Econometric Theory, **23**, 2007, 930–951. Printed in the United States of America.
DOI: 10.1017/S0266466607070399

A LIMIT THEOREM FOR QUADRATIC FORMS AND ITS APPLICATIONS

Wei Biao Wu
University of Chicago

Xiaofeng Shao
University of Illinois at Urbana-Champaign

We consider quadratic forms of martingale differences and establish a central limit theorem under mild and easily verifiable conditions. By approximating Fourier transforms of stationary processes by martingales, our central limit theorem is applied to the smoothed periodogram estimate of spectral density functions. Our results go beyond earlier ones by allowing a variety of nonlinear time series and by avoiding strong mixing and/or summability conditions on joint cumulants.

1. INTRODUCTION

Let $(X_k)_{k\in\mathbb{Z}}$ be a real stationary process with mean 0 and finite covariances $\gamma(k) = \mathbb{E}(X_0X_k)$, $k \in \mathbb{Z}$. Assume that the covariances are absolutely summable. Then the spectrum or spectral density function

$$f(\lambda) = \frac{1}{2\pi}\sum_{k\in\mathbb{Z}}\gamma(k)\cos(k\lambda), \qquad \lambda \in [0, 2\pi] \tag{1}$$

exists and is continuous and finite. A fundamental problem in time series analysis is to estimate *f*. Given the observations $X_1, \ldots, X_n$, let the periodogram

$$I_n(\lambda) = \frac{1}{2\pi n}|S_n(\lambda)|^2, \quad \text{where } S_n(\lambda) = \sum_{t=1}^{n} X_t e^{t\lambda\sqrt{-1}}. \tag{2}$$

Here $S_n(\lambda)$ is the Fourier transform of the sample $X_1, \ldots, X_n$ and $\sqrt{-1}$ is the imaginary unit. It is well known that $I_n(\lambda)$ is an asymptotic unbiased but inconsistent estimate of $f(\lambda)$ (see, e.g., Anderson, 1971; Brillinger, 1975; Brockwell and Davis, 1991; Hannan, 1970; Priestley, 1981). Denote by $\lambda_j = \lambda_{j,n} = 2\pi j/n$, $j \in \mathbb{Z}$, the Fourier frequencies. Under suitable conditions on the underlying process (X_k), $I_n(\lambda_j)$ are asymptotically independent at different Fourier fre-

We thank the two reviewers for their detailed comments, which led to substantial improvements. The work is supported in part by NSF grant DMS-0478704. Address correspondence to Wei Biao Wu, Department of Statistics, The University of Chicago, 5734 S. University Ave., Chicago, IL, 60637, USA; e-mail: wbwu@galton.uchicago.edu.

  0266-4666/07 $15.00

quencies. To obtain a consistent estimate of the spectral density at a given frequency $\theta \in [0, \pi]$, one can naturally smooth the periodograms over Fourier frequencies near θ. Namely, the following weighted periodogram estimate can be used:

$$\hat{f}_n(\theta) = \frac{1}{K_m} \sum_{k=-m}^{m} K(k/m) I_n(\lambda_{k+k_0}), \tag{3}$$

where $k_0 = \lfloor n\theta/(2\pi) \rfloor$, K is a nonnegative kernel function, and $K_m = \sum_{k=-m}^{m} K(k/m)$. Here $\lfloor x \rfloor$ denotes the integer part of the real number x, namely, $\lfloor x \rfloor = \max\{k \in \mathbb{Z} : k \le x\}$. Generically (3) is called the smoothed periodogram spectral density estimate. See Robinson (1983) for a review of various approaches to spectral density estimation. The smoothing parameter $m = m_n$ is chosen such that

$$m_n^{-1} + m_n n^{-1} \to 0. \tag{4}$$

Under appropriate conditions on (X_k), (4) ensures that $\hat{f}_n$ is a consistent estimate of f. As a significant merit, $\hat{f}_n$ can be quickly computed via the fast Fourier transforms. On the other hand, however, it has been a notoriously difficult problem to establish a central limit theorem of the estimate under the natural bandwidth condition (4). Such asymptotic results are certainly needed in the related statistical inference in the frequency domain, such as hypothesis testing and the construction of confidence intervals. In previous results the process (X_k) is assumed to have very special structures. For example, Lomnicki and Zaremba (1959) and Hannan (1970) deal with linear processes; Brillinger (1969) assumes that X_k has finite moment of all order and the joint cumulants of the process are summable of any order; also see Dahlhaus (1985). Rosenblatt (1984) considers strong mixing processes. See also Bentkus and Rudzkis (1982) for Gaussian processes and Henry (2001) for linear processes with martingale difference innovations.

An important problem in econometrics is to estimate the long-run variance, which basically corresponds to the spectral density evaluated at zero frequency. A closely related problem in the multivariate case is to study the so-called heteroskedasticity and autocorrelation consistent (HAC) covariance matrix estimate. Asymptotic results have been obtained by Andrews (1991), Hansen (1992), de Jong and Davidson (2000), and Jansson (2002) among others. The HAC estimation plays an important role in various econometrics problems, such as unit root tests, robust hypothesis testing, and cointegration estimation. For recent developments see Phillips, Sun, and Jin (2006a, 2006b).

In this paper we obtain a central limit theorem of $\hat{f}_n$ for a very general class of stationary processes under conditions stronger than (4) (see expressions (38) and (39) in Section 3.2). In particular, our results allow linear processes and nonlinear processes including threshold, bilinear, and exponential autoregressive processes among others. The imposed condition on (X_k) is directly related to the data-generating mechanism of the process and hence is easily verifiable. Additionally, we do not need the summability conditions on joint cumulants and/or strong mixing conditions. Conditions of the latter two types do not seem to be tractable in many applications. We shall study asymptotic properties of the quadratic form

$$Q_n = \sum_{t,t'=1}^{n} a_n(t,t')X_t X_{t'}, \tag{5}$$

where the weights $a_n(t,t')$, $1 \le t, t' \le n$, $n = 1,2,\dots$, are (deterministic) real coefficients. The form of Q_n is very general. An important special case of (5) is the smoothed periodogram spectral density estimate (3). Elementary calculations show that (3) is of the form (5) with $a_n(t,t') = a_n(t-t')$, where

$$a_n(j) = \frac{1}{2\pi n K_m} \sum_{k=-m}^{m} K(k/m)\cos(j\lambda_{k+k_0}). \tag{6}$$

It would certainly be impossible to obtain limit theorems for Q_n without imposing suitable conditions on $a_n(t,t')$ and (X_t). In Section 2 we assume that the process (X_t) is a stationary martingale difference sequence. Using the idea of martingale approximation, we show in Section 3 that, for stationary processes (X_t) that may not necessarily be martingale differences, the smoothed periodogram spectral density estimation (3) can be approximated by the quadratic form (5) of martingale differences under mild conditions.

The asymptotic problem of Q_n has a long history. See Whittle (1964), de Wet and Venter (1973), ten Vregelaar (1990), Varberg (1966), Mikosch (1991), Basalykas (1994), and Götze and Tikhomirov (1999), and references cited therein. Results of this sort have many applications in statistics. In all those works X_k are assumed to be independent and identically distributed (i.i.d.), and/or the weights $a_n(t,t')$ do not depend on n. Both assumptions are violated in our setting. In particular, in our problem the independence assumption is too restrictive, and we have to resort to other powerful and more versatile methods.

In a series of recent papers (see, e.g., Wu and Mielniczuk, 2002; Wu, 2005a, 2005b; Wu and Woodroofe, 2004; Hsing and Wu, 2004), we argue that the method of martingale approximation is quite useful in a variety of asymptotic problems. In particular, Hsing and Wu (2004) established an asymptotic theory for the weighted U-statistics

$$U_n = \sum_{t,t'=1}^{n} a_{|t-t'|} G(X_t, X_{t'}) \tag{7}$$

for a quite general class of stationary processes including linear processes and many widely used nonlinear time series. However, even though the bivariate function $G(u,v)$ in (7) can be specifically chosen to be the product uv, the result in Hsing and Wu (2004) is not directly applicable to Q_n. The major difficulty is that the weights $a_n(t,t')$ in (5) depend on n. The dependence of $a_n(t,t')$ on n makes the asymptotic problem of Q_n considerably more challenging. Nonetheless the method of martingale approximation used in Hsing and Wu (2004) sheds new light on asymptotic properties of Q_n. In this paper we shall apply a modified version of the method in Hsing and Wu (2004).

Throughout the paper let $\Rightarrow$ denote convergence in distribution and $N(\mu,\sigma^2)$ the Gaussian distribution with mean μ and variance σ^2. The notation C stands for a generic constant that may vary from line to line. For a random variable ξ write $\xi \in \mathcal{L}^p$ $(p > 0)$ if $\|\xi\|_p := [\mathbb{E}(|\xi|^p)]^{1/p} < \infty$ and $\|\xi\| = \|\xi\|_2$. Denote the real part of a complex number a by $\Re(a)$.

Suitable structural assumptions on the process (X_k) are certainly needed. Here we assume that (X_k) is a stationary causal process of the form

$$X_k = g(\ldots,\varepsilon_{k-1},\varepsilon_k), \tag{8}$$

where $\varepsilon_k, k \in \mathbb{Z}$, are i.i.d. random variables and g is a measurable function for which X_k is a properly defined random variable. The class of processes that (8) represents is huge; see Rosenblatt (1971), Kallianpur (1981), and Tong (1990, p. 204).

Let the shift process $\mathcal{F}_k = (\ldots,\varepsilon_{k-1},\varepsilon_k)$. For $\xi \in \mathcal{L}^1$ define the projection operator

$$\mathcal{P}_i\xi = \mathbb{E}(\xi|\mathcal{F}_i) - \mathbb{E}(\xi|\mathcal{F}_{i-1}), \qquad i \in \mathbb{Z}. \tag{9}$$

The projection operator plays an important role in the study of Q_n. Clearly $\mathcal{P}_i\xi$, $i \in \mathbb{Z}$, are martingale differences, and hence they are orthogonal in $\mathcal{L}^2$ if $\xi \in \mathcal{L}^2$.

The rest of the paper is structured as follows. Section 2 presents a central limit theorem for Q_n for martingale differences. Applications to the smoothed periodogram spectral density estimate (3) are given in Section 3. Some proofs are collected in the Appendix.

2. QUADRATIC FORMS FOR MARTINGALE DIFFERENCES

In this section we assume that $X_k, k \in \mathbb{Z}$, are martingale differences, namely, $\mathbb{E}(X_k|\mathcal{F}_{k-1}) = 0$ almost surely. We are interested in the asymptotic distribution of the quadratic form

$$T_n = \sum_{1 \le j < j' \le n} a_n(j, j') X_j X_{j'} = \sum_{t=2}^{n} X_t Z_{t-1,n}, \qquad Z_{t-1,n} = \sum_{j=1}^{t-1} a_n(j,t) X_j, \tag{10}$$

where, for each n, the weights $a_n(j, j')$, $1 \le j < j' \le n$, are real numbers. A particularly interesting special case is that the bivariate function $a_n(\cdot,\cdot)$ can be written in the univariate way, namely, $a_n(j, j') = a_n(j - j')$. For smoothed periodogram spectral density estimates, we present in Section 3 a martingale approximation scheme and show that (3) can be reduced to the form (10) even though the process (X_k) itself may not be a martingale difference sequence. Martingale approximations of this type act as a bridge that connects general stationary processes and martingales (Wu and Woodroofe, 2004).

Let

$$A_{t,n} = \sum_{j=1}^{t-1} a_n^2(j,t), \qquad B_{j,n} = \sum_{t=1+j}^{n} a_n^2(j,t), \qquad \text{and} \qquad \sigma_n^2 = \sum_{t=2}^{n} A_{t,n}. \tag{11}$$

Theorem 1, which follows, concerns the asymptotic distribution of T_n/σ_n under easily verifiable and mild conditions on $a_n(j, j')$, $1 \le j < j' \le n$.

THEOREM 1. *Let the process X_k of (8) be a martingale difference sequence and $X_t \in \mathcal{L}^q$ for some $q > 4$. Assume*

(i)

$$\max_{2 \le t \le n} A_{t,n} = o(\sigma_n^2); \tag{12}$$

(ii) for any fixed $J \in \mathbb{N}$,

$$\sum_{j=1}^{n-J} a_n^2(j, j+J) = o(\sigma_n^2); \tag{13}$$

(iii)

$$n \sum_{t'=1}^{n-1} B_{t',n}^2 = O(\sigma_n^4); \tag{14}$$

(iv)

$$\sum_{k=1}^{n-1} \sum_{t=1}^{k-1} \left[\sum_{j=1+k}^{n} a_n(k,j) a_n(t,j) \right]^2 = o(\sigma_n^4). \tag{15}$$

Then we have the asymptotic normality

$$\frac{T_n}{\sigma_n} \Rightarrow N(0, \|X_0\|^4). \tag{16}$$

We now briefly discuss conditions (12)–(15) of Theorem 1. Condition (12) is basically the Lindeberg-type condition. In the important special case $a_n(t', t) = a_n(t' - t)$, we have $A_{t,n} \le A_{n,n}$ for $2 \le t \le n$, and then (12) indicates that $A_{n,n}$ does not dominate σ_n^2. In this case, (12) is equivalent to the following claim: there exists a $\delta \in (0,1)$ such that $A_{\lfloor(1-\delta)n\rfloor,n} > \delta A_{n,n}$ for all sufficiently large n. Condition (13) means that the contribution of $a_n(j,j')X_j X_{j'}$ to T_n is negligible if $j' - j$ is small. Note that $\sigma_n^2 = \sum_{j=1}^{n-1} B_{j,n}$; by the Cauchy–Schwarz inequality, $n\sum_{t'=1}^{n-1} B_{t',n}^2 \ge \sigma_n^4$. So (14) indicates that the magnitudes of $n\sum_{t'=1}^{n-1} B_{t',n}^2$ and σ_n^4 are comparable.

Condition (15) is needed to exclude sequences $a_n(\cdot,\cdot)$ that may lead to non-Gaussian limiting distributions. For example, if $a_n(j,j') \equiv 1$, then $A_{t,n} = t - 1$, $\sigma_n^2 = \sum_{t=2}^{n}(t-1) = n(n-1)/2$, and (12)–(14) are fulfilled, whereas (15) is violated. In this case, if X_k are i.i.d. standard normal random variables, then T_n/σ_n has the non-Gaussian asymptotic distribution $(W^2 - 1)/\sqrt{2}$, where W is the standard Gaussian random variable. For a more general case for which (15) is violated, we let $a_n(k,j) = K[(k-j)/n]$, where K is a kernel function with support $[-1,1]$. This case is closely related to the spectral density estimation without truncation where the bandwidth is equal to the sample size. Kiefer and Vogelsang (2002) considered estimators of such type and obtained a non-Gaussian limiting distribution. See also Phillips et al. (2006a, 2006b) for more discussion. We are unclear as to what conditions on $a_n(\cdot,\cdot)$ are needed such that T_n/σ_n has a non-Gaussian limiting distribution.

With trigonometric identities, it is easy to verify (12)–(15) for $a_n(j)$ with form (6); see Section 3.2.

Proof of Theorem 1. Without loss of generality we assume $\|X_0\| = 1$. Note that $X_t Z_{t-1,n}$, $t \in \mathbb{Z}$, are martingale differences with respect to the filter $\mathcal{F}_t$; we shall apply the martingale central limit theorem (MCLT) (Chow and Teicher, 1988). By Lemma 1, which follows, $\|Z_{t-1,n}\|_q^2 \le C_q^2 A_{t,n}\|X_0\|_q^2$. By the Cauchy–Schwarz inequality, the Lindeberg condition is satisfied because

$$\sum_{t=2}^{n}\|X_t Z_{t-1,n}\|_{q/2}^{q/2} \le \sum_{t=2}^{n}\|X_t\|_q^{q/2}\|Z_{t-1,n}\|_q^{q/2}$$

$$\le \|X_0\|_q^{q/2} C_q^{q/2}\|X_0\|_q^{q/2}\sum_{t=2}^{n} A_{t,n}^{q/4} = o(\sigma_n^{q/2}).$$

The last step is due to (12) and $q > 4$. Applying Lemmas 2 and 3, which are given later in this section, with $W_t = \mathbb{E}(X_t^2|\mathcal{F}_{t-1})$, it is easily seen that the convergence of conditional variance

$$\frac{1}{\sigma_n^2}\sum_{t=2}^{n} Z_{t-1,n}^2 \mathbb{E}(X_t^2|\mathcal{F}_{t-1}) \to \|X_0\|^4 \tag{17}$$

in probability. By the MCLT, (16) holds. ∎

LEMMA 1. *Let $D_1, \ldots, D_n$ be a martingale difference sequence for which $D_i \in \mathcal{L}^p$, $p > 1$. Let $p' = \min(2, p)$. Then*

$$\left\| \sum_{i=1}^{n} D_i \right\|_p^{p'} \le C_p^{p'} \sum_{i=1}^{n} \|D_i\|_p^{p'},$$

where $C_p = 18p^{3/2}/(p-1)^{1/2}$.

Lemma 1 is a simple consequence of the Minkowski and the Burkholder inequalities (Chow and Teicher, 1988). The details of the proof are omitted.

LEMMA 2. *Assume that the process $W_t = w(\ldots, \varepsilon_{t-1}, \varepsilon_t) \in \mathcal{L}^{q/2}$ for some $q > 4$. Further assume that $a_n(j, j')$ satisfies (12) and (13). Then*

$$\frac{1}{\sigma_n^2} \sum_{t=2}^{n} [W_t - \mathbb{E}(W_0)] Z_{t-1,n}^2 \to 0 \quad \text{in } \mathcal{L}^1. \tag{18}$$

Proof. Without loss of generality let $\mathbb{E}(W_0) = 0$. For integer $I \ge 1$ introduce the truncated process $Z_{t-1,n,I} = \mathbb{E}(Z_{t-1,n} | \mathcal{F}_{t-I})$. Then $Z_{t-1,n,I} = 0$ if $t \le I$ and $Z_{t-1,n,I} = \sum_{j=1}^{t-I} a_n(j,t) X_j$. For $1 \le t \le n$, by Lemma 1,

$$\|Z_{t-1,n,I} - Z_{t-1,n}\|_q^2 \le C_q^2 \|X_0\|_q^2 \sum_{j=\max(1, t-I+1)}^{t-1} a_n^2(j,t).$$

Let $L_n(I) = \sum_{J=1}^{I} \ell_n(J)$, where $\ell_n(J)$ is the quantity on the left-hand side of (13). Let

$$V_n(I) = \sigma_n^{-2} \sum_{t=2}^{n} W_t Z_{t-1,n,I}^2 \quad \text{and} \quad T_n(I) = \sigma_n^{-2} \sum_{t=2}^{n} \mathbb{E}(W_t | \mathcal{F}_{t-I}) Z_{t-1,n,I}^2.$$

By the Cauchy–Schwarz inequality, (12), and (13),

$$\begin{aligned}
\mathbb{E}|V_n(1) - V_n(I)| &\le \sigma_n^{-2} \sum_{t=2}^{n} \mathbb{E}|W_t (Z_{t-1,n}^2 - Z_{t-1,n,I}^2)| \\
&\le \sigma_n^{-2} \sum_{t=2}^{n} \|W_t\| \, \|Z_{t-1,n,I} - Z_{t-1,n}\|_4 \|Z_{t-1,n,I} + Z_{t-1,n}\|_4 \\
&= O(\sigma_n^{-2}) \sum_{t=2}^{n} \|Z_{t-1,n,I} - Z_{t-1,n}\|_4 A_{t,n}^{1/2} \\
&= O(\sigma_n^{-2}) \left(\sum_{t=2}^{n} \|Z_{t-1,n,I} - Z_{t-1,n}\|_4^2 \right)^{1/2} \left(\sum_{t=2}^{n} A_{t,n} \right)^{1/2} \\
&= O(\sigma_n^{-1}) [L_n(I)]^{1/2} \to 0
\end{aligned} \tag{19}$$

as $n \to \infty$. For $0 \le j \le I - 1$ let

$$U_n(j) = \sigma_n^{-2} \sum_{t=2}^{n} (\mathcal{P}_{t-j} W_t) Z_{t-1,n,I}^2.$$

Then $V_n(I) - T_n(I) = \sum_{j=0}^{I-1} U_n(j)$. Note that the sequence $(\mathcal{P}_{t-j} W_t) Z_{t-1,n,I}^2$, $t = 2, \dots, n$, forms martingale differences. By Lemma 1 and (12), because $\|\mathcal{P}_{t-j} W_t\|_{q/2} \le 2\|W_t\|_{q/2}$,

$$\begin{aligned}
\limsup_{n\to\infty} \|U_n(j)\|_{q/4}^{q/4} &\le C_{q/4}^{q/4} \limsup_{n\to\infty} \sigma_n^{-q/2} \sum_{t=2}^{n} \|(\mathcal{P}_{t-j} W_t) Z_{t-1,n,I}^2\|_{q/4}^{q/4} \\
&\le C_{q/4}^{q/4} \limsup_{n\to\infty} \sigma_n^{-q/2} \|W_0\|_{q/2}^{q/4} \sum_{t=2}^{n} \|Z_{t-1,n,I}\|_q^{q/2} \\
&= \limsup_{n\to\infty} O(\sigma_n^{-q/2}) \sum_{t=2}^{n} A_{t,n}^{q/4} = 0.
\end{aligned}$$

Because $|V_n(1)| \le |V_n(1) - V_n(I)| + |V_n(I) - T_n(I)| + |T_n(I)|$, by (19),

$$\begin{aligned}
\limsup_{n\to\infty} \mathbb{E}|V_n(1)| &\le \limsup_{n\to\infty} \|V_n(I) - T_n(I)\|_{q/4} + \limsup_{n\to\infty} \mathbb{E}|T_n(I)| \\
&\le \limsup_{n\to\infty} \sum_{j=0}^{I-1} \|U_n(j)\|_{q/4} \\
&\quad + \limsup_{n\to\infty} \sigma_n^{-2} \sum_{t=2}^{n} \|\mathbb{E}(W_t|\mathcal{F}_{t-I})\| \, \|Z_{t-1,n,I}\|_4^2 \\
&\le \|\mathbb{E}(W_0|\mathcal{F}_{-I})\| C_q^2 \|X_0\|_q^2.
\end{aligned}$$

Therefore (18) follows because $\|\mathbb{E}(W_0|\mathcal{F}_{-I})\| \to 0$ as $I \to \infty$. ■

LEMMA 3. *Assume that (X_k) defined in (8) is a martingale difference sequence with $X_0 \in \mathcal{L}^4$. Then under (14) and (15), we have*

$$\frac{1}{\sigma_n^2} \sum_{t=2}^{n} Z_{t-1,n}^2 \to \|X_0\|^2 \quad \text{in } \mathcal{L}^2. \tag{20}$$

Proof. For notational convenience we omit the subscript n in $a_n^2(t', t)$, $B_{j,n}$, and $Z_{t-1,n}$. Write $\theta_j = \|\mathcal{P}_0 X_j^2\|$; then $\sum_{j=0}^{\infty} \theta_j^2 = \|X_0^2 - \mathbb{E}(X_0^2)\|^2 < \infty$. Recall that $B_{t'} = \sum_{t=1+t'}^{n} a^2(t', t)$, $1 \le t' \le n - 1$, and let $\Xi_n = \sum_{t'=1}^{n-1} B_{t'}^2$. To show (20), we shall apply the martingale decomposition method. By the orthogonality of $\mathcal{P}_k$, $k = \dots, n-2, n-1$, and because $\mathcal{P}_k Z_{t-1}^2 = 0$ if $1 \le t \le k$,

$$\left\| \sum_{t=2}^{n} Z_{t-1}^2 - \sigma_n^2 \|X_0\|^2 \right\|^2 = \sum_{k=-\infty}^{0} \left\| \mathcal{P}_k \sum_{t=2}^{n} Z_{t-1}^2 \right\|^2 + \sum_{k=1}^{n-1} \left\| \mathcal{P}_k \sum_{t=1+k}^{n} Z_{t-1}^2 \right\|^2. \tag{21}$$

Then it remains to show that both terms in the preceding display are of the order $o(\sigma_n^4)$. For the first one, let k be a nonpositive integer. Because (X_j) are martingale differences,

$$\mathcal{P}_k Z_{t-1}^2 = \sum_{t'=1}^{t-1} a^2(t',t)\mathcal{P}_k X_{t'}^2,$$

which by the triangle inequality implies

$$\|\mathcal{P}_k Z_{t-1}^2\| \le \sum_{t'=1}^{t-1} a^2(t',t)\|\mathcal{P}_k X_{t'}^2\| = \sum_{t'=1}^{t-1} a^2(t',t)\theta_{t'-k}. \tag{22}$$

Hence by the Cauchy–Schwarz inequality and (14),

$$\begin{aligned}\sum_{k=-\infty}^{0}\left[\sum_{t=2}^{n}\|\mathcal{P}_k Z_{t-1}^2\|\right]^2 &\le \sum_{k=-\infty}^{0}\left[\sum_{t'=1}^{n-1} B_{t'}\theta_{t'-k}\right]^2\\ &\le \Xi_n \sum_{k=-\infty}^{0}\left[\sum_{t'=1}^{n-1}\theta_{t'-k}^2\right]\\ &= o(n\Xi_n) = o(\sigma_n^4).\end{aligned} \tag{23}$$

It is slightly more complicated to deal with the second term on the right-hand side of (21). Let $1 \le k \le n-1$; then $\mathcal{P}_k Z_{t-1}^2 = 0$ if $t \le k$ and

$$\mathcal{P}_k Z_{t-1}^2 = \sum_{t'=k}^{t-1} a^2(t',t)\mathcal{P}_k X_{t'}^2 + 2a(k,t)X_k \sum_{t'=1}^{k-1} a(t',t)X_{t'} =: I_k(t) + II_k(t) \quad \text{(say)} \tag{24}$$

if $t > k$. By the Cauchy–Schwarz inequality, Lemma 1, and (15),

$$\begin{aligned}\sum_{k=1}^{n-1}\left\|\sum_{t=1+k}^{n} II_k(t)\right\|^2 &= 4\sum_{k=1}^{n-1}\left\|X_k \sum_{t=1}^{k-1}\sum_{t'=1+k}^{n} a(k,t')a(t,t')X_t\right\|^2\\ &\le 4\sum_{k=1}^{n-1}\|X_k\|_4^2\left\|\sum_{t=1}^{k-1}\sum_{t'=1+k}^{n} a(k,t')a(t,t')X_t\right\|_4^2\\ &\le 4\|X_0\|_4^2 C_4^2 \sum_{k=1}^{n-1}\sum_{t=1}^{k-1}\left[\sum_{t'=1+k}^{n} a(k,t')a(t,t')\right]^2 = o(\sigma_n^4).\end{aligned}$$

For the term $I_k(t)$, we have

$$\sum_{t=k+1}^{n}\|I_k(t)\| \le \sum_{t=1+k}^{n}\sum_{t'=k}^{t-1} a^2(t',t)\theta_{t'-k} = \sum_{t'=k}^{n-1}\theta_{t'-k}B_{t'}.$$

Let $\ell = \lfloor n^{1/4} \rfloor$. Then

$$\begin{aligned}\frac{1}{2}\sum_{k=1}^{n-1}\left(\sum_{t'=k}^{n-1}\theta_{t'-k}B_{t'}\right)^2 &\le \sum_{k=1}^{n-1-\ell}\left(\sum_{t'=k+\ell}^{n-1}\theta_{t'-k}B_{t'}\right)^2 \\ &\quad + \sum_{k=1}^{n-1-\ell}\left(\sum_{t'=k}^{k+\ell-1}\theta_{t'-k}B_{t'}\right)^2 + \sum_{k=n-\ell}^{n-1}\left(\sum_{t'=k}^{n-1}\theta_{t'-k}B_{t'}\right)^2 \\ &\le \sum_{k=1}^{n-1-\ell}\left(\sum_{t'=k+\ell}^{\infty}\theta_{t'-k}^2\right)\left(\sum_{t'=k+\ell}^{n-1}B_{t'}^2\right) + \Xi_n \ell \sum_{j=0}^{\ell-1}\theta_j^2 \\ &\le n\Xi_n \sum_{j=\ell}^{\infty}\theta_j^2 + \Xi_n \ell \sum_{j=0}^{\ell-1}\theta_j^2 = o(n\Xi_n). \qquad \textbf{(25)}\end{aligned}$$

By (24),

$$\begin{aligned}&\frac{1}{2}\sum_{k=1}^{n-1}\left[\sum_{t=1+k}^{n}\|\mathcal{P}_k Z_{t-1}^2\|\right]^2 \\ &\quad \le \sum_{k=1}^{n-1}\left(\sum_{t'=k}^{n-1}\theta_{t'-k}B_{t'}\right)^2 + \sum_{k=1}^{n-1}\left\|\sum_{t=1+k}^{n} II_k(t)\right\|^2 = o(\sigma_n^4). \qquad \textbf{(26)}\end{aligned}$$

Combining (23) and (26), we have (20). ■

3. SMOOTHED PERIODOGRAM ESTIMATES

The central limit theorem presented in Section 2 is only for martingale differences. To obtain asymptotic distribution of the smoothed periodogram spectral density estimate (3) for processes with general forms, we shall approximate $S_n(\theta) = \sum_{t=1}^{n} X_t e^{t\theta\sqrt{-1}}$ by martingales so that Theorem 1 is applicable. Such a martingale approximation scheme has been proposed in Wu (2005a). An explicit construction of approximating martingales is given in Section 3.1. Section 3.2 shows the asymptotic normality of the estimate $\hat{f}_n(\theta)$ in (3).

3.1. A Martingale Approximation Scheme

LEMMA 4. *Assume that the process (X_k) defined in (8) satisfies*

$$\sum_{k=0}^{\infty} k\|\mathcal{P}_0 X_k\|_q < \infty \qquad \textbf{(27)}$$

for some $q \ge 2$. Then for every $\theta \in \mathbb{R}$, the process

$$Y_k(\theta) = \sum_{t=0}^{\infty} \mathbb{E}(X_{t+k}|\mathcal{F}_k)\exp(t\theta\sqrt{-1}) \qquad \textbf{(28)}$$

exists and is in $\mathcal{L}^q$. Let $D_k(\theta) = \mathcal{P}_k Y_k(\theta) = Y_k(\theta) - \mathbb{E}[Y_k(\theta)|\mathcal{F}_{k-1}]$ and

$$M_n(\theta, \lambda) = \sum_{k=1}^{n} e^{\sqrt{-1}k(\lambda+\theta)} D_k(\theta). \tag{29}$$

Then there exists a constant $C > 0$, independent of n and λ, such that

$$\sup_{\theta\in\mathbb{R}} \|S_n(\theta + \lambda) - M_n(\theta, \lambda)\| \le C(\sqrt{n}|1 - e^{-\sqrt{-1}\lambda}| + 1). \tag{30}$$

Proof. Let $p_k = \|\mathcal{P}_0 X_k\|_q$ and $t \ge 0$. Note that $\mathbb{E}(X_t|\mathcal{F}_0) = \sum_{j=-\infty}^{0} \mathcal{P}_j X_t$. By Lemma 1, we have $\|\mathbb{E}(X_t|\mathcal{F}_0)\|_q^2 \le C_q^2 \sum_{j=-\infty}^{0} p_{t-j}^2$, which in conjunction with (27) implies that

$$\sum_{t=0}^{\infty} \|\mathbb{E}(X_t|\mathcal{F}_0)\|_q = C_q \sum_{t=0}^{\infty} \left(\sum_{k=t}^{\infty} p_k^2 \right)^{1/2}$$
$$\le C_q \sum_{t=0}^{\infty} \sum_{k=t}^{\infty} p_k = C_q \sum_{t=0}^{\infty} (t+1) p_t < \infty.$$

Hence $Y_k(\theta) \in \mathcal{L}^q$, and additionally $X_k = Y_k(\theta) - \mathbb{E}[Y_{k+1}(\theta)|\mathcal{F}_k] e^{\sqrt{-1}\theta}$. For notational convenience write $r = e^{\sqrt{-1}\theta}$, $s = e^{\sqrt{-1}(\theta+\lambda)}$, and $Z_k(\theta) = \mathbb{E}[Y_{k+1}(\theta)|\mathcal{F}_k]$. Then

$$S_n(\theta + \lambda) - M_n(\theta, \lambda) = \sum_{k=1}^{n} e^{\sqrt{-1}k\lambda} \{Y_k(\theta) r^k - r^{k+1} Z_k(\theta)\} - M_n(\theta, \lambda)$$
$$= (1 - e^{-\sqrt{-1}\lambda}) \sum_{k=1}^{n} s^k Z_{k-1}(\theta) + r Z_0(\theta) - r s^n Z_n(\theta). \tag{31}$$

Because $Z_{k-1}(\theta) = \sum_{j=1}^{\infty} \mathcal{P}_{k-j} Z_{k-1}(\theta)$, by the triangle inequality,

$$\left\| \sum_{k=1}^{n} s^k Z_{k-1}(\theta) \right\| \le \sum_{j=1}^{\infty} \left\| \sum_{k=1}^{n} s^k \mathcal{P}_{k-j} Z_{k-1}(\theta) \right\|$$
$$\le \sum_{j=1}^{\infty} \left[\sum_{k=1}^{n} \|\mathcal{P}_{k-j} Z_{k-1}(\theta)\|^2 \right]^{1/2}$$
$$= \sum_{j=1}^{\infty} \sqrt{n} \|\mathcal{P}_{1-j} Z_0(\theta)\|$$
$$= \sqrt{n} \sum_{j=1}^{\infty} \|\mathcal{P}_{1-j} Y_1(\theta)\|.$$

Note that

$$\sum_{j=1}^{\infty}\|\mathcal{P}_{1-j}Y_1(\theta)\| \le \sum_{j=1}^{\infty}\sum_{t=1}^{\infty}\|\mathcal{P}_{1-j}X_t\| \le \sum_{l=1}^{\infty} lp_l < \infty.$$

Then (30) follows from (31) with $C = 2\sum_{l=0}^{\infty}(l+1)p_l$. ∎

In Wu and Min (2005) the condition (27) implies that the process (X_k) is $\mathcal{L}^q$ *weakly dependent with order* 1. It also implies $\sum_{k\in\mathbb{Z}}|k\gamma(k)| < \infty$; thus the spectral density function is continuously differentiable; see Lemma A.1 in the companion paper in this issue (Shao and Wu, 2007b). Note that $D_k(\theta) = Y_k(\theta) - \mathbb{E}[Y_k(\theta)|\mathcal{F}_{k-1}]$ is a martingale difference sequence and it is related to the spectral density in an interesting way. By Wu (2005a), $\|D_k(\theta)\|^2 = 2\pi f(\theta)$. The latter identity gives a probabilistic representation of the spectral density.

3.2. Asymptotic Normality

Recall that $\lambda_j = 2\pi j/n$, $j \in \mathbb{Z}$, are Fourier frequencies. Let $n > 4$ and m be a positive integer with $m < n/2 - 1$. For a real sequence $s = (s_j)_{j=-m}^{m}$, let $\chi_m(s) = \sum_{j=-m}^{m} s_j^2$, $\tau_m(s) = \sum_{j=-m}^{m} s_j$, $\varpi_m(s) = \sum_{j=-m}^{m}|s_j|$, and $\omega_m(s) = |s_{-m}| + \sum_{j=1-m}^{m}|s_j - s_{j-1}|$. Crudely speaking the quantity ω_m measures the oscillation of the sequence s. For $\theta \in [0, \pi]$ let

$$\Delta_n(\theta) = \sum_{j=-m}^{m} s_j[I(\theta+\lambda_j) - \mathbb{E}\{I(\theta+\lambda_j)\}]. \tag{32}$$

Let $(\varepsilon_j')_{j\in\mathbb{Z}}$ be an i.i.d. copy of $(\varepsilon_j)_{j\in\mathbb{Z}}$. For $k \ge 0$ define

$$X_k' = g(\ldots, \varepsilon_{-2}, \varepsilon_{-1}, \varepsilon_0', \varepsilon_1, \ldots, \varepsilon_k). \tag{33}$$

Then X_k' is a coupled version of X_k with ε_0 replaced by ε_0'. Our weak dependence condition (cf. expression (34) in Theorem 2, which follows) is expressed in terms of the distance between X_k and X_k'.

THEOREM 2. *Assume that, for some* $q > 4$,

$$\sum_{k=1}^{\infty} k\|X_k - X_k'\|_q < \infty. \tag{34}$$

Further assume that $m = o(n^{2/3})$ *and the sequence* $(s_j)_{j=-m}^{m}$ *satisfies*

$$\omega_m^2(s)\log n = o(\chi_m(s)). \tag{35}$$

(i) If $\theta = 0$ *and* $s_j = 0$ *for* $j = -m, \ldots, -1$, *then as* $n \to \infty$,

$$\frac{2\pi\Delta_n(0)}{\sqrt{\chi_m(s)}} \Rightarrow N[0, \|D_0(0)\|^4], \tag{36}$$

where $D_0(\theta) = \sum_{t=0}^{\infty}\mathcal{P}_0 X_t \exp(\sqrt{-1}t\theta) \in \mathcal{L}^2$ *satisfies* $\|D_0(\theta)\|^2 = 2\pi f(\theta)$.

(ii) If $\theta \in (0,\pi)$, then (36) holds with $\Delta_n(0)$ and $D_0(0)$ replaced by $\Delta_n(\theta)$ and $D_0(\theta)$, respectively.

Remark 1. In case (i), s_j are assumed to be 0 for $-m \le j \le -1$. If otherwise, noting that $I_n(\lambda) = I_n(-\lambda)$, we can let $s_j' = s_j + s_{-j}$, $s_{-j}' = 0$ for $j = 1,\dots,m$ and $s_0' = s_0$. Then the central limit theorem (36) is still applicable.

As in Wiener (1958) and Priestley (1988), the causal process (8) can be interpreted as a physical system with $\dots,\varepsilon_{k-1},\varepsilon_k$ being the inputs, X_k the output, and g a transform or a filter. If g is a linear function, then (X_k) is a linear process. Otherwise (X_k) is a nonlinear process. The condition (34) has the following interesting interpretation. Note that X_k' is a coupled version of X_k by replacing ε_0 in X_k by an i.i.d. copy ε_0'. If the function $g(\dots,\varepsilon_{k-1},\varepsilon_k)$ does not depend on ε_0, then $X_k - X_k' = 0$. Hence the quantity $\|X_k - X_k'\|_q$ measures the contribution of ε_0 to X_k, in other words, the degree of dependence of X_k on ε_0. In this sense (34) means that the weighted cumulative contribution of ε_0 to all future values X_k, $k > 0$, is finite, and hence (34) ensures short-range dependence. See Wu (2005b) for a more detailed discussion on the dependence of stationary causal processes from the nonlinear system theory point of view. Conditions based on the quantity $\|X_k - X_k'\|_q$ are often easily verifiable because they are directly related to the data-generating mechanism of the process (X_k).

Our dependence condition (34) is very different from the classical strong mixing conditions, which may be too restrictive in certain applications (Andrews, 1984). On the other hand, we avoid summability conditions on joint cumulants that are commonly imposed in the large-sample spectral theory (Brillinger, 1975; Rosenblatt, 1984). The verification of the latter generally is not easy for processes that are non-Gaussian and nonlinear. In the companion paper in this issue (Shao and Wu, 2007b), we find an easily verifiable sufficient condition for the absolute summability of qth ($q \in \mathbb{N}$, $q \ge 2$) joint cumulants, which is $\sum_{k=0}^{\infty} k^{q-1}\|X_k - X_k'\|_q < \infty$; see Theorem 4.1 and Remark 4.1 therein. Unfortunately the latter sufficient condition appears to be overly strong. It is an open problem whether weaker versions such as $\sum_{k=0}^{\infty}\|X_k - X_k'\|_q < \infty$ suffice.

Clearly (34) implies (27) by the identity $\mathcal{P}_0 X_k = \mathbb{E}[(X_k - X_k')|\mathcal{F}_0]$ and the Jensen inequality. Theorem 2 has the following two immediate corollaries. The proofs are straightforward and hence the details are omitted. The first corollary gives a central limit theorem for $\hat{f}_n(\theta)$, and the second one has an application in Shao and Wu (2007b). The latter paper studies the local Whittle estimation of long memory index for fractionally integrated nonlinear processes.

COROLLARY 1. *Assume that (34) holds for some $q > 4$. Further assume that the kernel K is a nonnegative function with support $[-1,1]$, K has bounded variation, and*

$$\frac{1}{m}\sum_{i=-m}^{m} K^2(i/m) \to \int_{-1}^{1} K^2(u)\,du =: \kappa \quad \text{and} \quad \frac{1}{m}\sum_{i=-m}^{m} K(i/m) \to 1 \tag{37}$$

as $m \to \infty$. Additionally assume that the $m = m_n$ satisfies

$$\frac{\log n}{m} + \frac{m}{n^{2/3}} \to 0. \tag{38}$$

Then $m^{1/2}[\hat{f}_n(\theta) - \mathbb{E}\{\hat{f}_n(\theta)\}] \Rightarrow f(\theta)N(0, \kappa)$.

Using the Cramér–Wold device, we can show that, for different frequencies $\omega_1, \dots, \omega_J \in (0, \pi)$, $m^{1/2}[\hat{f}_n(\omega_j) - \mathbb{E}\{\hat{f}_n(\omega_j)\}]/f(\omega_j)$, $1 \le j \le J$, are asymptotically i.i.d. normal $N(0, \kappa)$. Because the proof of the latter claim is routine and it involves very lengthy and tedious calculations, we omit the details.

We conjecture that the central limit theorem in Corollary 1 still holds if (38) is weakened to $m^{-1}\log n + n^{-1}m \to 0$. A key difficulty toward this result is that our argument requires $m = o(n^{2/3})$ to ensure a satisfactory martingale approximation of $\Delta_n(\theta)$ by $\Lambda_n(\theta)$; see the proof of Theorem 2 in the Appendix. If Corollary 1 is valid under $m^{-1}\log n + n^{-1}m \to 0$, then we can allow m to be a multiple of $n^{4/5}$ that minimizes the mean squares error $\|\hat{f}_n(\theta) - f(\theta)\|^2 = \|\hat{f}_n(\theta) - \mathbb{E}\{\hat{f}_n(\theta)\}\|^2 + |\mathbb{E}\{\hat{f}_n(\theta)\} - f(\theta)|^2$ because the bias $\mathbb{E}\{\hat{f}_n(\theta)\} - f(\theta)$ is of the order $(m/n)^2$ under suitable conditions on K and f.

COROLLARY 2. *Let $s_j = 1 + \log(j/m)$ for $1 \le j \le m$ and $s_j = 0$ for $-m \le j \le 0$. Assume (34) for some $q > 4$ and*

$$\frac{(\log n)^3}{m} + \frac{m}{n^{2/3}} \to 0. \tag{39}$$

Then (36) holds.

Example 1 (Nonlinear time series)

Let $\varepsilon_k, k \in \mathbb{Z}$, be i.i.d. random variables and define X_n recursively by

$$X_n = R(X_{n-1}, \varepsilon_n), \qquad n \in \mathbb{Z}, \tag{40}$$

where R is a measurable function. Many popular nonlinear time series models, such as threshold autoregressive (TAR) models, bilinear autoregressive models, and autoregressive models with conditional heteroskedascity (ARCH) are of the form (40). Let

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

Assume that

$$\mathbb{E}(\log L_{\varepsilon_0}) < 0 \quad \text{and} \quad \mathbb{E}[L_{\varepsilon_0}^\alpha + |R(x_0, \varepsilon_0)|^\alpha] < \infty \tag{41}$$

for some $\alpha > 0$ and x_0. Then (40) has a stationary distribution (Diaconis and Freedman, 1999; Wu and Shao, 2004), and iterates of (40) give rise to (8). Additionally, Wu and Woodroofe (2000) show that (41) implies that there exist $\beta > 0$ and $r \in (0,1)$ for which

$$\|X_n - X_n^*\|_\beta = O(r^n), \quad \text{where } X_n^* = G(\ldots, \varepsilon'_{-1}, \varepsilon'_0, \varepsilon_1, \ldots, \varepsilon_n). \tag{42}$$

The preceding property is called geometric-moment contraction (GMC) in Hsing and Wu (2004). Furthermore, if $X_k \in \mathcal{L}^q$ for some $q > 4$, then (42) holds for all $\beta \in (0,q)$ (cf. Wu and Min, 2005, Lem. 2). It is easily seen that (42) implies $\|X_n - X_n'\|_\beta = O(r^n)$. Hence (34) holds. Recently, Shao and Wu (2007a) showed that GMC holds for various generalized autoregressive conditional heteroskedasticity (GARCH) models, including general asymmetric GARCH(r,s) and nonlinear GARCH$(1,1)$ models. For the GMC property of an exponential GARCH (EGARCH) model, see Min (2004).

Example 2 (Nonlinear transforms of linear processes)

Let ε_k be i.i.d. random variables with $\varepsilon_k \in \mathcal{L}^q$ for some $q > 4$; let $a_0, a_1, \ldots$ be a square summable real sequence and $U_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}$ be a linear process. Consider the process

$$X_t = |U_t| - \mathbb{E}|U_t|.$$

Let $(\varepsilon_k')_{k\in\mathbb{Z}}$ be an i.i.d. copy of $(\varepsilon_k)_{k\in\mathbb{Z}}$. Then $\big||U_t| - |U_t'|\big| \le |a_t|\,|\varepsilon_0 - \varepsilon_0'|$, and consequently $\|X_t - X_t'\|_q \le \big\||U_t| - |U_t'|\big\|_q \le |a_t|\,\|\varepsilon_0 - \varepsilon_0'\|_q$. Under the simple sufficient condition

$$\sum_{j=0}^{\infty} j|a_j| < \infty, \tag{43}$$

because $\|\varepsilon_0\|_q < \infty$, (34) holds. It is easily seen that, for any Lipschitz continuous function G, (43) implies (34) for $X_t = G(U_t) - \mathbb{E}\{G(U_t)\}$. The classical central limit theorems on spectral density estimates are not applicable here because they require strong mixing conditions and summability conditions on joint cumulants. As pointed out in Andrews (1984), the process X_t is not strong mixing if ε_k are i.i.d. with the distribution $\mathbb{P}(\varepsilon_k = 1) = \mathbb{P}(\varepsilon_k = -1) = \frac{1}{2}$ and $a_j = 2^{-j}$, $j \in \mathbb{N}$. On the other hand, if G is a nonlinear function, it seems very difficult to verify summability conditions on joint cumulants of X_t, because of the nonlinear nature. The central limit theorem in Hannan (1970, Thm. 5.11) is only for linear processes and hence is not applicable to our X_t. The argument in Hannan (1970) does not work either because it depends heavily on the linearity structure.

REFERENCES

Anderson, T.W. (1971) *The Statistical Analysis of Time Series*. Wiley.

Andrews, D.W.K. (1984) Non-strong mixing autoregressive processes. *Journal of Applied Probability* 21, 930–934.

Andrews, D.W.K. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–854.

Basalykas, A. (1994) Functional limit theorems for random multilinear forms. *Stochastic Processes and Their Applications* 53, 175–191.

Bentkus, R.Y. & R.A. Rudzkis (1982) On the distribution of some statistical estimates of a spectral density. *Theory of Probability and Its Applications* 27, 795–814.

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.

Brockwell, P.J. & R.A. Davis (1991) *Time Series: Theory and Methods*, 2nd ed. Springer-Verlag.

Chow, Y.S. & H. Teicher (1988) *Probability Theory*, 2nd ed. Springer-Verlag.

Dahlhaus, R. (1985) On a spectral density estimate obtained by averaging periodograms. *Journal of Applied Probability* 22, 592–610.

de Jong, R.M. & J. Davidson (2000) Consistency of kernel estimators of heteroskedastic and autocorrelated covariance matrices. *Econometrica* 68, 407–424.

de Wet, T. & J.H. Venter (1973) Asymptotic distributions for quadratic forms with applications to tests of fit. *Annals of Statistics* 1, 380–387.

Diaconis, P. & D. Freedman (1999) Iterated random functions. *SIAM Review* 41, 41–76.

Götze, F. & A.N. Tikhomirov (1999) Asymptotic distribution of quadratic forms. *Annals of Probability* 27, 1072–1098.

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

Hansen, B.E. (1992) Consistent covariance matrix estimation for dependent heterogenous processes. *Econometrica* 60, 967–972.

Henry, M. (2001) Average periodogram spectral estimation with long-memory conditional heteroscedasticity. *Journal of Time Series Analysis* 22, 431–459.

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

Jansson, M. (2002) Consistent covariance matrix estimation for linear processes. *Econometric Theory* 18, 1449–1459.

Kallianpur, G. (1981) Some ramifications of Wiener's ideas on nonlinear prediction. In N. Wiener & P. Masani (eds.), *Norbert Wiener, Collected Works with Commentaries*, pp. 402–424. MIT Press.

Kiefer, N.M. & T.J. Vogelsang (2002) Heteroskedasticity-autocorrelation robust testing using bandwidth equal to sample size. *Econometric Theory* 18, 1350–1366.

Lomnicki, Z.A. & S.C. Zaremba (1959) On some moments and distributions occurring in the theory of linear stochastic processes, part II. *Monatshefte für Mathematik* 63, 128–168.

Mikosch, T. (1991) Functional limit theorems for random quadratic forms. *Stochastic Processes and Their Applications* 37, 81–98.

Min, W. (2004) Inference on time series driven by dependent innovations. Ph.D. Dissertation, University of Chicago.

Phillips, P.C.B., Y. Sun, & S. Jin (2006a) Spectral density estimation and robust hypothesis testing using steep origin kernels without truncation. *International Economic Review* 47, 837–894.

Phillips, P.C.B., Y. Sun, & S. Jin (2006b) Long run variance estimation and robust regression testing using sharp origin kernels with no truncation. *Journal of Statistical Planning and Inference* 137, 985–1023.

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

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

Robinson, P.M. (1983) Review of various approaches to power spectrum estimation. In D.R. Brillinger & R.R. Krishnaiah (eds.), *Time Series in the Frequency Domain*, Handbook of Statistics 3, pp. 343–368. North-Holland.

Rosenblatt, M. (1971) *Markov Processes. Structure and Asymptotic Behavior.* Springer-Verlag.

Rosenblatt, M. (1984) Asymptotic normality, strong mixing and spectral density estimates. *Annals of Probability* 12, 1167–1180.

Shao, X. & W.B. Wu (2007a) Asymptotic spectral theory for nonlinear time series. *Annals of Statistics*, forthcoming.
Shao, X & W.B. Wu (2007b) Local Whittle estimation of fractional integration for nonlinear processes. *Econometric Theory* 23 (this issue).
ten Vregelaar, J.M. (1990) Note on the convergence to normality of quadratic forms in independent variables. *Theory of Probability and Its Applications* 35, 177–179.
Tong, H. (1990) *Non-linear Time Series: A Dynamical System Approach.* Oxford University Press.
Varberg, D.E. (1966) Convergence of quadratic forms in independent random variables. *Annals of Mathematical Statistics* 37, 567–576.
Whittle, P. (1964) On the convergence to normality of quadratic forms in independent variables. *Theory of Probability and Its Applications* 9, 103–108.
Wiener, N. (1958) *Nonlinear Problems in Random Theory.* MIT Press.
Wu, W.B. (2005a) Fourier transforms of stationary processes. *Proceedings of the American Mathematical Society* 133, 285–293.
Wu, W.B. (2005b) Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences* 102, 14150–14154.
Wu, W.B. (2007) Strong invariance principles for dependent random variables. *Annals of Probability*, forthcoming.
Wu, W.B. & J. Mielniczuk (2002) Kernel density estimation for linear processes. *Annals of Statistics* 30, 1441–1459.
Wu, W.B. & W. Min (2005) On linear processes with dependent innovations. *Stochastic Processes and Their Applications* 115, 939–958.
Wu, W.B. & X. Shao (2004) Limit theorems for iterated random functions. *Journal of Applied Probability* 41, 425–436.
Wu, W.B. & M. Woodroofe (2000) A central limit theorem for iterated random functions. *Journal of Applied Probability* 37, 748–755.
Wu, W.B. & M. Woodroofe (2004) Martingale approximations for sums of stationary processes. *Annals of Probability* 32, 1674–1690.

APPENDIX

Proof of Theorem 2. Recall the definitions of $\chi_m(s)$, $\tau_m(s)$, $\varpi_m(s)$, and $\omega_m(s)$ in Section 3.2. For simplicity we abbreviate them as χ, τ, ϖ, and ω, respectively. By Lemma 4, condition (34) implies that the martingale $M_n(\theta,\lambda)$ defined in (29) exists in $\mathcal{L}^q$. Recall that $D_k(\theta) = Y_k(\theta) - \mathbb{E}[Y_k(\theta)|\mathcal{F}_{k-1}]$. Let

$$\Lambda_n(\theta) = \sum_{j=-m}^{m} s_j[|M_n(\theta,\lambda_j)|^2 - \mathbb{E}\{|M_n(\theta,\lambda_j)|^2\}]. \tag{A.1}$$

Let $R_n(\theta,\lambda) = S_n(\theta+\lambda) - M_n(\theta,\lambda)$. By (30) of Lemma 4, for $-m \le j \le m$, $\|R_n(\theta,\lambda_j)\| = O(\sqrt{n}m/n + 1)$. Because $m = o(n^{2/3})$ and $\sum_{j=-m}^{m}|s_j| \le (2m+1)^{1/2}\chi^{1/2}$,

$$\begin{aligned}\mathbb{E}|\Delta_n(\theta) - \Lambda_n(\theta)| &\le 2\sum_{j=-m}^{m}|s_j| \times \mathbb{E}\Big||M_n(\theta,\lambda_j)|^2 - |S_n(\theta+\lambda_j)|^2\Big| \\ &\le 2\sum_{j=-m}^{m}|s_j|[2\|M_n(\theta,\lambda_j)\|\,\|R(\theta,\lambda_j)\| + \|R(\theta,\lambda_j)\|^2] \\ &= \sum_{j=-m}^{m}|s_j|O[\sqrt{n}(\sqrt{n}m/n+1) + (\sqrt{n}m/n+1)^2] \\ &= O(m+\sqrt{n})\sqrt{m\chi} = o(n\chi^{1/2}).\end{aligned}$$

It then remains to show (36) with $\Delta(\theta)$ replaced by $\Lambda_n(\theta)$.

(i) Let $\theta = 0$. Then (A.1) becomes

$$\Lambda_n(0) = 2 \sum_{1\le k<k'\le n} a_n(k'-k) D_k D_{k'} + \tau \sum_{k=1}^{n} (D_k^2 - \|D_0\|^2) =: 2T_n + \tau\Omega_n,$$

where

$$a_n(l) = \sum_{j=-m}^{m} s_j \cos(l\lambda_j).$$

Note that $\mathcal{P}_0 D_k^2 = \mathbb{E}(D_k^2|\mathcal{F}_0) - \mathbb{E}(D_k^2|\mathcal{F}_0') = \mathbb{E}[D_k^2 - (D_k')^2|\mathcal{F}_0]$. The second term $\tau\Omega_n$ in $\Lambda_n(0)$ is of order $O(\sqrt{m\chi})\sqrt{n} = o(n\chi^{1/2})$ because

$$\begin{aligned}\left\| \sum_{k=1}^{n} (D_k^2 - \|D_0\|^2) \right\| &\le \sqrt{n} \sum_{k=0}^{\infty} \|\mathcal{P}_0 D_k^2\| \le \sqrt{n} \sum_{k=0}^{\infty} \|D_k^2 - (D_k')^2\| \\ &\le \sqrt{n} \sum_{k=0}^{\infty} \|D_k - D_k'\|_4 \|D_k + D_k'\|_4 \\ &= O(\sqrt{n}) \sum_{k=0}^{\infty}\sum_{t=0}^{\infty} \|X_{t+k} - X_{t+k}'\|_4 = O(\sqrt{n}) \qquad \textbf{(A.2)}\end{aligned}$$

as a consequence of (34). We now shall apply Theorem 1 to T_n by verifying conditions (12)–(15). By Lemma 5, which is given later in this Appendix,

$$A_{t,n} = \sum_{j=1}^{t-1} a_n^2(t-j) \le \sum_{j=1}^{n} a_n^2(j) =: A_{n+1,n} = \frac{n}{2}(\chi + s_0^2)$$

and

$$\sigma_n^2 = \sum_{t=2}^{n} A_{t,n} = \sum_{j=1}^{n} (n-j) a_n^2(j) = \frac{n^2}{4}(\chi + s_0^2) - \frac{n}{2}\tau^2.$$

So (12) follows because $\tau^2 = O(m\chi)$ and $m = o(n^{2/3})$. Note that for any l, $a_n^2(l) = O(m\chi) = o(n^{2/3}\chi)$, (13) easily follows. Because $\sum_{t=1+t'}^{n} a_n^2(t-t') \le A_{n+1,n}$, we have (14). It is slightly more complicated to verify (15). To this end, for $\delta \in (0, 1/4)$ let

$$\Gamma_\delta = \{(k,t) \in \mathbb{N}\times\mathbb{N} : \delta n < k < (1-\delta)n, \delta n < t < (1-\delta)n, \delta n < k - t\}$$

and $c_n(k,t) = \sum_{j=1+k}^{n} a_n(j-k) a_n(j-t)$. By the Cauchy–Schwarz inequality, $|c_n(k,t)| \le A_{n+1,n}$. For $l \ge 1$, let $h_v = \sum_{j=-m}^{v} \cos(l\lambda_j)$, $v \ge -m$, and $h_{-m-1} = 0$. Then $\sup_{v\ge -m} |h_v| \le 2/|1 - \exp(\sqrt{-1}\, l\lambda_1)| = O[1/\sin(l\lambda_1/2)]$. Using the Abelian summation technique,

$$|a_n(l)| = \left| \sum_{j=-m}^{m} s_j (h_j - h_{j-1}) \right| \le O(n/l) \left[|s_{-m}| + \sum_{j=1-m}^{m} |s_j - s_{j-1}| \right] = O(\omega n/l).$$

Hence $|a_n(l)| = O(\omega)$ uniformly over $\delta n \le l \le (1-\delta)n$. By (35),

$$\frac{\sup_{k,t\in\Gamma_\delta}|c_n(k,t)|}{n\chi} \le \frac{1}{n\chi}\max_{\delta n\le l\le(1-\delta)n}|a_n(l)|\max_{\delta n\le k}\sum_{j=k+1}^{n}|a_n(j-k)|$$

$$\le \frac{O(\omega)}{n\chi}\sum_{l=1}^{n-\delta n}\omega n/l = \frac{O(\omega^2\log n)}{\chi} = o(1)$$

as $n\to\infty$. Therefore,

$$\limsup_{n\to\infty}\frac{1}{\sigma_n^4}\sum_{k=1}^{n-1}\sum_{t=1}^{k-1}c_n^2(k,t) \le \limsup_{n\to\infty}\frac{1}{\sigma_n^4}\sum_{k,t\in\Gamma_0-\Gamma_\delta} + \limsup_{n\to\infty}\frac{1}{\sigma_n^4}\sum_{k,t\in\Gamma_\delta}$$

$$\le \limsup_{n\to\infty}\frac{1}{\sigma_n^4}A_{n+1,n}^2\#(\Gamma_0-\Gamma_\delta) = 8\delta - 8\delta^2,$$

which completes the proof of (15) because $\delta > 0$ can be made arbitrarily small. Here $\#(\Gamma_0 - \Gamma_\delta)$ denotes the number of elements in the set $\Gamma_0 - \Gamma_\delta$.

(ii) *Outline of the proof in the case* $\theta \in (0,\pi)$. The proof for this case can be done in an analogous way as in the case $\theta = 0$, and it does not involve additional methodological difficulties. However, it does involve quite lengthy manipulations. Here we only provide an outline of the arguments. Let $D_k(\theta) = Y_k(\theta) - \mathbb{E}[Y_k(\theta)|\mathcal{F}_{k-1}] = A_k + \sqrt{-1}B_k$, where both (A_k) and (B_k) are real, stationary martingale difference sequences with $A_k, B_k \in \mathcal{L}^q$. Write

$$\sum_{j=-m}^{m} s_j\exp[\sqrt{-1}l(\theta+\lambda_j)] = a_n(l) + \sqrt{-1}b_n(l),$$

where $a_n(l)$ and $b_n(l)$ are real numbers. The quantity $\Lambda_n(\theta)$ in (A.1) can be rewritten as

$$\Lambda_n(\theta) = 2\sum_{1\le k<k'\le n}\Re\{[a_n(k'-k)+\sqrt{-1}b_n(k'-k)]D_k(\theta)\bar{D}_{k'}(\theta)\} + \tau\Omega_n$$

$$= 2\sum_{1\le k<k'\le n}\{[a_n(k'-k)A_k - b_n(k'-k)B_k]A_{k'}$$

$$+ [a_n(k'-k)B_k + b_n(k'-k)A_k]B_{k'}\} + \tau\Omega_n$$

$$= 2\sum_{k'=2}^{n}[Z_{1,n}(k'-1)A_{k'} + Z_{2,n}(k'-1)B_{k'}] + \tau\Omega_n,$$

where $\Omega_n = \sum_{k=1}^n |D_k(\theta)|^2 - n\|D_k(\theta)\|^2$,

$$Z_{1,n}(k'-1) = \sum_{k=1}^{k'-1}[a_n(k'-k)A_k - b_n(k'-k)B_k],$$

and

$$Z_{2,n}(k'-1) = \sum_{k=1}^{k'-1} [a_n(k'-k)B_k + b_n(k'-k)A_k].$$

Similar to (A.2), we have $\|\Omega_n\| = O(\sqrt{n})$, and consequently $\|\tau\Omega_n\| = O(|\tau|\sqrt{n}) = o(n\chi^{1/2})$. Because $Z_{1,n}(k'-1)A_{k'} + Z_{2,n}(k'-1)B_{k'}$, $2 \le k' \le n$, form martingale differences, we can again apply the martingale central limit theorem. Following the proof of Theorem 1, it is easily seen that the Lindeberg condition holds. It remains to verify the convergence of the sum of conditional variance

$$\begin{aligned}\frac{1}{\sigma_n^2}\sum_{k'=2}^{n} &\mathbb{E}\{[Z_{1,n}(k'-1)A_{k'} + Z_{2,n}(k'-1)B_{k'}]^2 | \mathcal{F}_{k'-1}\} \\ &= \frac{1}{\sigma_n^2}\sum_{k'=2}^{n} \{Z_{1,n}^2(k'-1)\mathbb{E}(A_{k'}^2|\mathcal{F}_{k'-1}) + Z_{2,n}^2(k'-1)\mathbb{E}(B_{k'}^2|\mathcal{F}_{k'-1}) \\ &\qquad + 2Z_{1,n}(k'-1)Z_{2,n}(k'-1)\mathbb{E}(A_0B_0|\mathcal{F}_{-1})\} \to \|D_0(\theta)\|^4 \quad \text{in } \mathcal{L}^1. \qquad \textbf{(A.3)}\end{aligned}$$

By (ii) of Lemma 5,

$$\lim_{n\to\infty}\frac{1}{\sigma_n^2}\sum_{k'=2}^{n}\sum_{k=1}^{k'-1} a_n^2(k'-k) = 1 \quad \text{and} \quad \lim_{n\to\infty}\frac{1}{\sigma_n^2}\sum_{k'=2}^{n}\sum_{k=1}^{k'-1} b_n^2(k'-k) = 1.$$

Therefore, because

$$\begin{aligned}\|Z_{1,n}(k'-1)\|^2 &= \sum_{k=1}^{k'-1} [a_n^2(k'-k)\|A_0\|^2 + b_n^2(k'-k)\|B_0\|^2 \\ &\qquad - 2a_n(k'-k)b_n(k'-k)\mathbb{E}(A_0B_0)], \\ \|Z_{2,n}(k'-1)\|^2 &= \sum_{k=1}^{k'-1} [a_n^2(k'-k)\|B_0\|^2 + b_n^2(k'-k)\|A_0\|^2 \\ &\qquad + 2a_n(k'-k)b_n(k'-k)\mathbb{E}(A_0B_0)], \\ \mathbb{E}[Z_{1,n}(k'-1)Z_{2,n}(k'-1)] &= \sum_{k=1}^{k'-1} \{[a_n^2(k'-k) - b_n^2(k'-k)]\mathbb{E}(A_0B_0) \\ &\qquad + a_n(k'-k)b_n(k'-k)(\|A_0\|^2 - \|B_0\|^2)\},\end{aligned}$$

and $\|D_0(\theta)\|^2 = \|A_0\|^2 + \|B_0\|^2$, we have

$$\begin{aligned}\frac{1}{\sigma_n^2}\sum_{k'=2}^{n} &\{\|Z_{1,n}(k'-1)\|^2\|A_{k'}\|^2 + \|Z_{2,n}(k'-1)\|^2\|B_{k'}\|^2 \\ &+ 2\mathbb{E}[Z_{1,n}(k'-1)Z_{2,n}(k'-1)]\mathbb{E}(A_{k'}B_{k'})\} \to \|D_0(\theta)\|^4. \qquad \textbf{(A.4)}\end{aligned}$$

An analogue of Lemma 2 indicates that (A.3) follows from

$$\frac{1}{\sigma_n^2}\sum_{k'=2}^{n}\{Z_{1,n}^2(k'-1)\|A_0\|^2 + Z_{2,n}^2(k'-1)\|B_0\|^2$$
$$+ 2Z_{1,n}(k'-1)Z_{2,n}(k'-1)\mathbb{E}(A_0B_0)\} \to \|D_0(\theta)\|^4 \quad \text{in } \mathcal{L}^2.$$

The preceding convergence is a consequence of (A.4) and a similar version of Lemma 3. ■

LEMMA 5. *Let m be a positive integer with $m < n/2 - 1$; let $s_j, -m \le j \le m$, be real numbers, $\tau_m(s) = \sum_{j=-m}^m s_j$, $\varpi_m(s) = \sum_{j=-m}^m |s_j|$, and $\chi_m(s) = \sum_{j=-m}^m s_j^2$.*

(i) Assume that $s_j = 0$ for $-m \le j \le -1$. For $l \in \mathbb{Z}$ let

$$a_n(l) = \sum_{j=-m}^{m} s_j \cos(l\lambda_j). \tag{A.5}$$

Then

$$\sum_{l=1}^{n} a_n^2(l) = \frac{n}{2}[\chi_m(s) + s_0^2] \tag{A.6}$$

and

$$\sum_{l=1}^{n} l a_n^2(l) = \frac{n^2}{4}[\chi_m(s) + s_0^2] + \frac{n}{2}\tau_m^2(s). \tag{A.7}$$

(ii) For a fixed $\theta \in (0,\pi)$ and $l \in \mathbb{Z}$ let

$$a_n(l) = \sum_{j=-m}^{m} s_j \cos[l(\lambda_j+\theta)] \quad \text{and} \quad b_n(l) = \sum_{j=-m}^{m} s_j \sin[l(\lambda_j+\theta)]. \tag{A.8}$$

Assume that $m = o(n)$. Then

$$\lim_{n\to\infty}\frac{\sum_{l=1}^{n} a_n^2(l)}{n\chi_m(s)} = \lim_{n\to\infty}\frac{\sum_{l=1}^{n} b_n^2(l)}{n\chi_m(s)} = \frac{1}{2} \tag{A.9}$$

and

$$\lim_{n\to\infty}\frac{\sum_{l=1}^{n} l a_n^2(l)}{n^2\chi_m(s)} = \lim_{n\to\infty}\frac{\sum_{l=1}^{n} l b_n^2(l)}{n^2\chi_m(s)} = \frac{1}{4}. \tag{A.10}$$

Proof. (i) For $-m \le j, j' \le m$ let

$$\mu_n(j,j') = \sum_{l=1}^{n}\cos(l\lambda_j)\cos(l\lambda_{j'}) \quad \text{and} \quad \nu_n(j,j') = \sum_{l=1}^{n} l\cos(l\lambda_j)\cos(l\lambda_{j'}).$$

Because $\lambda_j = j\lambda_1$, basic trigonometric manipulations imply that $\mu_n(j,j') = n$ if $j = j' = 0$; $\mu_n(j,j') = 0$ if $j \neq j'$; $\mu_n(j,j') = n/2$ if $j = j' \neq 0$; $\nu_n(j,j') = n(n+1)/2$ if $j = j' = 0$; $\nu_n(j,j') = n/2$ if $j \neq j'$; and $\nu_n(j,j') = n(n+2)/4$ if $j = j' \neq 0$. Using these trigonometric identities, (A.6) and (A.7) easily follow.

(ii) We only prove (A.9) because (A.10) can be proved in a similar way. As in (i), let

$$\begin{aligned}\mu_n(j,j';\theta) &= \sum_{l=1}^{n} \cos[l(\lambda_j+\theta)]\cos[l(\lambda_{j'}+\theta)] \\ &= \frac{1}{2}\sum_{l=1}^{n}\{\cos[l(\lambda_{j+j'}+2\theta)] + \cos(l\lambda_{j-j'})\}.\end{aligned}$$

Let $\beta \in (0, 2\pi)$. Then

$$\begin{aligned}\left|\sum_{l=1}^{n}\cos(l\beta)\right| &\leq \left|\sum_{l=1}^{n}\exp(\sqrt{-1}l\beta)\right| \\ &\leq \frac{2}{|1-\exp(\sqrt{-1}\beta)|} = \frac{1}{|\sin(\beta/2)|}.\end{aligned} \tag{A.11}$$

If $j = j'$, because for sufficiently large n, $\theta/2 < \lambda_j + \theta < \theta + (\pi - \theta)/2$ uniformly over $j = -m, \ldots, m$, (A.11) implies that $\mu_n(j,j;\theta) = n/2 + O(1)$. On the other hand, if $j \neq j'$, we similarly have $\mu_n(j,j';\theta) = O(1)$ uniformly over j and j'. Therefore,

$$\begin{aligned}\sum_{l=1}^{n} a_n^2(l) &= \sum_{j,j'=-m}^{m} s_j s_{j'}\,\mu_n(j,j';\theta) \\ &= \sum_{j=-m}^{m} s_j^2\,\mu_n(j,j;\theta) + O(1)\sum_{j\neq j'}|s_j s_{j'}| \\ &= \frac{n}{2}\chi_m(s) + \sum_{j=-m}^{m} O(s_j^2) + O[\varpi_m^2(s)] \\ &= \frac{n}{2}\chi_m(s)[1+o(1)]\end{aligned}$$

because $\varpi_m^2(s) \leq (2m+1)\chi_m(s) = o[n\chi_m(s)]$. ■