A Single-Pass Algorithm for Spectrum Estimation With Fast Convergence

Han Xiao and Wei Biao Wu

***Abstract*—We propose a single-pass algorithm for estimating spectral densities of stationary processes. Our algorithm is computationally fast in the sense that, when a new observation arrives, it can provide a real-time update within $O(1)$ computation. The proposed algorithm is probabilistically fast in that, for stationary processes whose auto-covariances decay geometrically, the estimates from the algorithm converge at a rate which is optimal up to a multiplicative logarithmic factor. We also establish asymptotic normality for the recursive estimate. A simulation study is carried out and it confirms the superiority over the classical batched mean estimates.**

***Index Terms*—Batched mean estimate, bias reduction, nonparametric estimation, physical dependence measure, recursive algorithm, spectral density, stochastic process.**

I. Introduction

LET $(X_i)_{i\in\mathbb{Z}}$ be a stationary process with $\mathbb{E}X_i^2 < \infty$. Let the mean $\mu = \mathbb{E}X_i$ and the covariance function $\gamma_k = \operatorname{cov}(X_0, X_k) = \mathbb{E}[(X_0-\mu)(X_k-\mu)],\ k \in \mathbb{Z}$. If

$$\sum_{k\in\mathbb{Z}} |\gamma_k| < \infty \tag{1}$$

then the spectral density

$$f(\theta) = \frac{1}{2\pi}\sum_{k\in\mathbb{Z}} \gamma_k e^{\sqrt{-1}k\theta} = \frac{1}{2\pi}\sum_{k\in\mathbb{Z}} \gamma_k \varrho^k, \quad 0 \le \theta < 2\pi \tag{2}$$

exists and is continuous, where $\sqrt{-1}$ is the imaginary unit. Throughout the paper, we write ϱ for the rotation $e^{\sqrt{-1}\theta}$. The spectral density function captures the frequency content of the underlying process. In the study of stationary processes, a fundamental problem is to estimate $f(\cdot)$ based on observations $X_i, 1 \le i \le n$. The problem of spectral density estimation has a long history (see [1]–[10] and [11] among others) and it appears in almost all scientific fields including astronomy, geoscience, economics, physics, and engineering.

This paper considers nonparametric estimation of spectral density functions. Traditionally, one can use the lag window estimate

$$\hat{f}_L(\theta) = \frac{1}{2\pi}\sum_{k=1-n}^{n-1} w\left(k/B_n\right)\hat{\gamma}_k \varrho^k \tag{3}$$

where $w(\cdot)$ is the lag window satisfying $w(0) = 1, w(u) = 0$ if $|u| > 1$ and $w(\cdot) \ge 0$, B_n is the lag size and

$$\hat{\gamma}_k = \frac{1}{n}\sum_{j=1+|k|}^{n} \left(X_j - \bar{X}_n\right)\left(X_{j-|k|} - \bar{X}_n\right)$$

for $1-n \le k \le n-1$, are the estimated covariances, where $\bar{X}_n$ is the sample mean; or [12]'s overlapping batched mean (OBM) estimate

$$\hat{f}_B(\theta) = \frac{b_n}{2\pi(n-b_n+1)} \times \sum_{k=1}^{n-b_n+1} \left| \frac{1}{b_n} \sum_{j=k}^{k+b_n-1} \left(X_j - \bar{X}_n\right)\varrho^j \right|^2 \tag{4}$$

where $b_n \to \infty$ and $b_n/n \to 0$; or the smoothed periodogram estimate

$$\hat{f}_S(\theta) = \int_0^{2\pi} I_n(u) K\left((\theta-u)/b_n\right) du \tag{5}$$

where $K(\cdot)$ is a kernel function and $I_n(\cdot)$ is the periodogram. Statistical properties of the above and other nonparametric spectral density estimates have been discussed in [7], [8], [11], and [13], among others.

All of the above estimates are nonrecursive in the sense that they cannot be updated within $O(1)$ computation once a new observation arrives. Specifically, if a new value X_{n+1} comes at time $n+1$, then the estimates in (3)–(5) which are based on $X_1, \ldots, X_n$, should be updated within at least $O(n)$ computing steps. Additionally, one has to store all the data $X_1, \ldots, X_n$ available up to time n, thus having $O(n)$ memory complexity. The latter two shortcomings are highly undesirable in situations in which one needs to process very long time series. In contemporary signal processing, with technological advances, extra long time series which are machine collected are now commonly seen and it then poses new challenges for spectrum estimation.

Of particular interest is the estimation of $f(0)$. Under suitable conditions on (X_i) (cf. [14], [15], and [16], among others), one has

$$\sqrt{n}(\bar{X}_n - \mu) \Rightarrow N(0, \sigma^2), \text{ where } \sigma^2 = \sum_{k\in\mathbb{Z}} \gamma_k. \tag{6}$$

Here, $\Rightarrow$ denotes convergence in distribution and σ^2 is called the time-average variance constant (TAVC), long-run variance or asymptotic variance parameter. Note that $2\pi f(0) = \sigma^2$. Estimation of σ^2 has been extensively studied. The method of batched means has been discussed in [17]–[19] (see also references therein). Based on the batched means $\sum_{i=k}^{k+b_n-1} X_i/b_n$

Manuscript received August 01, 2009; revised April 19, 2010; accepted November 18, 2010. Date of current version June 22, 2011.

The authors are with the Department of Statistics, The University of Chicago, Chicago, IL 60637 USA (e-mail: xiao@galton.uchicago.edu; wbwu@galton.uchicago.edu).

Communicated by E. Serpedin, Associate Editor for Signal Processing.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2011.2145610

for $k = 1, 2, \ldots, n - b_n + 1$, the OBM estimate of σ^2 is a version of (4) with $\theta = 0$. If we use nonoverlapping batched means $\sum_{i=kb_n+1}^{(k_n+1)b_n} X_i/b_n$ for $k = 0, 1, \ldots, s_n - 1$ (assume for simplicity $n = s_n b_n$), then the estimate

$$\hat{\sigma}_n^2 = \frac{b_n}{s_n} \sum_{k=0}^{s_n-1} \left(\frac{1}{b_n} \sum_{i=kb_n+1}^{(k_n+1)b_n} X_i - \bar{X}_n \right)^2 \tag{7}$$

is called nonoverlapping batched means (NBM) estimate. In this paper, the term "BM estimate" refers to both OBM and NBM estimates. Note that $\bar{X}_n$ can be computed recursively via $\bar{X}_{n+1} = (n\bar{X}_n + X_{n+1})/(n+1)$. As in [20], if b_n is fixed and does not depend on n, the OBM estimate (4) can be computed recursively. The latter is no longer true if b_n grows as n increases. Generally, consistency of $\hat{\sigma}_n^2$ requires $b_n \to \infty$ and $n/b_n \to \infty$. Reference [21] proposed a recursive algorithm for computing σ^2 via a modified OBM estimate.

The rest of this article is structured as follows. Section II introduces a recursive (one-pass or single-pass) algorithm for estimating spectral density functions. Our algorithm provides real time updates and is therefore useful for efficient and fast processing for extra long time series. The computational advantage becomes more attractive if one wants to compute values of spectral densities at multiple frequencies. We also present an improved, probabilistically faster estimate where we can have a better control on the bias and hence we get faster convergence with smaller mean squares error (MSE). Both estimates are given for two cases depending on whether the mean μ is known or not. In Section III, we investigate asymptotic statistical properties of our algorithms, present MSE bounds and central limit theorems and discuss the choice of batch sizes. We present in Section IV a simulation study and compare the performance of recursive and non-recursive estimates. The technical lemmas and proofs are gathered in Section V.

We now introduce some notation. For $p > 1$, we say a (complex) random variable $X \in \mathcal{L}^p$ if $\|X\|_p := (\mathbb{E}|X|^p)^{1/p} < \infty$. Write $\|X\| = \|X\|_2$. For two nonnegative sequences (a_n) and (b_n), write $a_n = O(b_n)$ (resp. $a_n \asymp b_n$) if there exists a constant $c > 0$ such that $\limsup_{n\to\infty} a_n/b_n \le c$ (resp. $1/c \le \liminf_{n\to\infty} a_n/b_n \le \limsup_{n\to\infty} a_n/b_n \le c$) and $a_n = o(b_n)$ if $a_n/b_n \to 0$. For two real numbers a and b, define $a \vee b = \max\{a, b\}$ and $a \wedge b = \min\{a, b\}$. We use C for a constant and use C_p to emphasize that the constant depends on p. The values of C and C_p may vary from place to place.

II. Recursive Spectral Density Estimation

That (4) is nonrecursive is due to the fact that the block size $b_n \to \infty$ and the summands have the same size b_n. Assuming at the outset that $\mu = 0$, Section II-A introduces a modified NBM estimate with varying block sizes and a corresponding algorithm, for which the important goal of algorithmic recursiveness can be achieved. The algorithm is improved in Section II-B where an estimate with bias correction is given, while the recursive property is retained. Section II-C deals with the case in which the mean μ is unknown and one only needs to slightly modify the recursive algorithm in Sections II-A and II-B.

A. Estimates With Varying Block Sizes

We first assume that μ is known and $\mu = 0$ (say). Let (a_k) be a sequence of strictly increasing positive integers such that:) $a_1 = 1$; ii) the differences $\Delta_k := a_{k+1} - a_k$ is nondecreasing in k; and iii) $\Delta_k \to \infty$ as $k \to \infty$. For $a_k \le i < a_{k+1}$, let $l_i = a_k$ and $u_i = a_{k+1}$. As an example, if $a_k = k^3$, then $l_i = \lfloor i^{1/3} \rfloor^3$ and $u_i = (\lfloor i^{1/3} \rfloor + 1)^3$, where $\lfloor u \rfloor = \max\{i \in \mathbb{Z}, i \le u\}$ is the integer part of $u \in \mathbb{R}$. For $n \in \mathbb{N}$, let s_n be such that $a_{s_n} \le n < a_{s_n+1}$. In some occasions, we omit the subscript n from s_n when there is no confusion. Let $\mathcal{B}_k = \{a_k, a_k + 1, \ldots, a_{k+1} - 1\}$.

Given $X_1, X_2, \ldots, X_n$, define

$$V_n(\theta) = \sum_{k=1}^{s_n-1} |B_k(\theta)|^2 + |R_n(\theta)|^2 \tag{8}$$

where the block sums

$$B_k(\theta) = \sum_{i=a_k}^{a_{k+1}-1} X_i \varrho^i \text{ and } R_n(\theta) = \sum_{i=a_{s_n}}^{n} X_i \varrho^i. \tag{9}$$

We propose to estimate the spectral density $f(\theta)$ by $V_n(\theta)/(2\pi n)$. In the sequel, since θ will be treated as fixed, we shall also omit θ and write $V_n(\theta)$ (resp. $B_k(\theta)$, $f(\theta)$ etc.) as V_n (resp. B_k, f etc).

In the expression of R_n, if $n + 1 \neq a_{s_n+1}$, then $n + 1$ still belongs to the block $\mathcal{B}_{s_n}$ and $R_{n+1} = R_n + X_{n+1}\varrho^{n+1}$. If $n+1 = a_{s_n+1}$, then $n+1$ belongs to the next block $\mathcal{B}_{s_n+1}$ and we start a new $R_{n+1} = X_{n+1}\varrho^{n+1}$. To summarize, we propose the following single-pass algorithm:

Algorithm 1 At step n, we store the vector $\{s_n, R_n, V_n\}$. At step $n + 1$, we update it as follows:

(i) if $n+1 \neq a_{s_n+1}$, let $R_{n+1} = R_n + X_{n+1}\varrho^{n+1}$, $s_{n+1} = s_n$ and $V_{n+1} = V_n - |R_n|^2 + |R_{n+1}|^2$;

(ii) if $n+1 = a_{s_n+1}$, let $R_{n+1} = X_{n+1}\varrho^{n+1}$, $s_{n+1} = s_n + 1$ and $V_{n+1} = V_n + |R_{n+1}|^2$;

and we give the output as $\hat{f}_{n+1} = V_{n+1}/(2\pi(n+1))$.

The block sizes in (8) are changing. The memory complexity of Algorithm 1 is $O(1)$ and the computational complexity scales linearly in n.

B. Estimates With Bias Reduction

Our quadratic sum V_n can be expressed as

$$V_n = \sum_{i=1}^{n} Q_i \tag{10}$$

where

$$Q_i = X_i^2 + \sum_{j=l_i}^{i-1} X_i X_j \left(\varrho^{i-j} + \varrho^{j-i} \right).$$

In view of (2), the bias corresponding to the term Q_i is

$$\mathbb{E}Q_i - 2\pi f = 2 \sum_{j > i - l_i} \gamma_j \cos(j\theta). \tag{11}$$

The bias can be large if $i - l_i$ is small. Therefore, in order to reduce the bias, we modify V_n by only including those Q_i for which $i - l_i$ is large. Specifically, we pick a sequence of increasing thresholds $(d_k) \in \mathbb{N}$ and for each $i \in \mathcal{B}_k = \{a_k, a_k + 1, \ldots, a_{k+1} - 1\}$, we include Q_i only if $i - l_i \geq d_k$. This leads to the new estimate V_n° / v_n, where V_n° is given by

$$V_n^\circ = \sum_{i=1}^{n} Q_i \mathbf{1}\{i - l_i \geq d_{s_i}\} \tag{12}$$

and

$$v_n = n - \sum_{k=1}^{s-1} d_k - [(n - a_s) \wedge d_s]. \tag{13}$$

Algorithm 1 can be modified and one can compute V_n° recursively.

Algorithm 2 At step n, we store $\{s_n, R_n, V_n^\circ, v_n\}$. At step $n+1$, we update it as follows:

(i) if $n + 1 \neq a_{s_n+1}$, let $s_{n+1} = s_n$
 — if $n + 1 - a_{s_n} \geq d_{s_n}$, let $V_{n+1}^\circ = V_n^\circ + X_{n+1}^2 + X_{n+1}(\varrho^{n+1}\bar{R}_n + \varrho^{-n-1}R_n)$, $R_{n+1} = R_n + X_{n+1}\varrho^{n+1}$ and $v_{n+1} = v_n + 1$;
 — if $n + 1 - a_{s_n} < d_{s_n}$, let $V_{n+1}^\circ = V_n^\circ$, $R_{n+1} = R_n + X_{n+1}\varrho^{n+1}$ and $v_{n+1} = v_n$;

(ii) if $n+1 = a_{s_n+1}$, let $R_{n+1} = X_{n+1}\varrho^{n+1}$, $s_{n+1} = s_n + 1$, $V_{n+1}^\circ = V_n^\circ$ and $v_{n+1} = v_n$;

and we give the output as $\hat{f}_{n+1} = V_{n+1}^\circ / (2\pi v_{n+1})$.

By deleting Q_i in V_n with small $i - l_i$, we can have an estimate which converges more quickly in probability; see Theorem 2. On the other hand, however, our estimate $V_n^\circ/(2\pi v_n)$ may take negative values. To implement Algorithm 2, we need to choose the block sequence (a_k) and the threshold sequence (d_k). Theorem 2 also provides a guideline on how to choose them.

C. Estimates With Unknown Means

In practice, μ is often unknown. So B_k and R_n need to be centered. It is natural to estimate μ by the sample mean $\bar{X}_n = \sum_{i=1}^{n} X_i / n$. Let

$$V_n' = \sum_{k=1}^{s-1} |B_k'|^2 + |R_n'|^2 \tag{14}$$

where, similarly as (8) and (9)

$$B_k' = \sum_{i=a_k}^{a_{k+1}-1} (X_i - \bar{X}_n)\varrho^i \text{ and } R_n' = \sum_{i=a_s}^{n} (X_i - \bar{X}_n)\varrho^i.$$

We propose to estimate f by $V_n'/(2\pi n)$. Let $\Delta_k(\theta) = \sum_{i=a_k}^{a_{k+1}-1} \varrho^i$ and $\delta_n(\theta) = \sum_{i=a_s}^{n} \varrho^i$, simple algebra shows that the difference

$$V_n' - V_n = -(F_n + \bar{F}_n)\bar{X}_n + q_n \bar{X}_n^2 \tag{15}$$

where

$$F_n = \sum_{k=1}^{s-1} \Delta_k(\theta)\bar{B}_k + \delta_n(\theta)\bar{R}_n$$

and

$$q_n = \sum_{k=1}^{s-1} |\Delta_k(\theta)|^2 + |\delta_n(\theta)|^2.$$

We use the following algorithm to compute V_n'/n recursively.

Algorithm 3 At step n, we store

$$\{\bar{X}_n, s_n, R_n, V_n, \delta_n, q_n, F_n\}.$$

At step $n+1$, we update it as follows:

(i) let $\bar{X}_{n+1} = (n\bar{X}_n + X_{n+1})/(n+1)$;

(ii) if $n + 1 \neq a_{s_n+1}$, let $R_{n+1} = R_n + X_{n+1}\varrho^{n+1}$, $s_{n+1} = s_n$, $V_{n+1} = V_n - |R_n|^2 + |R_{n+1}|^2$, $\delta_{n+1} = \delta_n + \varrho^{n+1}$, $q_{n+1} = q_n - |\delta_n|^2 + |\delta_{n+1}|^2$ and $F_{n+1} = F_n - \delta_n \bar{R}_n + \delta_{n+1}\bar{R}_{n+1}$;

(iii) if $n+1 = a_{s_n+1}$, let $R_{n+1} = X_{n+1}\varrho^{n+1}$, $s_{n+1} = s_n + 1$, $V_{n+1} = V_n + |R_{n+1}|^2$, $\delta_{n+1} = \varrho^{n+1}$, $q_{n+1} = q_n + |\delta_{n+1}|^2$ and $F_{n+1} = F_n + \delta_{n+1}\bar{R}_{n+1}$;

and as the output, we compute $\hat{f}_{n+1} = V_{n+1}'/(2\pi(n+1))$, where

$$V_{n+1}' = V_{n+1} - (F_{n+1} + \bar{F}_{n+1})\bar{X}_{n+1} + q_{n+1}\bar{X}_{n+1}^2.$$

A similar bias-corrected version of the estimate V_n' can be constructed in a straightforward way:

$$V_n'^\circ = \sum_{i=1}^{n} Q_i' \mathbf{1}\{i - l_i \geq d_{s_i}\} \tag{16}$$

where

$$Q_i' = (X_i - \bar{X}_n)^2 + \sum_{j=l_i}^{i-1} (X_i - \bar{X}_n)(X_j - \bar{X}_n)(\varrho^{i-j} + \varrho^{j-i}).$$

Let

$$\begin{aligned} F_n' &= \sum_{k=1}^{s-1} \left(\Delta_k(\theta)B_k - d_k(\theta)B_k^d\right) \\ &\quad + \left(\delta_n(\theta)R_n - d_s(\theta)B_s^d\right)\mathbf{1}\{n \geq a_s + d_s\} \\ q_n' &= \sum_{k=1}^{s-1} \left(|\Delta_k(\theta)|^2 - |d_k(\theta)|^2\right) \\ &\quad + \left(|\delta_n(\theta)|^2 - |d_s(\theta)|^2\right)\mathbf{1}\{n \geq a_s + d_s\} \end{aligned}$$

where $d_k(\theta) = \sum_{i=a_k}^{a_k+d_k-1} \varrho^i$ and $B_k^d = \sum_{i=a_k}^{a_k+d_k-1} X_i \varrho^i$, then

$$V_n'^\circ = V_n^\circ - (F_n' + \bar{F}_n')\bar{X}_n + q_n'\bar{X}_n^2. \tag{17}$$

It is clear that $V_n'^\circ$ can be computed recursively as well. The algorithm, which is a combination of Algorithms 2 and 3, can be easily worked out and the details are omitted here.

III. Asymptotic Theory

Asymptotic properties of the BM estimates of the long-run variance $\sigma^2 = 2\pi f(0)$ have been extensively studied. [22] and [23] obtained strong consistency. [24] and [25] derived MSE bounds which are used to choose batch sizes. Their results depend on restrictive moment conditions and strong mixing conditions which are not easily verifiable. [21] studied the long-run variance estimation problem for recursive OBM estimates.

We shall establish an asymptotic theory of our recursive spectral density estimates by implementing the dependence measure in [15]. Assume that (X_i) is a stationary causal process of the form

$$X_i = g(\ldots, \epsilon_{i-1}, \epsilon_i) \tag{18}$$

where $\epsilon_i, i \in \mathbb{Z}$, are iid random variables and g is a measurable function for which X_i is a properly defined random variable. Following [26] and [27], we interpret (18) as an input/output physical system with $\mathcal{F}_i = (\ldots, \epsilon_{i-1}, \epsilon_i)$ being the input, g being the filter and X_i being the output. We shall also write $X_i = g(\mathcal{F}_i)$. The class of process that (18) represents is huge and it includes linear processes, Volterra processes and many other time series models; see [28] and [29]. Applying the idea of coupling, [15] introduced the physical dependence measure. Let (ϵ'_i) be an iid copy of (ϵ_i), $\mathcal{F}_i = (\ldots, \epsilon_{-1}, \epsilon_0, \epsilon_1, \ldots, \epsilon_i)$, $\mathcal{F}'_i = (\ldots, \epsilon_{-1}, \epsilon'_0, \epsilon_1, \ldots, \epsilon_i)$, $X_i = g(\mathcal{F}_i)$ and $X'_i = g(\mathcal{F}'_i)$. For $p > 1, X_i \in \mathcal{L}^p$, define the *physical dependence measure*

$$\delta_p(i) = \|X_i - X'_i\|_p, \tag{19}$$

which quantifies the dependence of $X_i = g(\mathcal{F}_i)$ on ϵ_0 by measuring the distance between $g(\mathcal{F}_i)$ and its coupled (at the position ϵ_0) version $g(\mathcal{F}'_i)$. The physical dependence measure is directly related to the underlying data-generating mechanism. Our main results are based on $\delta_p(i)$. Define

$$\Theta_{n,p} = \sum_{j=n}^{\infty} \delta_p(j) \text{ and } \Psi_{n,p} = \left(\sum_{j=n}^{\infty} \delta_p(j)^{p'} \right)^{1/p'} \tag{20}$$

where $p' = \min(2, p)$.

In the sequel, we let $\sigma^2(\theta) = 2\pi f(\theta)$ and write σ^2 (resp. V_n) for $\sigma^2(\theta)$ (resp. $V_n(\theta)$) if there is no confusion caused.

A. Mean-Squared Error Bounds

With the physical dependence measure, we have the following convergence rates of the variances of the estimates V_n/n and V_n°/v_n. Define $\varpi(\theta) = 4$ if $\theta/\pi \in \mathbb{Z}$ and $\varpi(\theta) = 2$ if otherwise.

Theorem 1: Assume $\mathbb{E}X_0 = 0$, $X_0 \in \mathcal{L}^4$, and $\Theta_{0,4} < \infty$. Assume the sequence (Δ_t) satisfies that $\sum_{t=1}^{s} \Delta_t^2 \asymp \Delta_s \sum_{t=1}^{s} \Delta_s$ and $\Delta_{s_n} = o(n)$.

(i) Let $\nu_n^2 = \varpi(\theta) \sum_{i=1}^{n} (i - l_i)$, then

$$\|V_n - \mathbb{E}V_n\| = \nu_n(\sigma^2 + o(1)) \asymp \Delta_{s_n}^{1/2} n^{1/2}.$$

(ii) Let $(\nu_n^\circ)^2 = \varpi(\theta) \sum_{i=1}^{n} (i - l_i) \mathbf{1}\{i - l_i \geq d_{s_i}\}$. If $\Delta_k - d_k \asymp \Delta_k$, then

$$\|V_n^\circ - \mathbb{E}V_n^\circ\| = \nu_n^\circ(\sigma^2 + o(1)) \asymp \Delta_{s_n}^{1/2} n^{1/2}.$$

Theorem 2 concerns convergence rates of the MSE for some particular choices of (a_k) and (d_k). We need either of the following assumptions on the auto-covariances (γ_k).

A1

$$\sum_{k=0}^{\infty} k^q |\gamma_k| < \infty \text{ for some } q > 0. \tag{21}$$

A2

$$|\gamma_k| \leq C\rho^k \text{ for some } C > 0,\ 0 < \rho < 1. \tag{22}$$

Theorem 2: Let conditions of Theorem 1 be satisfied.

(i) Assume (21) with $0 < q \leq 1$, then

$$|\mathbb{E}V_n - n\sigma^2|^2 = \begin{cases} o(s_n^2 \Delta_{s_n}^{2(1-q)}), & \text{if } 0 < q < 1 \\ O(s_n^2), & \text{if } q = 1 \end{cases}.$$

Hence, if $a_k = \lfloor ck^p \rfloor$ for some $c > 0$ and $p > 1$, then $\|V_n - n\sigma^2\|^2 = O(b_n)$, where $b_n = n^{2-1/p} + n^{2/p+2(p-1)(1-q)/p}$ which reaches the smallest order of magnitude if $p = 1 + 1/(2q)$. In particular, if $q = 1$ and $p = 3/2$, then $\|V_n/n - \sigma^2\|^2 = O(n^{-2/3})$.

(ii) Assume (21) with $q > 1$. Let $a_k = \lfloor c_1 k^p \rfloor$ and $d_k = \lfloor \Delta_k / c_2 \rfloor$ with some $c_1 > 0$, $p > 1$ and $c_2 > 1$, then $\|V_n^\circ - \mathbb{E}V_n^\circ\|^2 = O(n^{2-1/p})$ and

$$|\mathbb{E}V_n^\circ - v_n\sigma^2|^2 = \begin{cases} o(n^{2(p-1)(1-q)/p+2/p}), & \text{if } (p-1)(1-q) > -1 \\ o(\log(n)^2), & \text{if } (p-1)(1-q) = -1 \\ O(1), & \text{if } (p-1)(1-q) < -1 \end{cases}.$$

In particular, if $p = 1 + 1/(2q)$, then the MSE $\|V_n^\circ/v_n - \sigma^2\|^2 = O(n^{-2q/(2q+1)})$ reaches the smallest order.

(iii) Assume (22). Let $a_k = \lfloor ck^p \rfloor$ and $d_k = \lfloor \lambda \log(k) \rfloor$, where $c > 0$, $p > 1$ and $\lambda > 0$, then $\|V_n^\circ - \mathbb{E}V_n^\circ\|^2 = O(n^{2-1/p})$ and

$$|\mathbb{E}V_n^\circ - v_n\sigma^2|^2 = \begin{cases} O(n^{2/p+2\lambda\log(\rho)/p}), & \text{if } \lambda\log(\rho) > -1 \\ O(\log(n)^2) & \text{if } \lambda\log(\rho) = -1 \\ O(1), & \text{if } \lambda\log(\rho) < -1 \end{cases}$$

If $2\lambda\log(\rho) \leq 2p - 3$, then the MSE $\|V_n^\circ/v_n - \sigma^2\|^2 = O(n^{-1/p})$.

(iv) Assume (22). Let $a_k = \lfloor \lambda_1 k \log(k) \rfloor$ and $d_k = \lfloor \lambda_2 \log(k) \rfloor$, $\lambda_1 > \lambda_2 > 0$. Then $\|V_n^\circ - \mathbb{E}V_n^\circ\|^2 = O(n\log(n))$ and

$$|\mathbb{E}V_n^\circ - v_n\sigma^2|^2 = \begin{cases} O(n^{2+2\lambda_2\log(\rho)}), & \text{if } \lambda_2\log(\rho) > -1 \\ O(\log(n)^2) & \text{if } \lambda_2\log(\rho) = -1 \\ O(1), & \text{if } \lambda_2\log(\rho) < -1 \end{cases}$$

If $2\lambda_2\log(\rho) \leq -1$, then $\|V_n^\circ/v_n - \sigma^2\|^2 = O(\log(n)/n)$.

When (21) holds with $q > 1$, the MSE of the lag window estimate has the optimal order $O\left(n^{-2q/(2q+1)}\right)$ with an appropriately chosen kernel [30, see for example Section 9.3]. Since the BM estimate corresponds roughly to the lag window estimate with Bartlett kernel, it cannot utilize the extra smoothness of the spectral density and the optimal order remains $O(n^{-2/3})$ even when $q > 1$ or (22) holds. Our bias-reduced estimates actually correspond to the lag window which is flat with value 1 around

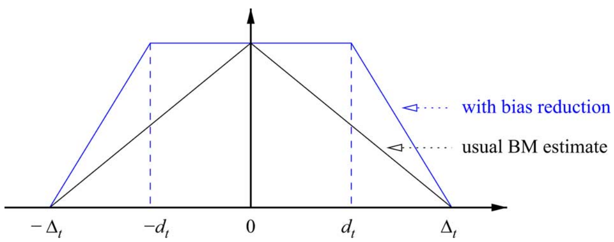

Fig. 1. Lag windows for the usual BM estimate and the one with bias reduction.

0 and then decreases linearly to 0 (see Fig. 1). Therefore, they can exploit the quick decay of the autocovariances and achieve the optimal convergence rate as the lag window estimates (see part (ii) of Theorem 2). We shall remark that the bias reduction works not only for the recursive estimates, but also for the usual non-recursive BM estimates.

Theorem 2 suggests how to choose the block sequence (a_k) and the threshold sequence (d_k). If (22) holds, we have a very good control on the bias term; see (11). Thus, the bias-correction is necessary. We can either choose $a_k = \lfloor ck^p \rfloor$ for p very close to one and $d_k = \lambda \log(k)$ for some $\lambda > 0$ large enough to get an order of $O(n^{-1/p})$ for the MSE $\|V^\circ/v_n - \sigma^2\|^2$, or choose $a_k = \lfloor \lambda_1 k \log(k) \rfloor$ and $d_k = \lfloor \lambda_2 \log(k) \rfloor$ for some $\lambda_1 > \lambda_2 > 0$ large enough so that the convergence rate of $\|V^\circ/v_n - \sigma^2\|^2$ is $O(\log(n)/n)$, which is nearly optimal. If (21) holds with $q > 1$, we can choose $a_k = \lfloor c_1 k^p \rfloor$ and $d_k = \lfloor \Delta_k / c_2 \rfloor$ with $p = 1 + 1/(2q)$ and $c_2 > 1$ to get the order $O(n^{-2q/(2q+1)})$. However, if we only have (21) with $0 < q \le 1$, the bias correction does no good and we are in case (i) of Theorem 2.

We now compare Theorem 2 with existing results on BM estimates of the TAVC $\sigma^2(0) = 2\pi f(0)$. [31] considered the special AR(1) process $X_i = \alpha X_{i-1} + \epsilon_i$, where $|\alpha| < 1$ and ϵ_i are iid standard normal random variables and showed that the MSE of the NBM estimates of the TAVC is asymptotically $3(2\alpha/(1-\alpha^2))^{2/3}(1-\alpha)^{-4} n^{-2/3}$. [25] showed that under some summability condition of the ϕ-mixing coefficients and the moment condition $\mathbb{E}|X_i|^{12} < \infty$, the optimal error bound of the MSE is $O(n^{-2/3})$ if the batch size is of order $O(n^{1/3})$. In comparison, our dependence measures and moment conditions are mild and natural. [21] obtained the same bound for recursive OBM estimates under our Assumption 1 with $q = 1$. By Theorem 2, we can obtain the same bound by choosing $a_k = \lfloor ck^{3/2} \rfloor$ when Assumption 1 holds with $q = 1$, whereas if Assumption 2 holds, we can get much better error bounds. The error bound $O(\log(n)/n)$ in case (iii) is almost optimal, since in the case where X_i's are independent and identically distributed (i.i.d.), the order is $O(1/n)$.

Remark 1: We provide sufficient conditions for Assumptions 1 and 2. Since projection operators

$$\mathcal{P}_k \cdot = \mathbb{E}(\cdot|\mathcal{F}_k) - \mathbb{E}(\cdot|\mathcal{F}_{k-1}), \quad k \in \mathbb{Z}, \tag{23}$$

generate martingale differences and $X_i = \sum_{j\in\mathbb{Z}} \mathcal{P}_j X_i$,

$$|\gamma_k| = |\mathbb{E}(X_0 X_k)| = \left|\mathbb{E}\sum_{j\in\mathbb{Z}} (\mathcal{P}_j X_0)(\mathcal{P}_j X_k)\right| \le \sum_{j\le 0} \|\mathcal{P}_j X_0\| \|\mathcal{P}_j X_k\| \le \sum_{j\le 0} \delta_2(-j)\delta_2(k-j).$$

Hence, if $\sum_{k=0}^{\infty} k^q \delta_\alpha(k) < \infty$ for some $0 < q \le 1$ and $\alpha \ge 2$, then Assumption 1 holds; if $\delta_\alpha(k) = O(\rho^k)$ for some $0 < \rho < 1$ and $\alpha \ge 2$; then Assumption 2 holds. The latter property is called the *geometric-moment contraction* and it holds for a wide class of nonlinear time series; see [32].

B. Central Limit Theorems

To construct confidence intervals for values of spectral density functions, one needs to have a central limit theorem (CLT) instead of MSE bounds. CLTs for quadratic forms have a long history. See [33], [34] and [35] and references therein for the case where X_i's are iid. For stationary processes in which dependence is an intrinsic nature, the central limit problem becomes very challenging. [8], [36] and [37] assumed strong mixing conditions. [38] made a recent breakthrough and they obtained a CLT for the lag window spectral density estimate. Here, with the physical dependence measure, we shall present a CLT for our recursive estimates.

Theorem 3: Let conditions of Theorem 1 be satisfied. Recall that $\varpi(\theta) = 4$ if $\theta/\pi \in \mathbb{Z}$ and $\varpi(\theta) = 2$ if otherwise.

(i) Let $\nu_n^2 = \varpi(\theta) \sum_{i=1}^n (i - l_i)$. Then
$$\frac{V_n - \mathbb{E}V_n}{\nu_n} \Rightarrow N(0, \sigma^4). \tag{24}$$

(ii) Let $(\nu_n^\circ)^2 = \varpi(\theta) \sum_{i=1}^n (i - l_i) \mathbf{1}\{i - l_i \ge d_{s_i}\}$. Then
$$\frac{V_n^\circ - \mathbb{E}V_n^\circ}{\nu_n^\circ} \Rightarrow N(0, \sigma^4). \tag{25}$$

To construct confidence intervals for spectral densities, we shall replace $\mathbb{E}V_n$ and $\mathbb{E}V_n^\circ$ by $n\sigma^2$ and $v_n\sigma^2$ in (24) and (25), respectively. For the choices of a_k and d_k in Theorem 2 (ii), (iii) and (iv), since the squared bias is of smaller order than the variance, we can do the replacement without changing the limiting distribution. It is a similar situation for Theorem 2 (i) when $q < 1$ and $p = 1 + 1/2q$. However, when $q = 1$, if we choose the optimal $p = 3/2$, a direct calculation shows that the limiting distribution has a nonzero mean if we do the replacement. To summarize, we have

Corollary 4: Assume the same conditions as Theorem 1. Then

(i) Assume (21) with $0 < q \le 1$, choose $a_k = \lfloor ck^p \rfloor$ for some $c > 0$ and $p = 1 + 1/(2q)$. Then if $0 < q < 1$, we still have (24) if $\mathbb{E}V_n$ is replaced by $n\sigma^2$. If $q = 1$, we have
$$\frac{V_n - n\sigma^2}{\nu_n} \Rightarrow N(4r(\theta)/3c, \sigma^4),$$
where $r(\theta) = \sum_{j=1}^{\infty} j\gamma_j \cos(j\theta)$.

(ii) Assume (21) with $q > 1$, choose $a_k = \lfloor c_1 k^p \rfloor$ and $d_k = \lfloor \Delta_k / c_2 \rfloor$ with some $c_1 > 0$, $c_2 > 1$ and $p = 1 + 1/2q$. Then (25) remains true if we replace $\mathbb{E}V_n^\circ$ by $v_n\sigma^2$.

(iii) Assume (22), choose $a_k = \lfloor ck^p \rfloor$ and $d_k = \lfloor \lambda \log(k) \rfloor$, where $c > 0$, $p > 1$ and $\lambda > 0$ is such that $2\lambda \log(\rho) < 2p - 3$. Then (25) still holds true if we replace $\mathbb{E}V_n^\circ$ by $v_n\sigma^2$.

(iv) Assume (22), choose $a_k = \lfloor \lambda_1 k \log(k) \rfloor$ and $d_k = \lfloor \lambda_2 \log(k) \rfloor$, where $\lambda_1 > \lambda_2 > 0$ and $2\lambda_2 \log(\rho) \le -1$. Then (25) is still true if we replace $\mathbb{E}V_n^\circ$ by $v_n\sigma^2$.

C. Centered Estimates

By (15), same results for the centered estimate V_n' can be obtained from the corresponding ones for V_n, if we have a good control on $F_n \bar{X}_n$ and $q_n \bar{X}_n^2$; and indeed it is true [see (47)]. The situation for $V_n'^{\circ}$ is similar in view of (17). Corollary 5 summarizes the results for those centered estimates.

Corollary 5: Assume $\Delta_{s_n} = o(n)$ and let conditions of Theorem 1 be satisfied. Then Theorems 1 and 3 and Corollary 4 still hold for the centered estimates V_n' and $V_n'^{\circ}$. Furthermore, the following results in Theorem 2 remain true.

(i) Assume (21) with $0 < q \le 1$. Let $a_k = \lfloor ck^p \rfloor$ for some $c > 0$ and $p = 1 + 1/(2q)$. Then $\|V_n'/n - \sigma^2\|^2 = O(n^{-2q/(1+2q)})$.

(ii) Assume (21) with $q > 1$. Let $a_k = \lfloor c_1 k^p \rfloor$ and $d_k = \lfloor \Delta_k / c_2 \rfloor$ with some $c_1 > 0$, $c_2 > 1$ and $p = 1 + 1/2q$. Then $\|V_n'^{\circ}/v_n - \sigma^2\|^2 = O(n^{-2q/(1+2q)})$.

(iii) Assume (22). Let $a_k = \lfloor ck^p \rfloor$ and $d_k = \lfloor \lambda \log(k) \rfloor$, where $c > 0$, $p > 1$ and $2\lambda \log(\rho) \le 2p - 3$. Then $\|V_n'^{\circ}/v_n - \sigma^2\|^2 = O(n^{-1/p})$.

(iv) Assume (22). Let $a_k = \lfloor \lambda_1 k \log(k) \rfloor$ and $d_k = \lfloor \lambda_2 \log(k) \rfloor$, where $\lambda_1 > \lambda_2 > 0$ and $2\lambda_2 \log(\rho) \le -1$. Then $\|V_n'^{\circ}/v_n - \sigma^2\|^2 = O(\log(n)/n)$.

IV. Simulation Results

This section presents a simulation study and compares the MSE of our recursive BM estimates with other popular estimates:

(a) Recursive NBM estimate (8) with $a_k = \lfloor k^{3/2} \rfloor$;
(b) Recursive NBM estimate (12) with $a_k = \lfloor 6k \log(k) \rfloor + 1$ and $d_k = \lfloor 2 \log(k) \rfloor$;
(c) Recursive OBM estimate in [21] with $a_k = \lfloor k^{3/2} \rfloor$;
(d) Nonrecursive NBM estimate, c.f [25] with batch size $\lfloor n^{1/3} \rfloor$;
(e) Lag window estimate with truncation kernel

$$\hat{f}(\theta) = \frac{1}{2\pi} \sum_{k=-B_n}^{B_n} \hat{\gamma}_k \varrho^k.$$

We abbreviate them by RNB1, RNB2, ROB, NB, and LW, respectively. The LW estimate in (e) is the analog of the one in (b) in the context of lag window estimates. To compare the LW estimate with (b), we choose $B_n = \lfloor 2 \log(k) \rfloor$, where k is the solution of the equation $a_k = n$. We consider linear and bilinear processes whose spectral densities have closed forms.

Example 1 (Linear Process): Consider the $\mathrm{ARMA}(p,q)$ process

$$X_i - \beta_1 X_{i-1} - \cdots - \beta_p X_{i-p} = \epsilon_i + \alpha_1 \epsilon_{i-1} + \cdots + \alpha_q \epsilon_{i-q},$$

where ϵ_i are iid with mean 0 and variance τ^2 and $\beta_1, \ldots, \beta_p, \alpha_1, \ldots, \alpha_q$ are real parameters. Assume all roots of the equation $z^p - \sum_{j=1}^p \beta_j z^{p-j} = 0$ lie inside the unit circle. Recall $\varrho = e^{\sqrt{-1}\theta}$. The spectral density function is

$$f(\theta) = \frac{\tau^2}{2\pi} \frac{\left|1 + \sum_{j=1}^q \alpha_j \varrho^j\right|^2}{\left|1 - \sum_{j=1}^p \beta_j \varrho^j\right|^2}.$$

TABLE I
MSE of Estimates of TAVC for the AR(1) Model

n	5e3	1e4	2e4	5e4	1e5	2e5	5e5
RNB1	.2114	.1294	.0834	.0452	.0297	.0192	.0104
	.2211	.1347	.0857	.0462	.0302	.0194	.0105
RNB2	.3187	.1731	.1005	.0468	.0249	.0148	.0063
	.3190	.1745	.1005	.0467	.0250	.0148	.0063
ROB	.3333	.2180	.1416	.0789	.0527	.0342	.0180
	.3447	.2244	.1443	.0780	.0533	.0345	.0181
NB	.1891	.1290	.0779	.0396	.0267	.0175	.0094
	.1972	.1338	.0801	.0405	.0271	.0177	.0095
LW	.1161	.0711	.0397	.0181	.0093	.0050	.0024
	.1169	.0708	.0396	.0181	.0093	.0050	.0024

For each estimate, the first line is obtained for the model with known $\mathbb{E}X_i = 0$, while the senond line is for the centered estimates.

TABLE II
MSE for Estimates of TAVC for the Bilinear Model

n	5e3	1e4	2e4	5e4	1e5	2e5	5e5
RNB1	2.966	1.744	1.081	.5349	.3248	.1907	.0924
	3.01	1.777	1.094	.5401	.3278	.1922	.0929
RNB2	4.181	2.209	1.221	.5131	.3119	.1594	.0707
	4.146	2.205	1.216	.5125	.312	.1595	.0706
ROB	3.892	2.621	1.724	.8611	.5388	.3383	.1706
	3.944	2.657	1.738	.8678	.5423	.34	.1712
NB	2.899	1.697	1.043	.5119	.3149	.1771	.0902
	2.941	1.726	1.055	.5167	.3175	.1783	.0906
LW	2.702	1.421	.8087	.3785	.1868	.0912	.0368
	2.673	1.421	.8077	.3774	.1866	.0912	.0367

The footnote after Table I applies here.

Example 2 (Bilinear Model): Let ϵ_i be iid $N(0, \tau^2)$. Consider the recursion

$$X_i = (a + b\epsilon_i) X_{i-1} + \epsilon_i, \tag{26}$$

where a and b are real parameters. If $\rho^2 := a^2 + b^2\tau^2 < 1$, then (26) has a stationary solution [29, Theorem 4.5] and $\delta_2(n) = O(\rho^n)$. Simple calculation shows that

$$\gamma_0 = \frac{\tau^2}{1 - a^2 - b^2\tau^2} \text{ and } \gamma_k = a^{|k|}\gamma_0.$$

By Remark 1, (22) is satisfied. The spectral density is

$$f(\theta) = \frac{\gamma_0}{2\pi}\left(1 + \frac{2a(\cos(\theta) - a)}{(1 - a\cos(\theta))^2 + (a\sin(\theta))^2}\right).$$

We consider the estimation of TAVC on two models: (i) AR(1) model $X_i = 0.5X_{i-1} + \epsilon_i$; (ii) bilinear model $X_i = (0.6 + 0.4\epsilon_i)X_{i-1} + \epsilon_i$. In both models we let (ϵ_i) be iid $N(0,1)$. For each estimate and each sample size (see the top row of Table I, where 5e5 means 5×10^5), we repeat the simulation 1000 times and record the average of the 1000 squared distances from the true TAVC. The results (multiplied by $(2\pi)^2$) are summarized in Tables I and II. The MSE of RNB1 is always roughly the same as NB. The biased corrected version RNB2 does worse at the beginning, but performs better when the sample size gets larger. The MSE of RNB2 decreases faster than the other three estimates RNB1, ROB and NB as the sample size increases, which confirms the fast convergence asserted in Theorem 2 (iii). Furthermore, the RNB2 is closer to LW than RNB1, ROB and NB when the series is long enough. Elementary calculations show that, by [38, Theorem 2], the variance of the LW estimate is $8n^{-1}\log(n)\sigma^4(1 + o(1))$,

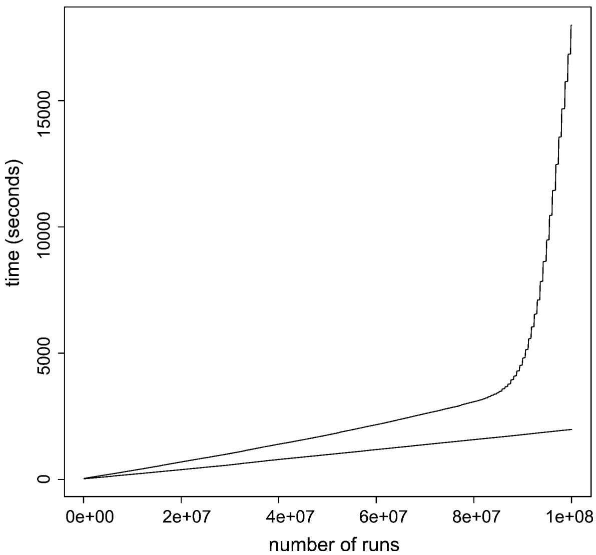

Fig. 2. Times needed for RNB1 and NB estimates. The straight line is for the recursive estimate. The computation is done on Intel Core Duo T2300 processor with 2G memory and Ubuntu 9.10 operating system.

TABLE III
MSE FOR SPECTRAL DENSITY ESTIMATES FOR THE AR(1) MODEL

	5e3	1e4	2e4	5e4	1e5	2e5	5e5
0	29962	17415	10254	5078	3009	1745	881.5
	43828	24730	12300	5183	2650	1487	607.6
	41831	25218	16029	8384	5180	3145	1725
	28576	17006	10059	4917	2927	1733	926.5
	29818	14745	8015	3672	1874	1039	404.8
$\pi/6$	2305	1244	657.1	301.5	170.4	94.29	44.7
	3659	2045	1051	439.2	248.4	119.9	54.36
	2795	1479	732.6	315.3	183.1	90.48	43.64
	2241	1223	627.9	267.8	161.1	85.61	44.6
	2145	1055	573.2	262.8	127.9	63.62	28.32
$\pi/3$	480.2	280.7	157.6	77.65	46.26	28.73	14.64
	485.6	297.6	154.2	67.69	35.56	18.05	7.783
	783.9	502.7	286	151.2	94.02	60.27	30.72
	483.1	282.9	154.3	79.6	45.3	28.11	14.12
	367.3	148.6	73.2	30.82	15.84	8.95	3.769
$\pi/2$	171.9	104.5	62.6	30.68	18.84	10.48	6.022
	157.6	83.31	42.84	19.23	10.05	5.122	2.321
	352.2	220	134.9	70.46	43.48	26.36	14.69
	184.5	103.7	63.12	31.64	18.84	10.96	5.928
	94.07	44.25	25.75	12.39	5.518	2.873	1.278
$2\pi/3$	92.39	55.15	33.12	17.35	10.45	6.481	3.291
	74.34	40.39	20.82	9.374	4.926	2.571	1.084
	193.7	118.3	74.04	39.44	24.2	15.21	8.148
	94.8	55.36	32.53	17.42	10.15	6.246	3.249
	34.12	27.94	15.18	6.465	3.109	1.143	0.6511
$5\pi/6$	63.38	38.49	22.6	11.42	7.042	4.173	2.228
	47.45	27.23	13.9	6.777	3.394	1.72	0.8407
	134.4	86.97	52.79	27.82	17.5	10.56	5.671
	62.47	39.14	21.89	11.95	6.875	4.2	2.176
	35.81	19.99	12.08	3.196	1.6	0.9712	0.5119
π	69.21	43.74	26.32	13.12	8.154	4.946	2.5
	71.43	42.32	23.27	9.917	4.959	2.891	1.418
	133.7	85.04	51.61	26.16	17.15	10.61	5.468
	67.4	42.65	24.44	12.55	7.735	4.911	2.503
	53.57	26.4	14.31	5.996	3.37	1.567	0.6711

We consider 7 frequencies evenly spaced on the interval $[0, \pi]$. There are five rows for each frequency, corresponding to RNB1, RNB2, ROB, NB and LW respectively. The average MSE is multiplied by $(2\pi)^2 \times 10^4$.

TABLE IV
MSE FOR SPECTRAL DENSITY ESTIMATES FOR THE BILINEAR MODEL

	5e3	1e4	2e4	5e4	1e5	2e5	5e5
0	2015	1323	859.4	456.4	284.2	184	104.8
	3020	1745	1003	430.6	243.9	131.1	59.92
	3362	2208	1435	811.6	505.5	315.7	177.9
	1927	1213	741.3	426.8	263	176.2	93.32
	1117	626.2	355.8	172.5	91.71	48.3	20.95
$\pi/6$	279.6	183.4	107.2	58.97	37.63	23.54	13.57
	733.8	428.2	229.6	102.2	57.19	30.9	15.21
	281.6	181.6	110.1	56.91	36.4	21.89	13.01
	261.5	159.9	98.12	52.66	31.92	20.8	12.09
	266.5	143.4	86.48	42.25	20.44	10.22	4.928
$\pi/3$	103.1	59.34	37.68	20.19	12.78	7.877	4.664
	187.3	114.6	61.88	29.84	16.41	8.385	3.529
	130.1	76.53	50.78	26.93	17.41	10.83	6.204
	91.63	56.2	33.19	18.98	10.75	7.202	3.883
	98.98	41.4	23.77	10.24	5.675	3.55	1.372
$\pi/2$	50.64	31.85	19.58	11.06	6.92	4.44	2.401
	74.49	42.12	21.71	10.31	5.747	2.948	1.275
	87.69	57.18	36.14	20.51	12.72	8.144	4.586
	47.02	30.34	18.38	10.08	6.576	3.739	2.186
	32.81	15.63	10.16	4.473	2.226	1.298	0.5403
$2\pi/3$	31.38	18.7	11.35	6.821	4.191	2.48	1.331
	37.27	20.91	12.1	5.484	2.727	1.488	0.6591
	61.3	37.5	24.06	14	8.809	5.405	2.913
	28.91	18.73	10.77	6.615	3.98	2.462	1.292
	13.63	10.21	5.628	2.512	1.391	0.5251	0.3339
$5\pi/6$	23.37	14.65	8.531	4.624	3.099	1.852	0.9928
	26.7	14.9	7.701	3.438	1.965	1.071	0.4596
	46.94	28.91	17.79	9.954	6.574	3.976	2.175
	23.04	13.86	8.606	4.526	2.822	1.774	0.9463
	12.17	8.139	4.228	1.374	0.6954	0.4284	0.2312
π	30.1	18.7	11.78	6.086	3.965	2.472	1.318
	47.67	25.72	14.39	6.774	3.7	1.925	0.8479
	50.3	31.53	20.27	10.54	6.982	4.35	2.327
	27.15	16.78	10.68	5.93	3.502	2.386	1.242
	22.84	11.02	6.3	2.952	1.596	0.7991	0.3708

The footnote after Table III applies here.

while by Theorem 1 our RNB2 estimate has a bigger variance $24n^{-1}\log(n)\sigma^4(1+o(1))$.

Now we compare the time used for computing different TAVC estimates. We run the AR(1) model with known $\mathbb{E}X_1 = 0$ for 10^8 steps, updating RNB1 and NB at each step. Fig. 2 plots the time used against the number of steps. We see that, as expected, the time needed for our recursive estimate increases linearly, while the usual batched means algorithm is much slower.

We also compare the spectral density estimates at other frequencies for the previous two models. Since the MSE for the centered estimates are very similar, we only consider the models with known $\mathbb{E}X_i = 0$. We report the average MSE (of 1000 repetitions) in Tables III and IV and observe similar behavior of these estimates at all seven frequencies equally spaced on $[0, \pi]$.

Finally, we demonstrate the effect of bias reduction on (i) AR(1) model $X_i = 0.8X_{i-1} + \epsilon_i$; and (ii) bilinear model $X_i = (0.8 + 0.4\epsilon_i)X_{i-1} + \epsilon_i$. Again we let (ϵ_i) be iid $N(0,1)$ in both models. For RNB2, we now use $\lambda_1 = 6$ and $\lambda_2 = 2.5$ so that the condition $2\lambda_2 \log(\rho) \le -1$ of Theorem 2 (iii) is satisfied. We obtain estimates for 24 frequencies equally spaced on the interval $[0, \pi]$. For each method, we repeat ten times and record the average. Fig. 3 suggests that RNB2 has the smallest bias, confirming the effect of bias reduction. From the second row of Fig. 3, we see the bias of RNB2 and LW are similar and they are smaller than the other ones.

V. PROOFS

We first provide a complete proof for the special case $\theta = 0$ or $\varrho = 1$. The argument for general θ is similar; and in Section V-D, we shall point out the necessary changes.

As in [38], we shall apply m-dependence and martingale approximations. For $m \in \mathbb{N}$, we can approximate functionals of

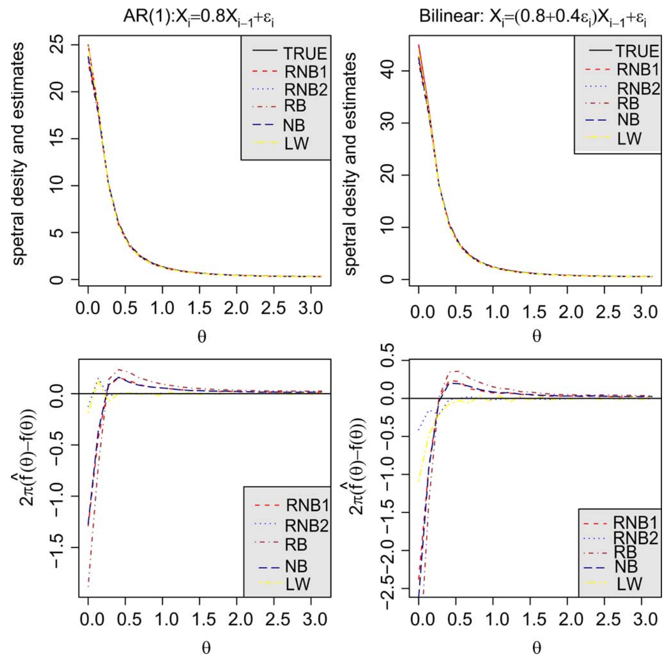

Fig. 3. Comparison of the centered spectral density estimates. The first (resp. second) column displays results for AR(1) model (resp. bilinear model). The first row gives the estimates and true spectral densities (all multiplied by 2π). The second row shows the distances from the true density.

the process (X_i) by the corresponding functionals of the m-dependence process

$$\tilde{X}_i = X_{i,m} := \mathbb{E}(X_i|\epsilon_{i-m},\ldots,\epsilon_i) = \mathbb{E}(X_i|\mathcal{F}_{i-m,i}) \qquad (27)$$

where $\mathcal{F}_{i-m,i} = \sigma(\epsilon_{i-m},\ldots,\epsilon_i)$ is the σ-field generated by $(\epsilon_{i-m},\ldots,\epsilon_i)$. For example, the functional

$$L_n := \frac{V_n - \sum_{i=1}^n X_i^2}{2} = \sum_{i=1}^n \left(X_i \sum_{j=l_i}^{i-1} X_j \right) \qquad (28)$$

can be approximated by

$$\tilde{L}_n = \sum_{i=1}^n \left(\tilde{X}_i \sum_{j=l_i}^{i-1} \tilde{X}_j \right).$$

Lemmas 6 and 7 provide error bounds of the m-dependence approximation. Lemma 8 gives a martingale approximation for $\tilde{L}_n$. Lemma 6 is essentially Lemma 1 in [38]. Lemmas 7 and 8 can be proved by using the method in the latter paper. For the sake of completeness, we present detailed proofs here.

As in the construction of X_i' in the definition of the physical dependence measure $\delta_p(i)$, we let $X_{i,\{k\}}$ be a coupled version of $X_i = g(\mathcal{F}_i)$ by replacing ε_k in $\mathcal{F}_i$ by ε_k'. If $k > j$, then $X_{j,\{k\}} = X_j$. Keep in mind that $X_{i,k}$ and $X_{i,\{k\}}$ represent different random variables. Similarly we define $\tilde{X}_{i,\{k\}}$.

Lemma 6: Assume $X_i \in \mathcal{L}^p$ for some $p > 1$ and $\mathbb{E}(X_i) = 0$. Let $\mathcal{C}_p = 18p^{3/2}(p-1)^{-1/2}$ and $p' = \min(2,p)$. Let $\beta_1, \beta_2, \ldots, \beta_n \in \mathbb{C}$. Then

$$\left\| \sum_{i=1}^n \beta_i (X_i - \tilde{X}_i) \right\|_p \le \mathcal{C}_p A_n \Theta_{m+1,p},$$

where

$$A_n = \left(\sum_{i=1}^n |\beta_i|^{p'} \right)^{1/p'}.$$

Remark 2: The lemma is still valid if we replace $\tilde{X}_i$ by 0 and the bound becomes $\mathcal{C}_p A_n \Theta_{0,p}$.

Lemma 7: Assume $\mathbb{E}X_i = 0$, $X_i \in \mathcal{L}^{2p}$ for some $p \ge 2$ and $\Theta_{0,2p} < \infty$. Then

$$\|L_n - \mathbb{E}L_n - (\tilde{L}_n - \mathbb{E}\tilde{L}_n)\|_p \le C_p \Theta_{0,2p} d_{m,2p} \Delta_s^{1/4} r_n^{1/2} \qquad (29)$$

where $s = s_n$ is the integer for which $a_s \le n < a_{s+1}$,

$$d_{m,q} = \sum_{i=0}^{\infty} \min(\delta_q(i), \Psi_{m+1,q}) \text{ and } r_n = \sum_{i=2}^n (i - l_i)^{1/2}.$$

Proof: Let $Z_i = \sum_{j=l_i}^{i-1} X_j$, $\tilde{Z}_{i-1} = \sum_{j=l_i}^{i-1} \tilde{X}_j$, then $L_n = \sum_{i=1}^n X_i Z_i$ and $\tilde{L}_n = \sum_{i=1}^n \tilde{X}_i \tilde{Z}_i$. Let $L_n^* = \sum_{i=2}^n X_i \tilde{Z}_{i-1}$. Similarly as $X_{i,\{k\}}$, we define $Z_{i,\{k\}}$ and $\tilde{Z}_{i,\{k\}}$ as the coupled version of Z_i and $\tilde{Z}_i$ respectively. By Minkowski's inequality

$$\begin{aligned}
&\|\mathcal{P}_k(L_n - L_n^*)\|_p \\
&\le \left\| \sum_{i=2}^n \left[X_i(Z_i - \tilde{Z}_i) - X_{i,\{k\}} \left(Z_{i,\{k\}} - \tilde{Z}_{i,\{k\}} \right) \right] \right\|_p \\
&\le \left\| \sum_{i=2}^n X_{i,\{k\}} \left[(Z_i - \tilde{Z}_i) - \left(Z_{i,\{k\}} - \tilde{Z}_{i,\{k\}} \right) \right] \right\|_p \\
&\quad + \sum_{i=2}^n \left\| (X_i - X_{i,\{k\}})(Z_i - \tilde{Z}_i) \right\|_p =: I_k + II_k. \qquad (30)
\end{aligned}$$

By [38, (3.3)], $\|\tilde{X}_i - \tilde{X}_{i,\{k\}}\|_{2p} = \tilde{\delta}_{2p}(i-k) \le \delta_{2p}(i-k)$. By Burkhölder's inequality, c.f [11, Lemma 1], $\|X_i - \tilde{X}_i\|_{2p} \le \mathcal{C}_{2p}\Psi_{m+1,2p}$. Thus $\|X_i - \tilde{X}_i - X_{i,\{k\}} + \tilde{X}_{i,\{k\}}\|_{2p} \le 2\mathcal{C}_{2p} \min(\delta_{2p}(i-k), \Psi_{m+1,2p})$. Here we let $\delta_{2p}(j) = 0$ if $j < 0$. By Remark 2, $\|\sum_{i=j+1}^{a_{t+1}-1} X_{i,\{k\}}\|_{2p} \le \mathcal{C}_{2p}(a_{t+1} - j - 1)^{1/2}\Theta_{0,2p}$. Therefore

$$\begin{aligned}
I_k &= \left\| \sum_{i=2}^n X_{i,\{k\}} \sum_{j=l_i}^{i-1} (X_j - \tilde{X}_j - X_{j,\{k\}} + \tilde{X}_{j,\{k\}}) \right\|_p \\
&= \left\| \sum_{j=1}^{n-1} (X_j - \tilde{X}_j - X_{j,\{k\}} + \tilde{X}_{j,\{k\}}) \sum_{i=j+1}^{(u_j-1)\wedge n} X_{i,\{k\}} \right\|_p \\
&= \left\| \sum_{t=1}^{s-1} \sum_{j=a_t}^{a_{t+1}-1} (X_j - \tilde{X}_j - X_{j,\{k\}} + \tilde{X}_{j,\{k\}}) \sum_{i=j+1}^{a_{t+1}-1} X_{i,\{k\}} \right. \\
&\quad \left. + \sum_{j=a_s}^{n} (X_j - \tilde{X}_j - X_{j,\{k\}} + \tilde{X}_{j,\{k\}}) \sum_{i=j+1}^{n} X_{i,\{k\}} \right\|_p \\
&\le \sum_{t=1}^{s-1} \sum_{j=a_t}^{a_{t+1}-1} \{ 2\mathcal{C}_{2p}^2 \min[\delta_{2p}(j-k), \Psi_{m+1,2p}] \\
&\qquad \times (a_{t+1} - j - 1)^{1/2} \Theta_{0,2p} \} \\
&\quad + \sum_{j=a_s}^{n-1} 2\mathcal{C}_{2p}^2 \min[\delta_{2p}(j-k), \Psi_{m+1,2p}] (n-j)^{1/2} \Theta_{0,2p}.
\end{aligned}$$

It follows that

$$\sum_{k=-\infty}^{n} I_k^2 \le C_p \Theta_{0,2p}^2 \Delta_s^{1/2} d_{m,2p} \times \sum_{k=-\infty}^{n} \left(\sum_{t=1}^{s-1} \sum_{j=a_t}^{a_{t+1}-1} \min(\delta_{2p}(j-k), \Psi_{m+1,2p})(a_{t+1}-j-1)^{1/2} + \sum_{j=a_s}^{n-1} \min(\delta_{2p}(j-k), \Psi_{m+1,2p})(n-j)^{1/2} \right)$$
$$\le C_p \Theta_{0,2p}^2 d_{m,2p}^2 \Delta_s^{1/2} \times \left(\sum_{t=1}^{s-1} \sum_{j=a_t}^{a_{t+1}-1} (a_{t+1}-j-1)^{1/2} + \sum_{j=a_s}^{n-1} (n-j)^{1/2} \right).$$

By Lemma 6, $\|Z_i - \tilde{Z}_i\|_{2p} \le \mathcal{C}_{2p}(i-l_i)^{1/2}\Theta_{m+1,2p}$. Therefore

$$\sum_{k=-\infty}^{n} II_k^2 \le \sum_{k=-\infty}^{n} \left(\sum_{i=2}^{n} \delta_{2p}(i-k) \mathcal{C}_{2p}(i-l_i)^{1/2} \Theta_{m+1,2p} \right)^2$$
$$\le C_p \Theta_{m+1,2p}^2 \sum_{k=-\infty}^{n} \Delta_s^{1/2} \Theta_{0,2p} \sum_{i=2}^{n} \left(\delta_{2p}(i-k)(i-l_i)^{1/2} \right)$$
$$\le C_p \Theta_{0,2p}^2 \Theta_{m+1,2p}^2 \Delta_s^{1/2} \sum_{i=2}^{n} (i-l_i)^{1/2}.$$

Putting these pieces together and noting that $\Theta_{m+1,2p} \le d_{m,2p}$, we have

$$\|L_n - \mathbb{E}L_n - (L_n^* - \mathbb{E}L_n^*)\|_p^2 \le \mathcal{C}_p^2 \sum_{k=-\infty}^{n} \|\mathcal{P}_k(L_n - L_n^*)\|_p^2 \le C_p \Theta_{0,2p}^2 d_{m,2p}^2 \Delta_s^{1/2} r_n.$$

Using a similar argument, the same upper bound can be derived for $\|L_n^* - \mathbb{E}L_n^* - (\tilde{L}_n - E\tilde{L}_n)\|_p^2$ and hence the proof is complete. ■

Remark 3: By the dominated convergence theorem, the condition $\Theta_{0,q} < \infty$ implies $\lim_{m\to\infty} d_{m,q} = 0$. This fact is useful in the proof of Theorem 3.

Lemma 8: Assume $\mathbb{E}X_i = 0$, $X_i \in \mathcal{L}^4$ and $\Theta_{0,4} < \infty$. Recall the projection operator defined in (23). Let

$$D_i = \sum_{t=0}^{\infty} \mathcal{P}_i \tilde{X}_{i+t} = \sum_{t=0}^{m} \mathcal{P}_i \tilde{X}_{i+t} \tag{31}$$

and $M_n = \sum_{i=1}^n D_i \sum_{j=l_i}^{i-1} D_j$. Then

$$\|\tilde{L}_n - \mathbb{E}\tilde{L}_n - M_n\| \le C m^{3/2} n^{1/2} \|X_0\|_4^2.$$

Proof: Observe that $\tilde{X}_k = \sum_{t=0}^m \mathcal{P}_{k-t}\tilde{X}_k$. We have

$$\left\| \sum_{j=a}^{b} (\tilde{X}_j - D_j) \right\|_4 = \left\| \sum_{j=a}^{b} \left(\sum_{t=0}^{m} \mathcal{P}_{j-t}\tilde{X}_j - \sum_{t=0}^{m} \mathcal{P}_j \tilde{X}_{j+t} \right) \right\|_4$$
$$= \left\| \sum_{t=0}^{m-1} \mathbb{E}(\tilde{X}_{a+t}|\mathcal{F}_{a-1}) - \sum_{t=0}^{m-1} \mathbb{E}(\tilde{X}_{b+t}|\mathcal{F}_{b-1}) \right\|_4$$
$$\le 2m\|X_0\|_4. \tag{32}$$

By Minkowski's inequality

$$\|\tilde{L}_n - \mathbb{E}\tilde{L}_n - M_n\| \le \left\| \sum_{i=1}^{n} (W_i - \mathbb{E}W_i) \right\| + \left\| \sum_{j=1}^{n} (W_j' - \mathbb{E}W_j') \right\|, \tag{33}$$

where

$$W_i = \tilde{X}_i \sum_{j=l_i}^{i-1} (\tilde{X}_j - D_j) \text{ and } W_j' = D_j \sum_{i=j+1}^{(u_j-1)\wedge n} (\tilde{X}_i - D_i).$$

We break the sum in W_i into two parts as $W_i = W_{1,i} + W_{2,i}$, where

$$W_{1,i} = \tilde{X}_i \sum_{j=l_i}^{i-2m} (\tilde{X}_j - D_j),$$
$$W_{2,i} = \tilde{X}_i \sum_{j=(i-2m+1)\vee l_i}^{i-1} (\tilde{X}_j - D_j).$$

Note that $W_{1,i}, W_{1,i+2m}, W_{1,i+4m}, \ldots$ are martingale differences. By (32), $\|W_{1,i}\| \le 2m\|X_0\|_4^2$. Therefore

$$\left\| \sum_{i=1}^{n} W_{1,i} \right\| \le \sum_{i=1}^{2m} \left\| \sum_{j=0}^{\lfloor (n-i)/(2m) \rfloor} W_{1,i+2mj} \right\| \le C m^{3/2} \|X_0\|_4^2 \sqrt{n}. \tag{34}$$

By (32), $\|W_{2,i}\| \le 2m\|X_0\|_4^2$. Since $(W_{2,i})$ is $(3m-1)$-dependent, similarly as (34), we have

$$\left\| \sum_{i=1}^{n} (W_{2,i} - \mathbb{E}W_{2,i}) \right\| \le C m^{3/2} \|X_0\|_4^2 \sqrt{n}. \tag{35}$$

By (34) and (35), $\|\sum_{i=1}^n (W_i - \mathbb{E}W_i)\| \le C m^{3/2} \|X_0\|_4^2 \sqrt{n}$. For the term W_j', write $W_j' = W_{1,j}' + W_{2,j}'$, where $W_{1,j}' = D_j \sum_{i=j+2m}^{u_j-1} (\tilde{X}_i - D_i)$ and $W_{2,j}' = D_j \sum_{i=j+1}^{j+2m-1} (\tilde{X}_i - D_i)$. Observe that $(W_{1,j-2km}')_{k\ge 0}$ is a martingale difference sequence with respect to the filtration $(\mathcal{F}_{j-(2k+1)m,\infty})_{k\ge 0}$. Similarly as (34) and (35), we have $\|\sum_{i=1}^n (W_i' - \mathbb{E}W_i')\| \le C m^{3/2} \|X_0\|_4^2 \sqrt{n}$. Hence by (33), Lemma 8 follows. ■

A. MSE

Proof of Theorem 1: (i) We first calculate the order of the variance of M_n. Note that M_n is a quadratic form of martingale differences and

$$\|M_n\|^2 = \sum_{i=1}^{n} \mathbb{E} \left(D_i \sum_{j=l_i}^{i-1} D_j \right)^2$$
$$= \sum_{i=1}^{n} \sum_{j=l_i}^{i-1} \mathbb{E}(D_i D_j)^2 + 2 \sum_{i=1}^{n} \sum_{l_i \le k < j \le i-1} \mathbb{E}(D_i^2 D_j D_k). \tag{36}$$

Since (D_k) is m-dependent, $\mathbb{E}(D_i^2 D_j D_k) = 0$ if $i - j > m$ or $j - k > m$ and $\mathbb{E}(D_i D_j)^2 = \|D_1\|^4$ if $i - j > m$. Hence

$$\sum_{i=1}^n (i - l_i)\|D_1\|^4 - \sum_{i=1}^n m\|D_1\|_4^4 - \sum_{i=1}^n 2m^2\|D_1\|_4^4 \le \|M_n\|_n^2 \le \sum_{i=1}^n (i - l_i)\|D_1\|^4 + \sum_{i=1}^n m\|D_1\|_4^4 + \sum_{i=1}^n 2m^2\|D_1\|_4^4. \quad (37)$$

Let $\hat{\gamma}_0 = \sum_{j=1}^n X_j^2/n$. Similarly as (A.1) in [11], $\|\hat{\gamma}_0 - \gamma_0\| \le Cn^{-1/2}\|X_0\|_4\Theta_{0,4}$. Now combine the results in Lemmas 7 and 8, observe that $\|D_1\| \le \Theta_{0,4}$, we have for each fixed m, when n is large enough

$$\left| \|V_n - \mathbb{E}V_n\| - \nu_n\sigma^2 \right| \le \nu_n \left| \|D_1\|^2 - \sigma^2 \right| + C\left(d_{m,4}\Delta_s^{1/4} r_n^{1/2} + \sqrt{n}m^{3/2}\right). \quad (38)$$

Note that $\|D_1\|^2 = \tilde{\sigma}^2$, where $\tilde{\sigma}^2$ is the TAVC of the sequence $(\tilde{X}_i)$ and that $\tilde{\sigma}^2 \to \sigma^2$ as $m \to \infty$. Since $r_n\Delta_s^{1/2}$ and ν_n^2 have the same order and $d_{m,2p} \to 0$ as $m \to \infty$, (i) follows by dividing both sides of (38) by $\nu_n\sigma^2$, then taking the limit with respect to n and then with respect to m.

(ii) The convergence rate of the variance of V_n° can be proved by the same argument as part (i). Details are omitted. ■

Proof of Theorem 2: (i) Let $\Gamma_k = \sum_{j=k}^\infty |\gamma_j|$. Again we write s for s_n. By (21), we know $\Gamma_k = o(k^{-q})$ and hence

$$|\mathbb{E}V_n - n\sigma^2| \le 2\sum_{t=1}^s \left(\Delta_s\Gamma_{\Delta_s} + \sum_{j=1}^{\Delta_s - 1} j|\gamma_j|\right) = \begin{cases} o(s\Delta_s^{1-q}), & \text{if } 0 < q < 1 \\ O(s), & \text{if } q = 1. \end{cases}$$

When $a_k = \lfloor ck^p \rfloor$, $\Delta_s = O(n^{1-1/p})$ and $s = O(n^{1/p})$. So the other two assertions of (i) follows from Theorem 1.

(ii) Let $\ell_k = \sum_{j=k}^\infty j^q\gamma_j$, then $\ell_k = o(1)$ and $\Gamma_k \le \ell_k k^{-q}$.

$$|\mathbb{E}V_n^\circ - v_n\sigma^2| \le 2\sum_{t=1}^s \left(\sum_{i=a_t+d_t}^{a_{t+1}-1} \Gamma_{i-a_t+1}\right) \le 2\sum_{t=1}^s \left(\sum_{k=d_t+1}^{\Delta_t} \ell_k k^{-q}\right) \le C\sum_{t=1}^s \ell_{d_t}(d_t)^{1-q}.$$

When $a_k = \lfloor c_1k^p \rfloor$ and $d_k = \lfloor \Delta_k/c_2 \rfloor$ with some $c_2 > 1$, we have $d_k \asymp k^{p-1}$ and therefore

$$|\mathbb{E}V_n^\circ - v_n\sigma^2| \le C\sum_{t=1}^s \ell_{d_t} t^{(p-1)(1-q)} = \begin{cases} o(n^{(p-1)(1-q)/p + 1/p}), & \text{if } (p-1)(1-q) > -1 \\ o(\log s) = o(\log n), & \text{if } (p-1)(1-q) = -1 \\ O(1) & \text{if } (p-1)(1-q) < -1 \end{cases} \quad (39)$$

(iii) Note that (22) also implies $\Gamma_k = O(\rho^k)$ and hence

$$|\mathbb{E}V_n^\circ - v_n\sigma^2| \le 2\sum_{t=1}^s \left(\sum_{i=a_t+d_t}^{a_{t+1}-1} \Gamma_{i-a_t+1}\right) \le 2C\sum_{t=1}^s \rho^{d_t+1}.$$

When $a_k = \lfloor ck^p \rfloor$ and $d_k = \lfloor \lambda\log(k) \rfloor$, we have $\Delta_k \asymp k^{p-1}$, which by Theorem 1 (ii) implies $\|V_n^\circ - \mathbb{E}V_n^\circ\|^2 = O(n^{2-1/p})$. Since $s = O(n^{1/p})$

$$|\mathbb{E}V_n^\circ - v_n\sigma^2|^2 \le \left(2C\sum_{t=1}^s \rho^{d_t+1}\right)^2 \le 4C^2\left(\sum_{t=1}^s t^{\lambda\log(\rho)}\right)^2 = \begin{cases} O(n^{2/p + 2\lambda\log(\rho)/p}), & \text{if } \lambda\log(\rho) > -1 \\ O(\log(s)^2) = O\left[\log(n)^2\right], & \text{if } \lambda\log(\rho) = -1 \\ O(1) & \text{if } \lambda\log(\rho) < -1 \end{cases} \quad (40)$$

(iv) For $a_k = \lfloor \lambda_1 k\log(k) \rfloor$ and $d_k = \lfloor \lambda_2\log(k) \rfloor$ with $\lambda_1 > \lambda_2 > 0$, we have $\Delta_k \asymp \lambda_1\log(k)$. So the condition of Theorem 1 (ii) is still satisfied. The bias can be calculated similarly as part (ii), with the only difference being that now $s_n = O(n)$. Hence, the conclusions in (iii) follow. ■

Remark 4: In general, the bounds on the bias terms in Theorem 2 cannot be improved. For example, consider an AR(1) process with $\gamma_0 = 1$ and $\gamma_1 = \rho$, where $0 < \rho < 1$. Elementary calculations show that

$$|\mathbb{E}V_n^\circ - v_n\sigma^2| \ge 2\sum_{t=1}^{s-1}\left(\sum_{i=a_t+d_t}^{a_{t+1}-1} \Gamma_{i-a_t+1}\right) \ge \frac{2\rho}{1-\rho}\sum_{t=1}^{s-1}\rho^{d_t} \ge \frac{2\rho}{1-\rho}\sum_{t=1}^{s-1} t^{\lambda\log(\rho)}.$$

So the bounds in part (iii) cannot be improved, see (40). Similar claims can be made for other cases of Theorem 2.

B. Central Limit Theorems

Proof of Theorem 3: As mentioned in the proof of Theorem 1, $\|\hat{\gamma}_0 - \gamma_0\| = O(n^{-1/2})$. Hence, by (28), it suffices to prove that $2L_n/\nu_n \Rightarrow N(0, \sigma^4)$. We shall apply the argument in [38]. By Lemma 7, Remark 3 and Lemma 8 (recall that D_1 and M_n depend implicitly on m),

$$\limsup_{m\to\infty}\limsup_{n\to\infty}\left\|\frac{L_n - M_n}{\nu_n}\right\| = 0.$$

Recall that $\|D_1\|^2 = \tilde{\sigma}^2 \to \sigma^2$ as $m \to \infty$, where $\tilde{\sigma}^2$ is the TAVC of the sequence $(\tilde{X}_i)$. Therefore it remains to show that, for every m,

$$\frac{2M_n}{\nu_n} \Rightarrow N(0, \|D_1\|^4). \quad (41)$$

Since $\|\sum_{i=1}^n D_i \sum_{j=(i-2m+1)\vee l_i}^{i-1} D_j\| = O(\sqrt{n})$, the problem is reduced to

$$\frac{2}{\nu_n}\sum_{i=1+2m}^n D_iY_i \Rightarrow N(0, \|D_1\|^4), \text{ where } Y_i = \sum_{j=l_i}^{i-2m} D_j. \quad (42)$$

Note that D_i and Y_i are independent. By Burkhölder's inequality

$$\sum_{i=1+2m}^{n} \|D_i Y_i\|_4^4 \le C \sum_{i=1}^{n} (i-l_i)^2 = o(\nu_n^4). \tag{43}$$

Hence, by the martingale central limit theorem c.f [40], it suffices to verify

$$\frac{4}{\nu_n^2} \sum_{i=1+2m}^{n} \mathbb{E}[(D_i Y_i)^2 | \mathcal{F}_{i-1}] \to \|D_1\|^4 \text{ in probability.} \tag{44}$$

For any $-m \le l \le -1$, since

$$\begin{aligned}\left\| \sum_{i=1+2m}^{n} \mathcal{P}_{i+l}(D_i Y_i)^2 \right\|^2 &= \sum_{i=1+2m}^{n} \|\mathcal{P}_{i+l}(D_i Y_i)^2\|^2 \\ &\le 2 \sum_{i=1+2m}^{n} \|D_i\|_4^4 \|Y_i\|_4^4.\end{aligned}$$

we have $\|\sum_{i=1+2m}^n \mathcal{P}_{i+l}(D_i Y_i)^2\| = o(\nu_n^2)$ by (43). Note that $\mathbb{E}(D_i^2 Y_i^2 | \mathcal{F}_{i-m-1}) = Y_i^2 \mathbb{E}(D_i^2)$, (44) is therefore reduced to

$$\frac{4}{\nu_n^2} \sum_{i=1+2m}^{n} Y_i^2 \to \|D_1\|^2 \text{ in probability.} \tag{45}$$

Let $a^+ = a \vee 0$ for any real number a. Since

$$\begin{aligned}\frac{4}{\nu_n^2} \sum_{i=1+2m}^{n} \mathbb{E} Y_i^2 &= \frac{4}{\nu_n^2} \sum_{i=1+2m}^{n} (i-2m-l_i)^+ \|D_1\|^2 \\ &\to \|D_1\|^2\end{aligned} \tag{46}$$

it suffices to show $\nu_n^{-2} \sum_{i=1+2m}^n (Y_i^2 - \mathbb{E} Y_i^2)$ converges to zero in probability. We calculate its $\mathcal{L}^2$ norm,

$$\begin{aligned}\left\| \sum_{i=1+2m}^{n} (Y_i^2 - \mathbb{E} Y_i^2) \right\| &\le \left\| \sum_{i=1+2m}^{n} \sum_{j=l_i}^{i-2m} \left(D_j^2 - \mathbb{E} D_j^2\right) \right\| \\ &+ \left\| 2 \sum_{i=1+2m}^{n} \sum_{l_i \le j < k \le i-2m} D_j D_k \right\| =: A_n + B_n.\end{aligned}$$

It is clear that $B_n = o(\nu_n^2)$. Using a similar argument as (36), we have $A_n = o(\nu_n^2)$. It follows that $\|\sum_{i=1+2m}^n (Y_i^2 - \mathbb{E} Y_i^2)\| = o(\nu_n^2)$ and the proof is complete.

(ii) The CLT for V_n° can be proved similarly. ■

C. Centered Estimates

Proof of Corollary 5: We only give a proof for the estimate V_n', since the one for $V_n'^\circ$ is similar. Write $F_n = \sum_{i=1}^n \alpha_i X_i$, where $\alpha_i = \Delta_k$ if $a_k \le i < a_{k+1}$. By Lemma 6

$$\|F_n\|_4 \le C \Theta_{0,4} \left(\sum_{i=1}^{n} \alpha_i^2 \right)^{1/2} = O(\sqrt{n} \Delta_s)$$

and $\|\bar{X}_n\|_{2p} = O(n^{-1/2})$. It follows that

$$\|F_n \bar{X}_n\| = O(\Delta_s) \text{ and } \|q_n \bar{X}_n^2\| = O(\Delta_s). \tag{47}$$

Since $\Delta_s = o(n)$, $\|F_n \bar{X}_n\|$ and $\|q_n \bar{X}_n^2\|$ have smaller orders than $\|V_n - \mathbb{E} V_n\|$. Therefore, Corollary 5 follows. ■

D. Proofs for Other Frequency $\theta \in (0, \pi)$

The basic idea is to replace every X_i by $X_i \varrho^i$ in the proofs for $\theta = 0$. Specifically, Lemma 7 still holds with

$$L_n = \sum_{i=1}^{n} \left(X_i \varrho^i \sum_{j=l_i}^{i-1} X_j \varrho^{-j} \right).$$

Lemma 8 holds with $M_n = \sum_{i=1}^n D_i \varrho^i \sum_{j=l_i}^{i-1} D_j \varrho^{-j}$, where $D_i = \sum_{t=0}^m \mathcal{P}_i \tilde{X}_{i+t} \varrho^t$. Observe that (D_i) is an m-dependent and stationary martingale difference sequence and $\|D_i\|_4 \le \Theta_{0,4}$. A careful check of the proof of Theorem 1 with $\theta = 0$ implies that it holds for general θ by noting that:

(i) $\mathbb{E}|D_1|^2 = 2\pi \tilde{f}(\theta)$ for any θ, where $\tilde{f}(\theta)$ is the spectral density of the sequence $(\tilde{X}_i)$.

(ii) For $\theta \ne 0$, $V_n = \sum_{i=1}^n X_n^2 + L_n + \bar{L}_n$, where $\bar{\cdot}$ denotes the complex conjugate. Thus we need to calculate $\|M_n + \bar{M}_n\|^2$ instead of $\|M_n\|^2$. Let $Y_i = \sum_{j=l_i}^{i-1} \varrho^{i-j} D_j$, similarly as (36) and (37), we have

$$\begin{aligned}\|M_n + \bar{M}_n\|^2 &= \sum_{i=1}^{n} \mathbb{E}\left(D_i \bar{Y}_i + \bar{D}_i Y_i\right)^2 \\ &= 2 \sum_{i=1}^{n} \|D_i\|^2 \|Y_i\|^2 + \sum_{i=1}^{n} \left(\mathbb{E} D_i^2 \mathbb{E} \bar{Y}_i^2 + \mathbb{E} \bar{D}_i^2 \mathbb{E} Y_i^2\right) \\ &\quad + nm^2 \|D_1\|_4^4 O(1).\end{aligned} \tag{48}$$

Observe that $\|Y_i\|^2 = \sum_{j=l_i}^{i-1} \|D_j\|^2$ and $\mathbb{E} Y_i^2 = \sum_{j=l_i}^{i-1} \varrho^{2(i-j)} \mathbb{E} D_j^2$. Since $|\sum_{j=l_i}^{i-1} \varrho^{2(i-j)}| \le 1/|\sin(\theta)|$, we have $\sup_i |\mathbb{E} Y_i^2| = O(1)$. So Theorem 1 (i) holds for any θ by the same argument as $\theta = 0$. Theorem 1 (ii) can be shown similarly. Theorem 2 follows from Theorem 1.

Remark 5: Observe that the bounds in Lemmas 7 and 8 are uniform over $\theta \in \mathbb{R}$. By (48)), $\|M_n + \bar{M}_n\| \le \sqrt{2} \nu_n \|D_1\|^2 + C\sqrt{n} m$. Then, by elementary manipulations and Lemmas 7 and 8, we have the uniform upper bound $\sup_{\theta \in \mathbb{R}} \|V_n - \mathbb{E} V_n\| = O(\nu_n)$. Since all the orders of the squared bias in Theorem 2 are uniform over $\theta \in \mathbb{R}$, all the upper bounds of the MSE in that theorem are also uniform. □

Now we consider the central limit theorem for $\theta \in (0, \pi]$. The proof for $\theta = 0$ works in general with the following modifications.

(i) Equation (41) becomes $(M_n + \bar{M}_n)/\nu_n \Rightarrow N(0, \|D_1\|^4)$.

(ii) Let $Y_i = \sum_{j=l_i}^{i-2m} \varrho^{i-j} D_j$ (46) becomes

$$\begin{aligned}\frac{1}{\nu_n^2} \sum_{i=1+2m}^{n} &\left(\mathbb{E} D_i^2 \mathbb{E} \bar{Y}_i^2 + \mathbb{E} \bar{D}_i^2 \mathbb{E} Y_i^2 + 2\mathbb{E}|D_i|^2 \mathbb{E}|Y_i|^2\right) \\ &\to \|D_1\|^4.\end{aligned} \tag{49}$$

Since $\mathbb{E} Y_i^2 = \sum_{j=l_i}^{i-1} \varrho^{2(i-j)} \mathbb{E} D_j^2$, (49) is obvious when $\theta/\pi \in \mathbb{Z}$. When $\theta/\pi \notin \mathbb{Z}$, (49) follows by noting that $|\sum_{j=l_i}^{i-1} \varrho^{2(i-j)}| \le 1/|\sin(\theta)|$.

Finally, since (47) holds uniformly for $\theta \in [0, \pi]$, the results for centered estimates of $f(0)$ also hold for $f(\theta)$.

ACKNOWLEDGMENT

We thank two anonymous referees and an Associate Editor for their valuable suggestions and comments.

REFERENCES

[1] M. S. Bartlett, “On the theoretical specification and sampling properties of autocorrelated time-series,” *Suppl. J. Roy. Statist. Soc.*, vol. 8, pp. 27–41, 1946.

[2] E. Parzen, “On consistent estimates of the spectral density of a stationary time series,” in *Proc. Nat. Acad. Sci.*, 1956, vol. 42, pp. 154–157.

[3] U. Grenander and M. Rosenblatt, *Statistical Analysis of Stationary Time Series*. New York: Wiley, 1957.

[4] R. B. Blackman and J. W. Tukey, *The Measurement of Power Spectra: From the Point of View of Communications Engineering*. New York: Dover, 1959.

[5] C. W. J. Granger, *Spectral Analysis of Economic Time Series*, ser. in Association With M. Hatanaka. Princeton Studies in Mathematical Economics. Princeton, NJ: Princeton University Press, 1964.

[6] G. M. Jenkins and D. G. Watts, *Spectral Analysis and its Applications*. San Francisco, CA: Holden-Day, 1968.

[7] M. B. Priestley, *Spectral Analysis and Time Series*. London, U.K.: Academic Press, vol. 1.

[8] M. Rosenblatt, *Stationary Sequences and Random Fields*. Boston, MA: Birkhäuser, 1985.

[9] D. Thomson, “Spectrum estimation and harmonic analysis,” *Proc. IEEE*, vol. 70, no. 9, pp. 1055–1096, Sep. 1982.

[10] B. Buttkus, *Spectral Analysis and Filter Theory in Applied Geophysics, ser.* Berlin, Germany: Technology Press Research Monographs, Springer, 2000.

[11] W. B. Wu and X. Shao, “A limit theorem for quadratic forms and its applications,” *Econom. Theory*, vol. 23, no. 5, pp. 930–951, 2007.

[12] P. D. Welch, “The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms,” *IEEE Trans. Audio and Electroacoust.*, vol. AU-15, pp. 70–73, 1967.

[13] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*, 2nd ed. New York: Springer-Verlag, ser. Springer Series in Statistics, 1991.

[14] I. A. Ibragimov and Y. V. Linnik, *Independent and Stationary Sequences of Random Variables*. Groningen, The Netherlands: Wolters-Noordhoff, 1971.

[15] W. B. Wu, “Nonlinear system theory: Another look at dependence,” *Proc. Nat. Acad. Sci. USA*, vol. 102, no. 40, pp. 14150–14154, 2005.

[16] R. C. Bradley, *Introduction to Strong Mixing Conditions*. Heber, UT: Kendrick Press, 2007, vol. 1.

[17] B. Schmeiser, “Batch size effects in the analysis of simulation output,” *Oper. Res.*, vol. 30, no. 3, pp. 556–568, 1982.

[18] C. Alexopoulos and D. Goldsman, “To batch or not to batch?,” *ACM Trans. Model. Comput. Simul.*, vol. 14, no. 1, pp. 76–114, 2004.

[19] C. Alexopoulos, D. Goldsman, and R. F. Serfozo, “Stationary processes: Statistical estimation,” in *Encyclopedia of Statistical Sciences*, B. V. N. Balakrishnan and C. B. Read, Eds., 2nd ed. Hoboken, NJ: Wiley Interscience, 2006, vol. 12, pp. 7991–8006.

[20] G. M. Clark, “Recursive estimation of the variance of the sample average,” *ACM Trans. Math. Softw.*, vol. 6, no. 1, pp. 58–67, 1980.

[21] W. B. Wu, “Recursive estimation of time-average variance constants,” *Ann. Appl. Probab.*, vol. 19, no. 4, pp. 1529–1552, 2009.

[22] H. Damerdji, “Strong consistency of the variance estimator in steady-state simulation output analysis,” *Math. Oper. Res.*, vol. 19, no. 2, pp. 494–512, 1994.

[23] G. L. Jones, M. Haran, B. S. Caffo, and R. Neath, “Fixed-width output analysis for Markov chain Monte Carlo,” *J. Amer. Statist. Assoc.*, vol. 101, no. 476, pp. 1537–1547, 2006.

[24] W. T. Song and B. W. Schmeiser, “Optimal mean-squared-error batch sizes,” *Manag. Sci.*, vol. 41, no. 1, pp. 110–123, 1995.

[25] C. Chien, D. Goldsman, and B. Melamed, “Large-sample results for batch means,” *Manag. Sci.*, vol. 43, no. 9, pp. 1288–1295, 1997.

[26] N. Wiener, *Nonlinear Problems in Random Theory*. New York: Technology Press: Wiley/M.I.T., 1958.

[27] M. B. Priestley, *Nonlinear and Nonstationary Time Series Analysis*. London, U.K.: Academic/ Harcourt Brace Jovanovich, 1988.

[28] M. Rosenblatt, *Markov Processes Structure and Asymptotic Behavior*. New York: Springer-Verlag, 1971.

[29] H. Tong, *Nonlinear Time Series, a Dynamical System Approach, ser. Oxford Statistical Science Series*. New York: Clarendon/Oxford University, 1990, vol. 6.

[30] T. W. Anderson, *The Statistical Analysis of Time Series*. New York: Wiley, 1971.

[31] E. Carlstein, “The use of subseries values for estimating the variance of a general statistic from a stationary sequence,” *Ann. Statist.*, vol. 14, no. 3, pp. 1171–1179, 1986.

[32] X. Shao and W. B. Wu, “Asymptotic spectral theory for nonlinear time series,” *Ann. Statist.*, vol. 35, no. 4, pp. 1773–1801, 2007.

[33] T. Mikosch, “Functional limit theorems for random quadratic forms,” *Stochast. Process. Appl.*, vol. 37, no. 1, pp. 81–98, 1991.

[34] A. Basalykas, “Functional limit theorems for random multilinear forms,” *Stochast. Process. Appl.*, vol. 53, no. 1, pp. 175–191, 1994.

[35] F. Götze and A. N. Tikhomirov, “Asymptotic distribution of quadratic forms,” *Ann. Probab.*, vol. 27, no. 2, pp. 1072–1098, 1999.

[36] D. N. Politis, J. P. Romano, and M. Wolf, *Subsampling*, ser. Statistics. New York: Springer-Verlag, 1999.

[37] S. N. Lahiri, *Resampling Methods for Dependent Data*, ser. Statistics. New York: Springer-Verlag, 2003.

[38] W. Liu and W. B. Wu, “Asymptotics of spectral density estimates,” *Econom. Theory*, vol. 26, no. 4, pp. 1218–1245, 2010.

[39] D. L. Burkholder, “Distribution function inequalities for martingales,” *Ann. Probabil.*, vol. 1, pp. 19–42, 1973.

[40] P. Hall and C. C. Heyde, *Martingale Limit Theory and its Application*. New York: Academic.

Han Xiao received the B.S. degree in mathematics from Peking University, Beijing, China, in 2003 and the M.Sc. degree in statistics from National University of Singapore in 2006. He is currently pursuing the Ph.D. degree at the Department of Statistics, The University of Chicago, Chicago, IL.

His research interests include nonlinear and nonstationary time series, algebraic statistics, and random matrix theory.

Wei Biao Wu received the Ph.D. degree from the University of Michigan, Ann Arbor, in 2001.

He was an Assistant Professor from 2001 to 2008 and currently is an Associate Professor of the Department of Statistics, The University of Chicago. His research interests include probability theory, statistics, and econometrics.