The University of Chicago
Department of Statistics

TECHNICAL REPORT SERIES

Asymptotic Spectral Theory for Nonlinear Time Series

Xiaofeng Shao and Wei Biao Wu

TECHNICAL REPORT NO. 559

April 18, 2005

The University of Chicago
Department of Statistics
5734 S. University Avenue
Chicago, IL 60637

ASYMPTOTIC SPECTRAL THEORY FOR NONLINEAR TIME SERIES[1]

By Xiaofeng Shao and Wei Biao Wu

April 18, 2005

The University of Chicago

Abstract: We consider asymptotic problems in spectral analysis of stationary causal processes. Limiting distributions of periodograms and smoothed periodogram spectral density estimates are obtained and applications to spectral domain bootstrap are made. Instead of the commonly used strong mixing conditions, in our asymptotic spectral theory we impose conditions only involving (conditional) moments, which are easily verifiable for a variety of nonlinear time series.

1 Introduction

The frequency domain approach to time series has been extensively used; see Grenander and Rosenblatt (1957), Anderson (1971), Brillinger (1975) and Priestley (1981) among others. An asymptotic distributional theory is needed, for example, in hypothesis testing and in the construction of confidence intervals. However, most of the asymptotic results in the literature were developed for strong mixing processes and processes with quite restrictive summability conditions on joint cumulants [Brillinger (1969, 1975) and Rosenblatt (1984, 1985)]. Such conditions seem restrictive and they are not easily verifiable. For example, Andrews (1984) showed that, for a simple autoregressive process with innovations being independent and identically distributed (iid) Bernoulli random variables, the process is not strong mixing. Other special processes discussed include Gaussian processes [Slutsky (1929,1934)] and linear processes [Anderson (1971)].

There has been a recent surge of interest in nonlinear time series [Tong (1990) and Fan and Yao (2003)]. It seems that a systematic asymptotic spectral theory for such processes is lacking [Chanda (2005)]. The primary goal of the paper is to establish an asymptotic spectral theory for stationary, causal processes. Let

---

$(\varepsilon_n)_{n\in\mathbb{Z}}$ be a sequence of iid random variables; let

$$X_n = G(\ldots, \varepsilon_{n-1}, \varepsilon_n), \tag{1}$$

where G is a measurable function such that X_n is a proper random variable. Then the process (X_n) is causal or not-anticipative in the sense that it only depends on $\mathcal{F}_n = (\ldots, \varepsilon_{n-1}, \varepsilon_n)$, not on the future innovations $\varepsilon_{n+1}, \varepsilon_{n+2}, \ldots$. The class of processes within the framework of (1) is quite large. For example, it includes linear casual processes and some widely used nonlinear processes; see Section 5 for some examples.

Assume throughout the paper that $(X_n)_{n\in\mathbb{Z}}$ has mean zero and finite covariance function $r(k) = \mathbb{E}(X_0 X_k)$, $k \in \mathbb{Z}$. Let $i = \sqrt{-1}$ be the imaginary unit. If (X_n) is short-range dependent, namely

$$\sum_{k=0}^{\infty} |r(k)| < \infty, \tag{2}$$

then the spectral density

$$f(\lambda) = \frac{1}{2\pi} \sum_{k\in\mathbb{Z}} r(k) e^{ik\lambda}, \quad \lambda \in \mathbb{R}, \tag{3}$$

is continuous and bounded. Given the observations $X_1, \ldots, X_n$, let

$$S_n(\theta) = \sum_{k=1}^{n} X_k e^{ik\theta} \text{ and } I_n(\theta) = \frac{1}{2\pi n} |S_n(\theta)|^2 \tag{4}$$

be the Fourier transform and the periodogram, respectively. Let $\theta_k = 2\pi k/n$, $1 \le k \le n$, be the Fourier frequencies. Major goals in spectral analysis include the estimation of the spectral density f and the asymptotic distribution of $S_n(\theta)$ and $I_n(\theta)$.

Now we introduce some notation. For a vector $x = (x_1, \cdots, x_q)' \in \mathbb{R}^q$, let $|x| = (\sum_{i=1}^{q} x_i^2)^{1/2}$. Let ξ be a random vector. Write $\xi \in \mathcal{L}^p$ $(p > 0)$ if $\|\xi\|_p := [\mathbb{E}(|\xi|^p)]^{1/p} < \infty$ and let $\|\cdot\| = \|\cdot\|_2$. For $\xi \in \mathcal{L}^1$ define projection operators $\mathcal{P}_k \xi = \mathbb{E}(\xi|\mathcal{F}_k) - \mathbb{E}(\xi|\mathcal{F}_{k-1})$, $k \in \mathbb{Z}$. For two sequences (a_n), (b_n), denote by $a_n \asymp b_n$ if there are constants c, c' such that $0 < c \le a_n/b_n \le c' < \infty$ for all large n and by $a_n \sim b_n$ if $a_n/b_n \to 1$ as $n \to \infty$. Let $C > 0$ denote a generic constant which may vary from line to line; let Φ and $\phi = \Phi'$ be the standard normal distribution and

density functions. Denote by "$\Rightarrow$" convergence in distribution. All asymptotic statements in the paper are with respect to $n \to \infty$ unless otherwise specified.

The paper is structured as follows. In Section 2 we shall establish a central limit theorem for the Fourier transform $S_n(\theta)$ and the periodogram $I_n(\theta)$ at Fourier frequencies. Asymptotic properties of smoothed periodogram estimates of f are discussed in Section 3. Section 4 considers consistency of the frequency domain bootstrap approximation to the sampling distribution of spectral density estimates for both linear and nonlinear processes. Section 5 gives some sufficient conditions for geometric moment contraction (see (12)), a basic dependence assumption used in this paper. Some examples are also presented in that section. Proofs are gathered in the appendix.

2   Fourier transforms

The periodogram is a fundamental quantity in the frequency-domain analysis. Its asymptotic analysis has a substantial history; see for example Rosenblatt (Theorem 5.3, p. 131, 1985) for mixing processes; Brockwell and Davis (Theorem 10.3.2, page 347, 1991), Walker (1965) and Terrin and Hurvich (1994) for linear processes. Other contributions can be found in Olshen (1967), Rootzen (1976), Yajima (1989), Walker (2000) and Lahiri (2003). Recently, in a general setting, Wu (2005) considered asymptotic distributions of $S_n(\theta)$ at a fixed θ. However, results in Wu (2005) do not apply to $S_n(\theta)$ at Fourier frequencies. Here we shall show that $S_n(\theta_k)$ for $k = 1, \cdots, n$ are asymptotically independent normals under mild conditions; see Theorem 2.1 below. The central limit theorem is applied to empirical distribution functions of normalized periodogram ordinates (cf. Corollary 2.2). In the literature the latter problem has been mainly studied for iid random variables [Freedman and Lane (1980, 1981), Kokoszka and Mikosch (2000)] and linear processes (Chen and Hannan, 1980).

Denote the real and imaginary parts of $S_n(\theta_j)/\sqrt{\pi n f(\theta_j)}$ by

$$Z_j = \frac{\sum_{k=1}^{n} X_k \cos(k\theta_j)}{\sqrt{\pi n f(\theta_j)}}, \quad Z_{j+m} = \frac{\sum_{k=1}^{n} X_k \sin(k\theta_j)}{\sqrt{\pi n f(\theta_j)}}, \quad j = 1, \ldots, m, \qquad (5)$$

where $m = m_n := \lfloor (n-1)/2 \rfloor$ and $\lfloor a \rfloor$ is the integer part of a; denote the unit sphere by $\Omega_p = \{ c \in \mathbb{R}^p : |c| = 1 \}$. For the set $J = \{ j_1, \ldots, j_p \}$ with $1 \le j_1 <$

3

$\ldots < j_p \leq 2m$ write the vector $Z_J = (Z_{j_1}, \ldots, Z_{j_p})'$. Let the class $\Xi_{m,p} = \{J \subset \{1, \ldots, 2m\} : \#J = p\}$, where $\#J$ is the cardinality of J.

THEOREM **2.1**. *Assume* $X_t \in \mathcal{L}^2$,

$$\kappa := \sum_{k=0}^{\infty} \|\mathcal{P}_0 X_k\| < \infty \tag{6}$$

and $f_* := \min_{\mathbb{R}} f(\theta) > 0$. *Then for any fixed* $p \in \mathbb{N}$, *we have*

$$\sup_{J \in \Xi_{m,p}} \sup_{c \in \Omega_p} \sup_{x \in \mathbb{R}} |\mathbb{P}(Z_J' c \leq x) - \Phi(x)| = o(1) \ \text{as} \ n \to \infty. \tag{7}$$

Theorem 2.1 asserts that the projection of any vector of p of the Z_js on any direction is asymptotically normal. The condition (6) was first proposed by Hannan (1973). In many situations it is easily verifiable since it only involves conditional moments. For generalizations see Wu and Min (2005). In the special case of linear processes $X_t = \sum_{i=0}^{\infty} a_i \varepsilon_{t-i}$, where ε_i are iid with mean 0 and finite variance and $\sum_{i=0}^{\infty} a_i^2 < \infty$, (6) becomes $\sum_{i=0}^{\infty} |a_i| < \infty$, indicating that (X_n) is short-range dependent. In the literature, central limit theorems are established for Fourier transforms of linear processes [Fan and Yao (2003, p. 63), Brockwell and Davis (1991, p. 347) among others]. The spectral density function may be unbounded if (6) is violated.

COROLLARY **2.1**. *Under the conditions of Theorem 2.1, we have for any fixed* $q \in \mathbb{N}$,

$$\left\{ \frac{S_n(\theta_{l_j})}{\sqrt{n\pi f(\theta_{l_j})}}, \ 1 \leq j \leq q \right\} \Rightarrow \{Y_{2j-1} + iY_{2j}, \ 1 \leq j \leq q\} \tag{8}$$

for integers $1 \leq l_1 < l_2 < \ldots < l_q \leq m$, *where* Y_k, $1 \leq k \leq 2q$, *are iid standard normal. Consequently, for* $\tilde{I}_n(\theta) := I_n(\theta)/f(\theta)$,

$$\left\{ \tilde{I}_n(\theta_{l_j}), \ 1 \leq j \leq q \right\} \Rightarrow \{E_j, \ 1 \leq j \leq q\}, \tag{9}$$

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

Corollary 2.1 easily follows from Theorem 2.1 via the Cramer-Wold device. For the empirical distribution function of $\tilde{I}_n(\theta_k)$, $F_{\tilde{I},m}(x) := m^{-1} \sum_{j=1}^{m} \mathbf{1}_{\tilde{I}_n(\theta_j) \leq x}$, $x \geq 0$, we have

4

COROLLARY **2.2**. *Let $F_E(x) := 1 - e^{-x}$. Under conditions of Theorem 2.1,*

$$\sup_{x \geq 0} |F_{\tilde{I},m}(x) - F_E(x)| \to 0 \quad \text{in probability.} \tag{10}$$

PROOF. Since $F_{\tilde{I},m}$ and F_E are non-decreasing, it suffices to show (10) for a *fixed* x. Let $p_j = p_j(x) = \mathbb{P}[\tilde{I}_n(\theta_j) \leq x]$ and $p_{i,j} = p_{i,j}(x) = \mathbb{P}[\tilde{I}_n(\theta_i) \leq x, \tilde{I}_n(\theta_j) \leq x]$; let U and V, independent of the process (X_i), be iid uniformly distributed over $\{1, \cdots, m\}$. By Corollary 2.1, $p_U \to F_E(x)$ and $p_{U,V} \to F_E(x)^2$ almost surely. By the Lebesgue dominated convergence theorem, $\mathbb{E}(p_U) \to F_E(x)$ and $\mathbb{E}(p_{U,V}) \to F_E(x)^2$. Notice that

$$\mathbb{E}(p_U) = m^{-1} \sum_{j=1}^{m} p_j \text{ and } \mathbb{E}(p_{U,V}) = m^{-2} \sum_{i=1}^{m} \sum_{j=1}^{m} p_{i,j}.$$

So $\|F_{\tilde{I},m}(x) - F_E(x)\|^2 = \mathbb{E}(p_{U,V}) - F_E^2(x) + 2F_E(x)\{F_E(x) - \mathbb{E}(p_U)\}$ and (10) follows. $\diamond$

REMARK **2.1**. The above argument also implies that, for any integer $k \geq 2$,

$$\sup_{x_1, \cdots, x_k \geq 0} \left| m^{-k} \sum_{j_1=1}^{m} \cdots \sum_{j_k=1}^{m} \mathbb{P}[\tilde{I}_n(\theta_{j_1}) \leq x_1, \cdots, \tilde{I}_n(\theta_{j_k}) \leq x_k] - \prod_{i=1}^{k} F_E(x_i) \right| \to 0.$$

3   Spectral density estimation

Given a realization $(X_i)_{i=1}^{n}$, the spectral density f can be estimated by

$$f_n(\lambda) = \int_{-\pi}^{\pi} W_n(\lambda - \mu) I_n(\mu) d\mu, \tag{11}$$

where $W_n(\lambda)$ is a smoothing weight function (cf (13)). Here we study asymptotic properties of the smoothed periodogram estimate f_n. Spectral density estimation is an important problem and there is a rich literature. However, restrictive structural conditions have been imposed in many earlier results. For example, Brillinger (1969) assumed that all moments exist and cumulants for all orders are summable. Anderson (1971) dealt with the special linear processes. Rosenblatt (1984) considered strong mixing processes and assumed the stringent summability condition of cumulants up to eighth order. Whether the weaker fourth order cumulants summability condition suffices is proposed in the latter paper as an open

problem. Due to those limitations, the classical results cannot be directly applied to nonlinear time series models. Recently, Chanda (2005) obtained asymptotic normality of f_n for a class of nonlinear processes. However, it seems that his formulation doesn't include popular nonlinear time series models such as GARCH, EXPAR and ARMA-GARCH; see Section 5 for examples.

To establish an asymptotic theory for f_n, we adopt the geometric-moment contraction (GMC) condition. Let $(\varepsilon'_k)_{k \in \mathbb{Z}}$ be an iid copy of $(\varepsilon_k)_{k \in \mathbb{Z}}$; let $X'_n = G(\cdots, \varepsilon'_{-1}, \varepsilon'_0, \varepsilon_1, \cdots, \varepsilon_n)$ be a coupled version of X_n. We say that X_n is GMC(α), $\alpha > 0$, if there exist C and $0 < \rho = \rho(\alpha) < 1$ such that for all $n \in \mathbb{N}$,

$$\mathbb{E}(|X'_n - X_n|^\alpha) \leq C\rho^n. \tag{12}$$

Inequality (12) indicates that the process (X_n) quickly "forgets" the past $\mathcal{F}_0 = (\cdots, \varepsilon_{-1}, \varepsilon_0)$. GMC has the following interesting property: If $X_n \in \mathcal{L}^p, p > 0$ and GMC(α_0) holds for some $\alpha_0 > 0$, then (12) holds for any $\alpha < p$ [Wu and Shao (2004), Lemma 1]. Note that under GMC(2), $|r(k)| = O(\rho^k)$ for some $\rho \in (0,1)$ and hence the spectral density function is infinitely many times differentiable.

Many nonlinear time series models satisfy GMC (cf Section 5). Moreover, the GMC condition provides a convenient framework for a limit theory for nonlinear time series; see Hsing and Wu (2004), Wu and Shao (2004) and Wu and Min (2005). In view of those features, instead of the widely used strong mixing condition, we employ the GMC as an underlying assumption for an asymptotic theory for spectral density estimates.

Let $r_k^{(n)} = n^{-1} \sum_{j=1}^{n-|k|} X_j X_{j+|k|}$, $|k| < n$, be the covariance estimates; let $a(\cdot)$ be an even, Lipschitz continuous function with support $[-1, 1]$ and $a(0) = 1$; let B_n be a sequence of positive integers such that $B_n \to \infty$ and $B_n/n \to 0$; let $b_n = 1/B_n$,

$$W_n(\lambda) = \frac{1}{2\pi} \sum_{k=-B_n}^{B_n} a(kb_n)e^{-ik\lambda} \text{ and } f_n(\lambda) = \frac{1}{2\pi} \sum_{k=-B_n}^{B_n} r_k^{(n)} a(kb_n)e^{-ik\lambda}. \tag{13}$$

THEOREM **3.1**. *Assume (12), $X_n \in \mathcal{L}^{4+\delta}$ and $B_n \asymp n^\eta$ for some $\delta > 0$ and $0 < \eta < 1$. Then*

$$\sqrt{nb_n}\{f_n(\lambda) - \mathbb{E}(f_n(\lambda))\} \Rightarrow N(0, \sigma^2(\lambda)), \tag{14}$$

where $\sigma^2 := \sigma^2(\lambda) = \{1 + \eta(2\lambda)\}f^2(\lambda) \int_{-1}^1 a^2(t)dt$ and $\eta(\lambda) = 1$ if $\lambda = 2k\pi$ for some integer k and $\eta(\lambda) = 0$ otherwise.

6

The moment condition $X_n \in \mathcal{L}^{4+\delta}$ in Theorem 3.1 together with GMC implies the absolute summability of cumulants up to the fourth order (cf Lemma 6.1). In the context of strong mixing processes, Rosenblatt (1985, page 138) imposed $X_n \in \mathcal{L}^8$. The latter paper posed the problem that whether the eighth order cumulants summability condition can be weakened. Theorem 3.1 partially solves the conjecture for nonlinear processes satisfying GMC. Additionally, Theorem 3.1 is applicable to a variety of nonlinear time series models (Section 5) that are not covered by Chanda (2005).

Joint asymptotic distributions of spectral density estimates at distinct frequencies (cf Corollary 3.1 below) follow from the arguments in Parzen (1957, Theorem 5A) and Rosenblatt (1984) since GMC(4) ensures the summability of the fourth cumulants; see Lemma 6.1.

COROLLARY **3.1**. *Let* $\lambda_1, \cdots, \lambda_s \in [0, \pi]$ *be s distinct frequencies. Then under the conditions of Theorem 3.1,* $\sqrt{nb_n}\{f_n(\lambda_j) - \mathbb{E}(f_n(\lambda_j))\}$, $j = 1, \ldots, s$, *are jointly asymptotically independent* $N(0, \sigma^2(\lambda_j))$, $j = 1, \ldots, s$.

The problem of maximum deviation of spectral density estimates has been studied by Woodroofe and Van Ness (1967) for linear processes and Rudzkis (1985) for Gaussian processes. For nonlinear processes, we have

THEOREM **3.2**. *Assume (12),* $X_n \in \mathcal{L}^{4+\delta}$ *for some* $\delta > 0$, $B_n \asymp n^\eta$, $0 < \eta < 1/2$ *and* $f_* := \min_{\mathbb{R}} f(\theta) > 0$. *Then*

$$\max_{\lambda \in [0,\pi]} \sqrt{nb_n}|f_n(\lambda) - \mathbb{E}(f_n(\lambda))| = O_{\mathbb{P}}((\log n)^{1/2}). \tag{15}$$

Under GMC(2), since $\|\mathcal{P}_0 X_k\| = O(\rho^k)$, we have (6). However, it is quite difficult to establish (14) under the weaker condition (6). Regarding (15), we are unable to obtain a distributional result as in Woodroofe and Van Ness (1967) for nonlinear processes.

4 Frequency domain bootstrap

Here we consider bootstrap approximations of the distribution of the lag window estimate (13). Bootstrapping in the frequency domain has recently received considerable interest. See Hurvich and Zeger (1987), Nordgaard (1992) and Theiler, Paul and Rubin (1994) for Gaussian processes and Franke and Härdle (1992,

FH hereafter), Paparoditis and Politis (1999) and Kreiss and Paparoditis (2003) for linear processes. For nonlinear processes we adopt the residual-based bootstrap procedure proposed by FH. A variant of it is discussed in Remark 4.4. Let $I_j = I(\omega_j)$, $\omega_j = 2\pi j/n, j \in F_n = \{-\lfloor (n-1)/2 \rfloor, \cdots, \lfloor n/2 \rfloor\}$. Note that $r_k^{(n)} = n^{-1} 2\pi \sum_{j \in F_n} I_j e^{ik\omega_j}$. Then the lag window estimate (13) can be written as

$$f_n(\lambda) = \frac{1}{2\pi} \sum_{k=-B_n}^{B_n} r_k^{(n)} a(kb_n) e^{-ik\lambda} = \frac{1}{n} \sum_{j \in F_n} I_j \sum_{k=-B_n}^{B_n} a(kb_n) e^{-ik(\lambda - \omega_j)}. \quad (16)$$

The bootstrap procedure consists of the following several steps.

1. Calculate periodogram ordinates $\{I_j\}$, $j = 1, \cdots, N := \lfloor n/2 \rfloor$.

2. Obtain an estimate $\tilde{f}$ of f. (e.g. a lag window estimate with bandwidth $\tilde{b}_n := \tilde{B}_n^{-1}$).

3. Let $\bar{\varepsilon}_j = \tilde{\varepsilon}_j / \bar{\varepsilon}$, where $\tilde{\varepsilon}_j = I_j / \tilde{f}_j$, $\tilde{f}_j = \tilde{f}(\omega_j)$, $\bar{\varepsilon} = N^{-1} \sum_{j=1}^{N} \tilde{\varepsilon}_j$.

4. Draw iid bootstrap samples $\{\varepsilon_j^*\}$ from the empirical distribution of $\bar{\varepsilon}_j$.

5. Let $I_j^* = \tilde{f}_j \varepsilon_j^*$ be the bootstrapped periodogram values; let $I_{-j}^* = I_j^*$ and $I_0^* = 0$.

The rescaling in step 3 avoids an unwanted bias at the resampling stage. Setting $I_0^* = 0$ in step 5 corresponds to the periodogram value at 0 taken from a mean-corrected sample. The sampling distribution of $g_n(\lambda) = \sqrt{nb_n}\{f_n(\lambda) - f(\lambda)\}$ is expected to be close to its bootstrap counterpart $g_n^*(\lambda) = \sqrt{nb_n}\{f_n^*(\lambda) - \tilde{f}(\lambda)\}$, where

$$f_n^*(\lambda) = \frac{1}{n} \sum_{j \in F_n} I_j^* \sum_{k=-B_n}^{B_n} a(kb_n) e^{-ik(\lambda - \omega_j)}$$

is the bootstrapped version of (16). Here the closeness is measured by Mallow's d_2 metric (Bickel and Freedman, 1981). For two probability measures P_1 and P_2 on $\mathbb{R}$ for which $\int_{\mathbb{R}} |x|^2 dP_i < \infty$, $i = 1, 2$, let $d_2(P_1, P_2) = \inf \|Y_1 - Y_2\|$, where the infimum is taken over all vectors (Y_1, Y_2) with marginal distributions P_1 and P_2. Write

$$d_2[g_n(\lambda), g_n^*(\lambda)] = d_2\{\mathbb{P}[g_n(\lambda) \in \cdot], \mathbb{P}[g_n^*(\lambda) \in \cdot | X_1, \cdots, X_n]\}.$$

The bootstrap procedure is said to be (weakly) consistent if $d_2[g_n(\lambda), g_n^*(\lambda)] = o_{\mathbb{P}}(1)$. Let $\mathcal{L}(\cdot | X_1, \cdots, X_n)$ denote the conditional distribution given the sample $X_1, \cdots, X_n$.

It seems that in the literature the theoretical investigation of the consistency has been limited to linear processes. Let $X_t = \sum_{i=-\infty}^{\infty} a_i \varepsilon_{t-i}$. FH proved the consistency of the residual-based procedure for kernel spectral estimates under the condition

$$\sup\{|\mathbb{E}(e^{iu\varepsilon_1})|; \; |u| \geq \delta\} < 1 \quad \text{for all } \delta > 0. \tag{17}$$

Condition (17) excludes many interesting cases. For example, it is violated if ε_1 is a Bernoulli random variable. FH (1992, page 126) conjectured that their results still hold without the condition (17). The main result in this section is Theorem 4.1 which is applicable to linear as well as nonlinear processes; see Corollaries 4.1 and 4.2 respectively. The former corollary deals with linear processes and (17) is removed at the expense of the stronger 8-th moment condition. Since our results hold under various combinations of conditions, it is convenient to label the more common ones:

(A1) $\lim_{x \to 0} x^{-2}\{1 - a(x)\} = c_2$, where c_2 is a nonzero constant.
(A2) $\min_{\lambda \in [0,\pi]} f(\lambda) > 0$.
(A3) $\max_{\lambda \in [0,\pi]} |\tilde{f}(\lambda) - f(\lambda)| = o_{\mathbb{P}}(b_n)$.
(A3′) $\max_{\lambda \in [0,\pi]} |\tilde{f}(\lambda) - f(\lambda)| = o_{\mathbb{P}}(1)$.
(A4) $\sum_{k \in \mathbb{Z}} |r(k)|k^2 < \infty$.
(A4′) $\sum_{k \in \mathbb{Z}} |r(k)k| < \infty$.
(A5) $\sum_{t_1,\cdots,t_{k-1} \in \mathbb{Z}} |\text{cum}(X(0), X(t_1), \cdots, X(t_{k-1}))| < \infty$ for $k = 3, 4$.
(A5′) $\sum_{t_1,\cdots,t_{k-1} \in \mathbb{Z}} |\text{cum}(X(0), X(t_1), \cdots, X(t_{k-1}))| < \infty$ for $k = 3, \cdots, 8$.
(A6) $\sqrt{nb_n}\{f_n(\lambda) - \mathbb{E}(f_n(\lambda))\} \Rightarrow N(0, \sigma^2(\lambda))$ and $nb_n \text{var}(f_n(\lambda)) \to \sigma^2(\lambda)$.

REMARK **4.1**. Condition (A1) says that $a(\cdot)$ is locally quadratic at zero and it is satisfied for many lag windows. It is related to bias. By Anderson (1971, Theorem 9.4.3) or Priestley (1981, page 459), under (A1), (A4) and $B_n^3 = o(n)$,

$$B_n^2\{\mathbb{E}(f_n(\lambda)) - f(\lambda)\} \to c_2 f''(\lambda), \text{ where } f''(\lambda) = -\frac{1}{2\pi}\sum_{k \in \mathbb{Z}} r(k)k^2 e^{-ik\lambda}. \tag{18}$$

Furthermore, if (A6) holds, then the optimal bandwidth b_n is of order $n^{-1/5}$ in the sense of minimizing mean square error asymptotically [Priestley (1981), Chapter 7.2].

REMARK **4.2**. The cumulant summability conditions (A5) and (A5′) are commonly imposed in spectral analysis [Brillinger (1975), Rosenblatt (1985)]. For a linear

9

process $X_t = \sum_{i=-\infty}^{\infty} a_i \varepsilon_{t-i}$ with $\sum_{i=-\infty}^{\infty} |a_i| < \infty$, (A5) [resp. (A5′)] holds if $\varepsilon_1 \in \mathcal{L}^4$ [resp. $\varepsilon_1 \in \mathcal{L}^8$]. By Lemma 6.1, stationary processes of form (1) satisfy (A5) [resp. (A5′)] under GMC(4) [resp. GMC(8)]. Zhurbenko and Zuev (1975) considered strong mixing processes.

Let $\mathbb{E}^*$ and var* denote the conditional expectation and variance given the original data. Let $V_n(\lambda) = \sqrt{nb_n}\{f_n(\lambda) - \mathbb{E}(f_n(\lambda))\}$, $V_n^*(\lambda) = \sqrt{nb_n}\{f_n^*(\lambda) - \mathbb{E}^* f_n^*(\lambda)\}$, $\beta_n(\lambda) = \sqrt{nb_n}\{\mathbb{E}(f_n(\lambda)) - f(\lambda)\}$ and $\beta_n^*(\lambda) = \sqrt{nb_n}\{\mathbb{E}^* f_n^*(\lambda) - \tilde{f}(\lambda)\}$. For the consistency of the bootstrap approximation, it is common to treat the variance and the bias part separately.

PROPOSITION **4.1**. *Assume $X_t \in \mathcal{L}^8$, (A2)-(A3), (A4′), (A5′) and (A6). Let $B_n^2 = o(n)$. Then $d_2[V_n(\lambda), V_n^*(\lambda)] \to 0$ in probability.*

PROPOSITION **4.2**. *Assume $X_t \in \mathcal{L}^4$, (A1) and (A4)-(A5). Let $b_n = o(\tilde{b}_n)$, $B_n^3 = o(n)$ and $\tilde{B}_n^5 = o(n)$. Then $B_n^2\{\mathbb{E}^* f_n^*(\lambda) - \tilde{f}(\lambda)\} \to c_2 f''(\lambda)$ in probability.*

REMARK **4.3**. The condition $b_n = o(\tilde{b}_n)$ is needed to ensure the consistency of the bias part in view of (18). Hence $\tilde{f}(\lambda)$ should be smoother than our lag window estimate $f_n(\lambda)$. Over-smoothing is common in the frequency domain bootstrap [Paparoditis and Politis (1999), Kreiss and Paparoditis (2003) and FH].

THEOREM **4.1**. *Assume $X_t \in \mathcal{L}^8$, (A1)-(A4), (A5′) and (A6). Let $b_n \asymp n^{-1/5}$ and $b_n = o(\tilde{b}_n)$. Then $d_2[g_n(\lambda), g_n^*(\lambda)] \to 0$ and $d_2[g_n(\lambda)/f(\lambda), g_n^*(\lambda)/\tilde{f}(\lambda)] \to 0$ in probability.*

PROOF. In the proof λ is suppressed and we write g_n etc for $g_n(\lambda)$ etc. Since $d_2^2(g_n, g_n^*) = d_2^2(V_n, V_n^*) + d_2^2(\beta_n, \beta_n^*)$ (Lemma 8.8, Bickel and Freedman, 1981), by Propositions 4.1, 4.2 and (18), $d_2(g_n, g_n^*) = o_{\mathbb{P}}(1)$. The second assertion follows similarly. By (A2), (A3) and Proposition 4.2, $\beta_n^*/\tilde{f} - \beta_n/f = (\beta_n^* - \beta_n)/\tilde{f} + (\tilde{f}^{-1} - f^{-1})\beta_n = o_{\mathbb{P}}(1)$. It remains to show that $d_2(V_n/f, V_n^*/\tilde{f}) = o_{\mathbb{P}}(1)$. By Lemma 8.3 in Bickel and Freedman (1981), it suffices in view of (A6) to show that var$^*(V_n^*/\tilde{f}) \to \sigma^2/f^2$ and $\mathcal{L}(V_n^*/\tilde{f}|X_1, \cdots, X_n) \Rightarrow N(0, \sigma^2/f^2)$ in probability. By (A2), (A3), these two assertions follow from relation (51) in the proof of Proposition 4.1. $\diamond$

REMARK **4.4**. Since the residuals $\{I_n(\omega_j)/f(\omega_j)\}$ are asymptotically iid exp(1) (Corollary 2.1), a modified procedure is to replace the bootstrapped residuals ε_j^*

by iid standard exponential variables. For this modified bootstrap procedure, Theorem 4.1 holds with the assumption (A5′) replaced by (A5) and 8-th moment condition weakened to $X_t \in \mathcal{L}^4$; see the proof of Proposition 4.1.

COROLLARY **4.1**. *Let $X_t = \sum_{i=-\infty}^{\infty} a_i \varepsilon_{t-i}$, where $|a_k| = O(k^{-(1+\beta)})$, $\beta > 1/5$ and $\varepsilon_1 \in \mathcal{L}^8$. Assume (A1)-(A2), (A4), $b_n \asymp n^{-1/5}$ and $\tilde{b}_n \asymp n^{-\eta_1}$, $\eta_1 \in (1/10, 1/5)$. Then the conclusions in Theorem 4.1 hold.*

PROOF. By Theorem 4.1, it suffices to verify (A3), (A5′) and (A6). (A6) follows from Theorem 9.3.4 and 9.4.1 in Anderson (1971). The assumption (A5′) is satisfied under $\mathbb{E}(\varepsilon_1^8) < \infty$ and $|a_k| = O(k^{-(1+\beta)})$, $\beta > 1/5$ (see Remark 4.2). Note that

$$\max_{\lambda \in [0,\pi]} |\tilde{f}(\lambda) - f(\lambda)| \leq \max_{\lambda \in [0,\pi]} |\tilde{f}(\lambda) - \mathbb{E}(\tilde{f}(\lambda))| + \max_{\lambda \in [0,\pi]} |\mathbb{E}(\tilde{f}(\lambda)) - f(\lambda)|, \quad (19)$$

which is of order $O_{\mathbb{P}}((\log n)^{1/2}/(n\tilde{b}_n)^{1/2}) + O_{\mathbb{P}}(\tilde{b}_n^2) = o_{\mathbb{P}}(b_n)$ by Theorem 2.1 in Woodroofe and Van Ness (1967) and (18). So (A3) follows. $\diamond$

COROLLARY **4.2**. *Let the process (1) satisfy GMC(8). Assume (A1)-(A2), $b_n \asymp n^{-1/5}$ and $\tilde{b}_n \asymp n^{-\eta_2}$, $\eta_2 \in (1/10, 1/5)$. Then the conclusions in Theorem 4.1 hold.*

PROOF. We shall apply Theorem 4.1. By Lemma 6.1, GMC(8) implies (A4) and (A5′), while (A6) [resp. (A3)] follows from Theorem 3.1 [resp. Theorem 3.2 and (19)]. $\diamond$

5 Applications

There are two popular criteria to check the stationarity of nonlinear time series models: drift-type conditions [Tweedie (1975, 1976, 1988), Chan and Tong (1985), Feigin and Tweedie (1985), Meyn and Tweedie (1993) etc] and contraction conditions [Elton (1990), Diaconis and Freedman (1999), Jarner and Tweedie (2001) and Wu and Shao (2004) etc]. It turns out that contraction conditions typically imply GMC under some extra mild assumptions, and are thus quite useful in proving limit theorems [Hsing and Wu (2004), Wu and Min (2005)]. In this section we consider nonlinear autoregressive models and present sufficient conditions for GMC so that our asymptotic spectral theory is applicable.

Let ε_n be iid random elements, $p, d \geq 1$; let $X_n \in \mathbb{R}^d$ be recursively defined by

$$X_{n+1} = R(X_n, \cdots, X_{n-p+1}; \varepsilon_{n+1}), \tag{20}$$

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

THEOREM **5.1**. *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 non-negative constants $a_1, \cdots, a_p$ with $\sum_{i=1}^p a_i < 1$ such that*

$$\|R(y; \varepsilon) - R(y'; \varepsilon)\|_\alpha^{\alpha'} \leq \sum_{i=1}^p a_i |x_i - x_i'|^{\alpha'} \tag{21}$$

for all $y = (x_1, \ldots, x_p)$ and $y' = (x_1', \ldots, x_p')$. Then X_n satisfies GMC(α). In particular, if there exist functions H_i such that $|R(y; \varepsilon) - R(y'; \varepsilon)| \leq \sum_{i=1}^p H_i(\varepsilon)|x_i - x_i'|$ for all y and y' and $\sum_{i=1}^p \|H_i(\varepsilon)\|_\alpha^{\alpha'} < 1$, then we can let $a_i = \|H_i(\varepsilon)\|_\alpha^{\alpha'}$.

We omit the proof of Theorem 5.1 since it easily follows from Lemma 6.2.10 and Proposition 6.3.22 in Duflo (1997). Duflo assumed $\alpha \geq 1$ and called (21) Lipschitz mixing condition. In our result $\alpha < 1$ is allowed. Conditions of a similar type are given in Götze and Hipp (1994). An important special case of Theorem 5.1 is $p = 1$, which is called an *iterated random function* in Elton (1990) and Diaconis and Freedman (1999).

THEOREM **5.2**. *Assume that (η_t) satisfies GMC(α) (12) and that the ARMA(p, q) process*

$$X_t - \theta_1 X_{t-1} - \cdots - \theta_p X_{t-p} = \eta_t - \phi_1 \eta_{t-1} - \cdots - \phi_q \eta_{t-q} \tag{22}$$

is driven by the dependent innovations η_t. Further assume that all the roots of the polynomial $\lambda^p - \sum_{k=1}^p \theta_k \lambda^{p-k}$ lie inside the unit circle. Then X_t is also GMC(α).

Theorem 5.2 shows that the GMC property is preserved in ARMA modelling (Min, 2004) and it is an easy consequence of the representation $X_t = \sum_{k=0}^\infty b_k \eta_{t-k}$ with $|b_k| \leq C r^k$ for some $r \in (0, 1)$. Min (2004) considered the case $\alpha \geq 1$. Theorem 5.2 implies that the ARMA-ARCH and ARMA-GARCH models (Li, Ling and McAleer, 2002) are GMC; see Examples 5.4 and 5.5.

Near-epoch dependence (NED) is widely used in econometrics for central limit theorems [Davidson (1994, 2002)]. The process (1) is geometrically NED (G-NED(α)) on (ε_s) in L_α, $\alpha > 0$, if there exist $C < \infty$ and $\rho \in (0, 1)$ such that, for all $m \in \mathbb{N}$,

$$\|X_t - \mathbb{E}(X_t | \varepsilon_{t-m}, \varepsilon_{t-m+1}, \ldots, \varepsilon_t)\|_\alpha \leq C \rho^m. \tag{23}$$

12

It is easily seen that, for $\alpha \geq 1$, GMC(α) is equivalent to G-NED(α). In some situations GMC is more convenient to handle; see Remark 5.1. Additionally, GMC has the nice property that X'_t is identically distributed as X_t, while in NED, the distribution of $\mathbb{E}(X_t | \varepsilon_{t-m}, \dots, \varepsilon_t)$ typically differs. Davidson (2002) showed that a variety of nonlinear models are G-NED(2), and hence GMC(2). From Davidson's argument, it seems harder to verify G-NED(p) for $p > 2$ while this is not the case for GMC. Here we list some examples that are not covered by Davidson (2002).

EXAMPLE **5.1**. Amplitude-dependent exponential autoregressive (EXPAR) models have been studied by Jones (1976). Let $\varepsilon_i \in \mathcal{L}^\alpha$ be iid innovations and

$$X_n = [\alpha_1 + \beta_1 \exp(-aX_{n-1}^2)]X_{n-1} + \varepsilon_n, \quad a > 0.$$

Then $H_1(\varepsilon) = |\alpha_1| + |\beta_1|$. By Theorem 5.1, X_n is GMC(α) if $|\alpha_1| + |\beta_1| < 1$. $\quad\diamond$

EXAMPLE **5.2**. Consider the AR(2) model with ARCH(2) errors [Engle (1982)]

$$X_n = \theta_1 X_{n-1} + \theta_2 X_{n-2} + \varepsilon_n \sqrt{\theta_3^2 + \theta_4^2 X_{n-1}^2 + \theta_5^2 X_{n-2}^2}.$$

Theorem 5.1 is applicable here: we can chose $H_1(\varepsilon) = |\theta_1| + |\varepsilon\theta_4|$ and $H_2(\varepsilon) = |\theta_2| + |\varepsilon\theta_5|$. Then GMC($\alpha$) holds if $\sum_{i=1}^2 \|H_i(\varepsilon)\|_\alpha^{\alpha'} < 1$ and $\varepsilon_1 \in \mathcal{L}^\alpha$ for some $\alpha > 0$. $\quad\diamond$

EXAMPLE **5.3**. Let A_t be $p \times p$ random matrices and B_t be $p \times 1$ random vectors. The generalized random coefficient autoregressive process (X_t) is defined by

$$X_{t+1} = A_{t+1}X_t + B_{t+1}, \ t \in \mathbb{Z}. \tag{24}$$

Assume that (A_t, B_t) are iid. Bilinear and GARCH models fall within the framework of (24). The stationarity, geometric ergodicity and β-mixing properties of (24) have been investigated by Pham (1986) and Carrasco and Chen (2002). Their results require that innovations have a density, which is not needed for GMC.

For a $p \times p$ matrix A, let $|A|_\alpha = \sup_{z \neq 0} |Az|_\alpha / |z|_\alpha$, $\alpha \geq 1$, be the matrix norm induced by the vector norm $|z|_\alpha = (\sum_{i=1}^p |z_i|^\alpha)^{1/\alpha}$. It is easily seen that X_t is GMC(α) if $\mathbb{E}(|A_0|_\alpha) < 1$ and $\mathbb{E}(|B_0|_\alpha) < \infty$. For an application, consider the subdiagonal bilinear model [Granger and Anderson (1978), Subba Rao and Gabr (1984)]:

$$X_t = \sum_{j=1}^p a_j X_{t-j} + \sum_{j=0}^q c_j \varepsilon_{t-j} + \sum_{j=0}^P \sum_{k=1}^Q b_{jk} X_{t-j-k} \varepsilon_{t-k} \tag{25}$$

13

Let $s = \max(p, P+q, P+Q)$, $r = s - \max(q, Q)$ and $a_{p+j} = 0 = c_{q+j} = b_{P+i,Q+j} = 0$ for all $i, j \geq 1$; let H be an $1 \times s$ 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 $1, a_1 + c_1, \cdots, a_{s-r} + c_{s-r}$, d be an $s \times 1$ vector with the first r elements 0 followed by $b_{01}, \cdots, b_{0,s-r}$ and

$$
A = \begin{pmatrix}
0 & 1 & & & & \\
& & \ddots & & 0 & \\
& & & 1 & & \\
& 0 & & a_1 & & \ddots \\
& & & \vdots & & 1 \\
a_s & \cdots & \cdots & a_{s-r} & & 0
\end{pmatrix}_{s \times s}, \quad
B = \begin{pmatrix}
0 & \cdots & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & \cdots & 0 & 0 & \cdots & 0 \\
b_{r1} & \cdots & b_{01} & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
b_{r,s-r} & \cdots & b_{0,s-r} & 0 & \cdots & 0
\end{pmatrix}_{s \times s}.
$$

Let Z_t be an $s \times 1$ vector with X_{t-r+i} as its ith component for $i = 1, \cdots, r$ and

$$
\sum_{k=i}^{r} a_k X_{t+i-k} + \sum_{k=i}^{s-r} [c_k + \sum_{j=0}^{P} b_{jk} X_{t+i-k-j}] \varepsilon_{t+i-k}
$$

as its $(r+i)$-th element, $1 \leq i \leq s - r$. Pham (1985, 1993) gave the representation

$$
X_t = H Z_{t-1} + \varepsilon_t, \quad Z_t = (A + B\varepsilon_t) Z_{t-1} + c\varepsilon_t + d\varepsilon_t^2. \tag{26}
$$

By (26), X_t is GMC(α) if $\varepsilon_1 \in \mathcal{L}^{2\alpha}$ and $\mathbb{E}(|A + B\varepsilon_1|_\alpha) < 1$. $\diamondsuit$

REMARK **5.1**. Davidson (2002) considered the bilinear model (25) with $q = 0$ and $Q = 1$. He commented that it is not easy to show G-NED(2) for general cases due to the complexity of moment expressions. In comparison, our argument is simpler.

EXAMPLE **5.4**. Ding et al. (1993) proposed the asymmetric GARCH(r, s) model

$$
X_t = \varepsilon_t \sqrt{h_t}, \ h_t^{\varsigma/2} = \alpha_0 + \sum_{i=1}^{r} \alpha_i (|X_{t-i}| - \gamma X_{t-i})^\varsigma + \sum_{i=1}^{s} \beta_i h_{t-i}^{\varsigma/2}, \tag{27}
$$

where $\alpha_0 > 0$, $\alpha_j \geq 0$ $(j = 1, \cdots, r)$ with at least one $\alpha_j > 0$, $\beta_i \geq 0$ $(i = 1, \cdots, s)$, $\varsigma \geq 0$ and $|\gamma| < 1$. The linear GARCH(r, s) model is a special case of (27) with $\varsigma = 2, \gamma = 0$. Wu and Min (2005) showed GMC for linear GARCH(r, s) models. Let $Z_t = (|\varepsilon_t| - \gamma\varepsilon_t)^\varsigma$, $\xi_{\varsigma t} = (\alpha_0 Z_t, 0, \cdots, \alpha_0, 0, \cdots, 0)'_{(r+s) \times 1}$, of which the $(r+1)$-th

14

element is α_0 and

$$
A_{\varsigma t} = \left(\begin{array}{ccc|ccc}
\alpha_1 Z_t & \cdots & \alpha_r Z_t & \beta_1 Z_t & \cdots & \beta_s Z_t \\
I_{(r-1)\times(r-1)} & O_{(r-1)\times 1} & & O_{(r-1)\times s} & & \\
\hline
\alpha_1 & \cdots & \alpha_r & \beta_1 & \cdots & \beta_s \\
O_{(s-1)\times r} & & & I_{(s-1)\times(s-1)} & O_{(s-1)\times 1} &
\end{array} \right).
$$

Ling and McAleer (2002a) showed that $\mathbb{E}|X_t|^{m\varsigma} < \infty$ for some $m \in \mathbb{N}$ if and only if

$$
\rho\{\mathbb{E}(A_{\varsigma t}^{\otimes m})\} < 1, \tag{28}
$$

where $\otimes$ is the usual Kronecker product. It turns out that (28) also implies GMC($m\varsigma$).

PROPOSITION **5.1**. *For the asymmetric GARCH(r,s) model (27), let $\varepsilon_t \in \mathcal{L}^{m\varsigma}$, $\varsigma \geq 1$, then X_t is GMC($m\varsigma$) if (28) holds.*

Proof. Let $Y_t = [(|X_t| - \gamma X_t)^\varsigma, \cdots, (|X_{t-r+1}| - \gamma X_{t-r+1})^\varsigma, h_t^{\varsigma/2}, \cdots, h_{t-s}^{\varsigma/2}]'$. Then $Y_t = A_{\varsigma t} Y_{t-1} + \xi_{\varsigma t}$ (Ling and McAleer, 2002a). Let Y_0', independent of $\{\varepsilon_t, \ t \in \mathbb{Z}\}$, be an iid copy of Y_0 and we recursively define $Y_t' = A_{\varsigma t} Y_{t-1}' + \xi_{\varsigma t}$, $t \geq 1$; let $\tilde{Y}_t = Y_t - Y_t'$. Then $\tilde{Y}_t = A_{\varsigma t} \tilde{Y}_{t-1}$. Applying the argument of Proposition 3 in Wu and Min (2005), we have

$$
\tilde{Y}_t^{\otimes m} = A_{\varsigma t}^{\otimes m} \tilde{Y}_{t-1}^{\otimes m} = \cdots = A_{\varsigma t}^{\otimes m} \cdots A_{\varsigma 1}^{\otimes m} \tilde{Y}_0^{\otimes m}.
$$

Thus $\mathbb{E}(\tilde{Y}_t^{\otimes m}) = [\mathbb{E}(A_{\varsigma 1}^{\otimes m})]^t \mathbb{E}(\tilde{Y}_0^{\otimes m})$ since $A_{\varsigma t}, \cdots, A_{\varsigma 1}$ are iid. By (28), $|\mathbb{E}(\tilde{Y}_t^{\otimes m})| \leq C\rho^t$ for some $\rho \in (0,1)$. In particular, $\mathbb{E}(|h_t^{\varsigma/2} - (h_t')^{\varsigma/2}|^m)$ is also bounded by $C\rho^t$. So

$$
\mathbb{E}(|X_t - X_t'|^{m\varsigma}) = \mathbb{E}(\varepsilon_t^{m\varsigma})\mathbb{E}(|\sqrt{h_t} - \sqrt{h_t'}|^{m\varsigma}) \leq C\mathbb{E}(|h_t^{\varsigma/2} - (h_t')^{\varsigma/2}|^m) \leq C\rho^t,
$$

where the inequality $|a - b|^\varsigma \leq |a^\varsigma - b^\varsigma|$, $a \geq 0, b \geq 0$, $\varsigma \geq 1$, is applied. $\diamond$

EXAMPLE **5.5**. Let ε_t be iid with mean 0 and variance 1. Consider the signed volatility model (Yao, 2004)

$$
X_t = \varepsilon_t |s_t|^{1/\varsigma}, \ s_t = g(\varepsilon_{t-1}) + c(\varepsilon_{t-1})s_{t-1}, \ \varsigma > 0, \tag{29}
$$

When $s_t = h_t^\varsigma > 0$, (29) reduces to the general GARCH$(1,1)$ model [He and Teräsvirta (1999) and Ling and McAleer (2002b)]

$$X_t = \varepsilon_t h_t, \; h_t^\varsigma = g(\varepsilon_{t-1}) + c(\varepsilon_{t-1}) h_{t-1}^\varsigma, \; \varsigma > 0, \tag{30}$$

We shall show that the model (29) satisfies GMC under some mild conditions.

PROPOSITION **5.2**. *For the signed volatility model (29), suppose that for some $\alpha > 0$, $\mathbb{E}|\varepsilon_1|^{\alpha\varsigma} < 1$, $\mathbb{E}|c(\varepsilon_1)|^\alpha < 1$ and $\mathbb{E}|g(\varepsilon_1)|^\alpha < \infty$. Let $\varsigma \geq 1$, then X_t is GMC($\varsigma\alpha$).*

Proof. By Theorem 5.1, s_t is GMC(α). Since $\mathbb{E}\{(|s_t|^{1/\varsigma} - |s_t'|^{1/\varsigma})^{\varsigma\alpha}\} \leq \mathbb{E}(|s_t - s_t'|^\alpha)$ and $X_t = \varepsilon_t |s_t|^{1/\varsigma}$, X_t is GMC($\varsigma\alpha$). $\diamondsuit$

EXAMPLE **5.6**. Let $\{\varepsilon_t\}$ be iid nonnegative random variables with mean 1. Consider Engle and Russell's (1998) autoregressive conditional duration (ACD) model

$$X_t = \varepsilon_t \Phi_t, \; \Phi_t = \omega + \sum_{i=1}^q \alpha_i X_{t-i} + \sum_{j=1}^p \beta_j \Phi_{t-j}, \tag{31}$$

where $\omega > 0$, $\alpha_i \geq 0, i = 1, \cdots, q$, $\beta_j \geq 0$, $j = 1, \cdots, p$. Let $P = \max(p, q)$, $\alpha_i = 0, i > q$ and $\beta_j = 0, j > p$. Carrasco and Chen (2002) consider the existence of stationary solution in the special case $\alpha = 1$ under $\sum_{i=1}^P (\alpha_i + \beta_i) < 1$.

PROPOSITION **5.3**. *For the ACD(p, q) model (31), suppose that $\varepsilon_t \in \mathcal{L}^\alpha$, $\alpha \geq 1$. Then X_t is GMC(α) if $\sum_{i=1}^P \|\alpha_i \varepsilon_1 + \beta_i\|_\alpha < 1$.*

Proof. Write $\Phi_t = \omega + \sum_{i=1}^P (\alpha_i \varepsilon_{t-i} + \beta_i) \Phi_{t-i}$. Let $(\Phi_j')_{j\leq 0}$, independent of $\{\varepsilon_t, t \in \mathbb{Z}\}$, be an iid copies of $(\Phi_j)_{j\leq 0}$. Define recursively $\Phi_t' = \omega + \sum_{i=1}^P (\alpha_i \varepsilon_{t-i}' + \beta_i) \Phi_{t-i}'$, $\varepsilon_t' = \varepsilon_t$, $t \geq 1$. Let $\tilde{\Phi}_t = \Phi_t - \Phi_t'$. Then for $t \geq P$, $\tilde{\Phi}_t = \sum_{i=1}^P (\alpha_i \varepsilon_{t-i} + \beta_i) \tilde{\Phi}_{t-i}$. So $\|\tilde{\Phi}_t\|_\alpha \leq \sum_{i=1}^P \|\alpha_i \varepsilon_1 + \beta_i\|_\alpha \|\tilde{\Phi}_{t-i}\|_\alpha$. Since $\sum_{i=1}^P \|\alpha_i \varepsilon_1 + \beta_i\|_\alpha < 1$, by Lemma 6.2.10 in Duflo (1997), $\|\tilde{\Phi}_t\|_\alpha \leq C\rho^t, t \in \mathbb{N}$ for some $\rho \in (0, 1)$. In other words, Φ_t is GMC(α). Finally, $\|X_t - X_t'\|_\alpha = \|\varepsilon_t\|_\alpha \|\tilde{\Phi}_t\|_\alpha \leq C\rho^t$. $\diamondsuit$

6 Appendix

We now give the proofs of the results in Sections 2-4.

6.1 Proof of Theorem 2.1.

PROOF. For presentational clarity we restrict $J = \{j_1, \ldots, j_p\} \subset \{1, \ldots, m\}$ and hence Z_{j_l} corresponds to real parts of $S_n(\theta_{j_l})$. The argument easily extends to general cases. Let

$$T_n = \sum_{k=1}^{n} \mu_k X_k, \text{ where } \mu_k = \mu_k(c, J) = \sum_{l=1}^{p} \frac{c_l \cos(k\theta_{j_l})}{\sqrt{\pi f(\theta_{j_l})}}, \quad 1 \leq k \leq n.$$

Since $f_* := \min_{\mathbb{R}} f(\theta) > 0$, there exists μ_* such that $|\mu_k| \leq \mu_*$ for all $c \in \Omega_p$ and $J \in \Xi_{m,p}$. Let $d_n(h) = n^{-1} \sum_{k=1+h}^{n} \mu_k \mu_{k-h}$ if $0 \leq h \leq n-1$ and $d_n(h) = 0$ if $h \geq n$. Note that

$$\sum_{k=1}^{n} \cos(k\theta_{j_l}) \cos[(k+h)\theta_{j_{l'}}] = \frac{n}{2} \cos(h\theta_{j_l}) \mathbf{1}_{j_l = j_{l'}}.$$

Then it is easily seen that there exists a constant $K_0 > 0$ such that for all $h \geq 0$,

$$\tau_n(h) = \sup_{J \in \Xi_{m,p}} \sup_{c \in \Omega_p} \left| d_n(h) - \sum_{l=1}^{p} c_l^2 \frac{\cos(h\theta_{j_l})}{2\pi f(\theta_{j_l})} \right| \leq \frac{K_0 h}{n}.$$

Clearly $\tau_n(h) \leq \mu_* + (2\pi f_*)^{-1} =: K_1$. So we have uniformly over J and c that

$$\left| \frac{\|T_n\|^2}{n} - 1 \right| = \left| d_n(0)\gamma_n(0) + 2\sum_{h=1}^{\infty} d_n(h)\gamma(h) - 1 \right|$$

$$\leq 2\sum_{h=0}^{\infty} \tau_n(h)\gamma(h) \leq \sum_{h=0}^{\infty} K_2 \min(h/n, 1)\gamma(h) \to_{n \to \infty} 0 \quad (32)$$

by the Lebesgue dominated convergence theorem, where $K_2 = 2(K_0 + K_1)$.

Let $\tilde{T}_n = \sum_{k=1}^{n} \mu_k \tilde{X}_k$, where $\tilde{X}_k = \mathbb{E}(X_k | \varepsilon_{k-\ell+1}, \ldots, \varepsilon_k)$ are ℓ-dependent and $\delta_\ell = \|X_0 - \tilde{X}_0\|$. Then $\lim_{\ell \to \infty} \delta_\ell = 0$. If $k < \ell$, then $\mathcal{P}_0 \tilde{X}_k = \mathbb{E}(\mathcal{P}_0 X_k | \varepsilon_{k-\ell+1}, \ldots, \varepsilon_0)$. By Jensen's inequality $\|\mathcal{P}_0 \tilde{X}_k\| \leq \|\mathcal{P}_0 X_k\|$. If $k \geq \ell$, then $\mathcal{P}_0 \tilde{X}_k = 0$. Clearly $\|\mathcal{P}_0(X_k - \tilde{X}_k)\| \leq 2\delta_\ell$. By the Lebesgue dominated convergence theorem, (6) entails that

$$\frac{\|T_n - \tilde{T}_n\|}{\sqrt{n}} = \left[\frac{1}{n} \sum_{j=-\infty}^{n} \|\mathcal{P}_j(T_n - \tilde{T}_n)\|^2 \right]^{1/2} \leq \mu_* \sum_{k=0}^{\infty} \|\mathcal{P}_0(X_k - \tilde{X}_k)\|$$

$$\leq \mu_* \sum_{k=0}^{\infty} 2\min(\|\mathcal{P}_0 X_k\|, \delta_\ell) \to_{\ell \to \infty} 0. \quad (33)$$

17

Let $g_n(r) = r^2 \mathbb{E}[\tilde{X}^2 \mathbf{1}(|\tilde{X}| \geq \sqrt{n}/r)]$. Since $\mathbb{E}(\tilde{X}^2) < \infty$, $\lim_{n\to\infty} g_n(r) = 0$ for any fixed $r > 0$. Note that g_n is nondecreasing in r. Then there exists a sequence $r_n \uparrow \infty$ such that $g_n(r_n) \to 0$. Let $Y_k = \tilde{X}_k \mathbf{1}(|\tilde{X}_k| \leq \sqrt{n}/r_n)$ and $T_{n,Y} = \sum_{k=1}^n \mu_k Y_k$. Then $\|Y_k - \tilde{X}_k\| = o(1/r_n)$. Since $Y_k - \tilde{X}_k$ are ℓ-dependent,

$$\|T_{n,Y} - \tilde{T}_n\| \leq \sum_{a=1}^{\ell} \left\| \sum_{b \leq n, \, \ell|(b-a)} \mu_b(Y_b - \tilde{X}_b) \right\| = o(\sqrt{n}/r_n), \tag{34}$$

where $\ell|(b-a)$ means that ℓ is a divisor of $b - a$. Let $p_n = \lfloor r_n^{1/4} \rfloor$ and blocks $B_t = \{a \in \mathbb{N} : 1 + (t-1)(p_n + \ell) \leq a \leq p_n + (t-1)(p_n + \ell)\}$, $1 \leq t \leq t_n := \lfloor 1 + (n - p_n)/(p_n + \ell) \rfloor$. Define $U_t = \sum_{a \in B_t} \mu_a Y_a$, $V_n = \sum_{t=1}^{t_n} U_t$, $R_n = T_{n,Y} - V_n$, $W = (V_n - \mathbb{E}(V_n))/\sqrt{n}$ and $\Delta = \tilde{T}_n/\sqrt{n} - W$. Then U_t are independent and $\|R_n\| = O(\sqrt{t_n})$ since Y_a are ℓ-dependent. Note that $|\mathbb{E}(V_n)| = O(n)|\mathbb{E}(Y_k)| = o(\sqrt{n}/r_n)$. Then by (34),

$$\sqrt{n}\|\Delta\| \leq |\mathbb{E}(V_n)| + \|V_n - \tilde{T}_n\| = o(\sqrt{n}/r_n) + O(\sqrt{t_n} + \sqrt{n}/r_n) = O(\sqrt{t_n}). \tag{35}$$

Since $|U_t|^3 \leq \mu_*^3 p_n^2 \sum_{a \in B_t} |Y_a|^3$ and $\mathbb{E}(Y_a^2) \leq \mathbb{E}(X_k^2)$, $\mathbb{E}(|U_t|^3) = O(p_n^3 \sqrt{n}/r_n)$. By the Berry-Esseen Theorem (cf Chow and Teicher, 1988),

$$\sup_x |\mathbb{P}(W \leq x) - \Phi(x/\|W\|)| \leq C \frac{\sum_{t=1}^{t_n} \mathbb{E}(|U_t|^3)}{\|V_n - \mathbb{E}(V_n)\|^3} = \frac{O(t_n p_n^3 \sqrt{n}/r_n)}{\sqrt{n}^3} = O(p_n^{-2}). \tag{36}$$

Let $\delta = \delta_n = p_n^{-1/4}$. By (35), (36) and

$$\mathbb{P}(W \leq w - \delta) - \mathbb{P}(|\Delta| \geq \delta) \leq \mathbb{P}(W + \Delta \leq w) \leq \mathbb{P}(W \leq w + \delta) + \mathbb{P}(|\Delta| \geq \delta), \tag{37}$$

we have $\sup_x |\mathbb{P}(\tilde{T}_n \leq \sqrt{n}x) - \Phi(\sqrt{n}x/\|\tilde{T}_n\|)| = O[p_n^{-2} + \mathbb{P}(|\Delta| \geq \delta) + \delta + \delta^2] = O(\delta)$ since $\sup_x |\Phi(x/\sigma_1) - \Phi(x/\sigma_2)| \leq |\sigma_1/\sigma_2 - 1| \sup_x |x\phi(x)|$.

Let $W_1 = \tilde{T}_n/\sqrt{n}$, $\Delta_1 = (T_n - \tilde{T}_n)/\sqrt{n}$ and $\eta = \eta_{\ell,n} = (\|T_n - \tilde{T}_n\|/\sqrt{n})^{1/2}$. We apply (37) with W, Δ replaced by W_1, Δ_1,

$$\sup_x \left| \mathbb{P}\left(\frac{T_n}{\sqrt{n}} \leq x\right) - \Phi\left(\frac{\sqrt{n}x}{\|T_n\|}\right) \right| = O(\mathbb{P}(|\Delta_1| \geq \eta) + \delta + \eta + \eta^2) \tag{38}$$

Thus the conclusion follows from (32) and (33) as we first let $n \to \infty$ and then $\ell \to \infty$.

18

6.2 Proof of Theorem 3.1.

To prove Theorem 3.1, we need to have the following two lemmas.

LEMMA **6.1**. *(Wu and Shao, 2004). Assume (12) with $\alpha = k$ for some $k \in \mathbb{N}$. Then there exists a constant $C > 0$ such that for all $0 \leq m_1 \leq \ldots \leq m_{k-1}$,*

$$|\text{cum}(X_0, X_{m_1}, \ldots, X_{m_{k-1}})| \leq C\rho^{m_{k-1}/[k(k-1)]}. \tag{39}$$

LEMMA **6.2**. *Let the sequence $s_n \in \mathbb{N}$ satisfy $s_n \leq n$ and $B_n = o(s_n)$; let*

$$Y_u = (2\pi)^{-1} \sum_{k=-B_n}^{B_n} X_u X_{u+k} a(kb_n) \cos(k\lambda) \tag{40}$$

Then under (12) and $X_n \in \mathcal{L}^{4+\delta}$, $\delta > 0$, we have $\|\sum_{u=1}^{s_n}\{Y_u - \mathbb{E}(Y_u)\}\|^2 \sim s_n B_n \sigma^2$.

PROOF. By Lemma 6.1, we have the following summability condition

$$\sum_{m_1, m_2, m_3 \in \mathbb{Z}} |\text{cum}(X_0, X_{m_1}, \ldots, X_{m_3})| = \sum_{s=0}^{\infty} O(s^2 \rho^{s/[4(4-1)]}) < \infty.$$

See also Remark 3 in Wu and Shao (2004). Then the lemma easily follows from equations (3.9)-(3.12) in Rosenblatt (1984, page 1174). $\diamondsuit$

PROOF OF THEOREM 3.1. Let $\rho = \rho(4)$, $\alpha_k = a(kb_n)\cos(k\lambda)$ and

$$h_n(\lambda) := (2\pi)^{-1}(nB_n)^{-1/2}\left(\sum_{k=0}^{B_n}\sum_{u=n-k+1}^{n} X_u X_{u+k}\alpha_k + \sum_{k=-B_n}^{-1}\sum_{u=n+k+1}^{n} X_u X_{u+k}\alpha_k\right),$$

then

$$\sqrt{nb_n}\{f_n(\lambda) - \mathbb{E}(f_n(\lambda))\} = (nB_n)^{-1/2}\sum_{u=1}^{n}\{Y_u - \mathbb{E}(Y_u)\} + h_n(\lambda) - \mathbb{E}(h_n(\lambda)). \tag{41}$$

It suffices to show that $(nB_n)^{-1/2}\sum_{u=1}^{n}\{Y_u - \mathbb{E}(Y_u)\} \Rightarrow N(0, \sigma^2)$ by noting that $\|h_n(\lambda)\| = (nB_n)^{-1/2}O(B_n) = o(1)$ which follows from the summability of cumulants for order 2 and 4 [cf Rosenblatt 1985, page 139].

For $k \in \mathbb{Z}$ let $\tilde{X}_k = \mathbb{E}(X_k|\varepsilon_{k-l+1}, \ldots, \varepsilon_k)$, where $l = l_n = \lfloor c \log n \rfloor$ and $c = -4/\log \rho$. Recall (40) for the definition of Y_u and let $\tilde{Y}_u$ be the corresponding sum with X_k replaced by $\tilde{X}_k$. Observe that $\tilde{X}_n$ and $\tilde{X}_m$ are iid if $|n - m| \geq l$ and $\tilde{Y}_u$

19

and $\tilde{Y}_v$ are iid if $|u - v| \geq 2B_n + l$. The independence plays an important role in establishing the asymptotic normality of $\tilde{g}_n = \sum_{u=1}^n \tilde{Y}_u$. Note that

$$\|Y_u - \tilde{Y}_u\| \leq (2\pi)^{-1} \sum_{k=-B_n}^{B_n} \|X_u X_{u+k} - \tilde{X}_u \tilde{X}_{u+k}\| |\alpha_k| = O(B_n \rho^{l/4}). \qquad (42)$$

Now we claim that

$$(nB_n)^{-1/2} \{\tilde{g}_n - \mathbb{E}(\tilde{g}_n)\} \Rightarrow N(0, \sigma^2). \qquad (43)$$

Let q_n, p_n be two sequences of positive integers such that

$$p_n, q_n \to \infty, \ q_n = o(p_n), \ 2B_n + l = o(q_n) \text{ and } k_n = \lfloor n/(p_n + q_n) \rfloor \to \infty. \qquad (44)$$

Define the blocks $\mathcal{L}_r = \{j \in \mathbb{N} : (r-1)(p_n + q_n) + 1 \leq j \leq r(q_n + p_n) - q_n\}$, $1 \leq r \leq k_n$, $\mathcal{S}_r = \{j \in \mathbb{N} : r(p_n + q_n) - q_n + 1 \leq j \leq r(q_n + p_n)\}$, $1 \leq r \leq k_n - 1$ and $\mathcal{S}_{k_n} = \{j \in \mathbb{N} : k_n(p_n + q_n) - q_n + 1 \leq j \leq n\}$. Let $U_r = \sum_{j \in \mathcal{L}_r} \tilde{Y}_j$ and $V_r = \sum_{j \in \mathcal{S}_r} \tilde{Y}_j$. Observe that $U_1, \dots, U_{k_n}$ are iid and $V_1, \dots, V_{k_n-1}$ are also iid. By Lemma 6.2 and (42),

$$\|U_1 - \mathbb{E}(U_1)\| = \left\| \sum_{i=1}^{p_n} \{Y_i - \mathbb{E}(Y_0)\} \right\| + O(p_n \|Y_0 - \tilde{Y}_0\|) \sim (p_n B_n \sigma^2)^{1/2} + O(p_n B_n \rho^{l/4}). \qquad (45)$$

Similarly, $\|V_1 - \mathbb{E}(V_1)\| \sim (q_n B_n \sigma^2)^{1/2} + O(q_n B_n \rho^{l/4})$. By (44) and the independence of the V_i's,

$$\left\| \sum_{i=1}^{k_n} \{V_i - \mathbb{E}(V_i)\} \right\|^2 = (k_n - 1) \|V_1 - \mathbb{E}(V_1)\|^2 + \|V_{k_n} - \mathbb{E}(V_{k_n})\|^2$$
$$= O(k_n q_n B_n) + O[(p_n + q_n) B_n] = o(nB_n).$$

Then (43) follows if $(nB_n)^{-1/2} \sum_{r=1}^{k_n} \{U_r - \mathbb{E}(U_1)\} \Rightarrow N(0, \sigma^2)$. To this end, since the U_r's are iid, by the central limit theorem, it suffices in view of (45) to verify the Liapounov condition. Let $\tau = 2 + \delta/2$. By the triangle and Rosenthal's inequalities

$$\left\| \sum_{u=1}^{p_n} \sum_{k=-B_n}^{-l} \tilde{X}_u \tilde{X}_{u+k} \alpha_k \right\|_\tau \leq \sum_{i=1}^{l} \left\| \sum_{j=1}^{\lfloor (p_n - i)/l \rfloor} \sum_{k=-B_n}^{-l} \tilde{X}_{i+(j-1)l} \tilde{X}_{i+(j-1)l+k} \alpha_k \right\|_\tau$$
$$\leq O(l) \sqrt{p_n/l} \left\| \sum_{k=-B_n}^{-l} \tilde{X}_k \alpha_k \right\|_\tau$$

20

$$\leq O(\sqrt{p_n l}) \sum_{i=0}^{l-1} \left\| \sum_{j=0}^{\lfloor (B_n - l - i)/l \rfloor} \tilde{X}_{-B_n + i + jl} \alpha_{-B_n + i + jl} \right\|_\tau$$
$$= O[(\sqrt{p_n l}) l \sqrt{B_n/l}] = O(\sqrt{p_n B_n} l).$$

On the other hand, since $\tilde{X}_{i+3jl} \tilde{X}_{i+3jl+k}$, $0 \leq j \leq \lfloor (p_n - i)/(3l) \rfloor$, are iid,

$$\left\| \sum_{u=1}^{p_n} \sum_{k=1-l}^{0} \tilde{X}_u \tilde{X}_{u+k} \alpha_k \right\|_\tau \leq \sum_{k=1-l}^{0} \left\| \sum_{u=1}^{p_n} \tilde{X}_u \tilde{X}_{u+k} \alpha_k \right\|_\tau$$
$$\leq \sum_{k=1-l}^{0} \sum_{i=1}^{3l} \left\| \sum_{j=0}^{\lfloor (p_n - i)/(3l) \rfloor} \tilde{X}_{i+3jl} \tilde{X}_{i+3jl+k} \alpha_k \right\|_\tau$$
$$= O(l^2 \sqrt{p_n/l}).$$

Therefore we have $\|U_1\|_\tau = O(\sqrt{p_n B_n} l + l^2 \sqrt{p_n/l})$, implying $\|U_1 - \mathbb{E}(U_1)\|_\tau = O(\sqrt{p_n B_n} l)$. Let $p_n = \lfloor n^{(2+\eta)/3} \rfloor$ and $q_n = \lfloor n^{(1+2\eta)/3} \rfloor$. Then it is easily seen that the Liapounov condition $\|U_1 - \mathbb{E}(U_1)\|_\tau = o[(nB_n)^{1/2} k_n^{-1/\tau}]$ holds. $\diamond$

6.3 Proof of Theorem 3.2.

We adopt the block method. Let $U_r(\lambda), r = 1, \cdots, k_n$ be iid blocks with block length $p = p_n = \lfloor n^{1-\eta}(\log n)^{-2} \rfloor$, $V_r(\lambda), r = 1, \cdots, k_n - 1$ are iid blocks with block length $q = q_n = p_n$. The last block $V_{k_n}(\lambda)$ is negligible. Let $l = l_n = \lfloor -4 \log n / \log \rho(4) \rfloor$ as in the proof of Theorem 3.1. Define $U_r(\lambda)' := U_r(\lambda) \times \mathbf{1}(|U_r(\lambda)| \leq d_n)$ for $r = 1, \cdots, k_n$, where $d_n = \lfloor n^{(1+\eta)/2}(\log n)^{-3} \rfloor$. Before we prove the theorem, we first state a lemma.

LEMMA **6.3**. *Under the assumptions in Theorem 3.2, we have*

$$\sup_{\lambda \in [0,\pi]} \|V_{k_n}(\lambda)\|_1 = O(\sqrt{p_n l} B_n), \tag{46}$$

$$\sup_{\lambda \in [0,\pi]} \|h_n(\lambda)\|_1 = o(1), \tag{47}$$

$$\sup_{\lambda \in [0,\pi]} \mathrm{var}(U_1(\lambda)) = O(p_n B_n), \tag{48}$$

$$\mathrm{var}(U_1(\lambda)') = \mathrm{var}(U_1(\lambda))[1 + o(1)], \tag{49}$$

where the relation $o(1)$ holds uniformly over $[0, \pi]$.

21

PROOF. Let $z = k_n(p+q) + 1 - q$ and $\tau = 2 + \delta/2$. For any $\lambda \in [0, \pi]$,

$$\|V_{k_n}(\lambda)\|_1 \leq C \sum_{j=-B_n}^{B_n} \mathbb{E} \left| \sum_{u=z}^{n} \tilde{X}_u \tilde{X}_{u+j} \right|.$$

For $|j| \leq l$, $\|\sum_{u=z}^{n} \tilde{X}_u \tilde{X}_{u+j}\| = O(\sqrt{p_n l})$ since $\tilde{X}_u \tilde{X}_{u+j}$ is $2l$-dependent. When $|j| > l$, $\|\sum_{u=z}^{n} \tilde{X}_u \tilde{X}_{u+j}\|^2 = \sum_{u,u'=z}^{n} \mathbb{E}(\tilde{X}_u \tilde{X}_{u+j} \tilde{X}_{u'} \tilde{X}_{u'+j}) = O(p_n l)$ since the sum vanishes if $|u - u'| > l$. So $\sup_{\lambda \in [0,\pi]} \|V_{k_n}(\lambda)\|_1 = O(\sqrt{p_n l} B_n)$. Let $\tilde{h}_n(\lambda)$ be the corresponding sum of $h_n(\lambda)$ with $X_u X_{u+k}$ replaced by $\tilde{X}_u \tilde{X}_{u+k}$. As (42), we have $\sup_{\lambda \in [0,\pi]} \|h_n(\lambda) - \tilde{h}_n(\lambda)\|_1 = o(1)$. To show (47), it suffices to show $\sup_{\lambda \in [0,\pi]} \|\tilde{h}_n(\lambda)\|_1 = o(1)$ which follows from a similar argument as in the proof of (46). Regarding (48), we have

$$\begin{aligned}
\operatorname{var}(U_1(\lambda)) &= \left\| \sum_{u=1}^{p} \sum_{k=-B_n}^{B_n} \{X_u X_{u+k} - r(k)\} \alpha_k \right\|^2 \\
&= \sum_{u,u'=1}^{p} \sum_{k,k'=-B_n}^{B_n} \{r(u - u')r(u - u' + k - k') + r(u' - u + k') \\
&\quad \times r(u' - u - k) + \operatorname{cum}(X_0, X_k, X_{u'-u}, X_{u'-u+k'})\} \alpha_k \alpha_{k'} \\
&=: I_1 + I_2 + I_3.
\end{aligned}$$

Note that I_1 is bounded by $\sum_{h=1-p}^{p-1}(p - |h|)|r(h)| \sum_{g=-2B_n}^{2B_n}(2B_n + 1 - |g|)|r(h+g)|$ which is less than $p(2B_n + 1)(\sum_{k=-\infty}^{\infty} |r(k)|)^2$. Similarly, smaller bounds can be obtained for I_2 and I_3 due to the summability of the second and fourth cumulants. Thus $\sup_{\lambda \in [0,\pi]} \operatorname{var}(U_1(\lambda)) = O(p_n B_n)$. For (49), let $v = \operatorname{var}\{U_1(\lambda) - U_1(\lambda)'\}$ and $c = \mathbb{E}(U_1(\lambda)')\mathbb{E}\{U_1(\lambda) - U_1(\lambda)'\}$. Then $\operatorname{var}(U_1(\lambda)') = \operatorname{var}(U_1(\lambda)) - v + 2c$. By Markov's inequality and the order of $\|U_1(\lambda)\|_\tau$ verified in the proof of Theorem 3.1, we have

$$v \leq \|U_1(\lambda)\|_\tau^\tau / d_n^{\tau-2} = O((\sqrt{p_n B_n} l)^\tau (\log n)^{3(\tau-2)} / n^{(1+\eta)(\tau-2)/2}) = o(p_n B_n)$$

and similarly $c \leq \|U_1(\lambda)\|_\tau^{\tau+1} / d_n^{\tau-1} = o(p_n B_n)$, where $o(p_n B_n)$-relation holds uniformly over $\lambda \in [0, \pi]$. By Lemma 6.2 and since f is everywhere positive, (49) follows. $\diamond$

PROOF OF THEOREM 3.2. Let $H_n(\lambda) = \sum_{r=1}^{k_n}[U_r(\lambda) - \mathbb{E}\{U_r(\lambda)\}]$ and $H_n(\lambda)' = \sum_{r=1}^{k_n}[U_r(\lambda)' - \mathbb{E}\{U_r(\lambda)'\}]$. Let $\lambda_j = \pi j / t_n, j = 0, \cdots, t_n, t_n = \lfloor B_n \log(B_n) \rfloor$. By

22

(48) and (49), there exists a finite constant $C_1 > 1$ such that $\sup_{\lambda \in [0,\pi]} \text{var}(U_1(\lambda)') \leq C_1 p_n B_n$. Let $\alpha_n = (C_1 n B_n \log n)^{1/2}$. By Bernstein's inequality, we have

$$
\begin{aligned}
\mathbb{P}\left(\max_{0 \leq j \leq t_n} H_n(\lambda_j)' \geq 2\alpha_n\right) &\leq \sum_{j=0}^{t_n} \mathbb{P}\left(H_n(\lambda_j)' \geq 2\alpha_n\right) \\
&\leq (1+t_n)\exp\left(\frac{-4\alpha_n^2}{2k_n C_1 p_n B_n + 8 d_n \alpha_n}\right) \\
&= o(t_n n^{-1}) = o(1).
\end{aligned}
$$

By Corollary 2.1 in Woodroofe and Van Ness (1967), $\sup_{\lambda \in [0,\pi]} H_n(\lambda) = O_{\mathbb{P}}(\alpha_n)$ holds since

$$
\|U_1(\lambda) - U_1(\lambda)'\|_1 \leq \|U_1(\lambda)\|_\tau^\tau / d_n^{\tau-1} = O((\sqrt{p_n B_n}l)^\tau / d_n^{\tau-1}) = o(\sqrt{nB_n}k_n^{-1}),
$$

consequently $\sum_{r=1}^{k_n} \|U_r(\lambda) - U_r(\lambda)'\|_1 = o((nB_n)^{1/2})$ uniformly over $\lambda \in [0,\pi]$.

Similarly $\sup_{\lambda \in [0,\pi]} \sum_{r=1}^{k_n-1}\{V_r(\lambda) - \mathbb{E}(V_r(\lambda))\} = O_{\mathbb{P}}(\alpha_n)$. Then the conclusion follows from (46), (47) and, by (42), $\sup_{\lambda \in [0,\pi]} \|\tilde{g}_n - \sum_{u=1}^n Y_u\|_1 = o((nB_n)^{1/2})$; see (41). $\diamond$

6.4 Proof of Propositions 4.1 and 4.2

LEMMA **6.4**. (i) Assume (A4'), (A5') and $X_t \in \mathcal{L}^8$. Then $\max_{j,k \leq m} |\text{cov}(I_j^2, I_k^2) - 4f_j^4 \delta_{j,k}| = O(1/n)$, where $m = \lfloor (n-1)/2 \rfloor$, $\delta_{j,k} = 1$ if $j = k$ and 0 otherwise. (ii) Assume (A4'), (A5) and $X_t \in \mathcal{L}^4$. Then $\max_{j,k \leq m} |\text{cov}(I_j, I_k) - f_j^2 \delta_{j,k}| = O(1/n)$.

PROOF. We only show (i) since (ii) can be handled similarly. Note that

$$
\begin{aligned}
\text{cov}(I_j^2, I_k^2) &= \frac{1}{16\pi^4 n^4} \sum_{t_i, s_i \in \{1,\cdots,n\}, i=1,\cdots,4} e^{i(t_1-t_2+t_3-t_4)\lambda_j - i(s_1-s_2+s_3-s_4)\lambda_k} \\
&\quad \times \text{cov}(X_{t_1}X_{t_2}X_{t_3}X_{t_4}, X_{s_1}X_{s_2}X_{s_3}X_{s_4}). \quad (50)
\end{aligned}
$$

By Theorem II.2 in Rosenblatt (1985), we have

$$
\text{cov}(X_{t_1}X_{t_2}X_{t_3}X_{t_4}, X_{s_1}X_{s_2}X_{s_3}X_{s_4}) = \sum_v \text{cum}(X_{i_j}; i_j \in v_1)\cdots\text{cum}(X_{i_j}; i_j \in v_p),
$$

where $\sum_v$ is over all indecomposable partitions $v = v_1 \cup \cdots \cup v_p$ of the two-way table,

$$
\begin{array}{cccc}
X_{t_1}(+) & X_{t_2}(-) & X_{t_3}(+) & X_{t_4}(-) \\
X_{s_1}(-) & X_{s_2}(+) & X_{s_3}(-) & X_{s_4}(+).
\end{array}
$$

23

The sign in the above table is from the exponential term in the sum (50). Since $\mathbb{E}(X_t) = 0$, only partitions v with $\#v_i > 1$ for all i contribute. One of the many indecomposable partitions consisting only of pairs with $+$ in t matched to $-$ in s (say, $\{(t_1, s_1), (t_2, s_2), (t_3, s_3), (t_4, s_4)\}$) leads to the sum $[A(\lambda_j, \lambda_k)]^4$, where

$$A(\lambda_j, \lambda_k) = \frac{1}{2\pi n} \sum_{t_1, s_1 = 1}^{n} r(t_1 - s_1) e^{it_1 \lambda_j - is_1 \lambda_k} = f(\lambda_j) \mathbf{1}_{j=k} + O(1/n).$$

The other indecomposable partitions consisting entirely of pairs (with $+$ in t matched to $-$ in s) are $\{(t_1, s_3), (t_2, s_2), (t_3, s_1), (t_4, s_4)\}$, $\{(t_1, s_1), (t_2, s_4), (t_3, s_3), (t_4, s_2)\}$ and $\{(t_1, s_3), (t_2, s_4), (t_3, s_1), (t_4, s_2)\}$. It is easily seen after some calculations that partitions containing entirely pairs but with at least one $+$ in t matched to one $+$ in s result in a term of order $O(1/n)$ for any j, k. All other partitions that are not all pairs will give a quantity of order $O(1/n)$ due to the summability of cumulants up to the eighth order. Finally, it is not hard to see that $O(1/n)$ does not depend on (j, k). Thus the conclusion is established. $\qquad \diamondsuit$

LEMMA **6.5**. *Assume $X_t \in \mathcal{L}^8$, (A2), (A3'), (A4') and (A5'). Then $\mathrm{var}^*(\varepsilon_1^*) \to 1$ in probability and $\mathbb{E}^*(|\varepsilon_1^*|^4) = O_\mathbb{P}(1)$.*

PROOF. By (A3'), $\tilde{f}$ is a uniformly consistent estimate of f. It then suffices to show

$$\frac{1}{N} \sum_{j=1}^{N} \frac{I_j}{f_j} \to 1, \quad \frac{1}{N} \sum_{j=1}^{N} \frac{I_j^2}{f_j^2} \to 2 \quad \text{in probability and} \quad \frac{1}{N} \sum_{j=1}^{N} \frac{I_j^4}{f_j^4} = O_\mathbb{P}(1).$$

By Proposition 10.3.1 in Brockwell and Davis (1991) and Lemma 6.4, we have $\mathbb{E}(I_j) = f_j + o(1)$ and $\mathbb{E}(I_j^2) = 2f_j^2 + o(1)$ uniformly in j. Thus the first two assertions follow from Lemma 6.4 since their variances go to 0 as $n \to \infty$. By Lemma 6.4, $\mathbb{E}(I_j^4) = \mathrm{cov}(I_j^2, I_j^2) + (\mathbb{E}I_j^2)^2 = 8f_j^4 + o(1)$ uniformly in j, the last assertion holds. $\qquad \diamondsuit$

REMARK **6.1**. For linear processes, FH remarked that their consistency result strongly depends on the asymptotic normality of f_n and the weak convergence of $F_{\tilde{I},m}(x)$ (see Corollary 2.2). The latter condition holds under $\varepsilon_1 \in \mathcal{L}^5$ and (17) by Chen and Hannan (1980). FH further conjectured that the result is presumably correct, assuming only $\mathbb{E}(\varepsilon_1^4) < \infty$, under which the weak convergence of $F_{\tilde{I},m}(x)$

24

might be true. However, it seems from our argument (see the proof of Proposition 4.1) that it is not the weak convergence of $F_{\tilde{I},m}(x)$ but the following two conditions that play key roles; compare Proposition A1 in FH:

$$\frac{1}{N}\sum_{j=1}^{N}\frac{I_j}{f_j}\to 1 \quad\text{and}\quad \frac{1}{N}\sum_{j=1}^{N}\frac{I_j^2}{f_j^2}\to 2 \text{ in probability.}$$

The proof of the second assertion above (see Lemma 6.4, 6.5) in a general setting needs the stronger eighth moment assumption. $\diamond$

Let $\tilde{r}_2(k) = \int_0^{2\pi}\tilde{f}^2(\lambda)e^{ik\lambda}d\lambda$, $r_2(k) = \int_0^{2\pi}f^2(\lambda)e^{ik\lambda}d\lambda$, $\tilde{r}(k) = \int_0^{2\pi}\tilde{f}(\lambda)e^{ik\lambda}d\lambda$ and $F_n^+ = \{1,\cdots,\lfloor n/2\rfloor\}$. By (A3), $\max_{k\in\mathbb{Z}}|\tilde{r}_2(k) - r_2(k)| \le 2\pi\max_\lambda|\tilde{f}^2(\lambda) - f^2(\lambda)| = o_{\mathbb{P}}(b_n)$.

PROOF OF PROPOSITION 4.1. By Lemma 8.3 of Bickel and Freedman (1981), the convergence under the d_2 metric is equivalent to weak convergence and convergence of the first two moments. By (A6), it suffices to show that

$$nb_n\text{var}^*(f_n^*(\lambda)) \to \sigma^2(\lambda) \text{ and } \mathcal{L}(V_n^*(\lambda)|X_1,\cdots,X_n) \Rightarrow N(0,\sigma^2(\lambda)) \text{ in probability.}$$
(51)

Let $\Delta_j = \sum_{k=-B_n}^{B_n}a(kb_n)e^{-ik\lambda}(e^{ik\omega_j} + e^{-ik\omega_j})$. Since the re-sampled residuals $\{\varepsilon_j^*\}$ are iid given $X_1,\cdots,X_n$, we have $\text{var}^*(I_j^*) = \tilde{f}_j^2\text{var}^*(\varepsilon_1^*)$, and, since $I_0^* = 0$, $nb_n\text{var}^*(f_n^*(\lambda)) = \text{var}^*(\varepsilon_1^*)R_n(\lambda) + o_{\mathbb{P}}(1)$, where

$$
\begin{aligned}
R_n(\lambda) &= \frac{nb_n}{n^2}\sum_{j\in F_n^+}\tilde{f}_j^2\Delta_j^2 \\
&= \frac{1}{nB_n}\sum_{k,k'=-B_n}^{B_n}a(kb_n)a(k'b_n)e^{-i\lambda(k-k')}\sum_{j\in F_n}\tilde{f}_j^2\{e^{i\omega_j(k-k')} + e^{i\omega_j(k+k')}\} \\
&\quad + o_{\mathbb{P}}(1) \\
&= \frac{1}{2\pi B_n}\sum_{k,k'=-B_n}^{B_n}a(kb_n)a(k'b_n)e^{-i\lambda(k-k')}\{\tilde{r}_2(k-k') + \tilde{r}_2(k+k')\} + o_{\mathbb{P}}(1) \\
&= \frac{1}{2\pi B_n}\sum_{k,k'=-B_n}^{B_n}a(kb_n)a(k'b_n)e^{-i\lambda(k-k')}\{r_2(k-k') + r_2(k+k')\} + o_{\mathbb{P}}(1) \\
&= R_n^{(1)}(\lambda) + R_n^{(2)}(\lambda) + o_{\mathbb{P}}(1) \text{ (say) .}
\end{aligned}
$$
(52)

25

Let $\beta_n(k) := \int_0^{2\pi} R_n^{(1)}(\lambda)e^{ik\lambda}d\lambda$, $\beta(k) := \int_0^{2\pi} \int_{-1}^1 a^2(u)f^2(\lambda)e^{ik\lambda}dud\lambda$. Then, for each k,

$$\beta_n(k) = \frac{1}{B_n}r_2(k) \sum_{i=\max(-B_n,-B_n+k)}^{\min(B_n,B_n+k)} a(ib_n)a((i-k)b_n) \to r_2(k)\int_{-1}^1 a^2(u)du.$$

Since $|\beta_n(k)| \le C|r_2(k)|$ and $\sum_{k\in\mathbb{Z}}|r_2(k)| < \infty$, by the Lebesgue dominated convergence theorem, $R_n^{(1)}(\lambda) \to f^2(\lambda)\int_{-1}^1 a^2(u)du$. For $R_n^{(2)}(\lambda)$, $\lambda \neq 0, \pm\pi$, we have

$$
\begin{aligned}
R_n^{(2)}(\lambda) &= \frac{1}{2\pi B_n} \sum_{h=-2B_n}^{2B_n} r_2(h)e^{ih\lambda} \sum_{k=\max(-B_n,-B_n+h)}^{\min(B_n,B_n+h)} a(kb_n)a((k-h)b_n)e^{-2ik\lambda} \\
&= \frac{1}{2\pi B_n} \sum_{h=-2B_n}^{2B_n} r_2(h)e^{ih\lambda}O(1) \to 0.
\end{aligned}
$$

It is easily seen that $R_n^{(1)}(\lambda) = R_n^{(2)}(\lambda)$ when $\lambda = 0, \pm\pi$. Hence by (52) and Lemma 6.5, $nb_n\text{var}^*(f_n^*(\lambda)) \to \sigma^2(\lambda)$ in probability.

Finally, since $\{\varepsilon_j^*\}$ are iid conditional on $\{X_1, \cdots, X_n\}$, by the Berry-Esseen Theorem and Lemma 6.5, we have

$$
\begin{aligned}
\sup_x \left| \mathbb{P}^*(V_n^*(\lambda) \le x) - \Phi\left(\frac{x}{nb_n\text{var}^*(f_n^*(\lambda))}\right) \right| &\le \frac{C\sum_{j\in F_n^+} \tilde{f}_j^4 \mathbb{E}^*|\varepsilon_1^*|^4 \Delta_j^4}{[\sum_{j\in F_n^+} \tilde{f}_j^2 \text{var}^*(\varepsilon_1^*)\Delta_j^2]^2} \\
&= O_{\mathbb{P}}\left(\frac{nB_n^4}{n^2 B_n^2}\right),
\end{aligned}
$$

which implies $\mathcal{L}(V_n^*(\lambda)|X_1, \cdots, X_n) \Rightarrow N(0, \sigma^2(\lambda))$ in probability since $B_n^2 = o(n)$ and $\sup_x |\Phi(x/\sigma_1) - \Phi(x/\sigma_2)| \le |\sigma_1/\sigma_2 - 1| \sup_x |x\phi(x)|$. Here $\mathbb{P}^*$ denotes the conditional probability given the original sample. $\diamond$

PROOF OF PROPOSITION 4.2. Since $\tilde{r}(k) = a(k\tilde{b}_n)r_k^{(n)}$, $|k| \le \tilde{B}_n$ and 0 otherwise, for

$$J_n(\lambda) = \frac{B_n^2}{2\pi} \sum_{k=-\tilde{B}_n}^{\tilde{B}_n} a(k\tilde{b}_n)r_k^{(n)}e^{-ik\lambda}(a(kb_n)-1),$$

we have $\mathbb{E}^* f_n^*(\lambda) - \tilde{f}(\lambda) = J_n(\lambda)B_n^{-2} + o_{\mathbb{P}}(B_n^{-2})$ in view of $B_n/n = o(B_n^{-2})$. It remains to show that $\mathbb{E}(J_n(\lambda)) \to c_2 f''(\lambda)$ and $\text{var}(J_n(\lambda)) \to 0$. Under (A1),

(A4)-(A5),

$$
\begin{aligned}
\mathbb{E}(J_n(\lambda)) &= \frac{B_n^2}{2\pi} \sum_{k=-\tilde{B}_n}^{\tilde{B}_n} a(k\tilde{b}_n)e^{-ik\lambda}(1-|k|/n)r(k)(a(kb_n)-1) \\
&= -\frac{B_n^2}{2\pi} \sum_{k=-\tilde{B}_n}^{\tilde{B}_n} a(k\tilde{b}_n)e^{-ik\lambda}r(k)k^2b_n^2c_2(1+o(1)) \to c_2 f''(\lambda)
\end{aligned}
$$

and

$$
\begin{aligned}
\mathrm{var}(J_n(\lambda)) &= \frac{B_n^4}{4\pi^2} \sum_{k,k'=-\tilde{B}_n}^{\tilde{B}_n} a(k\tilde{b}_n)a(k'\tilde{b}_n)(a(kb_n)-1)(a(k'b_n)-1) \\
&\quad \times e^{-i(k-k')\lambda}\mathrm{cov}(r_k^{(n)}, r_{k'}^{(n)}) \\
&= \frac{(1+o(1))c_2^2}{4\pi^2 n^2} \sum_{k,k'=-\tilde{B}_n}^{\tilde{B}_n} a(k\tilde{b}_n)a(k'\tilde{b}_n)k^2 k'^2 e^{-i(k-k')\lambda} \\
&\quad \times \sum_{t=1}^{n-|k|} \sum_{t'=1}^{n-|k'|} \mathrm{cov}(X_t X_{t+|k|}, X_{t'} X_{t'+|k'|}) \\
&= O(\tilde{B}_n^4/n^2) \sum_{k,k'=0}^{\tilde{B}_n} \sum_{t=1}^{n-k} \sum_{t'=1}^{n-k'} |\mathrm{cov}(X_t X_{t+k}, X_{t'} X_{t'+k'})|. \quad (53)
\end{aligned}
$$

Note that $\mathrm{cov}(X_t X_{t+k}, X_{t'} X_{t'+k'}) = r(t-t')r(t+k-t'-k')+r(t-t'-k')r(t'-t-k)+\mathrm{cum}(X_t, X_{t+k}, X_{t'}, X_{t'+k'})$. The contribution of the first term $r(t-t')r(t+k-t'-k')$ to (53) is $O(\tilde{B}_n^5/n) \sum_{h=-\tilde{B}_n}^{\tilde{B}_n} \sum_{s=-2n}^{2n} |r(h)r(h+s)| = O(\tilde{B}_n^5/n) = o(1)$ since $\sum_{k \in \mathbb{Z}} |r(k)| < \infty$. Similarly, the contribution of the second term to (53) approaches zero as $n \to \infty$. The third term is $O(\tilde{B}_n^4/n) = o(1)$ due to the summability of the fourth cumulants. $\diamond$

ACKNOWLEDGMENTS

REFERENCES

ANDREWS, D. W. K. (1984). Non-strong mixing autoregressive processes. *J. Appl. Probab.* **21** 930-934.

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

27

BICKEL, P. J. and FREEDMAN, D. A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* **9** 1196-1217.

BRILLINGER, D. R. (1969). Asymptotic properties of spectral estimates of second order. *Biometrika* **56** 375-390.

BRILLINGER, D. R. (1975). *Time Series: Data Analysis and Theory.* Holden-Day, San Francisco.

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

CARRASCO, M. and CHEN, X. (2002). Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory* **18** 17-39.

CHAN, K. S. and TONG, H. (1985). On the use of the deterministic lyapunov function for the ergodicity of stochastic difference equations. *Adv. Appl. Prob.* **17** 666-678.

CHANDA, K. C. (2005). Large sample properties of spectral estimators for a class of stationary nonlinear processes. *J. Time Ser. Anal.* **26** 1-16.

CHEN, Z. G. and HANNAN, E. J. (1980). The distribution of the periodogram ordinates. *J. Time Ser. Anal.* **1** 73-82.

CHOW, Y. S. and TEICHER, H. (1988). *Probability Theory*, 2nd ed. Springer, New York.

DAVIDSON, J. (1994). *Stochastic limit Theory*, Oxford University Press, Oxford.

DAVIDSON, J. (2002). Establishing conditions for the functional central limit theorem in nonlinear and semiparametric time series processes. *J. Econometrics.* **106** 243-269.

DIACONIS, P. and FREEDMAN, D. (1999). Iterated random functions. *SIAM Rev.* **41** 41–76.

DING, Z., GRANGER, C. and ENGLE, R. (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance* **1** 83-106.

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

ELTON, J. H. (1990). A multiplicative ergodic theorem for Lipschitz maps. *Stochastic Process. Appl.* **34** 39-47.

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

ENGLE, R. F. and RUSSELL, J. (1998). Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica.* **66** 1127-1162.

FAN, J. and YAO, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods.* Springer-Verlag, New York.

FEIGIN, P. D. and TWEEDIE, R. L. (1985). Random coefficient autoregressive processes: A Markov chain analysis of stationarity and finiteness of moments. *J. Time Ser. Anal.* **6** 1-14.

FRANKE, J. and HÄRDLE, W. (1992). On bootstrapping kernel spectral estimates. *Ann. Statist.* **8** 121-145.

FREEDMAN, D. and LANE, D. (1980). The empirical distribution of the Fourier coefficients. *Ann. Statist.* **8** 1244-1251.

FREEDMAN, D. and LANE, D. (1981). The empirical distribution of the Fourier coefficients of a sequence of independent, identically distributed long-tailed random variables. *Z. Wahrsch. Verw. Geb* **58** 21-39.

GÖTZE, F. and HIPP, C. (1994). Asymptotic distribution of statistics in time series. *Ann. Statist.* **22** 2062-2088.

GRANGER, C. W. J. and ANDERSON, A. P. (1978). *An Introduction to Bilinear Time Series Models.* Gottinger: Vandenhoek and Ruprecht.

GRENANDER, U. and ROSENBLATT, M. (1957). *Statistical Analysis of Stationary Time Series.* Wiley, New York.

HANNAN, E. J. (1973). Central limit theorems for time series regression. *Z. Wahrsch. Verw. Geb* **26** 157-170.

HE, C. and TERÄSVIRTA, C. (1999). Properties of moments of a family of GARCH processes. *J. Econometrics.* **92** 173-192.

HSING, T. and WU, W. B. (2004). On weighted U-statistics for stationary processes. *Ann. Probab.* **32** 1600-1631.

HURVICH, C. M. and ZEGER, S. L. (1987). Frequency domain bootstrap methods for time series. Technical Report 87-115, Graduate School of Business Administration, New York Univ.

JARNER, S. and TWEEDIE, R. (2001). Local contracting iterated random functions and stability of Markov chains. *J. Appl. Prob.* **38** 494-507.

JONES, D. A. (1976). Non-linear autoregressive processes. Unpublished Ph.D.

Thesis, University of London.

KOKOSZKA, P. and MIKOSCH, T. (2000). The periodogram at the Fourier frequencies. *Stochastic Process. Appl.* **86** 49-79.

KREISS, J. P. and PAPARODITIS, E. (2003). Autoregressive-aided periodogram bootstrap for time series. *Ann. Statist.* **31** 1923-1955.

LAHIRI, S. N. (2003). A necessary and sufficient condition for asymptotic independence of discrete Fourier transforms under short- and long-range dependence. *Ann. Statist.* **31** 613-641.

LI, W. K., LING, S. and MCALEER, M. (2002). Recent theoretical results for time series models with GARCH errors. *Journal of Economic Surveys* **16** 245-269.

LING, S. and MCALEER, M. (2002a). Necessary and sufficient moment conditions for the GARCH(r, s) and asymmetric power GARCH(r, s) models. *Econometric Theory* **18** 722-729.

LING, S. and MCALEER, M. (2002b). Stationarity and the existence of moments of a family of GARCH processes. *J. Econometrics.* **106** 109-117.

MEYN, S. P. and TWEEDIE, R. L. (1993). *Markov chains and stochastic stability.* Springer-Verlag, London.

MIN, W. (2004). *Inference on time series driven by dependent innovations.* Unpublished Ph.D. Thesis, University of Chicago.

NORDGAARD, A. (1992). Resampling stochastic processes using a bootstrap approach. In *Lecture Notes in Econom. and Math. Systems* **376** 181-185.

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

PAPARODITIS, E. and POLITIS, D. N. (1999). The local bootstrap for periodogram statistics. *J. Time Ser. Anal.* **20** 193-222.

PARZEN, E. (1957). On consistent estimates of the spectrum of a stationary time series. *Ann. Math. Statist.* **28** 329-348.

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

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

PHAM, D. T. (1993). Bilinear time series models. In *Dimension Estimation and*

Models (H.Tong, ed.). World Scientific, Singapore.

PRIESTLEY, M. B. (1981). *Spectral Analysis and Time Series 1.* Academic, New York.

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

ROSENBLATT, M. (1984). Asymptotic normality, strong mixing, and spectral density estimates. *Ann. Probab.* **12** 1167-1180.

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

RUDZKIS, R. (1985). Distribution of the maximum deviation of an estimate of the spectral density of a gaussian stationary sequence. *Lit. Matem. Sb.* **25** 118-130 (In Russian).

SLUTSKY, E. (1929). Sur l'extension de la théorie de periodogrammes aux suites des quantités dépendentes. *Comptes Rendues.* **189** 722-733.

SLUTSKY, E. (1934). Alcuni applicazioni di coefficienti di Fourier al analizo di sequenze eventuali coherenti stazionarii. *Giorn. d. Instituto degli Atuari.* **5** 435-482.

SUBBA RAO, T. and GABR, M. M. (1984). *An Introduction to Bispectral Analysis and Bilinear Time Series Models.* Lecture Notes in Statistics, 24. New York: Springer-Verlag.

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.

THEILER, J., PAUL, L. S. and RUBIN, D. M. (1994). Detecting nonlinearity in data with long coherence times. In *Time Series Prediction* (A. Weigend and N. Gershenfeld, eds.). Addison-Wesley, Reading, MA.

TONG, H. (1990). *Non-linear Time Series: A Dynamical System Approach.* Oxford University Press.

TWEEDIE, R. L. (1975). Sufficient conditions for ergodicity and recurrence of Markov chains on a general state space. *Stochatic Process. Appl.* **3** 385-403.

TWEEDIE, R. L. (1976). Criteria for classifying general Markov chains. *Adv. Appl. Prob.* **8** 737-771.

TWEEDIE, R. L. (1988). Invariant measures for Markov chains with no irreducibil-

ity assumptions. *J. Appl. Probab.* **25A** 275-285.

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

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.

WOODROOFE, M. and VAN NESS, J. W. (1967). The maximum deviation of sample spectral densities. *Ann. Math. Statist.* **38** 1558-1569.

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

WU, W. B. and MIN, W. (2005). On linear processes with dependent innovations. To appear *Stochastic Process. Appl.*

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

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

YAO, J.-F. (2004). Stationarity, tails and moments of univariate GARCH processes. Preprint.

ZHURBENKO, I. G. and ZUEV, N. M. (1975). On higher spectral densities of stationary processes with mixing. *Ukrain. Matem. Z.* **27** 452-46 (English translation).

DEPARTMENT OF STATISTICS
THE UNIVERSITY OF CHICAGO
5734 S. UNIVERSITY AVENUE, CHICAGO, IL 60637
E-MAIL: shao@galton.uchicago.edu, wbwu@galton.uchicago.edu.