptotic properties of nonlinear time series.

Inequalities and Limit Theorems

For (X_i) defined in Eq. **1**, let $S_u = S_n + (u - n)X_{n+1}$, $n \leq u \leq n + 1$, $n = 0, 1, \ldots$, be the partial sum process. Let $R_n(s) = \sqrt{n}[F_n(s) - F(s)]$, where $F(s) = \mathbb{P}(X_0 \leq s)$ is the distribution function of X_0. Primary goals in the limit theory of stationary processes include obtaining asymptotic properties of $\{S_u, 0 \leq u \leq n\}$ and $\{R_n(s), s \in \mathbb{R}\}$. Such results are needed in the related statistical inference. The physical and predictive dependence measures provide a natural vehicle for an asymptotic theory for S_n and R_n.

Partial Sums. Let $S_n^* = \max_{i \leq n} |S_i|$, $Z_n = S_n^*/\sqrt{n}$ and $B_p = p\sqrt{2p}/(p - 1)$, $p > 1$. Recall $\Omega_p = \sum_{k=0}^{\infty} \omega_p(k)$ and let

$$\Theta_p = \sum_{k=0}^{\infty} \|\mathcal{P}_0 X_k\|_p.$$

By *Theorem 1*, $\Theta_p \leq \Omega_p \leq 2\Theta_p$. Moment inequalities and limit theorems of S_n are given in *Theorems 2* and *3*, respectively. Denote by *IB* the standard Brownian motion. An interesting feature in the large deviation result in *Theorem 2(ii)* is that Ω_p and X_k do not need to be bounded.

Theorem 2. *Let $p \geq 2$. (i) We have $\|Z_n\|_p \leq B_p\Theta_p \leq B_p\Omega_p$. (ii) Let $0 < \alpha \leq 2$ and assume*

$$\gamma := \limsup_{p\to\infty} p^{1/2-1/\alpha}\Omega_p < \infty. \quad [14]$$

Then $m(t) := \sup_{n\in\mathbb{N}}\mathbb{E}[\exp(tZ_n^\alpha)] < \infty$ for $0 \leq t < t_0$, where $t_0 = (e\alpha\gamma^\alpha)^{-1}2^{-\alpha/2}$. Consequently, for $u > 0$, $\mathbb{P}(Z_n > u) \leq \exp(-tu^\alpha)m(t)$.

Proof: (*i*) It follows from W.B.W. (unpublished results) and theorem 2.5 in ref. 39. For completeness we present the proof here. Let $M_{k,j} = \sum_{i=1}^{j} \mathcal{P}_{i-k}X_i$, $k, j \geq 0$ and $M^*_{k,n} = \max_{j\leq n}|M_{k,j}|$. Then $S_n = \sum_{k=0}^{\infty} M_{k,n}$. By Doob's maximal inequality and theorem 2.5 in ref. 39 (or proposition 4 in ref. 16),

$$\|M^*_{k,n}\|_p \leq p(p-1)^{-1}\|M_{k,n}\|_p \leq B_p\sqrt{n}\|M_{k,1}\|_p.$$

Since $S_n^* \leq \sum_{k=0}^{\infty} M^*_{k,n}$, (*i*) follows. (*ii*) Let $Z = Z_n$ and $p_0 = [2/\alpha] + 1$. By Stirling's formula and Eq. **14**

$$\limsup_{p\to\infty} \frac{tB^\alpha_{\alpha p}\Omega^\alpha_{\alpha p}}{(p!)^{1/p}} = \limsup_{p\to\infty} \frac{tB^\alpha_{\alpha p}\Omega^\alpha_{\alpha p}}{(2\pi p)^{1/(2p)}p/e} = te\alpha\gamma^\alpha 2^{\alpha/2} < 1.$$

By (*i*), since $e^v = \sum_{p=0}^{\infty} v^p/(p!)$, (*ii*) follows from

$$\sum_{p=p_0}^{\infty} \frac{\mathbb{E}[(tZ^\alpha)^p]}{p!} \leq \sum_{p=p_0}^{\infty} \frac{t^p(B_{\alpha p}\Omega_{\alpha p})^{\alpha p}}{p!} < \infty.$$

Example 1: For the linear process **6**, assume that

$$\#\{i:|a_i| > \eta\} = O(\eta^{-1/2}) \quad \text{as} \quad \eta \downarrow 0, \quad [15]$$

and $A := \mathbb{E}(e^{|\varepsilon_0|}) < \infty$. We now apply (*ii*) of *Theorem 2* to the sum $n[F_n(u) - F(u)] = \sum_{i=1}^{n} \tilde{g}(\xi_i)$, where $\tilde{g}(\xi_i) = \mathbf{1}_{X_i\leq u} - F(u)$. To this end, we need to calculate the predictive dependence measure $\omega_p(n, \tilde{g})$ (say) of the process $\tilde{g}(\xi_n)$. Without loss of generality let $a_0 = 1$. Let F_ε and f_ε be the distribution and density functions of ε_0 and assume $c := \sup_u f_\varepsilon(u) < \infty$. Then Eq. **14** holds with $\alpha = 1$. To see this, let $Y_{n-1} = X_n - \varepsilon_n$, $Z_{n-1} = Y_{n-1} - a_n\varepsilon_0$ and $Y^*_{n-1} = Z_{n-1} + a_n\varepsilon_0'$. Let $n \geq 1$. Then $\mathbb{E}(\mathbf{1}_{X_n\leq u}|\xi_0) = \mathbb{E}[F_\varepsilon(u - Y_{n-1})|\xi_0]$ and $\mathbb{E}[F_\varepsilon(u - Z_{n-1})|\xi_0^*] = \mathbb{E}[F_\varepsilon(u - Z_{n-1})|\xi_0]$. By the triangle inequality,

$$\begin{aligned} Q_n &:= |\mathbb{E}[F_\varepsilon(u - Y_{n-1})|\xi_0] - \mathbb{E}[F_\varepsilon(u - Y^*_{n-1})|\xi^*_0]| \\ &\leq |\mathbb{E}[F_\varepsilon(u - Y_{n-1})|\xi_0] - \mathbb{E}[F_\varepsilon(u - Z_{n-1})|\xi_0]| \\ &\quad + |\mathbb{E}[F_\varepsilon(u - Z_{n-1})|\xi^*_0] - \mathbb{E}[F_\varepsilon(u - Y^*_{n-1})|\xi^*_0]| \\ &\leq \mathbb{E}[c|Y_{n-1} - Z_{n-1}\|\xi_0] + \mathbb{E}[c|Z_{n-1} - Y^*_{n-1})|\xi^*_0] \\ &= c|a_n|(|\varepsilon_0| + |\varepsilon_0'|). \end{aligned}$$

Hence, $\omega_p(n, \tilde{g}) = \|Q_n\|_p \leq 2c|a_n|\|\varepsilon_0\|_p$. Since $A = \mathbb{E}(e^{|\varepsilon_0|})$, we have $\mathbb{E}(|\varepsilon_0|^p) \leq p!A$, $\|\varepsilon_0\|_p \leq pA^{1/p}$. Clearly, $0 \leq Q_n \leq 1$. So $\omega_p(n, \tilde{g}) \leq \min(1, C|a_n|p)$, where $C = 2cA$. For $\eta > 0$ let the set $J(\eta) = \{i \geq 0 : \eta/2 \leq |a_i| < \eta\}$. By Eq. **15**

$$\begin{aligned} \Omega_p &\leq \sum_{i=0}^{\infty} \min(1, C|a_i|p) \\ &= \sum_{i:|a_i|\geq p^{-1}} \min(1, C|a_i|p) \\ &\quad + \sum_{k=0}^{\infty} \sum_{i\in J((p2^k)^{-1})} \min(1, C|a_i|p) \\ &= O(\sqrt{p}) + \sum_{k=0}^{\infty} O[(p2^{k+1})^{1/2}(p2^k)^{-1}Cp] \\ &= O(\sqrt{p}). \end{aligned}$$

Condition **15** holds if $a_i = O(i^{-2})$.

Theorem 3. *(i) Assume that $\Omega_2 < \infty$. Then*

$$\{S_{nt}/\sqrt{n}, 0 \leq t \leq 1\} \Rightarrow \{\sigma IB(t), 0 \leq t \leq 1\}, \quad [16]$$

where $\sigma = \|\sum_{i=0}^{\infty} \mathcal{P}_0X_i\| \leq \Omega_2$. (ii) Let $2 < p \leq 4$ and assume that $\sum_{i=0}^{\infty} i\delta_p(i) < \infty$. Then on a possibly richer probability space, there exists a Brownian motion IB such that

$$\sup_{u\in[0,n]} |S_u - \sigma IB(u)| = O[n^{1/p}l(n)] \text{ almost surely}, \quad [17]$$

where $l(n) = (\log n)^{1/2+1/p}(\log\log n)^{2/p}$.

The proof of the strong invariance principle (*ii*) is given by W.B.W. (unpublished results). *Theorem 3*(*i*) follows from corollary 3 in ref. 40, and the expression $\sigma = \|\sum_{i=0}^{\infty} \mathcal{P}_0X_i\|$ is a consequence of the martingale approximation: let $D_k = \sum_{i=k}^{\infty} \mathcal{P}_kX_i$ and $M_n = D_1 + \ldots + D_n$, then $\|S_n - M_n\| = o(\sqrt{n})$ and $\|S_n\|/\sqrt{n} = \sigma + o(1)$ (see theorem 6 in ref. 41). *Theorem 3*(*i*) also can be proved by using the argument in ref. 42. The invariance principle in the latter paper has a slightly different form. We omit the details. See refs. 43 and 44 for some related works.

Empirical Distribution Functions. Let $H_i(u|\xi_0) = \mathbb{P}(X_i \leq u|\xi_0)$, $u \in \mathbb{R}$, be the conditional distribution function of X_i given ξ_0. By *Definition 2*, the predictive dependence measure for $\tilde{g}(\xi_i) = \mathbf{1}_{X_i\leq u} - F(u)$, at a fixed u, is $\|H_i(u|\xi_0) - H_i(u|\xi^*_0)\|_p$. To study the asymptotic properties of R_n, it is certainly necessary to consider the whole range $u \in (-\infty, \infty)$. To this end, we introduce the integrated predictive dependence measure

$$\phi_p^{(j)}(i) = \left[\int_{\mathbb{R}} \|H_i^{(j)}(u|\xi_0) - H_i^{(j)}(u|\xi^*_0)\|_p^p du\right]^{1/p}, \quad [18]$$

and the uniform predictive dependence measure

$$\varphi_p^{(j)}(i) = \sup_u \|H_i^{(j)}(u|\xi_0) - H_i^{(j)}(u|\xi^*_0)\|_p, \quad [19]$$

where $H_i^{(j)}(u|\xi_0) = \partial^j H_i(u|\xi_0)/\partial u^j$, $j = 0, 1, \ldots, i \geq 1$. Let $h_i(t|\xi_0) = H_i^{(1)}(t|\xi_0)$. *Theorem 4* below concerns the weak convergence of R_n based on $\phi_2^{(j)}(i)$. It follows from corollary 1 by W.B.W. (unpublished results).

Theorem 4. *Assume that $X_1 \in \mathcal{L}^\tau$ and $\sup_u h_1(u|\xi_0) \leq c_0$ for some positive constants τ, $c_0 < \infty$. Further assume that*

$$\sum_{i=1}^{\infty} [\phi_2^{(0)}(i) + \phi_2^{(1)}(i) + \phi_2^{(2)}(i)] < \infty. \quad [20]$$

Then $R_n \Rightarrow \{W(s), s \in \mathbb{R}\}$, where W is a centered Gaussian process.

Kernel Density Estimation. An important problem in nonparametric inference of stochastic processes is to estimate the marginal density function f (say) given the data $X_1, \ldots, X_n$. A popular method is the kernel density estimation (45,46). Let K be a bounded kernel function for which $\int_{\mathbb{R}} K(u)du = 1$ and $b_n > 1$ be a sequence of bandwidths satisfying

$$b_n \to 0 \quad \text{and} \quad nb_n \to \infty. \qquad \mathbf{[21]}$$

Let $K_b(x) = K(x/b)$. Then f can be estimated by

$$f_n(x) = \frac{1}{nb_n} \sum_{i=1}^{n} K_{b_n}(x - X_i). \qquad \mathbf{[22]}$$

If X_i are iid, Parzen (46) proved a central limit theorem for $f_n(x) - \mathbb{E}[f_n(x)]$ under the natural condition **21**. There has been a substantial literature on generalizing Parzen's result to time series (47,48). Wu and Mielniczuk (49) solved the open problem that, for short-range dependent linear processes, Parzen's central limit theorem holds under Eq. **21**. See references therein for historical developments. Here, we shall generalize the result in ref. 49 to nonlinear processes. To this end, we shall adopt the uniform predictive dependence measure **19**. The asymptotic normality of f_n requires a summability condition of $\varphi^{(1)}(k) = \sup_t \|h_k(t|\xi_0) - h_k(t|\xi_0^*)\|$.

Theorem 5. *Assume that* $\sup_u h_1(u|\xi_0) \le c_0$ *for some constant* $c_0 < \infty$ *and that* $f = F'$ *is continuous. Let* $\kappa := \int_{\mathbb{R}} K^2(u)du < \infty$. *Then under Eq.* **21** *and*

$$\sum_{k=1}^{\infty} \varphi^{(1)}(k) < \infty, \qquad \mathbf{[23]}$$

we have $\sqrt{nb_n}\{f_n(x) - \mathbb{E}[f_n(x)]\} \Rightarrow N[0, f(x)\kappa]$ *for every* $x \in \mathbb{R}$.

Proof: Let m be a nonnegative integer. By the identity $\mathbb{E}[\mathbb{P}(X_{m+1} \le u|\xi_m)|\xi_0] = \mathbb{P}(X_{m+1} \le u|\xi_0)$ and the Lebesgue dominated convergence theorem, we have $\mathbb{E}[h_1(u|\xi_m)|\xi_0| = h_{m+1}(u|\xi_0)$ and h_{m+1} is also bounded by c_0. By *Theorem 1(ii)*, $\|\mathcal{P}_0 h_1(u|\xi_m)\| \le \varphi^{(1)}(m+1)$. Let $A_n(u) = \sum_{i=1}^n h_1(u|\xi_{i-1}) - nf(u)$. By *Theorem 2(i)* and Eq. **23**

$$\frac{\sup_u \|A_n(u)\|}{B_2\sqrt{n}} \le \sup_u \sum_{m=0}^{\infty} \|\mathcal{P}_0 h_1(u|\xi_m)\| < \infty.$$

Let $M_n = \sum_{i=1}^n \mathcal{P}_i K_{b_n}(x - X_i)$ and $N_n = \int_{\mathbb{R}} K(v) A_n(x - b_n v)dv$. Observe that

$$\mathbb{E}[K_{b_n}(x - X_i)|\xi_{i-1}] = b_n \int_{\mathbb{R}} K(v) h_1(x - b_n v|\xi_{i-1})dv.$$

Then $nb_n\{f_n(x) - \mathbb{E}[f_n(x)]\} = M_n + b_n N_n$. Following the argument of lemma 2 in ref. 49, $M_n/\sqrt{nb_n} \Rightarrow N[0, f(x)\kappa]$, which finishes the proof since $\mathbb{E}|N_n| = O(n^{1/2})$ and $b_n \to 0$.

I thank J. Mielniczuk, M. Pourahmadi, and X. Shao for useful comments. I am very grateful for the extremely helpful suggestions of two reviewers. This work was supported by National Science Foundation Grant DMS-0448704.

1. Wiener, N. (1958) *Nonlinear Problems in Random Theory* (MIT Press, Cambridge, MA).
2. Rosenblatt, M. (1959) *J. Math. Mech.* **8,** 665–681.
3. Rosenblatt, M. (1971) *Markov Processes. Structure and Asymptotic Behavior* (Springer, New York).
4. Kallianpur, G. (1981) in *Norbert Wiener, Collected Works with Commentaries* eds. Wiener, N. & Masani, P. (MIT Press, Cambridge, MA) pp. 402–424.
5. Rosenblatt, M. (1956) *Proc. Natl. Acad. Sci. USA* **42,** 43–47.
6. Bradley, R. C. (2005) *Introduction to Strong Mixing Conditions* (Indiana Univ. Press, Bloomington, IN).
7. Blum, J. R. & Rosenblatt, M. (1956) *Proc. Natl. Acad. Sci. USA* **42,** 412–413.
8. Doukhan, P. & Louhichi, S. (1999) *Stochastic Process. Appl.* **84,** 313–342.
9. Dedecker, J. & Prieur, C. (2005) *Probab. Theor. Rel. Fields* **132,** 203–236.
10. Rosenblatt, M. (1964) *J. Res. Natl. Bureau Standards Sect. D* **68D,** 933–936.
11. Rosenblatt, M. (1980) *J. Appl. Prob.* **17,** 265–270.
12. Andrews, D. W. K. (1984) *J. Appl. Prob.* **21,** 930–934.
13. Priestley, M. B. (1988). *Nonlinear and Nonstationary Time Series Analysis* (Academic, London).
14. Pourahmadi, M. (2001) *Foundations of Time Series Analysis and Prediction Theory* (Wiley, New York).
15. Chow, Y. S. & Teicher, H. (1988) *Probability Theory* (Springer, New York).
16. Dedecker, J. & Doukhan, P. (2003) *Stochastic Process. Appl.* **106,** 63–80.
17. Box, G. E. P., Jenkins, G. M. & Reinsel, G. C. (1994) *Time Series Analysis: Forecasting and Control* (Prentice–Hall, Englewood Cliffs, NJ).
18. Rugh, W. J. (1981) *Nonlinear System Theory: The Volterra/Wiener Approach* (Johns Hopkins Univ. Press, Baltimore).
19. Schetzen, M. (1980) *The Volterra and Wiener Theories of Nonlinear Systems* (Wiley, New York).
20. Casti, J. L. (1985) *Nonlinear System Theory* (Academic, Orlando, FL).
21. Bendat, J. S. (1990) *Nonlinear System Analysis and Identification from Random Data* (Wiley, New York).
22. Mathews, V. J. & Sicuranza, G. L. (2000) *Polynomial Signal Processing* (Wiley, New York).
23. Wu, W. B. (2006) *Econometric Theory* **22,** in press.
24. Dittmann, I. & Granger, C. W. J. (2002) *J. Econometrics* **110,** 113–133.
25. Sun, T. C. (1963) *J. Math. Mech.* **12,** 945–978.
26. Taqqu, M. S. (1975) *Z. Wahrscheinlichkeitstheorie Verw. Geb.* **31,** 287–302.
27. Ho, H. C. & Hsing, T. (1997) *Ann. Prob.* **25,** 1636–1669.
28. Feller, W. (1971) *An Introduction to Probability Theory and its Applications* (Wiley, New York) Vol. II.
29. Avram, F. & Taqqu, M. (1987) *Ann. Prob.* **15,** 767–775.
30. Ho, H. C. & Hsing, T. (1996) *Ann. Stat.* **24,** 992–1024.
31. Wu, W. B. (2003) *Bernoulli* **9,** 809–831.
32. Wu, W. B. (2003) *Statistica Sinica* **13,** 1259–1267.
33. Tong, H. (1990) *Non-linear Time Series: A Dynamical System Approach* (Oxford Univ. Press, Oxford).
34. Haggan, V. & Ozaki, T. (1981) *Biometrika* **68,** 189–196.
35. Engle, R. F. (1982) *Econometrica* **50,** 987–1007.
36. Diaconis, P. & Freedman, D. (1999) *SIAM Rev.* **41,** 41–76.
37. Wu, W. B. & Shao, X. (2004) *J. Appl. Prob.* **41,** 425–436.
38. Hsing, T. & Wu, W. B. (2004) *Ann. Prob.* **32,** 1600–1631.
39. Rio, E. (2000) *Theorie Asymptotique des Processus Aleatoires Faiblement Dependants* (Springer, Berlin).
40. Dedecker, J. & Merlevède, P. (2003) *Stochastic Process. Appl.* **108,** 229–262.
41. Volný, D. (1993) *Stochastic Process. Appl.* **44,** 41–74.
42. Hannan, E. J. (1979) *Stochastic Process. Appl.* **9,** 281–289.
43. Hannan, E. J. (1973) *Z. Wahrscheinlichkeitstheorie Verw. Geb.* **26,** 157–170.
44. Heyde, C. C. (1974) *Z. Wahrscheinlichkeitstheorie Verw. Geb.* **30,** 315–320.
45. Rosenblatt, M. (1956) *Ann. Math. Stat.* **27,** 832–837.
46. Parzen, E. (1962) *Ann. Math. Stat.* **33,** 1965–1976.
47. Robinson, P. M. (1983) *J. Time Ser. Anal.* **4,** 185–207.
48. Bosq, D. (1996) *Nonparametric Statistics for Stochastic Processes. Estimation and Prediction* (Springer, New York).
49. Wu, W. B. & Mielniczuk, J. (2002) *Ann. Stat.* **30,** 1441–1459.