The Annals of Statistics
2014, Vol. 42, No. 1, 87–114
DOI: 10.1214/13-AOS1184

INFERENCE OF WEIGHTED V-STATISTICS FOR NONSTATIONARY TIME SERIES AND ITS APPLICATIONS

BY ZHOU ZHOU[1]

University of Toronto

We investigate the behavior of Fourier transforms for a wide class of nonstationary nonlinear processes. Asymptotic central and noncentral limit theorems are established for a class of nondegenerate and degenerate weighted V-statistics through the angle of Fourier analysis. The established theory for V-statistics provides a unified treatment for many important time and spectral domain problems in the analysis of nonstationary time series, ranging from nonparametric estimation to the inference of periodograms and spectral densities.

1. Introduction. Consider the following weighted V-statistics:

$$V_n = \sum_{k=1}^{n} \sum_{j=1}^{n} W_n(t_k, t_j) H(X_k, X_j), \tag{1}$$

where $\{X_k\}_{k=1}^n$ is a nonstationary time series, $t_k = k/n$, the kernel $H(\cdot, \cdot)$ and the weights $W_n(\cdot, \cdot)$ are symmetric and Borel measurable functions. Many important time and spectral domain problems in the analysis of nonstationary time series boil down to the investigation of weighted V-statistics in the form of (1). For instance, in various situations one may be interested in estimating parameter functions

$$\theta(t) = \int_{\mathbb{R}} \int_{\mathbb{R}} H(u, v)\, dF(t, u)\, dF(t, v)$$

over time t, where $F(t, \cdot)$ is the marginal distribution of $\{X_j\}$ at time t. In this case many nonparametric estimators of $\theta(t)$ are asymptotically equivalent to the weighted V-statistics

$$V_n = \sum_{k=1}^{n} \sum_{j=1}^{n} K\big((t_k - t)/b_n, (t_j - t)/b_n\big) H(X_k, X_j),$$

where $K(\cdot, \cdot)$ is two-dimensional kernel function, and b_n is a bandwidth that restricts the estimation in a neighborhood of t. Additionally, after the nonparametric

Received March 2013; revised October 2013.

[1]Supported by NSERC of Canada.

MSC2010 subject classifications. 62E20, 60F05.

Key words and phrases. V-statistics, Fourier transform, nondegeneracy, degeneracy, locally stationary time series, nonparametric inference, spectral analysis.

fitting one may want to specify or test whether the parameter function is of certain parametric forms. In this case many $\mathcal{L}^2$ distance based test statistics are of the form (1) with degenerate kernels; that is, kernels $H(\cdot,\cdot)$ such that $\mathbb{E}H[X_k, x] = 0$ for every k and x. See, for instance, the $\mathcal{L}^2$ distance based quantile specification test in Zhou (2010). Furthermore, in spectral analysis of the nonstationary time series $\{X_k\}$, both the periodogram

$$I_n(\lambda) = \frac{1}{2\pi n}\left|\sum_{j=1}^{n} X_j \exp(ij\lambda)\right|^2, \qquad 0 \le \lambda \le \pi,$$

and the classic smoothed periodogram estimate of the spectral density

$$\tilde{f}_n(\lambda) = \int_{\mathbb{R}} \frac{1}{m} K\left(\frac{u}{m}\right) I_n(\lambda + 2\pi u/n)\, du$$

are in the form of (1) with kernel $H(x, y) = xy$. Here $i = \sqrt{-1}$ stands for the imaginary unit, $K(\cdot)$ is a kernel function and $m = m_n$ is a window size satisfying $m \to \infty$ with $m/n \to 0$. Note that in every example above, $W_n(\cdot,\cdot)$ involves a tuning parameter (either bandwidth or window size) which varies with the sample size n. Hence it is important to write the weights as a function of n in (1).

The purpose of the paper is to establish an asymptotic theory for (1) through the angle of Fourier analysis. To illustrate the main idea, suppose that $H^*(x, y) = H(x, y)/(L(x)L(y))$ is absolutely integrable on $\mathbb{R}^2$ for some function $L(\cdot)$. Then under mild conditions $H^*(\cdot,\cdot)$ admits the Fourier representation $H^*(x, y) = \int_{\mathbb{R}^2} g(u, v)e^{i(xu+yv)}\, du\, dv$. Consequently V_n can be written as

$$V_n = \int_{\mathbb{R}^2} g(x, y) \sum_{k,j=1}^{n} W_n(t_k, t_j)\beta_k(x)\beta_j(y)\, dx\, dy, \tag{2}$$

where $\beta_k(x) = L(X_k)\exp(ixX_k)$. Note that in (2) the complex structure of V_n is reduced to a process of quadratic forms $\{\sum_{k,j} W_n(t_k, t_j)\beta_k(x)\beta_j(y)\}_{x,y\in\mathbb{R}}$ through the aid of Fourier transformation. The multiplicative structure of the quadratic forms makes in-depth asymptotic investigations possible for a wide class of nonstationary time series. Furthermore, the continuous structure of $\beta_k(x)$ in x makes a stochastic equi-continuity and continuous mapping argument possible which is shown to be powerful compared to the discrete spectral decomposition methods used in the literature. With the aid of the above structural simplifications, in this paper we are able to establish a uniform approximation scheme of $\{\sum_{k,j} W_n(t_k, t_j)\beta_k(x)\beta_j(y)\}_{x,y\in\mathbb{R}}$ by a process of Gaussian quadratic forms. As a consequence we establish a unified asymptotic theory for a class of nondegenerate and degenerate weighted V-statistics with reflexible weight functions. Both central and noncentral limit theorems are developed for a wide class of nonstationary time series with both smoothly and abruptly changing data generating mechanisms over time. The established theory can be applied to many problems in the study of nonstationary time series, including topics such as nonparametric estimation and specification, periodogram and spectral density inferences discussed above.

In (1), if the summation is taken over indices $1 \le k \ne j \le n$, then the statistics are commonly called weighted U-statistics in the literature. For many problems that are of practical importance, statistical theory for the U- and V-statistics can be established with essentially the same techniques. Since the seminal papers of von Mises (1947) and Hoeffding (1948), the analysis of V- and U-statistics has attracted much attention in the statistics and probability literature. It seems that most efforts have been put in un-weighted U- or V-statistics with $W_n(\cdot,\cdot) \equiv 1$ for stationary data. See, for instance, Yoshihara (1976), Dehling and Taqqu (1989), Huskova and Janssen (1993), Dehling and Wendler (2010), Leucht (2012) and Beutner and Zähle (2012, 2013) for various approaches for nondegenerate and degenerate un-weighted U- and V-statistics. We also refer to the monographs of Denker (1985), Lee (1990) and Dehling (2006) for more references. As we observe from the examples in the beginning of this Introduction, it is important to consider weighted V-statistics with sample size dependent weights in the study of nonstationary time series.

There are a small number of papers discussing weighted V- or U-statistics in the literature. See, for instance, de Jong (1987), O'Neil and Redner (1993), Major (1994) and Rifi and Utzet (2000) for weighted U- and V-statistics of independent data and Hsing and Wu (2004) for weighted nondegenerate U-statistics of stationary time series. For most of the above discussions, the weights are not allowed to be sample size dependent. Exceptions include de Jong (1987) who discovered a very deep result that $\theta_{n,1} \to 0$ implies asymptotic normality of a very wide class of weighted degenerate V-statistics for independent data, where $\theta_{n,1}$ is the eigenvalue of the matrix $\{W_n(t_j,t_k)/\sqrt{\sum_{u,v=1}^n W_n^2(t_u,t_v)}\}_{j,k=1}^n$ with the maximum absolute value. There the proof heavily depended on the martingale structure of degenerate V-statistics of independent data and is hard to generalize to the time series setting. In this paper, from a Fourier analysis angle, we generalize the result of de Jong (1987) and show that, for many temporally dependent processes, $\theta_{n,1} \to 0$ implies asymptotic normality of V_n for a class of degenerate kernels and weight functions.

Quadratic forms are special cases of (1) with $H(x,y) = xy$. There are many papers in the literature devoted to the analysis of such statistics. See, for instance, de Wet and Venter (1973), Fox and Taqqu (1987), Götze and Tikhomirov (1999), Gao and Anh (2000) and Bhansali, Giraitis and Kokoszka (2007), among others. It seems that most of the results are on independent or stationary data. Exceptions include Lee and Subba Rao (2011) who recently studied asymptotic normality of a class of quadratic forms with banded weight matrix for α-mixing nonstationary time series. For independent data, Götze and Tikhomirov (1999), among others, established deep theoretical results indicating that distributions of generic quadratic forms can be approximated by those of corresponding Gaussian quadratic forms. In this paper, we generalize this type of result and show that the laws of a wide class of quadratic forms for nonstationary time series can be well approximated by the distributions of corresponding quadratic forms of independent Gaussian random variables. Consequently, central and noncentral limit theorems are established for the latter class of quadratic forms for nonstationary time series.

The rest of the paper is organized as follows. In Section 2 we shall introduce the class of absolutely convergent Fourier transformations and the piece-wise locally stationary time series models used in this paper. Sections 3 and 4 establish the asymptotic theory for nondegenerate and degenerate V-statistics, respectively. Theory for quadratic forms will be covered in Section 4 as a special case of degenerate V-statistics. In Section 5, we shall apply our theory to the problems of nonparametric estimation as well as spectral analysis of nonstationary processes. Several examples will be discussed in detail. Finally, the theoretical results are proved in Section 6.

2. Preliminaries. We first introduce some notation. For a vector $\mathbf{v} = (v_1, v_2, \ldots, v_p) \in \mathbb{R}^p$, let $|\mathbf{v}| = (\sum_{i=1}^p v_i^2)^{1/2}$. Let $i = \sqrt{-1}$ be the imaginary unit. For a complex number $z = x + yi \in \mathbb{C}$, write $|z| = \sqrt{x^2 + y^2}$. For $q > 0$, denote by $L^q(\mathbb{R}^p)$ the collection of functions $f : \mathbb{R}^p \to \mathbb{C}$ such that $\int_{\mathbb{R}^p} |f(\mathbf{x})|^q \, d\mathbf{x} < \infty$. For a function $f \in L^1(\mathbb{R}^p)$, denote by $\hat{f}$ its Fourier transform, that is, $\hat{f}(\mathbf{v}) = \int_{\mathbb{R}^p} f(\mathbf{x}) e^{-i\langle \mathbf{x}, \mathbf{v}\rangle} \, d\mathbf{x}$. For a Borel set A in $\mathbb{R}^p$, denote by $\mathfrak{B}(A)$ the collection of all Borel sets in A. For a random vector $\mathbf{V}$, write $\mathbf{V} \in \mathcal{L}^q$ $(q > 0)$ if $\|\mathbf{V}\|_q := [\mathbb{E}(|\mathbf{V}|^q)]^{1/q} < \infty$ and $\|\mathbf{V}\| = \|\mathbf{V}\|_2$. Denote by $\Rightarrow$ the weak convergence. The symbol C denotes a generic finite constant which may vary from place to place.

2.1. *Absolutely convergent Fourier transforms.* Following the classic notation, a function $f : \mathbb{R}^2 \to \mathbb{R}$ is said to belong to $W_0(\mathbb{R}^2)$ if there exists a function g, $\mathbb{R}^2 \to \mathbb{C}$, such that $g \in L^1(\mathbb{R}^2)$ and

$$f(x, y) = \int_{\mathbb{R}^2} g(t, s) e^{itx + isy} \, dt \, ds. \tag{3}$$

The class $W_0(\mathbb{R}^2)$ is called the Wiener ring or Wiener algebra, naming after Norbert Wiener for his fundamental contributions in the study of absolutely convergent Fourier integrals. Due to its importance in various problems, the possibility to represent a function as an absolutely convergent Fourier integral has been intensively studied in mathematics, and various sufficient conditions are continuously being discovered in recent years. For a relatively comprehensive survey, we refer to Liflyand, Samko and Trigub (2012).

For the study of V-statistics of dependent data, we need to further define the subclass $W_0^\delta(\mathbb{R}^2)$ of $W_0(\mathbb{R}^2)$ as follows. We call $f \in W_0^\delta(\mathbb{R}^2)$ for some $\delta \ge 0$ if $f \in W_0(\mathbb{R}^2)$ with representation (3) and

$$\int_{\mathbb{R}^2} \left|(t, s)\right|^\delta \left|g(t, s)\right| dt \, ds < \infty, \tag{4}$$

where $0^0 := 1$. Note that $W_0^0(\mathbb{R}^2) = W_0(\mathbb{R}^2)$.

In order for $f \in W_0(\mathbb{R}^2)$, it is necessary that f is uniformly continuous on $\mathbb{R}^2$ and vanishes at ∞. However, the above conditions are generally not sufficient. Generally speaking, the Fourier transform of a smoother function tends to have a lighter tail at infinity. The latter fact implies that stronger smoothness conditions are needed to insure $f \in W_0(\mathbb{R}^2)$.

PROPOSITION 1. *Suppose that a symmetric function* $f \in L^1(\mathbb{R}^2)$. (i) [*Liflyand*, *Samko and Trigub* (2012), *Theorem* 10.11]. *Assume that f is uniformly continuous on $\mathbb{R}^2$ and vanishes at ∞. Let f and its partial derivative $(\partial/\partial x)f$ be locally absolutely continuous on $(\mathbb{R}\setminus\{0\})^2$ in each variable. Further let each partial derivative $(\partial/\partial x)f$ and $[\partial^2/(\partial x\,\partial y)]f$ exist and belong to $L^p(\mathbb{R}^2)$ for some $p \in (1,\infty)$. Then $f \in W_0(\mathbb{R}^2)$.* (ii) *If f satisfies*

$$
\begin{aligned}
&\left|\frac{\partial^2}{\partial x^2} f(x+\epsilon, y) - \frac{\partial^2}{\partial x^2} f(x,y)\right| + \left|\frac{\partial^2}{\partial x^2} f(x, y+\epsilon) - \frac{\partial^2}{\partial x^2} f(x,y)\right| \\
&\quad + \left|\frac{\partial^2}{\partial x\,\partial y} f(x+\epsilon, y) - \frac{\partial^2}{\partial x\,\partial y} f(x,y)\right| \qquad (5)\\
&\quad \le C|\epsilon|^{\gamma} k(x,y)
\end{aligned}
$$

for sufficiently small ϵ, where $\gamma > 0$ and $k \in L^1(\mathbb{R}^2)$, then $f \in W_0^{\delta}(\mathbb{R}^2)$ for any $\delta \in [0,\gamma)$.

Proposition 1 gives some easily checkable sufficient conditions for $f \in W_0(\mathbb{R}^2)$ and $W_0^{\delta}(\mathbb{R}^2)$ based on the smoothness of its partial derivatives. In the literature, numerous other sufficient conditions based on various notions of smoothness or variation are available; see, for instance, the review of Liflyand, Samko and Trigub (2012). We point out here that the conditions in Proposition 1 are not minimal sufficient. For instance, the function $f = \exp(-|\mathbf{x}|) \in W_0^{\delta}(\mathbb{R}^2)$ for any $\delta \in [0, 1/2)$. But the latter function is not differentiable on $\mathbb{R}^2$. Thanks to the fast computation of Fourier transforms, in practice when facing specific choice of f, condition (4) with $g = \hat{f}$ can also be checked via numerical computation.

2.2. *Nonstationary time series models.* For an observed process $\{X_j\}_{j=1}^n$, consider the class of nonstationary time series models of the form [Zhou (2013), Definition 1]

$$
X_k = \sum_{j=0}^{r} G_j(t_k, \mathcal{F}_k) I_{(b_j, b_{j+1}]}(t_k), \qquad k = 1, 2, \ldots, n, \tag{6}
$$

where $b_1 < b_2 < \cdots < b_r$ are r unknown (but nonrandom) break points with $b_0 = 0$, and $b_{r+1} = 1$, $I_{\cdot}(\cdot)$ is the indicator function, $G_j : (b_j, b_{j+1}] \times \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ are $(\mathfrak{B}((b_j, b_{j+1}]) \times \mathfrak{B}(\mathbb{R})^{\mathbb{N}}, \mathfrak{B}(\mathbb{R}))$-measurable functions, $j = 0, \ldots, r$, $\mathcal{F}_k =$

$(\ldots, \varepsilon_{k-1}, \varepsilon_k)$ and ε_k's are i.i.d. random variables. Recall that $t_k = k/n$. We observe from (6) that the data generating mechanism could break at the points b_k, $k = 1, \ldots, r$ and hence lead to structural breaks of $\{X_j\}$ at the latter points. On the other hand, note that if $G_k(t, \cdot)$ are smooth functions of t for each k, then the data generating mechanisms change smoothly between adjacent break points. As a consequence $\{X_j\}$ is approximately stationary in any small temporal interval between adjacent break points. By the above discussion, we shall call the class of time series in the form of (6) piece-wise locally stationary (PLS) processes.

Time series in the form of (6) constitute a relatively large class of nonstationary time series models which allow the data generating mechanism to change flexibly over time. In particular, when the number of the break points $r = 0$, then (6) reduces to the locally stationary time series models in Zhou and Wu (2009). On the other hand, if for each $k = 0, 1, \ldots, r$, $G_k(t, \mathcal{F}_j)$ does not depend on t, then (6) is a piece-wise stationary time series which is studied, for instance, in Davis, Lee and Rodriguez-Yam (2006), among others. Process (6) can be viewed as a time-varying nonlinear system with ε_k's being the inputs and X_k's being the outputs. The functions $G_k(t, \ldots)$ can be viewed as time-varying filters of the system. From this point of view, we adapt the following dependence measures of $\{X_k\}$ in Zhou (2013):

$$\delta(j, p) = \sup_t \max_k \| G_k(t, \mathcal{F}_k) - G_k(t, \mathcal{F}_{k,k-j}) \|_p, \tag{7}$$

where $\mathcal{F}_{j,k}$ is a coupled version of $\mathcal{F}_j$ with ε'_k in $\mathcal{F}_j$ replaced by an i.i.d. copy ε'_k; that is,

$$\mathcal{F}_{j,k} = (\ldots, \varepsilon_{k-1}, \varepsilon'_k, \varepsilon_{k+1}, \ldots, \varepsilon_{j-1}, \varepsilon_j),$$

and $\{\varepsilon'_j\}$ is an i.i.d. copy of $\{\varepsilon_j\}$.

We observe from (7) that $\delta(j, p)$ measures the impact of the system's inputs j-steps before on the current output of the system. When $\delta_{j,p}$ decays fast to zero as j tends to infinity, we have short memory of the series as the system tends to fast "forget about" past inputs. We refer to Section 2 on page 728 of Zhou (2013) and Section 4 on pages 2706–2708 of Zhou and Wu (2009) for more examples of linear and nonlinear nonstationary time series of the form (6) and detailed calculations of the dependence measures (7).

3. Nondegenerate V-statistics. Suppose that, for some function $L(\cdot)$, $H^*(s, t) := H(s, t)/[L(s)L(t)] \in W_0(\mathbb{R}^2)$ and $\max_j \|L(X_j)\| < \infty$. Then V_n defined in (1) admits the representation (2) with $\int_{\mathbb{R}^2} |g(x, y)|\, dx\, dy < \infty$. Now define

$$H_j(x) = \mathbb{E}[H(x, X_j)] = \mathbb{E}[H(X_j, x)] = \int_{\mathbb{R}^2} g(t, s) L(x) \exp(itx) \mathbb{E}[\beta_j(s)]\, dt\, ds$$

and $\gamma_j(x) = \beta_j(x) - \mathbb{E}[\beta_j(x)]$. Then elementary calculations using the Hoeffding's decomposition show that V_n can be decomposed as

$$V_n - \mathbb{E}V_n = 2N_n + D_n - \mathbb{E}[D_n], \tag{8}$$

where

$$N_n = \sum_{k=1}^{n} \sum_{j=1}^{n} W_n(t_k, t_j)\{H_j(X_k) - \mathbb{E}[H_j(X_k)]\},$$

$$D_n = \int_{\mathbb{R}^2} g(x, y) \sum_{k,j} W_n(t_k, t_j)\gamma_k(x)\gamma_j(y)\, dx\, dy.$$

Here N_n and D_n are the nondegenerate and degenerate part of V_n, respectively.

To investigate the limiting behavior of N_n, we need the following conditions:

(A1) For some $\eta \in (0, 1]$, $H^*(t, s) := H(t, s)/[L(t)L(s)] \in W_0^{\eta}(\mathbb{R}^2)$ for some function $L(\cdot)$.

(A2) The function L is differentiable with derivative L'. $\max_{a \le t \le b} |L'(t)| \le C(|L'(a)| + |L'(b)| + 1)$ for all a, b, where C is a finite constant independent of a and b. $\max_j \|L(X_j)\|_{4+2\epsilon} < \infty$ and $\max_j \|L'(X_j)\|_{4+2\epsilon} < \infty$ for some $\epsilon > 0$.

(A3) Define $W_{j,\cdot} = \sum_{r=1}^{n} |W_n(t_j, t_r)|$, $j = 1, 2, \dots, n$. Let $W_{j,\cdot} = 0$ for $j > n$. Let $W^{(n)} = \sum_{j=1}^{n} W_{j,\cdot}^2$. For sequences l_n, m_n and $s_n = l_n + m_n$, define

$$A_j = \sum_{k=1}^{l_n} W_{s_n(j-1)+k,\cdot}^2 \quad \text{and} \quad a_j = \sum_{k=l_n+1}^{s_n} W_{s_n(j-1)+k,\cdot}^2, \qquad j = 1, 2, \dots, \lceil n/s_n \rceil.$$

Assume that there exist sequences $m_n/\log n \to \infty$ with $l_n/n \to 0$ and $m_n/l_n \to 0$, such that

$$\sum_j a_j/W^{(n)} \to 0 \quad \text{and} \quad \max_j A_j/W^{(n)} \to 0.$$

(A4) $\phi_n := \mathrm{Var}(N_n)/W^{(n)} \ge c$ for some $c > 0$ and sufficiently large n.

(A5) The dependence measures $\delta_X(k, 4+2\epsilon) = O(\rho^k)$ for some $\rho \in [0, 1)$ and $\epsilon > 0$.

A few comments on the regularity conditions are in order. The role of $L(t)$ in (A1) is to lighten the tail of $H(s, t)$ and hence make it absolutely integrable on $\mathbb{R}^2$. Some typical choices of $L(t)$ are $(1 + t^2)^p$ for kernels $H(\cdot, \cdot)$ with algebraically increasing tails and $\exp[(1 + t^2)^p]$ for kernels $H(\cdot, \cdot)$ with exponentially increasing tails. Since the function $L(t)$ controls the tail behavior of the kernel $H(\cdot, \cdot)$, condition (A1) essentially poses some smoothness requirement on $H(\cdot, \cdot)$. In practice, Proposition 1 or direct numerical computations can be used to check $H^*(t, s) \in W_0^{\delta}(\mathbb{R}^2)$.

Condition (A2) posts some moment restrictions on $L(X_j)$ and $L'(X_j)$. The requirement $\max_{a \le t \le b} |L'(t)| \le C(|L'(a)| + |L'(b)| + 1)$ is mild and in particular it is always satisfied when $|L'(\cdot)|$ is piece-wise monotone or piece-wise convex with finite many pieces. Consequently, the latter inequality holds for functions

$L(t) = (1 + t^2)^p$ or $\exp[(1 + t^2)^p]$ listed above. Condition (A3) requires that $\{1, 2, \dots, n\}$ can be divided into big blocks and small blocks such that the sum of squares of $W_{j,\cdot}$ in the small blocks is negligible and $W_{j,\cdot}$ in the big blocks satisfy a Lindeberg type condition. Condition (A3) can be checked easily in practice. Examples 1–3 below verify (A3) for some frequently used weight functions. By the proof of Theorem 1 in Section 6, $\mathrm{Var}(N_n) = O(W^{(n)})$. Condition (A4) avoids degenerate kernels with $N_n \equiv 0$ and other degenerate cases with non-Gaussian limits. For instance, if $H(x, y) = x + y$, $W_n(x, y) \equiv 1/n$ and $X_j = Y_j - Y_{j-1}$, where Y_j is a weakly stationary time series, then we have that $N_n = Y_n - Y_0$ fails to be asymptotically normal. Note in this case $W^{(n)} = n$ and $\mathrm{Var}(N_n)/W^{(n)} \to 0$. Finally (A5) requires that the dependence measures of $\{X_j\}$ decay exponentially fast to zero. The theoretical results of the paper can be derived with $\delta_X(k, p)$ decaying polynomially fast to zero at the expense of much more complicated conditions and proofs. For presentational simplicity and clarity we assume exponentially decaying dependence measures throughout the paper.

THEOREM 1. *Under conditions* (A1)–(A5), *we have*

$$N_n/\sqrt{\mathrm{Var}(N_n)} \Rightarrow N(0, 1).$$

Theorem 1 establishes a central limit result for the nondegenerate part of V_n. Due to nonstationarity, $H_j(x) = \mathbb{E}[H(x, X_j)]$ is j-dependent. Consequently asymptotic investigations for N_n are much more difficult than the stationary case. Thanks to the structural simplification by the Fourier transformation, we are able to control the dependence structure of N_n and establish the above central limit results. By Theorem 1, if the degenerate part D_n is asymptotically negligible compared to N_n, then CLT for V_n can be derived. To this end, the following Theorem 2 is crucial.

(A6) Assume that there exist functions f_n and a $p \in (0, \infty)$, such that

$$\left|W_n(t_l, t_m) - W_n(t_k, t_j)\right| \le f_n(t_k, t_j)\left|(l, m) - (k, j)\right|^p$$

for all integers $k, j, l, m \in \{1, 2, \dots, n\}$.

Condition (A6) requires that, for each fixed n, the weight function $W_n(x, y)$ grows algebraically fast at any (x, y). (A6) is mild and is satisfied by most frequently used weight functions. In particular, if $W_n(\cdot, \cdot)$ is Lipschitz continuous in the sense that $|W_n(t, s) - W_n(t_1, s)| \le q_n|t_1 - t|$ for every t, t_1 and s, then $f_n(\cdot, \cdot)$ can be chosen as Cq_n/n.

THEOREM 2. *Assume that conditions* (A2) *and* (A5) *hold with* $4 + 2\epsilon$ *therein replaced by* 8. *Further assume* (A1) *and* (A6). *Then*

$$\|D_n - \mathbb{E}D_n\|^2 = O(W_{(n)} + \Delta_n) \qquad \text{where } W_{(n)} = \sum_{k=1}^{n}\sum_{j=1}^{n} W_n^2(t_k, t_j)$$

and $\Delta_n = \sum_{k=1}^n \sum_{j=1}^n |W_n(t_k, t_j)| f_n(t_k, t_j)$.

Theorem 2 investigates the order of D_n. Observe that if X_j's are independent with $H(x, y) = xy$, then simple calculations yield $\|D_n - \mathbb{E}D_n\|^2 = O(W_{(n)})$. For many important weight functions (see, e.g., Examples 1–3 below) $\Delta_n = O(W_{(n)})$ and the order established in Theorem 2 is sharp. For un-weighted U-statistics of stationary mixing data, deep theoretical results on the order of the degenerate part of U-statistics were obtained in Yoshihara (1976) and Dehling and Wendler (2010), among others. Our Theorem 2 extends the latter results to weighted V-statistics of nonstationary processes from a Fourier analysis angle.

COROLLARY 1. *Assume that conditions* (A2) *and* (A5) *hold with* $4 + 2\epsilon$ *therein replaced by* 8. *Further assume that* (A1), (A3), (A4) *and* (A6) *hold and* $W^{(n)}/[W_{(n)} + \Delta_n] \to \infty$. *Then we have* $(V_n - \mathbb{E}V_n)/\sqrt{\operatorname{Var}(V_n)} \Rightarrow N(0, 1)$.

Corollary 1 is an immediate consequence of Theorems 1 and 2. Corollary 1 establishes a CLT for V_n. It is clear from the definitions of $W^{(n)}$ and $W_{(n)}$ that $W^{(n)} \geq W_{(n)}$. Examples 1–3 below verify $W^{(n)}/[W_{(n)} + \Delta_n] \to \infty$ and condition (A3) for some frequently used weight functions.

EXAMPLE 1. Consider the case where $W_n(t_j, t_k) = f(t_j, t_k)/n$ for some symmetric function f on $[0, 1] \times [0, 1]$ such that $|f(x_1, y) - f(x_2, y)| \leq C|x_1 - x_2|$ for all x_1, x_2 and y in $[0, 1]$. Note that the classic un-weighted V-statistics are contained in this case with $f(\cdot, \cdot) \equiv 1$. Elementary calculations yield $W_{j,\cdot} = f(t_j, \cdot) + O(1/n)$, where $f(t, \cdot) = \int_0^1 |f(t, x)|\,dx$, $W^{(n)} = n \int_0^1 (\int_0^1 |f(x, y)|\,dy)^2\,dx + O(1)$, $W_{(n)} = O(1)$ and $\Delta_n = O(1/n)$. Hence $W^{(n)}/[W_{(n)} + \Delta_n] \to \infty$ provided that f is not a constant zero function. Additionally, it is elementary to check that (A3) is satisfied for every sequence $m_n/\log n \to \infty$, $l_n/n \to 0$ and $m_n/l_n \to 0$ provided that f is not always 0.

EXAMPLE 2. In this example we investigate weight functions in the form $W_n(t_j, t_k) = g((t_j - t)/b_n, (t_k - t)/b_n)/(nb_n)$, where $t \in [0, 1]$, $g(\cdot, \cdot)$ is a continuously differentiable function and $b_n \to 0$ with $nb_n \to \infty$. This type of weights may appear in nonparametric estimation of nonstationary time series. Assume that $g(\cdot, \cdot)$ is absolutely integrable and its first-order partial derivatives are bounded on $\mathbb{R}^2$. Then elementary calculations show that $W^{(n)} = (nb_n) \int_{\mathbb{R}} (\int_{\mathbb{R}} |g(x, y)|\,dy)^2\,dx + O(1)$, $W_{(n)} = O(1)$ and $\Delta_n = O(1/(nb_n))$ for any $t \in (0, 1)$. Therefore $W^{(n)}/[W_{(n)} + \Delta_n] \to \infty$ provided that g is not always zero and (A3) is satisfied for every sequence $m_n/\log n \to \infty$, $l_n/(nb_n) \to 0$ and $m_n/l_n \to 0$ provided that g is not always 0. Similar results hold for $t = 0$ or 1.

EXAMPLE 3. Consider the class of weights $W_n(t_j, t_k) = \sqrt{m_n} h(|t_j - t_k| m_n)/n$, where $h(\cdot)$ is a continuously differentiable and nonconstant function on $[0, \infty)$

and $m_n \to \infty$ with $m_n/n \to 0$. This type of weight functions may appear, for instance, in nonparametric specification tests and spectral analysis of $\{X_j\}$. Further assume that h is absolutely integrable on $[0, \infty)$ with bounded derivatives. After some simple algebra, we have $W^{(n)} = \frac{n}{m_n}[4(\int_0^\infty |h(x)|\,dx)^2 + o(1)]$, $W_{(n)} = O(1)$ and $\Delta_n = O(m_n/n)$. Therefore $W^{(n)}/[W_{(n)} + \Delta_n] \to \infty$ and (A3) is satisfied for every sequence $m_n/\log n \to \infty$, $l_n/n \to 0$ and $m_n/l_n \to 0$.

4. Degenerate V-statistics. Without loss of generality, throughout Section 4 we assume $c \le \sum_{k=1}^n \sum_{j=1}^n W_n^2(t_k, t_j) \le C$ for some constants $0 < c \le C < \infty$. In this section we shall investigate the class of degenerate V-statistics for which $\mathbb{E}[H(x, X_j)] = 0$ for every j and x. Then it is clear that $N_n = 0$ in (8) and $V_n = D_n$. Before we state the theoretical results, we need to post the following regularity conditions:

(A7) Define

$$\mathcal{V}_n = \sum_{j=1}^{n-1}\left[\sum_{k=1}^{n}\big(W_n(t_j, t_k) - W_n(t_{j+1}, t_k)\big)^2\right]^{1/2} + \left[\sum_{k=1}^{n} W_n^2(1, t_k)\right]^{1/2}.$$

Assume that $n^{1/4}\log^2 n\mathcal{V}_n = o(1)$. Further assume that for some $\delta > 0$,

$$\sum_{j=1}^{n}\left[\sum_{|k-j|\le \log^{1+\delta} n}\big(W_n(t_j, t_j) - W_n(t_j, t_k)\big)^2\right]^{1/2} = o(1). \tag{9}$$

(A8) If $|m - l| = O(\log n)$, then $\sum_{k=m}^{l}[\sum_{|j-k|\le \log^{1+\delta} n} W_n^2(t_k, t_j)]^{1/2} = o(1)$ for some $\delta > 0$.

(A9) For $s = (s_1, s_2, \dots, s_m)^\top \in \mathbb{R}^m$ and $t \in [0, 1]$, define

$$\beta_{k,j}(t, s) = L\big(G_k(t, \mathcal{F}_j)\big)\big(\cos\big(s_1 G_k(t, \mathcal{F}_j)\big), \sin\big(s_1 G_k(t, \mathcal{F}_j)\big), \dots, \cos\big(s_m G_k(t, \mathcal{F}_j)\big), \sin\big(s_m G_k(t, \mathcal{F}_j)\big)\big)^\top.$$

Let $\beta^*_{k,j}(t, s) = (L(G_k(t, \mathcal{F}_j)), \beta^\top_{k,j}(t, s))^\top$. For $k = 0, 1, \dots, r$, define the long-run covariances $\Sigma_k(t, s) = \sum_{j=-\infty}^{\infty} \mathrm{Cov}[\beta_{k,0}(t, s), \beta_{k,j}(t, s)]$ and $\Sigma^*_k(t, s) = \sum_{j=-\infty}^{\infty} \mathrm{Cov}[\beta^*_{k,0}(t, s), \beta^*_{k,j}(t, s)]$. Assume that for all $m \in \mathbb{N}$ and all $s \in \mathbb{R}^m$ with $s_1 < s_2 < \cdots < s_m$ and $s_j \ne 0$, $\Sigma_k(t, s)$ is positive definite for $t \in [b_k, b_{k+1}]$, $k = 0, 1, \dots, r$ if $L(x) \equiv C$. Replace $\Sigma_k(t, s)$ by $\Sigma^*_k(t, s)$ in the above assumption for all other functions $L(x)$.

(A10) $\|G_k(t, \mathcal{F}_0) - G_k(s, \mathcal{F}_0)\|_4 \le C|t - s|$ for $t, s \in [b_k, b_{k+1}]$ and $k = 0, 1, \dots, r$.

Condition (A7) posts some restrictions on the smoothness of the weight function $W_n(\cdot,\cdot)$. Condition (A8) is a mild technical condition. In particular, elementary calculations show that (A7) and (A8) are satisfied by weight functions in Example 1. We have (A7) and (A8) hold if $b_n \gg n^{-1/2}\log^4 n$ under the additional assumption that $\int_{\mathbb{R}}(\int_{\mathbb{R}}(\frac{\partial g(x,y)}{\partial y})^2\,dy)^{1/2}\,dx < \infty$ in Example 2. For weights considered in Example 3, we have (A7) and (A8) hold when $m_n \ll n^{1/4}/\log^2 n$ under the extra assumption that $\int_0^{\infty}(h'(x))^2\,dx < \infty$.

By definition, $\Sigma_k(t,s)$ and $\Sigma_k^*(t,s)$ are the spectral density matrices of the time series $\{\beta_{k,j}(t,s)\}_{j=0}^{\infty}$ and $\{\beta_{k,j}^*(t,s)\}_{j=0}^{\infty}$ at frequency 0, respectively. Hence it is clear that the latter spectral density matrices are positive semi-definite. Condition (A9) is mild, and it requires that the latter spectral density matrices are nonsingular. Finally, condition (A10) means that the data generating mechanism $G_k(t,\cdot)$ changes smoothly between adjacent break points. The following theorem investigates the asymptotic behavior of degenerate weighted V-statistics:

THEOREM 3. *Write $L^*(x) = xL(x)$. Let condition* (A2′) *be condition* (A2) *when $L(x)$ therein is replaced by $L^*(x)$. Assume that conditions* (A2), (A2′) *and* (A5) *hold with $4+2\epsilon$ therein replaced by $8+4\epsilon$ for some $\epsilon > 0$. Assume* (A1) *and* (A7)–(A10). *Then there exist constants $\alpha_{n,1}, \alpha_{n,2}, \dots$ with $\sum_{k=1}^{\infty}\alpha_{n,k}^2 = O(1)$ and i.i.d. standard normal random variables $Z_1, Z_2, \dots,$ such that for any bounded and continuous function $h(\cdot)$*

$$\left|\mathbb{E}\big[h(D_n - \mathbb{E}D_n)\big] - \mathbb{E}\left[h\left(\sum_{j=1}^{\infty}\alpha_{n,j}(Z_j^2 - 1)\right)\right]\right| \to 0. \tag{10}$$

Let $\Gamma(t,s) = (\Gamma_1(t,s), \Gamma_2(t,s))^{\top}$ be a centered two-dimensional Gaussian process defined on $[0,1]\times\mathbb{R}$ with the covariance function

$$\operatorname{Cov}\big[\Gamma(t_1,s_1), \Gamma(t_2,s_2)\big] = \int_0^{\min(t_1,t_2)} \Xi_{\zeta(t)}\big(t, (s_1,s_2)^{\top}\big)\,dt,$$

where $\zeta(t) = k$ if $b_k < t \le b_{k+1}$, $k = 0, 1, \dots, r$, $\zeta(0) = 0$ and

$$\Xi_k\big(t, (s_1,s_2)^{\top}\big) = \sum_{j=-\infty}^{\infty} \operatorname{Cov}\big[\beta_{k,0}(t,s_1), \beta_{k,j}(t,s_2)\big].$$

Define the complex-valued Gaussian process $\Gamma^*(t,s) = \Gamma_1(t,s) + i\Gamma_2(t,s)$ and let

$$\Gamma_n^*(t,s) = \sqrt{n}\left[\Gamma^*(t,s) - \Gamma^*\left(t - \frac{1}{n}, s\right)\right], \qquad t \ge 1/n. \tag{11}$$

Let $\varpi_n(x,y) = \sum_{k,j=1}^{n} W_n(t_k,t_j)\Gamma_n^*(t_k,x)\Gamma_n^*(t_j,y)$. Then by the classic Gaussian process theory [see, for instance, Kuo (1975), Chapter 1.2], the real part of $\int_{\mathbb{R}^2} g(x,y)\varpi_n(x,y)\,dx\,dy$ is a quadratic form of i.i.d. Gaussian random variables $Z_1, Z_2, \dots.$ The coefficients $\alpha_{n,j},\ j = 1, 2, \dots,$ in Theorem 3 correspond to the eigenvalues of the latter Gaussian quadratic form. Theorem 3 establishes a general asymptotic result for degenerate V-statistics with smooth weight functions. Define the Lévy–Prokhorov metric

$$\pi(\mu,\nu) = \inf\big\{\epsilon > 0 \,|\, \mu(A) \le \nu(A^{\epsilon}) + \epsilon,\ \nu(A) \le \mu(A^{\epsilon}) + \epsilon \text{ for every Borel set } A\big\},$$

where A^ϵ is the ϵ-neighborhood of A. Then (10) is equivalent to

$$\pi\left(\text{law}(D_n - \mathbb{E}D_n), \text{law}\left(\sum_{j=1}^{\infty} \alpha_{n,j}(Z_j^2 - 1)\right)\right) \to 0.$$

In other words, the distribution of $D_n - \mathbb{E}D_n$ can be well approximated asymptotically by that of a weighted sum of i.i.d. centered $\chi^2(1)$ random variables. An important observation from (10) is that if $\max_{1\le j<\infty} |\alpha_{n,j}| \to 0$, then $\pi(\text{law}(\sum_{j=0}^{\infty} \alpha_{n,j}(Z_j^2 - 1)), \text{law}(N(0, 2\sum_{j=0}^{\infty} \alpha_{n,j}^2))) \to 0$. Consequently $D_n - \mathbb{E}D_n$ is asymptotically normal. On the other hand, if $\alpha_{n,j} \to \alpha_j$ uniformly in j as $n \to \infty$, then $\sum_{j=1}^{\infty} \alpha_{n,j}(Z_j^2 - 1)) \to \sum_{j=1}^{\infty} \alpha_j(Z_j^2 - 1))$ and therefore $D_n - \mathbb{E}D_n$ converges to a mixture of i.i.d. centered $\chi^2(1)$ random variables. In the following, Corollaries 2 to 4 explore the above discussions in detail.

COROLLARY 2. *Let $\theta_{n,1}, \theta_{n,2}, \ldots, \theta_{n,n}$ be the eigenvalues of the matrix $\{W_n(t_j, t_k)\}_{j,k=1,\ldots,n}$ with $|\theta_{n,1}| \ge |\theta_{n,2}| \ge |\theta_{n,n}|$. Assume that $\theta_{n,1} \to 0$. Then under the conditions of Theorem 3, we have for any bounded and continuous function $h(\cdot)$*

$$|\mathbb{E}h[D_n - \mathbb{E}D_n] - \mathbb{E}h\{N(0, \text{Var}[D_n])\}| \to 0. \tag{12}$$

Note that by Lemma 4 in Zhou (2013), $\text{Var}[D_n] = O(1)$. Corollary 2 asserts that $|\theta_{n,1}| \to 0$ implies $\max_{1\le j<\infty} |\alpha_{n,j}| \to 0$ and hence the asymptotical normality of D_n. In the literature, de Jong (1987) derived that $\theta_{n,1} \to 0$ implies asymptotic normality of a very wide class of weighted degenerate V-statistics of independent data based on very deep martingale techniques. The martingale arguments depended heavily on the independence assumption and are hard to generalize to dependent data. From a Fourier analysis point of view, Corollary 2 generalizes the latter result to a class of weighted degenerate V-statistics of nonstationary time series with smooth weights. A particular case of this type is Example 3 where the weight matrix is Toeplitz. From standard Toeplitz matrix theory we have $|\theta_{n,1}| \le \sum_{j=1}^{n} \sqrt{m_n}|h(t_j m_n)|/n = O(1/\sqrt{m_n}) = o(1)$, and hence (12) holds for this type of weight matrices.

COROLLARY 3. *Suppose that* (a): *$W_n(t, s) = Q_1(t, s)/n$ and $Q_1(t, s)$ satisfies* (5) *with $\mathbb{R}^2$ therein replaced by* $[0, 1] \times [0, 1]$; *or* (b): *$W_n(t, s) = \sum_{j=1}^{\infty} a_j \times f_{1,j}(t) f_{2,j}(s)/n$, where $\sum_{j=1}^{\infty} |a_j| < \infty$ and $f_{1,j}(\cdot)$ and $f_{2,j}(\cdot)$ are continuous functions defined on* $[0, 1]$. *Then under the conditions of Theorem* 3, *there exist constants $\alpha_1, \alpha_2, \ldots$ with $\sum_{j=1}^{\infty} \alpha_j^2 < \infty$ and i.i.d. standard normal random variables $Z_1, Z_2, \ldots$, such that*

$$D_n - \mathbb{E}D_n \Rightarrow \sum_{j=1}^{\infty} \alpha_j(Z_j^2 - 1).$$

COROLLARY 4. *Suppose* (a): $W_n(t,s) = Q_2((t-a)/b_n, (s-b)/b_n)/(nb_n)$ *for some* $a, b \in [0,1]$, *where* $Q_2(t,s)$ *has support* $[-1,1] \times [-1,1]$ *and satisfies* (5) *with* $\mathbb{R}^2$ *therein replaced by* $[-1,1] \times [-1,1]$ *and* $b_n \to 0$; *or* (b): $W_n(t,s) = \sum_{j=1}^{\infty} a_j g_{1,j}((t-a)/b_n) g_{2,j}((s-b)/b_n)/(nb_n)$ *for some* $a, b \in [0,1]$, *where* $\sum_{j=1}^{\infty} |a_j| < \infty$, $b_n \to 0$ *and* $g_{1,j}(\cdot)$ *and* $g_{2,j}(\cdot)$ *are continuous functions on* $\mathbb{R}$ *with support* $[-1,1]$. *Then under the conditions of Theorem* 3, *there exist constants* $\alpha_1, \alpha_2, \dots$ *with* $\sum_{j=1}^{\infty} \alpha_j^2 < \infty$ *and i.i.d. standard normal random variables* $Z_1, Z_2, \dots,$ *such that*

$$D_n - \mathbb{E}D_n \Rightarrow \sum_{j=1}^{\infty} \alpha_j (Z_j^2 - 1).$$

Corollaries 3 and 4 are proved in the online supplement of the paper, Zhou (2014). Corollaries 3 and 4 establish that D_n converges to a mixture of i.i.d. centered $\chi^2(1)$ random variables for four classes of smooth weight functions which are absolutely integrable on $\mathbb{R}^2$. Note that the classic un-weighted V-statistics belong to cases (a) and (b) in Corollary 3. In the literature, Leucht (2012), among others, derived asymptotic distributions of un-weighted U-statistics for stationary time series. Corollary 3 generalizes the latter results to a class of weighted V-statistics of nonstationary data. Weight functions in Corollary 4 may appear, for instance, in nonparametric estimation of nonstationary time series.

Quadratic forms of a centered nonstationary process $\{X_j\}$ are of the form

$$Q_n = \sum_{j=1}^{n} \sum_{k=1}^{n} W_n(t_j, t_k) X_j X_k. \tag{13}$$

Clearly Q_n is a special case of the degenerate V-statistics with $H(x,y) = xy$. Hence the theory established above applies to this class of statistics. However, due to the special multiplicative structure, the asymptotic theory for Q_n can be established with weaker conditions. The following proposition follows from the corresponding proofs of Theorem 3 and Corollaries 2 to 4.

PROPOSITION 2. *Assume that conditions* (A7), (A8) *and* (A10) *hold and* (A5) *holds with* $4 + 2\epsilon$ *therein replaced by* 4. *Further assume that* $\tilde{\sigma}^2(k,t) := \sum_{j=-\infty}^{\infty} \mathrm{Cov}[G_k(t, \mathcal{F}_0), G_k(t, \mathcal{F}_j)] > 0$ *for* $k = 0, 1, \dots, r$ *and* $t \in [b_k, b_{k+1}]$. *Write* $\tilde{\sigma}^2(t) = \tilde{\sigma}^2(\zeta(t), t)$. *Then we have that conclusions of Theorem* 3 *and Corollaries* 2 *to* 4 *hold with* D_n *therein replaced by* Q_n.

By the proof of Theorem 3, on a possibly richer probability space, there exist i.i.d. standard normal random variables $Z_1, Z_2, \dots, Z_n$, such that

$$Q_n - Q_n^o = o_{\mathbb{P}}(1), \qquad \text{where } Q_n^o = \sum_{j,k=1}^{n} W_n(t_j, t_k) \tilde{\sigma}(t_j) Z_j \tilde{\sigma}(t_k) Z_k. \tag{14}$$

The above equation asserts that Q_n can be well approximated a quadratic form of independent Gaussian random variables. In the literature, Götze and Tikhomirov (1999), among others, established deep theoretical results showing that distributions of quadratic forms of independent data can be approximated by those of corresponding Gaussian quadratic forms. In (14), we generalize these results to a class of quadratic forms of nonstationary time series with smooth weights.

5. Applications.

5.1. *Nonparametric estimation of nonstationary time series.* Let $F(t,\cdot)$ be the marginal distribution function of $\{X_j\}$ at time t; namely $F(t,\cdot)$ is the distribution of $G_{\zeta(t)}(t,\mathcal{F}_0)$. Under various situations one is interested in estimating the quantity

$$\theta(t)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}H(x,y)\,dF(t,x)\,dF(t,y) \tag{15}$$

for all $t\in[0,1]$. Here $H(\cdot,\cdot)$ is assumed to be a symmetric function. For instance, if $H(x,y)=(x-y)^2/2$, then $\theta(t)$ is the time-varying variance function of the process $\{X_j\}$. In the statistics literature, enormous efforts have been put on nonparametric estimation of parameter functions $\tau(t)$ in the form $\tau(t)=\int_{-\infty}^{\infty}M(x)\,dF(t,x)=E[M[G_{\zeta(t)}(t,\mathcal{F}_0)]]$; see, for instance, the monographs of Fan and Gijbels (1996) and Fan and Yao (2003) and the citations therein. Note that $\tau(t)$ is a special case of (15) with $H(x,y)=[M(x)+M(y)]/2$. On the other hand, however, it seems that there are few results on nonparametric inference of general parameter functions in the form of $\theta(t)$ in (15). One of the major difficulties, especially in the case of time series applications, lies in the lack of corresponding theoretical results on weighted V-statistics for dependent data. Define

$$\theta(t,s)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}H(x,y)\,dF(t,x)\,dF(s,y). \tag{16}$$

Assume $\theta(t,s)$ is smooth at (t^*,t^*) for some $t^*\in(0,1)$. By the first-order local Taylor expansion of $\theta(t,s)$, $\theta(t^*)$ can be estimated by $\hat{\theta}_{b_n}(t^*)$, where

$$\begin{aligned}(\hat{\theta}_{b_n}(t^*),\hat{\eta}_1,\hat{\eta}_2) &= \operatorname*{argmin}_{(\eta_0,\eta_1\eta_2)\in\mathbb{R}^3}\sum_{j,k=1}^{n}\big(H(X_j,X_k)-\eta_0-\eta_1(t_j-t^*)\\ &\qquad -\eta_2(t_k-t^*)\big)^2W_n(t_j,t_k).\end{aligned} \tag{17}$$

Here for presentational simplicity we assume that

$$W_n(t_j,t_k)=K\big((t_j-t^*)/b_n\big)K\big((t_k-t^*)/b_n\big)/(nb_n),$$

where $K\in\mathcal{K}$ and $\mathcal{K}$ is the collection of continuously differentiable and symmetric density functions with support $[-1,1]$. Furthermore, the bandwidth b_n satisfies

$b_n \to 0$ with $nb_n \to \infty$. Estimator (17) is an extension of the classic local linear kernel methods [Fan and Gijbels (1996)] to second-order parameter functions of the form (15). Meanwhile, if higher-order Taylor expansions of $\theta(t,s)$ are used in (17), then one obtains local polynomial estimations of $\theta(t^*)$.

It is easy to see that the asymptotic behavior of $\hat{\theta}_{b_n}(t^*)$ is decided by that of the V-statistics $V_n = \sum_{j,k=1}^n H(X_j, X_k) W_n(t_j, t_k)$. The following proposition, which is proved in Zhou (2013), investigates the limiting distribution of $\hat{\theta}_{b_n}(t^*)$:

PROPOSITION 3. *Assume that conditions* (A2) *and* (A5) *hold with* $4+2\epsilon$ *therein replaced by* 8. *Further assume that* (A1) *and* (A4) *hold and* $\theta(t,s)$ *is* $\mathcal{C}^2$ *in a neighborhood of* (t^*, t^*). *Then under the above assumptions of* $K(\cdot)$ *and* b_n, *we have*

$$\frac{\sqrt{nb_n}}{\sqrt{4\phi_n \int_{-1}^1 K^2(x)\,dx}}[\hat{\theta}_{b_n}(t^*) - \theta(t^*) - B_n(t^*)] \Rightarrow N(0,1), \tag{18}$$

where $B_n(t^*) = b_n^2 \frac{\partial^2 \theta(t^*,t^*)}{\partial t^2} \int_{-1}^1 x^2 K(x)\,dx$.

EXAMPLE 4 (Estimating the time-varying variance function). Consider the kernel $H(x,y) = (x-y)^2/2$. Then $\theta(t)$ in (15) equals $\operatorname{Var}[G_{\zeta(t)}(t,\mathcal{F}_0)] = \mathbb{E}[G_{\zeta(t)}(t,\mathcal{F}_0) - \mathbb{E}G_{\zeta(t)}(t,\mathcal{F}_0)]^2$. In particular, $\theta(t_i) = \operatorname{Var}[X_i]$.

For this variance kernel H, we can choose $L(x) = (1+x^2)^2$ and assume that $\mathbb{E}[X_i^{32}] < \infty$, $i = 1,2,\dots,n$. Then $L(x)$ satisfies condition (A2) with $4+2\epsilon$ therein replaced by 8. By Proposition 1, condition (A1) is satisfied with $\eta = 1$. Furthermore, condition (A3) is satisfied by Example 2 and the assumption that $K \in \mathcal{K}$. Note that $H_j(x) = \mathbb{E}[H(x,X_j)] = (x^2 - 2x\mathbb{E}[X_j] + \mathbb{E}[X_j^2])/2$ does not always equal 0. Hence the kernel is nondegenerate. Meanwhile,

$$2\theta(t,s) = \mathbb{E}[G^2_{\zeta(t)}(t,\mathcal{F}_0)] + \mathbb{E}[G^2_{\zeta(s)}(s,\mathcal{F}_0)] - 2\mathbb{E}[G_{\zeta(t)}(t,\mathcal{F}_0)]\mathbb{E}[G_{\zeta(s)}(s,\mathcal{F}_0)].$$

Assuming that $\mu(t) := \mathbb{E}[G_{\zeta(t)}(t,\mathcal{F}_0)]$ and $v(t) := \mathbb{E}[G^2_{\zeta(t)}(t,\mathcal{F}_0)]$ are $\mathcal{C}^2$ in a neighborhood of t^*, then $\theta(t,s)$ is $\mathcal{C}^2$ in a neighborhood of (t^*,t^*). Further assume that t^* is not a break point of the time series and condition (A10). By the local stationarity of $\{X_j\}$ in the neighborhood of t^*, we have that $4\phi_n \to \sigma^2(t^*)$, where

$$\sigma^2(t) = \sum_{j=-\infty}^{\infty} \operatorname{Cov}\{G^2_{\zeta(t)}(t,\mathcal{F}_0) - 2\mu(t)G_{\zeta(t)}(t,\mathcal{F}_0), G^2_{\zeta(t)}(t,\mathcal{F}_j) - 2\mu(t)G_{\zeta(t)}(t,\mathcal{F}_j)\}.$$

Note that $\sigma^2(t)$ is the spectral density of the stationary sequence $\{G^2_{\zeta(t)}(t,\mathcal{F}_j) - 2\mu(t)G_{\zeta(t)}(t,\mathcal{F}_j)\}_{j=-\infty}^{\infty}$ at frequency 0. Hence condition (A4) holds provided

$\sigma^2(t^*) > 0$. Finally, we have under the other regularity assumptions of Proposition 3 that

$$\frac{\sqrt{nb_n}}{\sqrt{\sigma^2(t^*)\int_{-1}^{1} K^2(x)\,dx}}[\hat{\theta}_{b_n}(t^*) - \theta(t^*) - \tilde{B}_n(t^*)] \Rightarrow N(0,1),$$

where $\tilde{B}_n(t^*) = \frac{b_n^2}{2}[v''(t^*) - 2\mu(t^*)\mu''(t^*)]\int_{-1}^{1} x^2 K(x)\,dx$.

5.2. *Spectral analysis.* Consider a PLS time series $\{X_j\}$ defined in (6). Assume further that $\mathbb{E}[X_j] = 0$ and $\|X_j\| < \infty$, $j = 1, 2, \dots, n$. Then we can define its spectral density at time t as

$$f(t,\lambda) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty} \gamma(t,k)\cos(k\lambda), \qquad \lambda \in [0, 2\pi],$$

where $\gamma(t,k) = \mathrm{Cov}[G_{\zeta(t)}(t,\mathcal{F}_0), G_{\zeta(t)}(t,\mathcal{F}_k)]$ is the kth-order auto covariance of $\{X_j\}$ at time t. Write the classic periodogram of the series $\{X_j\}$

$$I_n(\lambda) = \frac{1}{2\pi n}|S_n(\lambda)|^2, \qquad \text{where } S_n(\lambda) = \sum_{j=1}^{n} X_j \exp(ij\lambda), 0 \le \lambda \le \pi.$$

Consider also the classic smoothed periodogram estimate of the spectral density

$$\tilde{f}_n(\lambda) = \int_{-m}^{m} \frac{1}{m} K\left(\frac{u}{m}\right) I_n(\lambda + 2\pi u/n)\,du,$$

where $K(\cdot) \in \mathcal{K}$ is an even function, and m is a block size satisfying $m \to \infty$ with $n/m \to \infty$. The analysis of $\tilde{f}_n(\lambda)$ depends heavily on the theory of quadratic forms for nonstationary processes. For strictly stationary time series, the asymptotic behaviors of the periodogram and spectral density estimates have been intensively studied in the literature. See, for instance, Brillinger (1969), Priestley (1981), Rosenblatt (1984), Walker (2000) and Shao and Wu (2007) among others. On the other hand, however, there are few corresponding results for nonstationary time series. Exceptions include, among others, Dwivedi and Subba Rao (2011) who studied the asymptotic behavior of the periodogram for short memory locally stationary linear processes and Dette, Preuss and Vetter (2011) who studied the behavior of the averaged spectral density estimates for locally stationary Gaussian linear processes. In this section we shall investigate the behaviors of $I_n(\lambda)$ and $\tilde{f}_n(\lambda)$ for linear and nonlinear PLS time series. The following is a key theorem which establishes a Gaussian approximation result for Fourier transforms of nonstationary time series. Theorem 4 could be of separate interest in spectral analysis of nonstationary processes.

Let $\lambda = \lambda_n$ be a sequence of frequencies of interest. For $1 \le a < b \le n$, define $S^*_{a,b,\lambda} = \sum_{j=a}^{b} X_j \cos(j\lambda)$ and $S^o_{a,b,\lambda} = \sum_{j=a}^{b} X_j \sin(j\lambda)$. Write $S^*_{a,\lambda} := S^*_{1,a,\lambda}$ and $S^o_{a,\lambda} := S^o_{1,a,\lambda}$.

THEOREM 4. *Assume that* $\gamma(t,h)$ *is* $\mathcal{C}^p$ *in* t *on* $(b_k, b_{k+1}]$ *for any* h, $k = 0, 1, \dots, r$, $p \geq 1$. *Further assume that conditions* (A5) *and* (A10) *hold with* $4 + 2\epsilon$ *therein replaced by* 4 *and* $\inf_{t\in[0,1]} f(t,\lambda) > 0$. (i): *If* $0 < \lambda_* \leq \lambda \leq \lambda^* < \pi$ *for some constants* λ_* *and* λ^*, *then on a possibly richer probability space, there exist i.i.d. two-dimensional standard normal random vectors* $(G_{1,1}, G_{1,2})^\top, \dots, (G_{n,1}, G_{n,2})^\top$, *such that*

$$\max_{1\leq j\leq n}\left|(S^*_{j,\lambda}, S^o_{j,\lambda})^\top - \sum_{k=1}^{j}\sqrt{\pi}\big(f^{1/2}(t_k,\lambda)G_{k,1}, f^{1/2}(t_k,\lambda)G_{k,2}\big)^\top\right|$$
$$= o_{\mathbb{P}}\big(n^{p^*}\log^2 n\big),$$

where $p^* = (p+4)/[4(p+2)]$. (ii): *If* $\lambda = 0$ *or* π, *then on a possibly richer probability space, there exist i.i.d. standard normal random variables* $G^*_1, \dots, G^*_n$, *such that*

$$\max_{1\leq j\leq n}\left|S^*_{j,\lambda} - \sqrt{2\pi}\sum_{k=1}^{j} f^{1/2}(t_k,\lambda)G^*_k\right| = o_{\mathbb{P}}\big(n^{1/4}\log^2 n\big).$$

Based on Theorem 4, we have the following corollary on the behavior of the periodogram for nonstationary time series.

COROLLARY 5. *Under the conditions of Theorem* 4, *we have* (i): *if the frequency* λ *satisfies* $0 < \lambda_* \leq \lambda \leq \lambda^* < \pi$, *then*

$$I_n(\lambda)\Big/\int_0^1 f(t,\lambda)\,dt \Rightarrow \mathrm{Exp}(1),$$

where Exp(1) *stands for the exponential distribution with mean* 1. (ii) *If* $\lambda = 0$ *or* π, *then* $I_n(\lambda)/\int_0^1 f(t,\lambda)\,dt$ *converges in distribution to a* $\chi^2(1)$ *random variable.*

REMARK 1. The condition $0 < \lambda_* \leq \lambda \leq \lambda^* < \pi$ is important for the validity of Theorem 4 and Corollary 5. By (ii) of Lemma 5 in Zhou (2013), we have if $1 \leq k \leq C$ for some finite constant C, then

$$\left|\mathrm{Cov}\big(S^*_{n,\lambda_k}, S^o_{n,\lambda_k}\big) - \pi\sum_{j=1}^{n} f(t_j,\lambda_k)\sin(4\pi k t_j)\right| = O\big(\log^2 n\big), \tag{19}$$

where $\lambda_k = 2\pi k/n$. Therefore the real and imaginary parts of $S_n(\lambda_k)$ are no longer uncorrelated, and the periodogram does not converge to an Exp(1) distribution. Similar results hold for frequencies near π. This is drastically different from the stationary case where it is well known that the real and imaginary parts of $S_n(\lambda_k)$ are asymptotically independent and $I_n(\lambda_k)/f(\lambda_k) \Rightarrow \mathrm{Exp}(1)$. Indeed, note that if $f(t,\lambda_k)$ does not change with t, then we have $\sum_{j=1}^{n} f(t_j,\lambda_k)\sin(4\pi k t_j) = 0$ in (19). Due to the time-varying nature of $f(t,\lambda_k)$, the behavior of Fourier transforms near frequency 0 or π is complicated for nonstationary time series.

The following proposition investigates the asymptotic behavior of $\tilde{f}_n(\lambda)$ for PLS time series.

PROPOSITION 4. *Assume that $K \in \mathcal{K}$ is even. Then under the conditions of Theorem* 4 *and the assumption that $m \to \infty$ with $m/(n^{p/[4(p+2)]} \log^2 n) \to 0$, we have* (i): *if $0 < \lambda_* \le \lambda \le \lambda^* < \pi$, then*

$$\sqrt{m}\big(\tilde{f}_n(\lambda) - \mathbb{E}\tilde{f}_n(\lambda)\big) \Rightarrow N\left(0, \int_{-1}^{1} [K(t)]^2\,dt \int_0^1 f^2(t,\lambda)\,dt\right);$$

and (ii): *if $\lambda = 0$ or π, then*

$$\sqrt{m}\big(\tilde{f}_n(\lambda) - \mathbb{E}\tilde{f}_n(\lambda)\big) \Rightarrow N\left(0, 2\int_{-1}^{1} [K(t)]^2\,dt \int_0^1 f^2(t,\lambda)\,dt\right).$$

Proposition 4, which is proved in Zhou (2013), establishes the asymptotic normality of $\tilde{f}_n(\lambda)$ for a class of nonstationary nonlinear processes. Simple calculations show that $\mathbb{E}\tilde{f}_n(\lambda) = \int_0^1 f(t,\lambda)\,dt + o(1)$. Hence $\tilde{f}_n(\lambda)$ is a consistent estimator of the averaged energy at frequency λ over time. An important observation from Proposition 4 is that the asymptotic variance of $\tilde{f}_n(\lambda)$ is determined by $\int_0^1 f^2(t,\lambda)\,dt$, the averaged squared spectral density over time. The latter quantity should be estimated if one wishes to construct confidence intervals for $\int_0^1 f(t,\lambda)\,dt$.

6. Proofs.

PROOF OF THEOREM 1. Let $Z_k = \sum_{j=1}^n W_n(t_k,t_j)\{H_j(X_k) - \mathbb{E}[H_j(X_k)]\}$, $k = 1,2,\ldots,n$. Note that $N_n = \sum_{j=1}^n Z_j$. To prove the theorem, we need to deal with the dependence structure of $\{Z_j\}$ first. According to (2),

$$\begin{aligned} Z_k &= \int_{\mathbb{R}^2} g(t,s)\left\{\sum_{j=1}^n W_n(t_k,t_j)\mathbb{E}\big[\beta_j(s)\big]\right\}\gamma_k(t)\,dt\,ds \\ &:= \int_{\mathbb{R}^2} g(t,s)\,\Xi_k(s)\gamma_k(t)\,dt\,ds. \end{aligned}$$

Let $Z_{k,r} = \sum_{j=1}^n W_n(t_k,t_j)\{H_j(X_{k,r}) - \mathbb{E}[H_j(X_{k,r})]\}$. By Lemmas 1 and 2 in Zhou (2013), the dependence measures

$$\begin{aligned} &\|Z_k - Z_{k,r}\|_p \\ &\quad\le \int_{\mathbb{R}^2} |g(t,s)|\,|\Xi_k(s)|\,\|\gamma_k(t) - \gamma_{k,r}(t)\|_p\,dt\,ds \\ &\quad\le \int_{\mathbb{R}^2} |g(t,s)|\,|\Xi_k(s)|\big\{\|L(X_k)\|_{2p}|t|^\eta\big[\delta_X(r,\eta 2p)\big]^\eta + \delta_{L(X)}(r,p)\big\}\,dt\,ds \\ &\quad\le \int_{\mathbb{R}^2} C\big(1+|t|^\eta\big)|g(t,s)|\,|\Xi_k(s)|\rho_1^r \end{aligned}$$

for some $\rho_1 \in (0, 1)$, where $p = 2 + \epsilon$. On the other hand, note that

$$\Xi_k(s) \le \sum_{j=1}^{n} |W_n(t_k, t_j)| |\mathbb{E}[\beta_j(s)]| \le \sum_{j=1}^{n} |W_n(t_k, t_j)| \|L(X_j)\| \le C W_{k,\cdot}.$$

Hence $\|Z_k - Z_{k,r}\|_p \le C W_{k,\cdot} \rho_1^r$. As a second step, we shall approximate N_n by the sum of an m-dependent sequence. Define $Z_{k,\{m\}} = \mathbb{E}[Z_k | \tilde{\mathcal{F}}_{k,k-m}]$, where $\tilde{\mathcal{F}}_{k,k-m} = (\varepsilon_k, \varepsilon_{k-1}, \dots, \varepsilon_{k-m})$. For $j \in \mathbb{Z}$, define the projection operator

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

Elementary manipulations show that $\mathcal{P}_{k-r} Z_{k,\{m\}} = \mathbb{E}[\mathcal{P}_{k-r} Z_k | \tilde{\mathcal{F}}_{k,k-m}]$. Hence by Jensen's inequality,

$$\begin{aligned} \|\mathcal{P}_{k-r}[Z_k - Z_{k,\{m\}}]\|_p &\le \|\mathcal{P}_{k-r} Z_k\|_p + \|\mathcal{P}_{k-r} Z_{k,\{m\}}\|_p \le 2\|\mathcal{P}_{k-r} Z_k\|_p \\ &\le 2\|Z_k - Z_{k,r}\|_p \le C W_{k,\cdot} \rho_1^r. \end{aligned} \tag{20}$$

Note that $Z_{k,\{m\}} - Z_k = \sum_{j=m}^{\infty} \{\mathbb{E}[Z_k | \tilde{\mathcal{F}}_{k,k-j}] - \mathbb{E}[Z_k | \tilde{\mathcal{F}}_{k,k-j-1}]\}$ and the summands form a martingale difference sequence. By Burkholder's inequality,

$$\begin{aligned} \|Z_{k,\{m\}} - Z_k\|_p^2 &\le C \sum_{j=m}^{\infty} \|\mathbb{E}[Z_k | \tilde{\mathcal{F}}_{k,k-j}] - \mathbb{E}[Z_k | \tilde{\mathcal{F}}_{k,k-j-1}]\|_p^2 \\ &\le C \sum_{j=m}^{\infty} \|Z_k - Z_{k,j}\|_p^2 \le C W_{k,\cdot}^2 \rho_1^{2m}. \end{aligned}$$

Therefore

$$\|\mathcal{P}_{k-r}[Z_k - Z_{k,\{m\}}]\|_p \le \|Z_{k,\{m\}} - Z_k\|_p \le C W_{k,\cdot} \rho_1^m. \tag{21}$$

By (20), (21) and Burkholder's inequality, for any $r \ge 0$,

$$\left\| \sum_{k=1}^{n} \mathcal{P}_{k-r}[Z_k - Z_{k,\{m\}}] \right\|_p^2 \le \sum_{k=1}^{n} \left\| \mathcal{P}_{k-r}[Z_k - Z_{k,\{m\}}] \right\|_p^2 \le C W^{(n)} \rho_1^{2\max(m,r)}.$$

Observe that $\sum_{k=1}^{n} [Z_k - Z_{k,\{m\}}] = \sum_{r=0}^{\infty} \sum_{k=1}^{n} \mathcal{P}_{k-r}[Z_k - Z_{k,\{m\}}]$. Therefore

$$\begin{aligned} \left\| \sum_{k=1}^{n} [Z_k - Z_{k,\{m\}}] \right\|_p &\le \sum_{r=0}^{\infty} \left\| \sum_{k=1}^{n} \mathcal{P}_{k-r}[Z_k - Z_{k,\{m\}}] \right\|_p \\ &\le C \sqrt{W^{(n)}} m \rho_1^m. \end{aligned} \tag{22}$$

Inequality (22) shows that N_n can be well approximated by the sum of the m-dependent sequence $\{Z_{k,\{m\}} - \mathbb{E}[Z_{k,\{m\}}]\}$. In particular, let $m = C \log n$. Then clearly approximation error in (22) can be made as $O(n^{-p})\sqrt{W^{(n)}}$ for any $p > 0$. In the final step we shall prove a central limit theorem for $N_{n,\{m\}} := \sum_{k=1}^{n}\{Z_{k,\{m\}} - \mathbb{E}[Z_{k,\{m\}}]\}$. Define the big blocks and small blocks

$$R_j = \sum_{k=1}^{l_n}\{Z_{(j-1)m+k,\{m\}} - \mathbb{E}[Z_{(j-1)m+k,\{m\}}]\} \quad \text{and}$$

$$r_j = \sum_{k=l_n+1}^{s_n}\{Z_{(j-1)m+k,\{m\}} - \mathbb{E}[Z_{(j-1)m+k,\{m\}}]\},$$

$j = 1, 2, \ldots, \lceil n/s_n \rceil$. Note that R_j's are independent and r_j's are also independent. Then similar to the proof of (22), we can obtain $|\operatorname{Var}[N_n] - \operatorname{Var}[N_{n,\{m\}}]| = o(W^{(n)})$, $\|R_j\|_p^2 = O(A_j)$,

$$\Big\|\sum_j R_j\Big\|_p^2 = O\Big(\sum_j A_j\Big) \quad \text{and} \quad \Big\|\sum_j r_j\Big\|_p^2 = O\Big(\sum_j a_j\Big). \tag{23}$$

Therefore by condition (A3),

$$\Big\|N_{n,\{m\}} - \sum_j R_j\Big\|_p = \Big\|\sum_j r_j\Big\|_p = o\big(\sqrt{W^{(n)}}\big). \tag{24}$$

By (22), $\|N_{n,\{m\}} - N_n\|_p = o(\sqrt{W^{(n)}})$. By condition (A4), (23) and (24),

$$\operatorname{Var}\Big[\sum_j R_j\Big]\Big/W^{(n)} \ge c/2 \qquad \text{for sufficiently large } n.$$

Now by (23),

$$\frac{\sum_j \|R_j\|_p^p}{(\operatorname{Var}[\sum_j R_j])^{p/2}} \le C\frac{\sum_j A_j^{p/2}}{[W^{(n)}]^{p/2}} \le \left\{\frac{\max A_j}{[W^{(n)}]}\right\}^{p/2-1}\frac{\sum_j A_j}{W^{(n)}}.$$

Hence by condition (A3), $\sum_j \|R_j\|_p^p/\operatorname{Var}[\sum_j R_j])^{p/2} \to 0$. Therefore by the Lyapunov CLT, $\sum_j R_j/\sqrt{\operatorname{Var}[\sum_j R_j]} \Rightarrow N(0,1)$. Now by (22)–(24), the theorem follows. □

PROOF OF THEOREM 2. For any fixed $(x, y) \in \mathbb{R}^2$, define

$$\varrho_n(x, y) = \sum_{k,j} W_n(t_k, t_j)\gamma_k(x)\gamma_j(y).$$

We shall first determine the order of magnitude of $\varrho_n(x, y)$. For complex-valued random variables X and Y, define $V(X) = \mathbb{E}|X - \mathbb{E}X|^2$ and $\operatorname{Cov}[X, Y] = \mathbb{E}[(X -$

$\mathbb{E}X)(\bar{Y}-\mathbb{E}\bar{Y})]$. Note that, by (A6) and the symmetry of $W(\cdot,\cdot)$,

$$\begin{aligned} V(\varrho_n(x,y)) &= \sum_{k,j}\sum_{l,m} W_n(t_l,t_m)W_n(t_k,t_j)\operatorname{Cov}[\gamma_l(x)\gamma_m(y),\gamma_k(x)\gamma_j(y)] \\ &\le \sum_{k,j}\sum_{l,m} W_n^2(t_k,t_j)\big|\operatorname{Cov}[\gamma_l(x)\gamma_m(y),\gamma_k(x)\gamma_j(y)]\big| \\ &\quad + 2^{p/2}\sum_{k,j}\sum_{l,m} f_n(t_k,t_j)\big|W_n(t_k,t_j)\big|\rho_{k,j}^p(l,m) \\ &\qquad\qquad \times \big|\operatorname{Cov}[\gamma_l(x)\gamma_m(y),\gamma_k(x)\gamma_j(y)]\big| \\ &:= \sum_{k,j} W_n^2(t_k,t_j)\varrho_{k,j}(x,y;0) \\ &\quad + 2^{p/2}\sum_{k,j} f_n(t_k,t_j)\big|W_n(t_k,t_j)\big|\varrho_{k,j}(x,y;p), \end{aligned}$$

where $\rho_{k,j}(l,m)=\min\{\max(|l-k|,|m-j|),\max(|l-j|,|m-k|)\}$ and

$$\varrho_{k,j}(x,y;p)=\sum_{l,m}\rho_{k,j}^p(l,m)\big|\operatorname{Cov}[\gamma_l(x)\gamma_m(y),\gamma_k(x)\gamma_j(y)]\big|.$$

We will omit the subscripts k, j in $\rho_{k,j}(l,m)$ in the sequel for simplicity. Let

$$\begin{aligned} \rho^*(l,m) = \max\{&\min(|k-l|,|k-m|,|j-l|,|j-m|), \\ &\min(|l-k|,|l-j|,|l-m|),\min(|m-k|,|m-j|,|m-l|), \\ &\min(|k-l|,|k-m|,|k-j|),\min(|j-l|,|j-m|,|j-k|)\}. \end{aligned}$$

We shall first show that

$$\rho(l,m)\le 2\rho^*(l,m) \qquad \text{for all } l,m. \tag{25}$$

By the symmetry of $\rho(l,m)$ and $\rho^*(l,m)$, we only need to consider the case $l\le m$ and $k\le j$. Now if $l\le k$, then

$$\begin{aligned} 2\min(|l-k|,|l-j|,|l-m|) &\ge |k-l| &&\text{if } m\ge (l+k)/2 \quad\text{and} \\ 2\min(|k-l|,|k-m|,|j-l|,|j-m|) &\ge |l-k| &&\text{if } m\le (l+k)/2. \end{aligned}$$

If $l\ge k$, then

$$\begin{aligned} 2\min(|k-l|,|k-m|,|k-j|) &\ge |k-l| &&\text{if } j\ge (l+k)/2 \quad\text{and} \\ 2\min(|k-l|,|k-m|,|j-l|,|j-m|) &\ge |l-k| &&\text{if } j\le (l+k)/2. \end{aligned}$$

In summary, $2\rho^*(l,m)\ge|l-k|$. Similarly, $2\rho^*(l,m)\ge|j-m|$. Hence (25) follows. Now by Lemmas 1, 2 and 7 in Zhou (2013), elementary calculations using condition (A5) show that

$$\begin{aligned} \big|\operatorname{Cov}[\gamma_l(x)\gamma_m(y),\gamma_k(x)\gamma_j(y)]\big| &\le C(|x|^\eta+1)(|y|^\eta+1)r_1^{\rho^*(l,m)} \\ &\le C(|(x,y)|^{2\eta}+1)r_1^{\rho^*(l,m)} \end{aligned} \tag{26}$$

for some $r_1 \in [0,1)$. Observe that for $r = 0, 1, \dots, n$, the number of pairs (l, m) such that $\rho(l,m) = r$ is at most $2r^2$. Now by (25) and (26), we obtain that, for $r_2 = \sqrt{r_1}$,

$$
\begin{aligned}
\varrho_{k,j}(x, y; p) &\le C \sum_{l,m} \rho^p(l,m)\big(\big|(x,y)\big|^{2\eta} + 1\big) r_2^{\rho(l,m)} \\
&\le C\big(\big|(x,y)\big|^{2\eta} + 1\big).
\end{aligned}
\tag{27}
$$

Note that the constant C does not depend on (k, j). Now by (27), we obtain

$$
\|\varrho_n(x,y) - \mathbb{E}\varrho_n(x,y)\| \le C\big(\big|(x,y)\big|^{\eta} + 1\big)\sqrt{W_{(n)} + \Delta_n}.
$$

Therefore by condition (A1),

$$
\begin{aligned}
\|D_n - \mathbb{E}D_n\| &\le \int_{\mathbb{R}^2} \big|g(x,y)\big| \|\varrho_n(x,y) - \mathbb{E}\varrho_n(x,y)\| \, dx\, dy \\
&\le C\sqrt{W_{(n)} + \Delta_n} \int_{\mathbb{R}^2} \big(\big|(x,y)\big|^{\eta} + 1\big)\big|g(x,y)\big| \, dx\, dy \\
&\le C\sqrt{W_{(n)} + \Delta_n}.
\end{aligned}
$$

The theorem follows. □

PROOF OF THEOREM 3. Let $\varpi_n(x,y) = \sum_{k,j=1}^{n} W_n(t_k, t_j)\Gamma_n^*(t_k, x)\Gamma_n^*(t_j, y)$. Recall the definition of $\Gamma_n^*(t,s)$ in (11). We shall show that the two processes $\varrho_n(x,y)$ and $\varpi_n(x,y)$ are close in the sense that

$$
\begin{aligned}
&\bigg|\mathbb{E}h\bigg[\int_{\mathbb{R}^2} g(x,y)\big[\varrho_n(x,y) - \mathbb{E}\varrho_n(x,y)\big] \, dx\, dy\bigg] \\
&\quad - \mathbb{E}h\bigg[\int_{\mathbb{R}^2} g(x,y)\big[\varpi_n(x,y) - \mathbb{E}\varpi_n(x,y)\big] \, dx\, dy\bigg]\bigg| \to 0
\end{aligned}
\tag{28}
$$

for any bounded and continuous h. For any $s > 0$, define the region $A(s) = \{(x,y) \in \mathbb{R}^2, |x| \le s, |y| \le s\}$, and let $\bar{A}(s) = \mathbb{R}^2 / A(s)$. Note that, by Lemma 4 in Zhou (2013), we have

$$
\begin{aligned}
&\bigg\|\int_{\bar{A}(s)} g(x,y)\big[\varrho_n(x,y) - \mathbb{E}\varrho_n(x,y)\big] \, dx\, dy\bigg\| \\
&\quad \le \int_{\bar{A}(s)} \big|g(x,y)\big| \|\varrho_n(x,y) - \mathbb{E}\varrho_n(x,y)\| \, dx\, dy \\
&\quad \le C \int_{\bar{A}(s)} \big|g(x,y)\big|\big(1 + \big|(x,y)\big|^{\eta}\big) \, dx\, dy.
\end{aligned}
$$

Observe that $\int_{\bar{A}(s)} |g(x,y)|(1 + |(x,y)|^{\eta}) \, dx\, dy$ is independent of n and converges to 0 as $s \to \infty$. Similar inequality holds for $\int_{\bar{A}(s)} g(x,y)[\varpi_n(x,y) -$

$\mathbb{E}\varpi_n(x, y)]\,dx\,dy$. Hence, to prove (28), one only need to show that, for each fixed s, (28) holds with $\mathbb{R}^2$ therein replaced by $A(s)$. To this end, we shall first show that, for any $(x_1, y_1), \dots, (x_m, y_m) \in A(s)$ and any bounded and continuous $h : \mathbb{R}^m \to \mathbb{R}$,

$$\begin{aligned} &\mathbb{E}h\{(\varrho_n(x_1, y_1) - \mathbb{E}\varrho_n(x_1, y_1), \dots, \varrho_n(x_m, y_m) - \mathbb{E}\varrho_n(x_m, y_m))\} \\ &\quad - \mathbb{E}h\{(\varpi_n(x_1, y_1) - \mathbb{E}\varpi_n(x_1, y_1), \dots, \varpi_n(x_m, y_m) - \mathbb{E}\varpi_n(x_m, y_m))\} \\ &\quad = o(1). \end{aligned} \tag{29}$$

We shall only prove the case $m = 1$ since similar arguments apply to general m. Consider the case $x, y \neq 0$. By Corollary 2 of Wu and Zhou (2011), we have, on a possibly richer probability space, a sequence a i.i.d. 4-dimensional standard normal random vectors $\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_n$, such that

$$\begin{aligned} &\max_{1 \le k \le n} \left| \sum_{j=1}^{k} \{(\gamma_{j,1}(x), \gamma_{j,2}(x), \gamma_{j,1}(y), \gamma_{j,2}(y))^\top - \Sigma_{j^*}^{1/2}(t_j, (x, y)^\top)\mathbf{Z}_j\} \right| \\ &\quad = o_{\mathbb{P}}(n^{1/4} \log^2 n). \end{aligned} \tag{30}$$

Define the complex-valued random variables $Z_j^* = [\Sigma_{j^*}^{1/2}(t_j, (x, y)^\top)\mathbf{Z}_j]_1 + i[\Sigma_{j^*}^{1/2}(t_j, (x, y)^\top)\mathbf{Z}_j]_2$ and $Z_j^{**} = [\Sigma_{j^*}^{1/2}(t_j, (x, y)^\top)\mathbf{Z}_j]_3 + i[\Sigma_{j^*}^{1/2}(t_j, (x, y)^\top) \times \mathbf{Z}_j]_4$, where $[\Sigma_{j^*}^{1/2}(t_j, (x, y)^\top)\mathbf{Z}_j]_r$ denotes the rth element of $\Sigma_{j^*}^{1/2}(t_j, (x, y)^\top) \times \mathbf{Z}_j$. Define the quadratic form

$$\varrho_n^{\diamondsuit}(x, y) = \sum_{k,j=1}^{n} W_n(t_k, t_j) Z_k^* Z_j^{**}. \tag{31}$$

Note that

$$\begin{aligned} |\varrho_n(x, y) - \varrho_n^{\diamondsuit}(x, y)| &\le \left| \sum_{k,j} W_n(t_k, t_j)[\gamma_k(x) - Z_k^*]\gamma_j(y) \right| \\ &\quad + \left| \sum_{k,j} W_n(t_k, t_j)[\gamma_j(y) - Z_j^{**}]Z_k^* \right|. \end{aligned}$$

Using the summation by parts technique and similar to the proof of inequality (12) in Zhou (2013), we have $|\varrho_n(x, y) - \varrho_n^{\diamondsuit}(x, y)| = o_{\mathbb{P}}(1)$. Now by the above inequality and Lemma 3 in Zhou (2013), we obtain

$$|\varrho_n(x, y) - \mathbb{E}\varrho_n(x, y) - [\varrho_n^{\diamondsuit}(x, y) - \mathbb{E}\varrho_n^{\diamondsuit}(x, y)]| = o_{\mathbb{P}}(1). \tag{32}$$

By condition (A10), we have that $\Sigma_k(t, (x, y)^\top)$ is continuous on $[b_k, b_{k+1}]$, $k = 0, 1, \dots, r$. Therefore elementary calculations show that, on a possibly richer probability space,

$$|\varrho_n^{\diamondsuit}(x, y) - \mathbb{E}\varrho_n^{\diamondsuit}(x, y) - [\varpi_n(x, y) - \mathbb{E}\varpi_n(x, y)]| = o_{\mathbb{P}}(1). \tag{33}$$

Hence (29) with $m=1$ follows from (32) and (33). Now consider the case $x=0$. If $L(\cdot)\equiv C$, then $\varrho_n(0,y)\equiv 0$ and $\varpi_n(0,y)\equiv 0$. Hence (29) trivially holds. If $L(\cdot)$ is not a constant, then (29) follows from similar and simpler arguments as above by considering the covariance matrix $\Sigma_k^*(y)$. In summary, (29) follows.

As a second step toward (28), we prove that $\{\varrho_n(x,y)-\mathbb{E}\varrho_n(x,y)\}$ is tight on $\mathcal{C}(A(s))$, where $\mathcal{C}(A(s))$ is the collection of all complex-valued continuous functions on $A(s)$ equipped with the uniform topology. Note that

$$\begin{aligned}
&\varrho_n(x_1,y_1)-\mathbb{E}\varrho_n(x_1,y_1)-\big[\varrho_n(x_2,y_2)-\mathbb{E}\varrho_n(x_2,y_2)\big]\\
&\quad= i\int_{y_2}^{y_1}\rho_n^{(1)}(x_2,y)\,dy+i\int_{x_2}^{x_1}\rho_n^{(2)}(x,y_2)\,dx\\
&\qquad-\int_{x_2}^{x_1}\int_{y_2}^{y_1}\rho_n^{(3)}(x,y)\,dx\,dy,
\end{aligned}\tag{34}$$

where $\rho_n^{(1)}(x,y)=\sum_{k,j}W_n(t_k,t_j)[\gamma_k(x)\gamma_j^{\diamond}(y)-\mathbb{E}\gamma_k(x)\gamma_j^{\diamond}(y)]$, $\rho_n^{(2)}(x,y)=\sum_{k,j}W_n(t_k,t_j)\times[\gamma_k^{\diamond}(x)\gamma_j(y)-\mathbb{E}\gamma_k^{\diamond}(x)\gamma_j(y)]$ and

$$\rho_n^{(3)}(x,y)=\sum_{k,j}W_n(t_k,t_j)\big[\gamma_k^{\diamond}(x)\gamma_j^{\diamond}(y)-\mathbb{E}\gamma_k^{\diamond}(x)\gamma_j^{\diamond}(y)\big]$$

with $\gamma_j^{\diamond}(x)=L^*(X_j)e^{ixX_j}-\mathbb{E}[L^*(X_j)e^{ixX_j}]$. By the proof of Lemma 4 in Zhou (2013), we have

$$\sup_{k=1,2,3,(x,y)\in A(s)}\big\|\rho_n^{(k)}(x,y)\big\|_{2+\epsilon}=O(1).\tag{35}$$

By (34) and (35), we have, for any fixed $(x_0,y_0)\in A(s)$ and $\delta>0$,

$$\begin{aligned}
&\Big\|\sup_{|(x,y)-(x_0,y_0)|\le\delta}\big|\varrho_n(x,y)-\mathbb{E}\varrho_n(x,y)-\big[\varrho_n(x_0,y_0)-\mathbb{E}\varrho_n(x_0,y_0)\big]\big|\Big\|_{2+\epsilon}\\
&\quad=O(\delta).
\end{aligned}\tag{36}$$

Define $\omega(\delta)=\sup_{|(x_1,y_1),(x_2,y_2)\in A(s),|(x_1,y_1)-(x_2,y_2)|\le\delta|}|\varrho_n(x_1,y_1)-\mathbb{E}\varrho_n(x_1,y_1)-[\varrho_n(x_2,y_2)-\mathbb{E}\varrho_n(x_2,y_2)]|$. By (36) and a standard chaining technique, we have for each fixed $\varepsilon>0$

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\mathbb{P}\big(\omega(\delta)>\varepsilon\big)=0.$$

Hence $\{\varrho_n(x,y)-\mathbb{E}\varrho_n(x,y)\}$ is tight on $A(s)$. By standard smooth Gaussian process techniques, it is easy to see that $\varpi_n(x,y)-\mathbb{E}\varpi_n(x,y)$ is tight on $A(s)$. Since both processes are relatively compact on $A(s)$ and the differences of their finite dimensional distributions converge in the sense of (29), we have for any bounded and continuous function h^*,

$$\mathbb{E}h^*\big(\varrho_n(x,y)-\mathbb{E}\varrho_n(x,y)\big)-\mathbb{E}h^*\big(\varpi_n(x,y)-\mathbb{E}\varpi_n(x,y)\big)\to 0.$$

Since $g(x, y) \in L^1(\mathbb{R}^2)$, we have $K(f) := \int_{A(s)} g(x, y) f(x, y)\,dx\,dy$ is continuous on $\mathcal{C}(A(s))$. Hence (28) follows.

Finally, note that $\Gamma_n^*(t_k, s)$'s are independent complex-valued Gaussian processes, $k = 0, 1, 2, \ldots, n$. By the classic Gaussian process theory [see, e.g., Kuo (1975), Chapter 1.2], we have $\Gamma_n^*(t_k, s)$ can be represented as (in the sense of equality in distribution)

$$\big(\Re(\Gamma_n^*(t_k, s)), \Im(\Gamma_n^*(t_k, s))\big)^\top = \sum_{j=1}^{\infty} A_{n,j}(t_k, s) B_{k,j}, \qquad k = 0, 1, 2, \ldots, n,$$

where $\Re(\cdot)$ and $\Im(\cdot)$ denotes real and imaginary parts of a complex number, respectively, $A_{n,j}(t, s)$'s are 2×2 matrix functions and $B_{k,j}$'s are independent 2-dimensional standard normal random vectors. Hence it is straightforward to see that $\int_{\mathbb{R}^2} g(x, y)[\varpi_n(x, y) - \mathbb{E}\varpi_n(x, y)]\,dx\,dy$ is a quadratic form of i.i.d. standard normal random variables $G_1, G_2, \ldots$. Moreover, by the arguments of Lemma 4 in Zhou (2013), we have $\| \int_{\mathbb{R}^2} g(x, y)[\varpi_n(x, y) - \mathbb{E}\varpi_n(x, y)]\,dx\,dy\| = O(1)$. Hence

$$\int_{\mathbb{R}^2} g(x, y)\big[\varpi_n(x, y) - \mathbb{E}\varpi_n(x, y)\big]\,dx\,dy = \sum_{j=1}^{\infty} \alpha_{n,j}\big(Z_j^2 - 1\big)$$

with $\sum_{j=1}^{\infty} \alpha_{n,j}^2 < \infty$. $\square$

REMARK 2. As we can see from the proof of (30), the positive-definiteness requirement on $\Sigma_k(t, s)$ and $\Sigma_k^*(t, s)$ in (A9) is to facilitate a Gaussian approximation result in Wu and Zhou (2011). We point out that the positive-definiteness requirement can be weakened to the assumption that certain block sums of the latter long-run covariance matrices are positive definite. See Remark 2 of Wu and Zhou (2011). For presentational simplicity, we shall stick to the everywhere positive definiteness assumption in this paper.

PROOF OF COROLLARY 2. We shall prove this corollary by showing that $\varrho_n(x, y) - \mathbb{E}\varrho_n(x, y)$ converges to a Gaussian measure on $\mathcal{C}(A(s))$. By the tightness of $\{\varrho_n(x, y) - \mathbb{E}\varrho_n(x, y)\}$ and the arguments in the proof of Theorem 3, it suffices to show that any finite dimensional distribution of the latter sequence of measures converges to a (multivariate) normal distribution. To this end, we will only show that $\rho_n^*(x, y) - \mathbb{E}\rho_n^*(x, y)$ converges to a Gaussian distribution for any $(x, y) \in A(s)$ since all other cases follow by similar arguments and the Cramer–Wold device. Here $\rho_n^*(x, y) = \sum_{k,j=1}^{n} W_n(t_k, t_j)\gamma_{k,1}(x)\gamma_{j,1}(y)$, $\gamma_{k,1}(x) = \Re(\gamma_k(x))$ and $\gamma_{k,2}(x) = \Im(\gamma_k(x))$. Consider the case $x, y \neq 0$. Then by the proof of (32), we have

$$\big|\rho_n^*(x, y) - \mathbb{E}\rho_n^*(x, y) - \big[\rho_{n,1}^{\diamond}(x, y) - \mathbb{E}\rho_{n,1}^{\diamond}(x, y)\big]\big| = o_{\mathbb{P}}(1),$$

where $\rho^{\diamond}_{n,1}(x, y) = \sum_{k,j=1}^{n} W_n(t_k, t_j)\Re(Z_k^*)\Re(Z_j^{**})$. Recall the definitions of Z_k^* and Z_j^{**} in (31). Note that $\rho^{\diamond}_{n,1}(x, y)$ is a quadratic form of i.i.d. normal random variables. More specifically, $\rho^{\diamond}_{n,1}(x, y)$ can be written as

$$2\rho^{\diamond}_{n,1}(x, y) = \mathbf{Z}^\top D^\top (W_n \otimes A) D\mathbf{Z}, \tag{37}$$

where $\mathbf{Z} = (\mathbf{Z}_1^\top, \dots, \mathbf{Z}_n^\top)^\top$ is a length $4n$ vector of i.i.d. standard normal random variables, $D = \mathrm{Diag}(\Sigma_{1*}^{1/2}(t_1, (x, y)^\top), \dots, \Sigma_{n*}^{1/2}(t_n, (x, y)^\top))$ is a $4n \times 4n$ block diagonal matrix, A is the 4×4 matrix with $(1, 3)$th and $(3, 1)$th elements equaling 1 and all other entries equaling 0 and $\otimes$ denotes the Kronecker product. Let $M_n = D^\top W_n \otimes AD$. By condition (A9), D is positive definite with eigenvalues bounded both above and below. Then it is easy to see that there exist constants $0 < c \le C < \infty$, such that $c \le \sum_{k,j=1}^{4n} M_n^2(k, j) \le C$. By the Lyapunov CLT, to prove the asymptotic normality of $\rho^{\diamond}_{n,1}(x, y) - \mathbb{E}\rho^{\diamond}_{n,1}(x, y)$, it suffices to show that the $|\theta^*_{n,1}| \to 0$, where $\theta^*_{n,1}$ is the eigenvalue of $W_n \otimes A$ with the maximum absolute value. By the basic property of Kronecker product, we have that the eigenvalues of $W_n \otimes A$ are the products of the eigenvalues of W_n and A. Hence it is clear that $|\theta^*_{n,1}| = |\theta_{n,1}| \to 0$. The case when $x = 0$ or $y = 0$ follows similarly. □

SKETCH OF PROOF OF THEOREM 4. Theorem 4 follows from Lemma 5 in the online supplement of the paper with $|b - a| = \lfloor c^{1/2} n^{1/2} \log^{-1} n \rfloor$ for some finite constant c together with a careful check of the proof of Theorem 1 in Wu and Zhou (2011) with $l = \lfloor c \log n \rfloor$ and $m = \lfloor l^{1/2} n^{1/2} \log^{-3/2} n \rfloor$ therein. □

Acknowledgments. The author is grateful to the two anonymous referees for their careful reading of the manuscript and many helpful comments.

SUPPLEMENTARY MATERIAL

Supplement for "Inference of weighted V-statistics for nonstationary time series and its applications" (DOI: 10.1214/13-AOS1184SUPP; .pdf). This supplementary material contains auxiliary lemmas and proofs of Propositions 1, 3, 4 and Corollaries 3, 4 of the paper.

REFERENCES

BEUTNER, E. and ZÄHLE, H. (2012). Deriving the asymptotic distribution of U- and V-statistics of dependent data using weighted empirical processes. *Bernoulli* **18** 803–822. MR2948902

BEUTNER, E. and ZÄHLE, H. (2013). Continuous mapping approach to the asymptotics of U- and V-statistics. *Bernoulli*. To appear.

BHANSALI, R. J., GIRAITIS, L. and KOKOSZKA, P. S. (2007). Approximations and limit theory for quadratic forms of linear processes. *Stochastic Process*. *Appl*. **117** 71–95. MR2287104

BRILLINGER, D. R. (1969). Asymptotic properties of spectral estimates of second order. *Biometrika* **56** 375–390. MR0246478

DAVIS, R. A., LEE, T. C. M. and RODRIGUEZ-YAM, G. A. (2006). Structural break estimation for nonstationary time series models. *J. Amer. Statist. Assoc.* **101** 223–239. MR2268041

DE JONG, P. (1987). A central limit theorem for generalized quadratic forms. *Probab. Theory Related Fields* **75** 261–277. MR0885466

DE WET, T. and VENTER, J. H. (1973). Asymptotic distributions for quadratic forms with applications to tests of fit. *Ann. Statist.* **1** 380–387. MR0353543

DEHLING, H. (2006). Limit theorems for dependent U-statistics. In *Dependence in Probability and Statistics. Lecture Notes in Statist.* **187** 65–86. Springer, New York. MR2283250

DEHLING, H. and TAQQU, M. S. (1989). The empirical process of some long-range dependent sequences with an application to U-statistics. *Ann. Statist.* **17** 1767–1783. MR1026312

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

DENKER, M. (1985). *Asymptotic Distribution Theory in Nonparametric Statistics*. Friedr. Vieweg & Sohn, Braunschweig. MR0889896

DETTE, H., PREUSS, P. and VETTER, M. (2011). A measure of stationarity in locally stationary processes with applications to testing. *J. Amer. Statist. Assoc.* **106** 1113–1124. MR2894768

DWIVEDI, Y. and SUBBA RAO, S. (2011). A test for second order stationarity based on the discrete Fourier transform. *J. Time Series Anal.* **32** 68–91.

FAN, J. and GIJBELS, I. (1996). *Local Polynomial Modelling and Its Applications. Monographs on Statistics and Applied Probability* **66**. Chapman & Hall, London. MR1383587

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

FOX, R. and TAQQU, M. S. (1987). Central limit theorems for quadratic forms in random variables having long-range dependence. *Probab. Theory Related Fields* **74** 213–240. MR0871252

GAO, J. and ANH, V. (2000). A central limit theorem for a random quadratic form of strictly stationary processes. *Statist. Probab. Lett.* **49** 69–79. MR1789666

GÖTZE, F. and TIKHOMIROV, A. N. (1999). Asymptotic distribution of quadratic forms. *Ann. Probab.* **27** 1072–1098. MR1699003

HOEFFDING, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statistics* **19** 293–325. MR0026294

HSING, T. and WU, W. B. (2004). On weighted U-statistics for stationary processes. *Ann. Probab.* **32** 1600–1631. MR2060311

HUŠKOVÁ, M. and JANSSEN, P. (1993). Consistency of the generalized bootstrap for degenerate U-statistics. *Ann. Statist.* **21** 1811–1823. MR1245770

KUO, H. H. (1975). *Gaussian Measures in Banach Spaces. Lecture Notes in Mathematics* **463**. Springer, Berlin. MR0461643

LEE, A. J. (1990). *U-statistics: Theory and Practice. Statistics: Textbooks and Monographs* **110**. Dekker, New York. MR1075417

LEE, J. and SUBBA RAO, S. (2011). A note on general quadratic forms of nonstationary time series. Preprint.

LEUCHT, A. (2012). Degenerate U- and V-statistics under weak dependence: Asymptotic theory and bootstrap consistency. *Bernoulli* **18** 552–585. MR2922461

LIFLYAND, E., SAMKO, S. and TRIGUB, R. (2012). The Wiener algebra of absolutely convergent Fourier integrals: An overview. *Anal. Math. Phys.* **2** 1–68. MR2891740

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

O'NEIL, K. A. and REDNER, R. A. (1993). Asymptotic distributions of weighted U-statistics of degree 2. *Ann. Probab.* **21** 1159–1169. MR1217584

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

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

ROSENBLATT, M. (1984). Asymptotic normality, strong mixing and spectral density estimates. *Ann. Probab.* **12** 1167–1180. MR0757774

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

VON MISES, R. (1947). On the asymptotic distribution of differentiable statistical functions. *Ann. Math. Statistics* **18** 309–348. MR0022330

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 Series Anal.* **21** 95–109. MR1766176

WU, W. B. and ZHOU, Z. (2011). Gaussian approximations for non-stationary multiple time series. *Statist. Sinica* **21** 1397–1413. MR2827528

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

ZHOU, Z. (2010). Nonparametric inference of quantile curves for nonstationary time series. *Ann. Statist.* **38** 2187–2217. MR2676887

ZHOU, Z. (2013). Heteroscedasticity and autocorrelation robust structural change detection. *J. Amer. Statist. Assoc.* **108** 726–740.

ZHOU, Z. (2014). Supplement to “Inference of weighted V-statistics for non-stationary time series and its applications.” DOI:10.1214/13-AOS1184SUPP.

ZHOU, Z. and WU, W. B. (2009). Local linear quantile estimation for nonstationary time series. *Ann. Statist.* **37** 2696–2729. MR2541444

DEPARTMENT OF STATISTICAL SCIENCES
UNIVERSITY OF TORONTO
100 ST. GEORGE ST.
TORONTO, ONTARIO M5S3G3
CANADA
E-MAIL: zhou@utstat.toronto.edu