The Annals of Statistics
2012, Vol. 40, No. 1, 466–493
DOI: 10.1214/11-AOS967

COVARIANCE MATRIX ESTIMATION FOR STATIONARY TIME SERIES[1]

BY HAN XIAO AND WEI BIAO WU

University of Chicago

We obtain a sharp convergence rate for banded covariance matrix estimates of stationary processes. A precise order of magnitude is derived for spectral radius of sample covariance matrices. We also consider a thresholded covariance matrix estimator that can better characterize sparsity if the true covariance matrix is sparse. As our main tool, we implement Toeplitz [*Math. Ann.* **70** (1911) 351–376] idea and relate eigenvalues of covariance matrices to the spectral densities or Fourier transforms of the covariances. We develop a large deviation result for quadratic forms of stationary processes using m-dependence approximation, under the framework of causal representation and physical dependence measures.

1. Introduction. One hundred years ago, in 1911, Toeplitz obtained a deep result on eigenvalues of infinite matrices of the form $\Sigma_\infty = (a_{s-t})_{s,t=-\infty}^{\infty}$. We say that λ is an eigenvalue of Σ_∞ if the matrix $\Sigma_\infty - \lambda \operatorname{Id}_\infty$ does not have a bounded inverse, where Id_∞ denotes the infinite-dimensional identity matrix. Toeplitz proved that, interestingly, the set of eigenvalues is the same as the image set $\{g(\theta), \theta \in [0, 2\pi]\}$, where

$$g(\theta) = \sum_{t\in\mathbb{Z}} a_t e^{it\theta} \qquad \text{with } i = \sqrt{-1}. \tag{1}$$

Note that $g(\theta)$ is the Fourier transform of the sequence $(a_t)_{t=-\infty}^{\infty}$. For a finite $T \times T$ matrix $\Sigma_T = (a_{s-t})_{1\le s,t\le T}$, its eigenvalues are approximately equally distributed (in the sense of Hermann Weyl) as $\{g(\omega_s), s = 0, \ldots, T-1\}$, where $\omega_s = 2\pi s/T$ are the Fourier frequencies. See the excellent monograph by Grenander and Szegö (1958) for a detailed account.

Covariance matrix is of fundamental importance in many aspects of statistics including multivariate analysis, principal component analysis, linear discriminant analysis and graphical modeling. One can infer dependence structures among variables by estimating the associated covariance matrices. In the context of stationary time series analysis, due to stationarity, the covariance matrix is Toeplitz in that,

Received May 2011; revised December 2011.

[1]Supported in part by NSF Grants DMS-09-06073 and DMS-11-06970.

MSC2010 subject classifications. Primary 62M10; secondary 62H12.

Key words and phrases. Autocovariance matrix, banding, large deviation, physical dependence measure, short range dependence, spectral density, stationary process, tapering, thresholding, Toeplitz matrix.

along the off-diagonals that are parallel to the main diagonal, the values are constant. Let $(X_t)_{t\in\mathbb{Z}}$ be a stationary process with mean $\mu = \mathbb{E}X_t$, and denote by $\gamma_k = \mathbb{E}[(X_0 - \mu)(X_k - \mu)]$, $k \in \mathbb{Z}$, its autocovariances. Then

$$\Sigma_T = (\gamma_{s-t})_{1\le s,t\le T} \tag{2}$$

is the autocovariance matrix of $(X_1, \dots, X_T)$. In the rest of the paper for simplicity we also call (2) the covariance matrix of $(X_1, \dots, X_T)$. In time series analysis it plays a crucial role in prediction [Kolmogoroff (1941), Wiener (1949)], smoothing and best linear unbiased estimation (BLUE). For example, in the Wiener–Kolmogorov prediction theory, one predicts X_{T+1} based on past observations $X_T, X_{T-1}, \dots$. If the covariances γ_k were known, given observations $X_1, \dots, X_T$, the coefficients of the best linear unbiased predictor $\hat{X}_{T+1} = \sum_{s=1}^{T} a_{T,s} X_{T+1-s}$ in terms of the mean square error $\|X_{T+1} - \hat{X}_{T+1}\|^2$ are the solution of the discrete Wiener–Hopf equation

$$\Sigma_T \mathbf{a}_T = \boldsymbol{\gamma}_T,$$

where $\mathbf{a}_T = (a_{T,1}, a_{T,2}, \dots, a_{T,T})^\top$ and $\boldsymbol{\gamma}_T = (\gamma_1, \gamma_2, \dots, \gamma_T)^\top$, and we use the superscript $\top$ to denote the transpose of a vector or a matrix. If γ_k are not known, we need to estimate them from the sample $X_1, \dots, X_T$, and a good estimate of Σ_T is required. As another example, suppose now $\mu = \mathbb{E}X_t \neq 0$ and we want to estimate it from $X_1, \dots, X_T$ by the form $\hat{\mu} = \sum_{t=1}^{T} c_t X_t$, where c_t satisfy the constraint $\sum_{t=1}^{T} c_t = 1$. To obtain the BLUE, one minimizes $\mathbb{E}(\hat{\mu} - \mu)^2$ subject to $\sum_{t=1}^{T} c_t = 1$, ensuring unbiasedness. Note that the usual choice $c_t \equiv 1/T$ may not lead to BLUE. The optimal coefficients are given by $(c_1, \dots, c_T)^\top = (\mathbf{1}^\top \Sigma_T^{-1} \mathbf{1})^{-1} \Sigma_T^{-1} \mathbf{1}$, where $\mathbf{1} = (1, \dots, 1)^\top \in \mathbb{R}^T$; see Adenstedt (1974). Again a good estimate of Σ_T^{-1} is needed.

Given observations $X_1, X_2, \dots, X_T$, assuming at the outset that $\mathbb{E}X_t = 0$, we can naturally estimate Σ_T via plug-in by the sample version

$$\hat{\Sigma}_T = (\hat{\gamma}_{s-t})_{1\le s,t\le T} \qquad \text{where } \hat{\gamma}_k = \frac{1}{T} \sum_{t=|k|+1}^{T} X_{t-|k|} X_t. \tag{3}$$

To judge the quality of a matrix estimate, we use the operator norm. The term "operator norm" usually indicates a class of matrix norms; in this paper it refers to the ℓ_2/ℓ_2 operator norm or spectral radius defined as $\lambda(A) := \max_{|\mathbf{x}|=1} |A\mathbf{x}|$ for any symmetric real matrix A, where $\mathbf{x}$ is a real vector, and $|\mathbf{x}|$ denotes its Euclidean norm. For the estimate $\hat{\Sigma}_T$ in (3), unfortunately, because too many parameters or autocovariances are estimated and the signal-to-noise ratios are too small at large lags, this estimate is not consistent. Wu and Pourahmadi (2009) showed that $\lambda(\hat{\Sigma}_T - \Sigma_T) \not\to 0$ in probability. In Section 2 we provide a precise order of magnitude of $\lambda(\hat{\Sigma}_T - \Sigma_T)$ and give explicit upper and lower bounds.

The inconsistency of sample covariance matrices has also been observed in the context of high-dimensional multivariate analysis, and this phenomenon is now well understood, thanks to the results from random matrix theory. See, among others, Marčenko and Pastur (1967), Bai and Yin (1993) and Johnstone (2001). Recently, there is a surge of interest on regularized covariance matrix estimation in high-dimensional statistical inference. We only sample a few works which are closely related to our problem. Cai, Zhang and Zhou (2010), Bickel and Levina (2008b) and Furrer and Bengtsson (2007) studied the banding and/or tapering methods, while Bickel and Levina (2008a) and El Karoui (2008) considered the regularization by thresholding. In most of these works, convergence rates of the estimates were established.

However, none of the techniques used in the aforementioned papers is applicable in our setting since their estimates require multiple independent and identically distributed (i.i.d.) copies of random vectors from the underlying multivariate distribution, though the number of copies can be far less than the dimension of the vector. In time series analysis, however, it is typical that only one realization is available. Hence we shall naturally use the sample autocovariances. In a companion paper, Xiao and Wu (2011) established a systematic theory for $\mathcal{L}^2$ and $\mathcal{L}^\infty$ deviations of sample autocovariances. Based on that, we adopt the regularization idea and study properties of the banded, tapered and thresholded estimates of the covariance matrices. Wu and Pourahmadi (2009) and McMurry and Politis (2010) applied the banding and tapering methods to the same problem, but here we shall obtain a better and optimal convergence rate. We shall point out that the regularization ideas of banding and tapering are not novel in time series analysis and they have been applied in nonparametric spectral density estimation.

In this paper we use the ideas in Toeplitz (1911) and Grenander and Szegö (1958) together with Wu's (2005) recent theory on stationary processes to present a systematic theory for estimates of covariance matrices of stationary processes. In particular, we shall exploit the connection between covariance matrices and spectral density functions and prove a sharp convergence rate for banded covariance matrix estimates of stationary processes. Using convergence properties of periodograms, we derive a precise order of magnitude for spectral radius of sample covariance matrices. We also consider a thresholded covariance matrix estimator that can better characterize sparsity if the true covariance matrix is sparse. As a main technical tool, we develop a large deviation type result for quadratic forms of stationary processes using m-dependence approximation, under the framework of causal representations and physical dependence measures.

The rest of this article is organized as follows. In Section 2 we introduce the framework of causal representation and physical dependence measures that are useful for studying convergence properties of covariance matrix estimates. We provide in Section 2 upper and lower bounds for the operator norm of the sample covariance matrices. The convergence rates of banded/tapered and thresholded sample covariance matrices are established in Sections 3 and 4, respectively. We also conduct a simulation study to compare the finite sample performances of banded and thresholded estimates in Section 5. Some useful moment inequalities are collected in Section 6. A large deviation result about quadratic forms of stationary processes, which is of independent interest, is given in Section 7. Section 8 concludes the paper.

We now introduce some notation. For a random variable X and $p > 0$, we write $X \in \mathcal{L}^p$ if $\|X\|_p := (\mathbb{E}|X|^p)^{1/p} < \infty$, and use $\|X\|$ as a shorthand for $\|X\|_2$. To express centering of random variables concisely, we define the operator $\mathbb{E}_0$ as $\mathbb{E}_0(X) := X - \mathbb{E}X$. Hence $\mathbb{E}_0(\mathbb{E}_0(X)) = \mathbb{E}_0(X)$. For a symmetric real matrix A, we use $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ for its smallest and largest eigenvalues, respectively, and use $\lambda(A)$ to denote its operator norm or spectral radius. For a real number x, $\lfloor x \rfloor := \max\{y \in \mathbb{Z} : y \le x\}$ denotes its integer part and $\lceil x \rceil := \min\{y \in \mathbb{Z} : y \ge x\}$. For two real numbers x, y, set $x \vee y = \max\{x, y\}$ and $x \wedge y := \min\{x, y\}$. For two sequences of positive numbers (a_T) and (b_T), we write $a_T \asymp b_T$ if there exists some constant $C > 1$ such that $C^{-1} \le a_T/b_T \le C$ for all T. The letter C denotes a constant, whose values may vary from place to place. We sometimes add symbolic subscripts to emphasize that the value of C depends on the subscripts.

2. Exact order of operator norms of sample covariance matrices.

Suppose Y is a $p \times n$ random matrix consisting of i.i.d. entries with mean 0 and variance 1, which could be viewed as a sample of size n from some p-dimensional population; then $YY^\top/n$ is the sample covariance matrix. If $\lim_{n\to\infty} p/n = c > 0$, then $YY^\top/n$ is not a consistent estimate of the population covariance matrix (which is the identity matrix) in term of the operator norm. This is a well-known phenomenon in random matrices literature; see, for example, Marčenko and Pastur (1967), Section 5.2 in Bai and Silverstein (2010), Johnstone (2001) and El Karoui (2005). However, the techniques used in those papers are not applicable here, because we have only one realization and the dependence within the vector can be quite complicated. Thanks to the Toeplitz structure of Σ_T, our method depends on the connection between its eigenvalues and the spectral density, defined by

$$f(\theta) = \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} \gamma_k \cos(k\theta). \tag{4}$$

The following lemma is a special case of Section 5.2 [Grenander and Szegö (1958)].

LEMMA 1. *Let h be a continuous symmetric function on $[-\pi, \pi]$. Denote by $\underline{h}$ and $\overline{h}$ its minimum and maximum, respectively. Define $a_k = \int_{-\pi}^{\pi} h(\theta) e^{-ik\theta}\, \mathrm{d}\theta$ and the $T \times T$ matrix $\Gamma_T = (a_{s-t})_{1 \le s,t \le T}$; then*

$$2\pi \underline{h} \le \lambda_{\min}(\Gamma_T) \le \lambda_{\max}(\Gamma_T) \le 2\pi \overline{h}.$$

Lemma 1 can be easily proved by noting that

$$\mathbf{x}^\top \Gamma_T \mathbf{x} = \int_{-\pi}^{\pi} |\mathbf{x}^\top \rho(\theta)|^2 h(\theta)\, \mathrm{d}\theta \qquad \text{where } \rho(\theta) = (e^{i\theta}, e^{i2\theta}, \ldots, e^{iT\theta})^\top. \tag{5}$$

The sample covariance matrix (3) is closely related to the periodogram

$$I_T(\theta) = T^{-1}\left|\sum_{t=1}^{T} X_t e^{it\theta}\right|^2 = \sum_{k=1-T}^{T-1} \hat{\gamma}_k e^{ik\theta}.$$

By Lemma 1, we have $\lambda(\hat{\Sigma}_T) \le \max_{-\pi\le\theta\le\pi} I_T(\theta)$. Asymptotic properties of periodograms have recently been studied by Peligrad and Wu (2010) and Lin and Liu (2009). To introduce the result in the latter paper, we assume that the process (X_t) has the causal representation

$$X_t = g(\varepsilon_t, \varepsilon_{t-1}, \ldots), \tag{6}$$

where g is a measurable function such that X_t is well defined, and $\varepsilon_t, t \in \mathbb{Z}$, are i.i.d. random variables. The framework (6) is very general [see, e.g., Tong (1990)] and easy to work with. Let $\mathcal{F}^t = (\varepsilon_t, \varepsilon_{t-1}, \ldots)$ be the set of innovations up to time t; we write $X_t = g(\mathcal{F}^t)$. Let $\varepsilon'_t, t \in \mathbb{Z}$, be an i.i.d. copy of $\varepsilon_t, t \in \mathbb{Z}$. Define $\mathcal{F}^t_* = (\varepsilon_t, \ldots, \varepsilon_1, \varepsilon'_0, \varepsilon_{-1}, \ldots)$, obtained by replacing ε_0 in $\mathcal{F}^t$ by ε'_0, and set $X'_t = g(\mathcal{F}^t_*)$. Following Wu (2005), for $p > 0$, define

$$\Theta_p(m) = \sum_{t=m}^{\infty} \delta_p(t), \qquad m \ge 0, \qquad \text{where } \delta_p(t) = \|X_t - X'_t\|_p. \tag{7}$$

In Wu (2005), the quantity $\delta_p(t)$ is called *physical dependence measure*. We make the convention that $\delta_p(t) = 0$ for $t < 0$. Throughout the article, we assume the *short-range dependence* condition $\Theta_p := \Theta_p(0) < \infty$. Under a mild condition on the tail sum $\Theta_p(m)$ (cf. Theorem 2), Lin and Liu (2009) obtained the weak convergence result

$$\max_{1\le s\le q}\left\{\frac{I_T(2\pi s/T)}{2\pi f(2\pi s/T)}\right\} - \log q \Rightarrow \mathcal{G}, \tag{8}$$

where $\Rightarrow$ denotes the convergence in distribution, $\mathcal{G}$ is the Gumbel distribution with the distribution function $e^{-e^{-x}}$, and $q = \lceil T/2\rceil - 1$. Using this result, we can provide explicit upper and lower bounds on the operator norm of the sample covariance matrix.

THEOREM 2. *Assume $X_t \in \mathcal{L}^p$ for some $p > 2$ and $\mathbb{E}X_t = 0$. If $\Theta_p(m) = o(1/\log m)$ and $\min_\theta f(\theta) > 0$, then*

$$\lim_{T\to\infty} P\left\{\frac{\pi[\min_\theta f(\theta)]^2 \log T}{12\Theta_2^2} \le \lambda(\hat{\Sigma}_T) \le 10\Theta_2^2 \log T\right\} = 1.$$

According to Lemma 1, we know $\lambda_{\max}(\Sigma_T) \le 2\pi \max_\theta f(\theta)$. As an immediate consequence of Theorem 2, there exists a constant $C > 1$ such that

$$\lim_{T\to\infty} P[C^{-1}\log T \le \lambda(\hat{\Sigma}_T - \Sigma_T) \le C\log T] = 1,$$

which means the estimate $\hat{\Sigma}_T$ is not consistent, and the order of magnitude of $\lambda(\hat{\Sigma}_T - \Sigma_T)$ is $\log T$. Earlier, Wu and Pourahmadi (2009) also showed that the plug-in estimate $\hat{\Sigma}_T = (\hat{\gamma}_{s-t})_{1\le s,t\le T}$ is not consistent, namely, $\lambda(\hat{\Sigma}_T - \Sigma) \not\to 0$ in probability. Our Proposition 2 substantially refines this result by providing an exact order of magnitude of $\lambda(\hat{\Sigma}_T - \Sigma)$.

An, Chen and Hannan (1983) showed that under suitable conditions, for linear processes with i.i.d. innovations,

$$\lim_{T\to\infty} \max_\theta \{I_T(\theta)/[2\pi f(\theta)\log T]\} = 1 \qquad \text{almost surely.} \tag{9}$$

A stronger version was found by Turkman and Walker (1990) for Gaussian processes. Based on (9), we conjecture that

$$\lim_{T\to\infty} \frac{\lambda(\hat{\Sigma}_T)}{2\pi \max_\theta f(\theta)\log T} = 1 \qquad \text{almost surely.}$$

Turkman and Walker (1984) established the following result on the maximum periodogram of a sequence of i.i.d. standard normal random variables:

$$\max_\theta I_T(\theta) - \log T - \frac{\log(\log T)}{2} + \frac{\log(3/\pi)}{2} \Rightarrow \mathcal{G}. \tag{10}$$

In view of (8) and (10), we conjecture that $\lambda(\hat{\Sigma}_T)$ also converges to the Gumbel distribution after proper centering and rescaling. Note that the Gumbel convergence (10), where the maximum is taken over the entire interval $\theta \in [-\pi, \pi]$, has a different centering term from the one in (8) which is obtained over Fourier frequencies.

If Y is a $p \times n$ random matrix consisting of i.i.d. entries, Geman (1980) and Yin, Bai and Krishnaiah (1988) proved a strong convergence result for the largest eigenvalues of $Y^\top Y$, in the paradigm where $n, p \to \infty$ such that $p/n \to c \in (0, \infty)$. See also Bai and Silverstein (2010) and references therein. If in addition the entries of Y are i.i.d. complex normal or normal random variables, Johansson (2000) and Johnstone (2001) presented an asymptotic distributional theory and showed that, after proper normalization, the limiting distribution of the largest eigenvalue follows the Tracy–Widom law [Tracy and Widom (1994)]. Again, their methods depend heavily on the setup that there are i.i.d. copies of a random vector with independent entries, and/or the normality assumption, so they are not applicable here. Bryc, Dembo and Jiang (2006) studied the limiting spectral distribution (LSD) of random Toeplitz matrices whose entries on different sub-diagonals are i.i.d. Solo (2010) considered the LSD of sample covariances matrices generated by a sample which is temporally dependent.

PROOF OF THEOREM 2. For notational simplicity we let $\underline{f} := \min_\theta f(\theta)$ and $\overline{f} := \max_\theta f(\theta)$. It follows immediately from (8) that for any $\delta > 0$

$$\lim_{T\to\infty} P\Big[\max_\theta I_T(\theta) \ge 2(1-\delta)\pi \underline{f} \log T\Big] = 1. \tag{11}$$

The result in (8) is not sufficient to yield an upper bound of $\max_\theta I_T(\theta)$. For this purpose we need to consider the maxima over a finer grid and then use Lemma 3 to extend to the maxima over the whole real line. Set $j_T = \lfloor T \log T \rfloor$ and $\theta_s := \theta_{T,s} = \pi s / j_T$ for $0 \le s \le j_T$. Define $m_T = \lfloor T^\beta \rfloor$ for some $0 < \beta < 1$, and $\tilde{X}_t = \mathcal{H}_{t-m_T} X_t = \mathbb{E}(X_t | \varepsilon_{t-m_T}, \varepsilon_{t-m_T+1}, \ldots)$. Let $S_T(\theta) = \sum_{t=1}^T X_t e^{it\theta}$ be the Fourier transform of $(X_t)_{1 \le t \le T}$, and $\tilde{S}_T(\theta) = \sum_{t=1}^T \tilde{X}_t e^{it\theta}$ for the m_T-dependent sequence $(\tilde{X}_t)_{1 \le t \le T}$. By Lemma 3.4 of Lin and Liu (2009), we have

$$\max_{0 \le s \le j_T} T^{-1/2} |S_T(\theta_s) - \tilde{S}_T(\theta_s)| = o_P((\log T)^{-1/2}). \tag{12}$$

Now partition the interval $[1, T]$ into blocks $\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_{w_T}$ of size m_T, where $w_T = \lfloor T / m_T \rfloor$, and the last block may have size $m_T \le |\mathcal{B}|_{w_T} < 2m_T$. Define the block sum $U_{T,k}(\theta) = \sum_{t \in \mathcal{B}_k} \tilde{X}_t e^{it\theta}$ for $1 \le k \le w_T$. Choose $\beta > 0$ small enough such that for some $0 < \gamma < 1/2$, the inequality

$$1 - \beta + \beta p - \gamma(p-1) - 1/2 < 0 \tag{13}$$

holds. We use truncation and define $\overline{U}_{T,k}(\theta) = \mathbb{E}_0[U_{T,k}(\theta) \mathbf{1}_{|U_{T,k}(\theta)| \le T^\gamma}]$. Using similar arguments as equation (5.5) [Lin and Liu (2009)] and (13), we have

$$\max_{0 \le s \le j_T} T^{-1/2} \left| \sum_{k=1}^{w_T} [U_{T,k}(\theta_s) - \overline{U}_{T,k}(\theta_s)] \right| = o_P((\log T)^{-1/2}). \tag{14}$$

Observe that $\overline{U}_{T,k_1}(\theta)$ and $\overline{U}_{T,k_2}(\theta)$ are independent if $|k_1 - k_2| > 1$. Let $\Re(z)$ denote the real part of a complex number z. Split the sum $\sum_{k=1}^{w_T} \overline{U}_{T,k}(\theta)$ into four parts

$$R_{T,1}(\theta) = \sum_{k \text{ odd}} \Re(\overline{U}_{T,k}(\theta)), \qquad R_{T,2}(\theta) = \sum_{k \text{ even}} \Re(\overline{U}_{T,k}(\theta))$$

and $R_{T,3}$, $R_{T,4}$ similarly for the imaginary part of $\overline{U}_{T,k}$. Since $\mathbb{E}|\overline{U}_{T,k}(\theta)|^2 \le \mathbb{E}|U_{T,k}(\theta)|^2 \le |\mathcal{B}_k| \Theta_2^2$, by Bernstein's inequality [cf. Freedman (1975)],

$$\sup_\theta P\left[|R_{T,l}(\theta)| \ge \frac{3\Theta_2}{2\sqrt{2}} \sqrt{T \log T} \right] \le 2 \exp\left\{ -\frac{(9/8) \log T}{1 + 3\Theta_2^{-1} \sqrt{2 \log T}\, T^{\gamma - 1/2}} \right\}$$

for $1 \le l \le 4$. It follows that

$$\lim_{T \to \infty} P\left[\max_{0 \le s \le j_T} \left| \sum_{k=1}^{w_T} \overline{U}_{T,k}(\theta_s) \right| \ge 3\Theta_2 \sqrt{T \log T} \right] = 0. \tag{15}$$

Combining (12), (14) and (15), we have

$$\lim_{T \to \infty} P\left[\max_{0 \le s \le j_T} I_T(\theta_s) \le 9.5 \Theta_2^2 \log T \right] = 1, \tag{16}$$

which together with Lemma 3 implies that

$$\lim_{T \to \infty} P\left[\max_\theta I_T(\theta) \le 10 \Theta_2^2 \log T \right] = 1. \tag{17}$$

The upper bound in Theorem 2 is an immediate consequence in view of Lemma 1. For the lower bound, we use the inequality

$$\lambda(\hat{\Sigma}_T) \geq \max_\theta \{T^{-1}\rho(\theta)^* \Sigma_T \rho(\theta)\},$$

where $\rho(\theta)$ is defined in (5), and $\rho(\theta)^*$ is its Hermitian transpose. Note that

$$\begin{aligned}
\rho(\theta)^* \Sigma_T \rho(\theta) &= \sum_{s,t=1}^{T} \hat{\gamma}_{s-t} e^{is\theta} e^{-it\theta} \\
&= \frac{1}{2\pi} \int_{-\pi}^{\pi} \sum_{s,t=1}^{T} I_T(\omega) e^{-i(s-t)\omega} e^{i(s-t)\theta}\, \mathrm{d}\omega \\
&= \frac{1}{2\pi} \int_{-\pi}^{\pi} I_T(\omega) \left| \sum_{t=1}^{T} e^{it(\omega-\theta)} \right|^2 \mathrm{d}\omega \\
&= \frac{1}{2\pi} \int_{-\pi-\theta}^{\pi-\theta} I_T(\omega+\theta) \left| \sum_{t=1}^{T} e^{it\omega} \right|^2 \mathrm{d}\omega.
\end{aligned}$$

By Bernstein's inequality on the derivative of trigonometric polynomials [cf. Zygmund (2002), Theorem 3.13, Chapter X], we have

$$\max_\theta |I_T'(\theta)| \leq T \cdot \max_\theta I_T(\theta).$$

Let $\theta_0 = \arg\max_\theta I_T(\theta)$. Set $c = (1-\delta)\pi \underline{f}/(10\Theta_2^2)$. By Lemma 1 and (40), we know $2\pi \overline{f} \leq \Theta_2^2$, and hence $c \leq 1/20$. If $I_T(\theta_0) \geq 2(1-\delta)\pi \underline{f} \log T$ and $\max_\theta I_T(\theta) \leq 10\Theta_2^2 \log T$, then for $|w| \leq c/T$, we have

$$I_T(\theta_0 + \omega) \geq [2(1-\delta)\pi \underline{f} - 10c\Theta_2^2] \log T = (1-\delta)\pi \underline{f} \log T.$$

Since $|\sum_{j=1}^{T} e^{ij\omega}|^2 \geq 10T^2/11$ when $|w| \leq c/T$, it follows that

$$\begin{aligned}
\rho(\theta_0)^* \Sigma_T \rho(\theta_0) &\geq \frac{1}{2\pi} \cdot (1-\delta)\pi \underline{f} \log T \cdot \frac{10T^2}{11} \cdot \frac{2c}{T} \\
&= \frac{\pi(1-\delta)^2 \underline{f}^2 T \log T}{11\Theta_2^2},
\end{aligned}$$

which implies that $\lambda(\hat{\Sigma}_T) \geq \pi(1-\delta)^2 \underline{f}^2 \log T/(11\Theta_2^2)$. The proof is completed by selecting δ small enough. □

REMARK 1. In the proof, as well as many other places in this article, we often need to partition an integer interval $[s,t] \subset \mathbb{N}$ into consecutive blocks $\mathcal{B}_1, \dots, \mathcal{B}_b$ with the same size m. Since $s - t + 1$ may not be a multiple of m, we make the convention that the last block $\mathcal{B}_b$ has the size $m \leq |\mathcal{B}_b| < 2m$, and all the other ones have the same size m.

3. Banded covariance matrix estimates. In view of Lemma 1, the inconsistency of $\hat{\Sigma}_T$ is due to the fact that the periodogram $I_T(\theta)$ is not a consistent estimate of the spectral density $f(\theta)$. To estimate the spectral density consistently, it is very common to use smoothing. In particular, consider the lag window estimate

$$\hat{f}_{T,B_T}(\theta) = \frac{1}{2\pi} \sum_{k=-B_T}^{B_T} K(k/B_T)\hat{\gamma}_k \cos(k\theta), \tag{18}$$

where B_T is the bandwidth satisfying natural conditions $B_T \to \infty$ and $B_T/T \to 0$, and $K(\cdot)$ is a symmetric kernel function satisfying

$$K(0) = 1, \qquad |K(x)| \le 1 \quad \text{and} \quad K(x) = 0 \qquad \text{for } |x| > 1.$$

Correspondingly, we define the tapered covariance matrix estimate

$$\hat{\Sigma}_{T,B_T} = [K((s-t)/B_T)\hat{\gamma}_{s-t}]_{1\le s,t\le T} = \hat{\Sigma}_T \star W_T,$$

where $\star$ is the Hadamard (or Schur) product, which is formed by element-wise multiplication of matrices. The term "tapered" is consistent with the terminology of the high-dimensional covariance regularization literature. However, the reader should not confuse this with the notion "data taper" that is commonly used in time series analysis. Our tapered estimate parallels a lag-window estimate of the spectral density with a tapered window. As a special case, if $K(x) = \mathbf{1}_{\{|x|\le 1\}}$ is the rectangular kernel, then $\hat{\Sigma}_{T,B_T} = (\hat{\gamma}_{s-t}\mathbf{1}_{\{|s-t|\le B_T\}})$ is the banded sample covariance matrix. However, this estimate may not be nonnegative definite. To obtain a nonnegative definite estimate, by the Schur product theorem in matrix theory [Horn and Johnson (1990)], since $\hat{\Sigma}_T$ is nonnegative definite, their Schur product $\hat{\Sigma}_{T,B_T}$ is also nonnegative definite if $W_T = [K((s-t)/B_T)]_{1\le s,t\le T}$ is nonnegative definite. The Bartlett or triangular window $K_B(u) = \max(0, 1-|u|)$ leads to a positive definite weight matrix W_T in view of

$$K_B(u) = \int_{\mathbb{R}} w(x)w(x+u)\,\mathrm{d}x, \tag{19}$$

where $w(x) = \mathbf{1}_{\{|x|\le 1/2\}}$ is the rectangular window. Any kernel function having form (19) must be positive definite.

Here we shall show that $\hat{\Sigma}_{T,B_T}$ is a consistent estimate of Σ_T and establish a convergence rate of $\lambda(\hat{\Sigma}_{T,B_T} - \Sigma)$. We first consider the bias. By the Geršgorin theorem [cf. Horn and Johnson (1990), Theorem 6.1.1], we have $\lambda(\mathbb{E}\hat{\Sigma}_{T,B_T} - \Sigma) \le b_T$, where

$$b_T = 2\sum_{k=1}^{B_T}\left[1 - K\left(\frac{k}{B_T}\right)\right]|\gamma_k| + \frac{2}{T}\sum_{k=1}^{B_T} k|\gamma_k| + 2\sum_{k=B_T+1}^{T-1} |\gamma_k|. \tag{20}$$

The first term on the right-hand side in (20) is due to the choice of the kernel function, whose order of magnitude is determined by the smoothness of $K(\cdot)$

at zero. In particular, this term vanishes if $K(\cdot)$ is the rectangular kernel. If $1 - K(u) = O(u^2)$ at $u = 0$ and $\gamma_k = O(k^{-\beta})$, $\beta > 1$, then $b_T = O(B_T^{1-\beta})$ if $1 < \beta < 2$, $b_T = O(B_T^{1-\beta} + T^{-1})$ if $2 < \beta < 3$ and $b_T = O(B_T^{-2} + T^{-1})$ if $\beta > 3$. Generally, if $\sum_{k=1}^{\infty} |\gamma_k| < \infty$, then $b_T \to 0$ as $B_T \to \infty$ and $B_T < T$.

It is more challenging to deal with $\lambda(\hat{\Sigma}_{T,B_T} - \mathbb{E}\hat{\Sigma}_{T,B_T})$. If $X_t \in \mathcal{L}^p$ for some $2 < p \le 4$ and $\mathbb{E}X_t = 0$, Wu and Pourahmadi (2009) obtained

$$\lambda(\hat{\Sigma}_{T,B_T} - \mathbb{E}\hat{\Sigma}_{T,B_T}) = O_P\left(\frac{B_T \Theta_p^2}{T^{1-2/p}}\right). \tag{21}$$

The key step of their method is to use the inequality

$$\lambda(\hat{\Sigma}_{T,B_T} - \mathbb{E}\hat{\Sigma}_{T,B_T}) \le 2\sum_{k=0}^{B_T} |K(k/B_T)||\hat{\gamma}_k - \mathbb{E}\hat{\gamma}_k|,$$

which is also obtained by the Geršgorin theorem. It turns out that the above rate can be improved by exploiting the Toeplitz structure of the autocovariance matrix. By Lemma 1,

$$\lambda(\hat{\Sigma}_{T,B_T} - \mathbb{E}\hat{\Sigma}_{T,B_T}) \le 2\pi \max_{\theta} |\hat{f}_{T,B_T}(\theta) - \mathbb{E}\hat{f}_{T,B_T}(\theta)|. \tag{22}$$

Since $\hat{f}_{T,B_T}(\theta)$ is a trigonometric polynomial of order B_T, we can bound its maximum by the maximum over a fine grid. The following lemma is adapted from Zygmund (2002), Theorem 7.28, Chapter X.

LEMMA 3. *Let $S(x) = \frac{1}{2}a_0 + \sum_{k=1}^{n}[a_k \cos(kx) + b_k \sin(kx)]$ be a trigonometric polynomial of order n. For any $x^* \in \mathbb{R}$, $\delta > 0$ and $l \ge 2(1+\delta)n$, let $x_j = x^* + 2\pi j/l$ for $0 \le j \le l$; then*

$$\max_x |S(x)| \le (1 + \delta^{-1}) \max_{0 \le j \le l} |S(x_j)|.$$

For $\delta > 0$, let $\theta_j = \pi j/(\lceil (1+\delta)B_T \rceil)$ for $0 \le j \le \lceil (1+\delta)B_T \rceil$; then by Lemma 3,

$$\max_{\theta} |\hat{f}_{T,B_T}(\theta) - \mathbb{E}\hat{f}_{T,B_T}(\theta)| \le (1+\delta^{-1}) \max_j |\hat{f}_{T,B_T}(\theta_j) - \mathbb{E}\hat{f}_{T,B_T}(\theta_j)|. \tag{23}$$

THEOREM 4. *Assume $X_t \in \mathcal{L}^p$ with some $p > 4$, $\mathbb{E}X_t = 0$, and $\Theta_p(m) = O(m^{-\alpha})$. Choose the banding parameter B_T to satisfy $B_T \to \infty$, and $B_T = O(T^{\gamma})$, for some*

$$0 < \gamma < 1, \qquad \gamma < \alpha p/2 \quad \text{and} \quad (1-2\alpha)\gamma < (p-4)/p. \tag{24}$$

Then for b_T defined in (20), *and $c_p = (p+4)e^{p/4}\Theta_4^2$,*

$$\lim_{T\to\infty} P\left[\lambda(\hat{\Sigma}_{T,B_T} - \Sigma_T) \le 12 c_p \sqrt{\frac{B_T \log B_T}{T}} + b_T\right] = 1. \tag{25}$$

In particular, if $K(x) = \mathbf{1}_{\{|x|\le 1\}}$ *and* $B_T \asymp (T/\log T)^{1/(2\alpha+1)}$, *then*

$$\lambda(\hat{\Sigma}_{T,B_T} - \Sigma_T) = O_P\left[\left(\frac{\log T}{T}\right)^{\alpha/(2\alpha+1)}\right]. \tag{26}$$

PROOF. In view of (20), to prove (25) we only need to show that

$$\lim_{T\to\infty} P\left[\lambda(\hat{\Sigma}_{T,B_T} - \mathbb{E}\hat{\Sigma}_{T,B_T}) \le 12c_p\sqrt{\frac{B_T \log B_T}{T}}\right] = 1. \tag{27}$$

By (22) and (23) where we take $\delta = 1$, the problem is reduced to

$$\lim_{T\to\infty} P\left[(2\pi)\cdot \max_j |\hat{f}_{T,B_T}(\theta_j) - \mathbb{E}\hat{f}_{T,B_T}(\theta_j)| \le 6c_p\sqrt{\frac{B_T \log B_T}{T}}\right] = 1. \tag{28}$$

By Theorem 10 (where we take $M = 2$), for any $\gamma < \beta < 1$, there exists a constant $C_{p,\beta}$ such that

$$\begin{aligned}
&\max_j P\left[(2\pi)\cdot |\hat{f}_{T,B_T}(\theta_j) - \mathbb{E}\hat{f}_{T,B_T}(\theta_j)| \ge 6c_p\sqrt{\frac{B_T \log B_T}{T}}\right]\\
&\quad\le C_{p,\beta}(TB_T)^{-p/4}(\log T)[(TB_T)^{p/4}T^{-\alpha\beta p/2} + TB_T^{p/2-1-\alpha\beta p/2} + T]\\
&\qquad + C_{p,\beta}B_T^{-2}.
\end{aligned} \tag{29}$$

If (24) holds, there exist a $0 < \beta < 1$ such that $\gamma - \alpha\beta p/2 < 0$ and $(p/4 - \alpha\beta p/2)\gamma - (p/4 - 1) < 0$. It follows that by (29),

$$\begin{aligned}
&P\left[\max_j |\hat{f}_{T,B_T}(\theta_j) - \mathbb{E}\hat{f}_{T,B_T}(\theta_j)| \ge 6c_p\sqrt{\frac{B_T \log B_T}{T}}\right]\\
&\quad\le C_{p,\beta}(\log T)\big[T^{\gamma-\alpha\beta p/2} + T^{1-p/4} + T^{(p/4-\alpha\beta p/2)\gamma-(p/4-1)}\big] + C_{p,\beta}B_T^{-1}\\
&\quad= o(1).
\end{aligned}$$

Therefore, (28) holds and the proof of (25) is complete. The last statement (26) is an immediate consequence. Details are omitted. □

REMARK 2. In practice, $\mathbb{E}X_1$ is usually unknown, and we estimate it by $\overline{X}_T = T^{-1}\sum_{t=1}^T X_t$. Let $\hat{\gamma}_k^c = T^{-1}\sum_{t=k+1}^T (X_{t-k} - \overline{X}_T)(X_t - \overline{X}_T)$, and Σ_{T,B_T}^c be defined as $\hat{\Sigma}_{T,B_T}$ by replacing $\hat{\gamma}_k$ therein by $\hat{\gamma}_k^c$. Since $\overline{X}_T - \mathbb{E}X_1 = O_P(T^{-1/2})$, it is easily seen that $\lambda(\hat{\Sigma}_{T,B_T} - \hat{\Sigma}_{T,B_T}^c) = O_P(B_T/T)$. Therefore, the results of Theorem 4 hold for $\hat{\Sigma}_{T,B_T}^c$ as well.

REMARK 3. In the proof of Theorem 4, we have shown that, as an intermediate step from (28) to (27),

$$\lim_{T\to\infty} P\Big[\max_{0\le\theta\le 2\pi} |\hat{f}_{T,B_T}(\theta) - \mathbb{E}\hat{f}_{T,B_T}(\theta)| \le 6\pi^{-1} c_p \sqrt{T^{-1} B_T \log B_T}\Big] = 1. \tag{30}$$

The above uniform convergence result is very useful in spectral analysis of time series. Shao and Wu (2007) obtained the weaker version

$$\max_{0\le\theta\le 2\pi} |\hat{f}_{T,B_T}(\theta) - \mathbb{E}\hat{f}_{T,B_T}(\theta)| = O_P\big(\sqrt{T^{-1} B_T \log B_T}\big)$$

under a stronger assumption that $\Theta_p(m) = O(\rho^m)$ for some $0 < \rho < 1$.

REMARK 4. For linear processes, Woodroofe and Van Ness (1967) derived the asymptotic distribution of the maximum deviations of spectral density estimates. Liu and Wu (2010) generalized their result to nonlinear processes and showed that the limiting distribution of

$$\max_{0\le j\le B_T} \sqrt{\frac{T}{B_T}} \frac{|\hat{f}_{T,B_T}(\pi j/B_T) - \mathbb{E}\hat{f}_{T,B_T}(\pi j/B_T)|}{f(\pi j/B_T)}$$

is Gumbel after suitable centering and rescaling, under stronger conditions than (24). With their result, and using similar arguments as Theorem 2, we can show that for some constant C_p,

$$\lim_{T\to\infty} P\left[C_p^{-1}\sqrt{\frac{B_T \log B_T}{T}} \le \lambda(\hat{\Sigma}_{T,B_T} - \mathbb{E}\hat{\Sigma}_{T,B_T}) \le C_p\sqrt{\frac{B_T \log B_T}{T}}\right] = 1,$$

which means that the convergence rate we have obtained in (27) is optimal.

REMARK 5. The convergence rate $\sqrt{T^{-1} B_T \log B_T} + b_T$ in Theorem 4 is optimal. Consider a process (X_t) which satisfies $\gamma_0 = 3$ and when $k > 0$,

$$\gamma_k = \begin{cases} A^{-\alpha j}, & \text{if } k = A^j \text{ for some } j \in \mathbb{N}, \\ 0, & \text{otherwise,} \end{cases}$$

where $\alpha > 0$ and $A > 0$ is an even integer such that $A^{-\alpha} \le 1/5$. Consider the banded estimate $\hat{\Sigma}_{T,B_T}$ with the rectangular kernel. As shown in the supplementary article [Xiao and Wu (2012)], there exists a constant $C > 0$ such that

$$\lim_{T\to\infty} P\left[\lambda(\hat{\Sigma}_{T,B_T} - \Sigma_T) \ge C\sqrt{\frac{B_T \log B_T}{T}} + b_T/5\right] = 1, \tag{31}$$

suggesting that the convergence rate given in (25) of Theorem 4 is optimal. This optimality property can have many applications. For example, it can allow one to derive convergence rates for estimates of $\mathbf{a}_T$ in the Wiener–Hopf equation, and the optimal weights $\mathbf{c}_T = (c_1, \dots, c_T)^\top$ in the best linear unbiased estimation problem mentioned in the Introduction.

REMARK 6. We now compare (21) and our result. For $p = 4$, (21) gives the order $\lambda(\hat{\Sigma}_{T,B_T} - \mathbb{E}\hat{\Sigma}_{T,B_T}) = O_P(B_T/\sqrt{T})$. Our result (27) is sharper by moving the bandwidth B_T inside the square root. We pay the costs of a logarithmic factor, a higher order moment condition ($p > 4$), as well as conditions on the decay rate of tail sum of physical dependence measures (24). Note that when $\alpha \geq 1/2$, the last two conditions of (24) hold automatically, so we merely need $0 < \gamma < 1$, allowing a very wide range of B_T. In comparison, for (21) to be useful, one requires $B_T = o(T^{1-2/p})$.

REMARK 7. The convergence rate (21) of Wu and Pourahmadi (2009) parallels the result of Bickel and Levina (2008b) in the context of high-dimensional multivariate analysis, which was improved in Cai, Zhang and Zhou (2010) by constructing a class of tapered estimates. Our result parallels the optimal minimax rate derived in Cai, Zhang and Zhou (2010), though the settings are different.

REMARK 8. Theorem 4 uses the operator norm. For the Frobenius norm see Xiao and Wu (2011) where a central limit theory for $\sum_{k=1}^{B_T} \hat{\gamma}_k^2$ and $\sum_{k=1}^{B_T} (\hat{\gamma}_k - \mathbb{E}\hat{\gamma}_k)^2$ is established.

4. Thresholded covariance matrix estimators. In the context of time series, the observations have an intrinsic temporal order and we expect that observations are weakly dependent if they are far apart, so banding seems to be natural. However, if there are many zeros or very weak correlations within the band, the banding method does not automatically generate a sparse estimate.

The rationale behind the banding operation is sparsity, namely autocovariances with large lags are small, so it is reasonable to estimate them as zero. Applying the same idea to the sample covariance matrix, we can obtain an estimate of Σ_T by replacing small entries in $\hat{\Sigma}_T$ with zero. This regularization approach, termed *hard thresholding*, was originally developed in nonparametric function estimation. It has recently been applied by Bickel and Levina (2008a) and El Karoui (2008) as a method of covariance regularization in the context of high-dimensional multivariate analysis. Since they do not assume any order of the observations, their sparsity assumptions are permutation-invariant. Unlike their setup, we still use $\Theta_p(m)$ [cf. (7)] and

$$\Psi_p(m) = \left(\sum_{t=m}^{\infty} \delta_p(t)^{p'}\right)^{1/p'}, \qquad \Delta_p(m) = \sum_{t=0}^{\infty} \min\{\mathcal{C}_p \Psi_p(m), \delta_p(t)\} \tag{32}$$

as our weak dependence conditions, where $p' = \min(2, p)$ and $\mathcal{C}_p$ is given in (38). This is natural for time series analysis.

Let $A_T = 2c'_p \sqrt{\log T/T}$, where c'_p is the constant given in Lemma 6. The thresholded sample autocovariance matrix is defined as

$$\hat{\Gamma}_{T,A_T} = \left(\hat{\gamma}_{s-t} \mathbf{1}_{|\hat{\gamma}_{s-t}| \geq A_T}\right)_{1 \leq s,t \leq T}$$

with the convention that the diagonal elements are never thresholded. We emphasize that the thresholded estimate may not be positive definite. The following result says that the thresholded estimate is also consistent in terms of the operator norm, and provides a convergence rate which parallels the banding approach in Section 3. In the proof we compare the thresholded estimate Γ_{T,A_T} with the banded one Σ_{T,B_T} for some suitably chosen B_T. This is merely a way to simplify the arguments. The same results can be proved without referring to the banded estimates.

THEOREM 5. *Assume $X_t \in \mathcal{L}^p$ with some $p > 4$, $\mathbb{E}X_t = 0$, and $\Theta_p(m) = O(m^{-\alpha})$, $\Delta_p(m) = O(m^{-\alpha'})$ for some $\alpha \geq \alpha' > 0$. If*

$$\alpha > 1/2 \quad or \quad \alpha' p > 2, \tag{33}$$

then

$$\lambda(\hat{\Gamma}_{T,A_T} - \Sigma_T) = O_P\left[\left(\frac{\log T}{T}\right)^{\alpha/(2(1+\alpha))}\right].$$

REMARK 9. Condition (33) is only required for Lemma 6. As commented by Xiao and Wu (2011), it can be reduced to $\alpha p > 2$ for linear processes. See Remark 2 of their paper for more details.

The key step for proving Theorem 5 is to establish a convergence rate for the maximum deviation of sample autocovariances. The following lemma is adapted from Theorem 3 of Xiao and Wu (2011), where the asymptotic distribution of the maximum deviation was also studied.

LEMMA 6. *Assume the conditions of Theorem 5. Then*

$$\lim_{T\to\infty} P\left(\max_{1\leq k<T} |\hat{\gamma}_k - \mathbb{E}\hat{\gamma}_k| \leq c'_p\sqrt{\frac{\log T}{T}}\right) = 1,$$

where $c'_p = 6(p+4)e^{p/4}\|X_0\|_4\Theta_4$.

PROOF OF THEOREM 5. Let $B_T = \lfloor (T/\log T)^{1/(2(1+\alpha))}\rfloor$, and $\hat{\Sigma}_{T,B_T}$ be the banded sample covariance matrix with the rectangular kernel. Recall that $b_T = (2/T)\sum_{k=1}^{B_T} k|\gamma_k| + 2\sum_{k=B_T+1}^{T-1}|\gamma_k|$ from (20). By Lemma 6, we have

$$\lambda(\hat{\Sigma}_{T,B_T} - \Sigma_T) = O_P\left(B_T\sqrt{\frac{\log T}{T}} + b_T\right). \tag{34}$$

Write the thresholded estimate $\hat{\Gamma}_{T,A_T} = \hat{\Gamma}_{T,A_T,1} + \hat{\Gamma}_{T,A_T,2}$, where

$$\hat{\Gamma}_{T,A_T,1} = \left(\hat{\gamma}_{s-t}\mathbf{1}_{|\hat{\gamma}_{s-t}|\geq A_T, |s-t|\leq B_T}\right)_{1\leq s,t\leq T}$$

and

$$\hat{\Gamma}_{T,A_T,2} = \big(\hat{\gamma}_{s-t}\mathbf{1}_{|\hat{\gamma}_{s-t}|\ge A_T, |s-t|>B_T}\big)_{1\le s,t\le T}.$$

By Geršgorin's theorem, it is easily seen that

$$\lambda(\hat{\Gamma}_{T,A_T,1} - \hat{\Sigma}_{T,B_T}) \le A_T B_T = O\left(B_T\sqrt{\frac{\log T}{T}}\right). \tag{35}$$

On the other hand,

$$\begin{aligned}\lambda(\hat{\Gamma}_{T,A_T,2}) \le 2\Bigg(&\sum_{k=B_T+1}^{T} |\hat{\gamma}_k - \mathbb{E}\hat{\gamma}_k|\mathbf{1}_{|\gamma_k|<A_T/2, |\hat{\gamma}_k|\ge A_T} \\ &+ \sum_{k=B_T+1}^{T} |\hat{\gamma}_k - \mathbb{E}\hat{\gamma}_k|\mathbf{1}_{|\gamma_k|\ge A_T/2, |\hat{\gamma}_k|\ge A_T} + \sum_{k=B_T+1}^{T} |\mathbb{E}\hat{\gamma}_k|\Bigg) \\ =:&\, 2(I_T + II_T + III_T).\end{aligned}$$

The term III_T is dominated by b_T. By Lemma 6, we know

$$\lim_{T\to\infty} P(I_T > 0) \le \lim_{T\to\infty} P\Big(\max_{1\le k\le T-1} |\hat{\gamma}_k - \mathbb{E}\hat{\gamma}_k| \ge A_T/2\Big) = 0. \tag{36}$$

For the remaining term II_T, note that the number of γ_k such that $k > B_T$ and $|\gamma_k| \ge A_T/2$ is bounded by $2\sum_{k=B_T+1}^{T} |\gamma_k|/A_T$; therefore by Lemma 6

$$II_T \le C(B_T^{-\alpha}/A_T) \cdot \max_{1\le k\le T-1} |\hat{\gamma}_k - \mathbb{E}\hat{\gamma}_k| = O_P(B_T^{-\alpha}). \tag{37}$$

Putting (34), (35), (36) and (37) together, the proof is complete. □

REMARK 10. If the mean $\mathbb{E}X_1$ is unknown, we need to replace $\hat{\gamma}_k$ by $\hat{\gamma}_k^c$ (Remark 2). Since Lemma 6 still holds when $\hat{\gamma}_k$ are replaced by $\hat{\gamma}_k^c$ [Xiao and Wu (2011)], Theorem 5 remains true for $\hat{\gamma}_k^c$.

5. A simulation study. The thresholded estimate is desirable in that it can lead to a better estimate when there are a lot of zeros or very weak autocovariances. Unfortunately, due to technical difficulties, the theoretical result (cf. Theorem 5) does not reflect this advantage. We show by simulations that thresholding does have a better finite sample performance over banding when the true autocovariance matrix is sparse.

Consider two linear processes $X_t = \sum_{s=0}^{\infty} a_s \varepsilon_{t-s}$ and $Y_t = \sum_{s=0}^{\infty} b_s \varepsilon_{t-s}$, where $a_0 = b_0 = 1$, and when $s > 0$

$$a_s = cs^{-(1+\alpha)}, \qquad b_s = c(s/2)^{-(1+\alpha)}\mathbf{1}_{s \text{ is even}}$$

for some $c > 0$ and $\alpha > 0$; and ε_s's are taken as i.i.d. $\mathcal{N}(0,1)$. Let γ_k^X, Σ_T^X, and γ_k^Y, Σ_T^Y denote the autocovariances and autocovariance matrices of the two processes, respectively. It is easily seen that $\gamma_k^Y = 0$ if k is odd. In fact, (Y_t) can be obtained by interlacing two i.i.d. copies of (X_t). For a given set of centered observations $X_1, X_2, \ldots, X_T$, assuming that its true autocovariance matrix is known, for a fair comparison we choose the optimal bandwidth B_T and threshold A_T as

$$\hat{A}_T^X = \mathop{\arg\min}_{l \in \{|\hat{\gamma}_0^X|, |\hat{\gamma}_1^X|, \ldots, |\hat{\gamma}_{T-1}^X|\}} \lambda(\hat{\Gamma}_{T,l}^X - \Sigma_T^X), \qquad \hat{B}_T^X = \mathop{\arg\min}_{1 \le k \le T} \lambda(\hat{\Sigma}_{T,k}^X - \Sigma_T^X).$$

The two parameters for the (Y_t) process are chosen in the same way. In all the simulations we set $c = 0.5$. For different combinations of the sample size T and the parameter α which controls the decay rate of autocovariances, we report the average distances in term of the operator norm of the two estimates $\hat{\Sigma}_{T,\hat{B}_T}$ and $\hat{\Gamma}_{T,\hat{A}_T}$ from Σ_T, as well as the standard errors based on 1000 repetitions. We also give the average bandwidth of $\hat{\Sigma}_{T,\hat{B}_T}$. Instead of reporting the average threshold for $\hat{\Gamma}_{T,\hat{A}_T}$, we provide the average number of nonzero autocovariances appearing in the estimates, which is comparable to the average bandwidth of $\hat{\Sigma}_{T,\hat{B}_T}$.

From Table 1, we see that for the process (X_t), the banding method outperforms the thresholding one, but the latter does give sparser estimates. For the process (Y_t), according to Table 2, we find that thresholding performs better than banding when the sample size is not very large ($T = 100, 200$), and yields sparser estimates as well. The advantage of thresholding in error disappears when the sample size

TABLE 1

Errors under operator norm for (X_t)

	$T = 100$		$T = 200$		$T = 500$	
	Error	**BW**	**Error**	**BW**	**Error**	**BW**
0.2	2.94 (1.17)	9.55 (6.60)	3.01 (1.22)	13.4 (7.67)	2.96 (1.23)	23.4 (13.1)
	3.66 (1.07)	5.40 (4.87)	3.88 (1.14)	7.39 (5.81)	4.08 (1.17)	12.5 (10.1)
	6.98 (2.63)		8.12 (2.85)		10.57 (3.93)	
0.5	1.52 (0.68)	6.31 (4.58)	1.38 (0.60)	8.46 (5.57)	1.15 (0.50)	11.9 (7.67)
	1.90 (0.64)	3.49 (2.56)	1.89 (0.59)	4.15 (3.07)	1.74 (0.54)	5.15 (3.27)
	5.55 (2.37)		6.73 (2.91)		8.88 (3.28)	
1	0.82 (0.39)	4.04 (2.33)	0.69 (0.32)	4.62 (2.47)	0.52 (0.24)	5.68 (3.06)
	1.03 (0.38)	2.24 (0.87)	0.95 (0.32)	2.29 (0.74)	0.81 (0.29)	2.58 (0.83)
	4.80 (2.14)		6.05 (2.25)		7.81 (2.64)	

"Error" refers to the average distance between the estimates and the true autocovariance matrix under the operator norm, and "BW" refers to the average bandwidth of the banded estimates, and the average number of nonzero sub-diagonals (including the diagonal) for the thresholded ones. The numbers 0.2, 0.5 and 1 in the first column are values of α. For each combination of T and α, three lines are reported, corresponding to banded estimates, thresholded ones and sample autocovariance matrices, respectively. Numbers in parentheses are standard errors.

TABLE 2
Error under operator norm for (Y_t)

	$T = 100$		$T = 200$		$T = 500$	
	Error	**BW**	**Error**	**BW**	**Error**	**BW**
0.2	3.33 (0.86)	9.87 (6.89)	3.54 (0.95)	13.7 (7.67)	3.61 (1.07)	24.7 (13.1)
	3.15 (0.93)	3.95 (3.50)	3.43 (1.00)	5.69 (4.72)	3.75 (1.08)	9.23 (8.04)
	7.21 (4.28)		8.69 (4.79)		11.1 (5.31)	
0.5	1.98 (0.61)	7.26 (5.32)	1.88 (0.59)	9.95 (6.44)	1.63 (0.53)	16.3 (10.1)
	1.81 (0.60)	2.93 (2.41)	1.81 (0.59)	3.44 (2.22)	1.71 (0.54)	4.64 (2.97)
	5.88 (3.27)		7.25 (3.59)		9.25 (3.72)	
1	1.19 (0.41)	5.31 (3.33)	1.01 (0.35)	6.20 (3.58)	0.79 (0.28)	8.28 (4.95)
	1.02 (0.39)	2.16 (0.65)	0.92 (0.32)	2.21 (0.57)	0.80 (0.28)	2.52 (0.77)
	5.09 (2.77)		6.39 (2.79)		8.18 (2.91)	

is 500. Intuitively speaking, banding is a way to threshold according to the truth (autocovariances with large lags are small), while thresholding is a way to threshold according to the data. Therefore, if the autocovariances are nonincreasing as for the process (X_t), or if the sample size is large enough, banding is preferable. If the autocovariances do not vary regularly as for the process (Y_t) and the sample size is moderate, thresholding is more adaptive. As a combination, in practice we can use a thresholding-after-banding estimate which enjoys both advantages.

Apparently our simulation is a very limited one, because we assume that the true autocovariance matrices are known. Practitioners would need a method to choose the bandwidth and/or threshold from the data. Although theoretical results suggest convergence rates of banding and thresholding parameters which lead to optimal convergence rates of the estimates, they do not offer much help for finite samples. The issue was addressed by Wu and Pourahmadi (2009) incorporating the idea of risk minimization from Bickel and Levina (2008b) and the technique of subsampling from Politis, Romano and Wolf (1999), and by McMurry and Politis (2010) using the rule introduced in Politis (2003) for selecting the bandwidth in spectral density estimation. An alternative method is to use the block length selection procedure in Bühlmann and Künsch (1999) which is designed for spectral density estimation. We shall study other data-driven methods in the future.

6. Moment inequalities. This section presents some moment inequalities that will be useful in the subsequent proofs. In Lemma 7, the case $1 < p \le 2$ follows from Burkholder (1988) and the other case $p > 2$ is due to Rio (2009). Lemma 8 is adopted from Proposition 1 of Xiao and Wu (2011).

LEMMA 7 [Burkholder (1988), Rio (2009)]. *Let* $p > 1$ *and* $p' = \min\{p, 2\}$; *let* D_t, $1 \le t \le T$, *be martingale differences, and* $D_t \in \mathcal{L}^p$ *for every* t. *Write* $M_T =$

$\sum_{t=1}^{T} D_t$. *Then*

$$\|M_T\|_p^{p'} \le \mathcal{C}_p^{p'} \sum_{t=1}^{T} \|D_t\|_p^{p'} \qquad \text{where } \mathcal{C}_p = \begin{cases} (p-1)^{-1}, & \text{if } 1 < p \le 2, \\ \sqrt{p-1}, & \text{if } p > 2. \end{cases} \tag{38}$$

It is convenient to use m-dependence approximation for processes with the form (6). For $t \in \mathbb{Z}$, define $\mathcal{F}_t = \langle \varepsilon_t, \varepsilon_{t+1}, \ldots \rangle$ be the σ-field generated by the innovations $\varepsilon_t, \varepsilon_{t+1}, \ldots$, and the projection operator $\mathcal{H}_t(\cdot) = \mathbb{E}(\cdot|\mathcal{F}_t)$ and $\mathcal{P}_t(\cdot) = \mathcal{H}_t(\cdot) - \mathcal{H}_{t+1}(\cdot)$. Observe that $(\mathcal{P}_{-t}(\cdot))_{t\in\mathbb{Z}}$ is a martingale difference sequence with respect to the filtration $(\mathcal{F}_{-t})$. For $m \ge 0$, define $\tilde{X}_t = \mathcal{H}_{t-m} X_t$; then $\|X_t - \tilde{X}_t\|_p \le \mathcal{C}_p \Psi_p(m+1)$ [see Proposition 1 of Xiao and Wu (2011) for a proof], and $(\tilde{X}_t)_{t\in\mathbb{Z}}$ is an $(m+1)$-dependent sequence.

LEMMA 8. *Assume* $\mathbb{E}X_t = 0$ *and* $p > 1$. *For* $m \ge 0$, *define* $\tilde{X}_t = \mathcal{H}_{t-m} X_t = \mathbb{E}(X_t|\mathcal{F}_{t-m})$. *Let* $\tilde{\delta}_p(\cdot)$ *be the physical dependence measure for the sequence* $(\tilde{X}_t)$. *Then*

$$\|\mathcal{P}_0 X_t\|_p \le \delta_p(t) \quad \text{and} \quad \tilde{\delta}_p(t) \le \delta_p(t), \tag{39}$$

$$|\gamma_k| \le \zeta_2(k) \qquad \text{where } \zeta_p(k) := \sum_{j=0}^{\infty} \delta_p(j)\delta_p(j+k), \tag{40}$$

$$\left\| \sum_{s,t=1}^{T} c_{s,t}(X_s X_t - \gamma_{s-t}) \right\|_{p/2} \le 4\mathcal{C}_{p/2}\mathcal{C}_p \Theta_p^2 \mathcal{B}_T \sqrt{T} \qquad \text{when } p \ge 4, \tag{41}$$

$$\left\| \sum_{t=1}^{T} c_t(X_t - \tilde{X}_t) \right\|_p \le \mathcal{C}_p \mathcal{A}_T \Theta_p(m+1) \qquad \text{when } p \ge 2, \tag{42}$$

where

$$\mathcal{A}_T = \left(\sum_{t=1}^{T} |c_t|^2 \right)^{1/2} \quad \text{and} \quad \mathcal{B}_T^2 = \max\left\{ \max_{1\le t\le T} \sum_{s=1}^{T} c_{s,t}^2, \max_{1\le s\le T} \sum_{t=1}^{T} c_{s,t}^2 \right\}.$$

7. Large deviations for quadratic forms. In this section we prove a result on probabilities of large deviations of quadratic forms of stationary processes, which take the form

$$Q_T = \sum_{1\le s\le t\le T} a_{s,t} X_s X_t.$$

The coefficients $a_{s,t} = a_{T,s,t}$ may depend on T, but we suppress T from subscripts for notational simplicity. Throughout this section we assume that $\sup_{s,t} |a_{s,t}| \le 1$, and $a_{s,t} = 0$ when $|s - t| > B_T$, where B_T satisfies $B_T \to \infty$, and $B_T = O(T^\gamma)$ for some $0 < \gamma < 1$.

Large deviations for quadratic forms of stationary processes have been extensively studied in the literature. Bryc and Dembo (1997) and Bercu, Gamboa and Rouault (1997) obtained the *large deviation principle* [Dembo and Zeitouni (1998)] for Gaussian processes. Gamboa, Rouault and Zani (1999) considered the functional large deviation principle. Bercu, Gamboa and Lavielle (2000) obtained a more accurate expansion of the tail probabilities. Zani (2002) extended the results of Bercu, Gamboa and Rouault (1997) to locally stationary Gaussian processes. In fact, our result is more relevant to the so-called *moderate deviations* according to the terminology of Dembo and Zeitouni (1998). Bryc and Dembo (1997) and Kakizawa (2007) obtained *moderate deviation principles* for quadratic forms of Gaussian processes. Djellout, Guillin and Wu (2006) studied moderate deviations of periodograms of linear processes. Bentkus and Rudzkis (1976) considered the Cramér-type moderate deviation for spectral density estimates of Gaussian processes; see also Saulis and Statulevičius (1991). Liu and Shao (2010) derived the Cramér-type moderate deviation for maxima of periodograms under the assumption that the process consists of i.i.d. random variables.

For our purpose, on one hand, we do not need a result that is as precise as the moderate deviation principle or the Cramér-type moderate deviation. On the other hand, we need an upper bound for the tail probability under less restrictive conditions. Specifically, we would like to relax the Gaussian, linear or i.i.d. assumptions which were made in the precedent works. Rudzkis (1978) provided a result in this fashion under the assumption of boundedness of the cumulant spectra up to a finite order. While this type of assumption holds under certain mixing conditions, the latter themselves are not easy to verify in general and many well-known examples are not strong mixing [Andrews (1984)]. We mean to impose alternative conditions through physical dependence measures, which are easy to use in many applications [Wu (2005)]. Furthermore, our result can be sharper; see Remark 11.

Our main tool is the m-dependence approximation. In the next lemma we use dependence measures to bound the $\mathcal{L}^p$ norm of the distance between Q_T and the m-dependent version $\tilde{Q}_T$. The proof and a few remarks on the optimality of the result are given in the supplementary article [Xiao and Wu (2012)].

LEMMA 9. *Assume $X_t \in \mathcal{L}^p$ with $p \geq 4$, $\mathbb{E}X_t = 0$ and $\Theta_p < \infty$. Let $\tilde{X}_t = \mathcal{H}_{t-m_T} X_t$ and $\tilde{Q}_T = \sum_{1\leq s\leq t\leq T} a_{s,t}\tilde{X}_s\tilde{X}_t$; then*

$$\begin{aligned}\|\mathbb{E}_0 Q_T - \mathbb{E}_0\tilde{Q}_T\|_{p/2} &\leq 4\Theta_p(m_T)^2 + 11(p-2)\Theta_p\sqrt{TB_T}\Theta_p(m_T)\\ &\quad + (p-2)\sqrt{TB_T}\big[3\Theta_p(\lfloor m_T/2\rfloor)\Delta_p(m_T) + 3\Theta_p(m_T)\Delta_p(\lfloor m_T/2\rfloor)\big].\end{aligned}$$

The following theorem is the main result of this section.

THEOREM 10. *Assume $X_t \in \mathcal{L}^p$, $p > 4$, $\mathbb{E}X_t = 0$, and $\Theta_p(m) = O(m^{-\alpha})$. Set $c_p = (p+4)e^{p/4}\Theta_4^2$. For any $M > 1$, let $x_T = 2c_p\sqrt{TMB_T\log B_T}$. Assume* *that $B_T \to \infty$ and $B_T = O(T^\gamma)$ for some $0 < \gamma < 1$. Then for any $\gamma < \beta < 1$, there exists a constant $C_{p,M,\beta} > 0$ such that*

$$\begin{aligned} &P(|\mathbb{E}_0 Q_T| \geq x_T) \\ &\quad \leq C_{p,M,\beta} x_T^{-p/2} (\log T)[(TB_T)^{p/4} T^{-\alpha\beta p/2} + TB_T^{p/2-1-\alpha\beta p/2} + T] \\ &\qquad + C_{p,M,\beta} B_T^{-M}. \end{aligned}$$

REMARK 11. Rudzkis (1978) proved that if $p = 4k$ for some $k \in \mathbb{N}$, then

$$P(|\mathbb{E}_0 Q_T| \geq x_T) \leq C x_T^{-p/2} (TB_T)^{p/4},$$

which can be obtained by using Markov inequality and (41) under our framework. The upper bound given in Theorem 10 has a smaller order of magnitude. We note that Rudzkis (1978) also proved a stronger exponential inequality under strong moment conditions. They required the existence of every moment and the absolute summability of cumulants of every order.

PROOF OF THEOREM 10. Without loss of generality, assume $B_T \leq T^\gamma$. For $\gamma < \beta < 1$, let $m_T = \lfloor T^{\sqrt{\beta}} \rfloor$, $\tilde{X}_t = \mathcal{H}_{t-m_T} X_t$ and

$$\tilde{Q}_T = \sum_{1 \leq s \leq t \leq T} a_{s,t} \tilde{X}_s \tilde{X}_t.$$

By Lemma 9 and (41), we have

$$\begin{aligned} &P\bigl[|\mathbb{E}_0(Q_T - \tilde{Q}_T)| \geq c_p M^{1/2} \sqrt{TB_T(\log B_T)}\bigr] \\ &\quad \leq C_{p,M} x_T^{-p/2} (TB_T)^{p/4} T^{-\alpha\sqrt{\beta} p/2}. \end{aligned} \tag{43}$$

Split $[1, T]$ into blocks $\mathcal{B}_1, \ldots, \mathcal{B}_{b_T}$ of size $2m_T$, and define

$$Q_{T,k} = \sum_{t \in \mathcal{B}_k} \sum_{1 \leq s \leq t} a_{s,t} \tilde{X}_s \tilde{X}_t.$$

By Corollary 1.7 of Nagaev (1979) and (41), we know for any $M > 1$, there exists a constant $C_{p,M,\beta}$ such that

$$\begin{aligned} &P\bigl[|\mathbb{E}_0 \tilde{Q}_T| \geq c_p \sqrt{TMB_T(\log B_T)}\bigr] \\ &\quad \leq \sum_{k=1}^{b_T} P\biggl(|\mathbb{E}_0 Q_{T,k}| \geq \frac{x_T}{C_{p,M,\beta}}\biggr) + \biggl[\frac{C_{p,M,\beta} T m_T^{-1} (m_T B_T)^{p/4}}{(TB_T)^{p/4}}\biggr]^{C_{p,M,\beta}} \\ &\qquad + C_\beta \exp\biggl\{\frac{c_p^2 (\log B_T)}{(p+4)^2 e^{p/2} \Theta_4^4}\biggr\} \\ &\quad \leq \sum_{k=1}^{b_T} P(|\mathbb{E}_0 Q_{T,k}| \geq x_T / C_{p,M,\beta}) + C_{p,M,\beta}(B_T^{-M} + T^{-M}). \end{aligned} \tag{44}$$

By Lemma 11, we have

$$
\begin{aligned}
&P(|\mathbb{E}_0 Q_{T,k}| \geq x_T / C_{p,M,\beta}) \\
(45)\qquad &\quad \leq C_{p,M,\beta} x_T^{-p/2} (\log T) \\
&\qquad \times \bigl[\bigl(T^{\sqrt{\beta}} B_T\bigr)^{p/4} T^{-\alpha\beta p/2} + T^{\sqrt{\beta}} B_T^{p/2-1-\alpha\beta p/2} + T^{\sqrt{\beta}}\bigr].
\end{aligned}
$$

Combining (43), (44) and (45), the proof is complete. □

LEMMA 11. *Assume $X_t \in \mathcal{L}^p$ with $p > 4$, $\mathbb{E}X_t = 0$, and $\Theta_p(m) = O(m^{-\alpha})$. If $x_T > 0$ satisfies $T^\delta \sqrt{T B_T} = o(x_T)$ for some $\delta > 0$, then for any $0 < \beta < 1$, there exists a constant $C_{p,\delta,\beta}$ such that*

$$
\begin{aligned}
P(|\mathbb{E}_0 Q_T| \geq x_T) &\leq C_{p,\delta,\beta} x_T^{-p/2} (\log T) \\
&\qquad \times [(T B_T)^{p/4} T^{-\alpha\beta p/2} + T B_T^{p/2-1-\alpha\beta p/2} + T].
\end{aligned}
$$

PROOF. For $j \geq 1$, define $m_{T,j} = \lfloor T^{\beta^j} \rfloor$, $X_{t,j} = \mathcal{H}_{t-m_{T,j}} X_t$ and

$$Q_{T,j} = \sum_{1 \leq s \leq t \leq T} a_{s,t} X_{s,j} X_{t,j}.$$

Let $j_T = \lceil -\log(\log T)/(\log \beta) \rceil$. Note that $m_{T,j_T} \leq e$. By Lemma 9 and (41),

$$(46)\qquad P[|\mathbb{E}_0(Q_T - Q_{T,1})| \geq x_T / j_T] \leq C_{p,\beta} (\log T)^{1/2} x_T^{-p/2} (T B_T)^{p/4} T^{-\alpha\beta p/2}.$$

Let j'_T be the smallest j such that $m_{T,j} < B_T/4$. For $1 \leq j < j'_T$, split $[1, T]$ into blocks $\mathcal{B}_1^{(j)}, \ldots, \mathcal{B}_{b_{T,j}}^{(j)}$ of size $B_T + m_{T,j}$. Define

$$R_{T,j,b} = \sum_{t \in \mathcal{B}_b^{(j)}} \sum_{1 \leq s \leq t} a_{s,t} X_{s,j} X_{t,j} \quad \text{and} \quad R'_{T,j,b} = \sum_{t \in \mathcal{B}_b^{(j)}} \sum_{1 \leq s \leq t} a_{s,t} X_{s,j+1} X_{t,j+1}.$$

By Corollary 1.6 of Nagaev (1979) and (41), we have for any $C > 2$

$$
\begin{aligned}
(47)\qquad P\biggl[|\mathbb{E}_0(Q_{T,j} - Q_{T,j+1})| > \frac{x_T}{2 j_T}\biggr] &\leq \sum_{b=1}^{b_{T,j}} P\biggl[|\mathbb{E}_0(R_{T,j,b} - R'_{T,j,b})| \geq \frac{x_T}{C j_T}\biggr] \\
(48)\qquad &\quad + 2\biggl[\frac{64 C e^2 \Theta_4^4 T B_T j_T^2}{x_T^2}\biggr]^{C/4}.
\end{aligned}
$$

It is clear that for any $M > 1$, there exists a constant $C_{M,\delta,\beta}$ such that the term in (48) is less than $C_{M,\delta,\beta} x_T^{-M}$. For (47), by Lemma 9 and (41)

$$
\begin{aligned}
&\sum_{b=1}^{b_{T,j}} P\biggl[|\mathbb{E}_0(R_{T,j,b} - R'_{T,j,b})| \geq \frac{x_T}{C j_T}\biggr] \\
&\qquad \leq C_{p,\beta} T (m_{T,j})^{-1} \cdot (\log T)^{1/2} \cdot x_T^{-p/2} \cdot (m_{T,j} B_T)^{p/4} \cdot m_{T,j+1}^{-\alpha p/2} \\
&\qquad \leq C_{p,\beta} x_T^{-p/2} \cdot (\log T)^{1/2} T B_T^{p/4} \cdot (m_{T,j})^{p/4-1-\alpha\beta p/2}.
\end{aligned}
$$

Depending on whether the exponent $p/4-1-\alpha\beta p/2$ is positive or not, the term $(m_{T,j})^{p/4-1-\alpha\beta p/2}$ is maximized when $j=1$ or $j=j'_T-1$, respectively, and we have

$$\sum_{b=1}^{b_{T,j}} P\left[|\mathbb{E}_0(R_{T,j,b}-R'_{T,j,b})| \ge \frac{x_T}{Cj_T}\right] \le C_{p,\beta} x_T^{-p/2} \cdot (\log T)^{1/2} \cdot [(TB_T)^{p/4} T^{-\alpha\beta p/2} + TB_T^{p/2-1-\alpha\beta p/2}]. \tag{49}$$

Combining (46), (47), (48) and (49), we have shown that

$$\begin{aligned} &P(|\mathbb{E}_0 Q_T| \ge x_T) \\ &\quad \le P(|\mathbb{E}_0 Q_{T,j'_T}| \ge x_T/2) + C_{p,M,\delta,\beta} x_T^{-M} \\ &\qquad + C_{p,M,\delta,\beta} x_T^{-p/2} (\log T)[(TB_T)^{p/4} T^{-\alpha\beta p/2} + TB_T^{p/2-1-\alpha\beta p/2}]. \end{aligned} \tag{50}$$

To deal with the probability concerning Q_{T,j'_T} in (50), we split $[1,T]$ into blocks $\mathcal{B}_1,\dots,\mathcal{B}_{b_T}$ with size $2B_T$, and define the block sums

$$R_{T,j'_T,b} = \sum_{t\in\mathcal{B}_b} \sum_{1\le s\le t} a_{s,t} X_{s,j'_T} X_{t,j'_T}.$$

Similarly as (47) and (48), there exists a constant $C_{p,M,\delta,\beta} > 2$ such that

$$P(|\mathbb{E}_0 Q_{T,j'_T}| \ge x_T/2) \le \sum_{b=1}^{b_T} P\left(|\mathbb{E}_0 R_{T,j'_T,b}| \ge \frac{x_T}{C_{p,M,\delta,\beta}}\right) + C_{p,M,\delta,\beta} x_T^{-M}.$$

By Lemma 12, we have

$$P(|\mathbb{E}_0 R_{T,j'_T,b}| \ge C_{p,M,\delta,\beta}^{-1} x_T) \le C_{p,M,\delta,\beta} x_T^{-p/2} (\log T)(B_T^{p/2-\alpha\beta p/2} + B_T);$$

and it follows that for some constant $C_{p,\delta,\beta} > 0$,

$$P(|\mathbb{E}_0 Q_{T,j'_T}| \ge x_T/2) \le C_{p,\delta,\beta} x_T^{-p/2} (\log T) T (B_T^{p/2-1-\alpha\beta p/2} + 1). \tag{51}$$

The proof is completed by combining (50) and (51). $\square$

In the next lemma we consider Q_T when the restriction $a_{s,t}=0$ for $|s-t|>B_T$ is removed. To avoid confusion, we use a new symbol. Let

$$R(T,m) = \sum_{1\le s\le t\le T} c_{s,t} (\mathcal{H}_{s-m} X_s)(\mathcal{H}_{t-m} X_t).$$

For $x_T > 0$, define

$$U(T,m,x_T) = \sup_{\{c_{s,t}\}} P[|\mathbb{E}_0 R(T,m)| \ge x_T],$$

where the supremum is taken over all arrays $\{c_{s,t}\}$ such that $|c_{s,t}| \le 1$. We use R_T and $U(T,x_T)$ as shorthands for $R(T,\infty)$ and $U(T,\infty,x_T)$, respectively.

LEMMA 12. *Assume $X_t \in \mathcal{L}^p$ with $p > 4$, $\mathbb{E}X_t = 0$, and $\Theta_p(m) = O(m^{-\alpha})$. If $x_T > 0$ satisfies $T^{1+\delta} = o(x_T)$ for some $\delta > 0$, then for any $0 < \beta < 1$, there exists a constant $C_{p,\delta,\beta}$ such that*

$$P(|\mathbb{E}_0 R_T| \geq x_T) \leq C_{p,\delta,\beta} x_T^{-p/2} (\log T)(T^{p/2-\alpha\beta p/2} + T).$$

PROOF. Let $m_T = \lfloor T^\beta \rfloor$ and $\tilde{R}_T := R(T, m_T)$. By Lemma 9 and (41),

$$P[|\mathbb{E}_0(R_T - \tilde{R}_T)| \geq x_T/2] \leq C_p x_T^{-p/2} T^{p/2-\alpha\beta p/2}.$$

We claim that there exists a constant $C_{p,\delta,\beta}$ such that

$$U(T, m_T, x_T/2) \leq C_{p,\delta,\beta} x_T^{-p/2} (T \log T)(m_T^{p/2-1-\alpha\beta p/2} + 1). \tag{52}$$

Therefore, the proof is complete by using

$$P(|\mathbb{E}_0 R_T| \geq x_T) \leq P[|\mathbb{E}_0(R_T - \tilde{R}_T)| \geq x_T/2] + U(T, m_T, x_T/2).$$

We need to prove the claim (52). Let z_T satisfy $T^{1+\delta} = o(z_T)$. Let $j_T = \lceil -\log(\log T)/(\log \beta) \rceil$, and note that $T^{\beta^{j_T}} \leq e$. Set $y_T = z_T/(2 j_T)$. We consider $U(T, m, z_T)$ for an arbitrary $1 < m < T/4$. Set $X_{t,1} := \mathcal{H}_{t-m} X_t$ and $X_{t,2} := \mathcal{H}_{t-\lfloor m^\beta \rfloor} X_t$. Define

$$Y_{t,1} = \sum_{s=1}^{t-3m-1} c_{s,t} X_{s,1} \quad \text{and} \quad Z_{t,1} = \sum_{s=1 \vee (t-3m)}^{t} c_{s,t} X_{s,1}$$

and $Y_{t,2}$, $Z_{t,2}$ similarly by replacing $X_{s,1}$ with $X_{s,2}$. Observe that $X_{t,k}$ and $Y_{t,l}$ are independent for $k, l = 1, 2$. We first consider $\sum_{t=1}^T (X_{t,1} Z_{t,1} - X_{t,2} Z_{t,2})$. Split $[1, T]$ into blocks $\mathcal{B}_1, \ldots, \mathcal{B}_{b_T}$ with size $4m$, and define $W_{T,b} = \sum_{t \in \mathcal{B}_b} (X_{t,1} Z_{t,1} - X_{t,2} Z_{t,2})$. Let y_T satisfy $y_T < z_T/2$ and $T^{1+\delta/2} = o(y_T)$. Since $W_{T,b}$ and $W_{T,b'}$ are independent if $|b - b'| > 1$, by Corollary 1.6 of Nagaev (1979), (41) and Lemma 9, similarly as (47) and (48), we know for any $M > 1$, there exists a constant $C_{p,M,\delta,\beta}$ such that

$$\begin{aligned} P\left[\left|\mathbb{E}_0\left(\sum_{t=1}^T X_{t,1} Z_{t,1} - X_{t,2} Z_{t,2}\right)\right| \geq y_T\right] & \\ \leq C_{p,M,\delta,\beta} y_T^{-M} + \sum_{b=1}^{b_T} P(|\mathbb{E}_0 W_{T,b}| \geq y_T / C_{M,\delta}) & \qquad (53) \\ \leq C_{p,M,\delta,\beta} y_T^{-M} + C_{p,M,\delta,\beta} y_T^{-p/2} T m^{p/2-1-\alpha\beta p/2}. & \end{aligned}$$

Now we deal with the term $\sum_{t=1}^T (X_{t,1} Y_{t,1} - X_{t,2} Y_{t,2})$. Split $[1, T]$ into blocks $\mathcal{B}_1^*, \ldots, \mathcal{B}_{b_T^*}^*$ with size m. Define $R_{T,b} = \sum_{t \in \mathcal{B}_b^*} (X_{t,1} Y_{t,1} - X_{t,2} Y_{t,2})$. Let ξ_b be the σ-fields generated by $\{\varepsilon_{l_b}, \varepsilon_{l_b - 1}, \ldots\}$, where $l_b = \max\{\mathcal{B}_b^*\}$. Observe that

$(R_{T,b})_{b \text{ is odd}}$ is a martingale sequence with respect to $(\xi_b)_{b \text{ is odd}}$, and so are $(R_{T,b})_{b \text{ is even}}$ and $(\xi_b)_{b \text{ is even}}$. By Lemma 1 of Haeusler (1984) we know for any $M > 1$, there exists a constant $C_{M,\delta}$ such that

$$\begin{aligned}
&P\left[\left|\sum_{t=1}^{T}(X_{t,1}Y_{t,1} - X_{t,2}Y_{t,2})\right| \geq y_T\right] \\
&\qquad \leq C_{M,\delta}y_T^{-M} + 4P\left[\sum_{b=1}^{b_T^*}\mathbb{E}(R_{T,b}^2|\xi_{b-2}) > \frac{y_T^2}{(\log y_T)^{3/2}}\right] \\
&\qquad\quad + \sum_{b=1}^{b_T^*}P\left[|R_{T,b}| \geq \frac{y_T}{\log y_T}\right] \\
&\qquad =: I_T + II_T + III_T.
\end{aligned} \tag{54}$$

Since $(X_{t,1}, X_{t,2})$ and $(Y_{t,1}, Y_{t,2})$ are independent, $R_{T,b}$ has finite pth moment. Using similar arguments as Lemma 9, we have

$$\|R_{T,b}\|_p \leq C_p(mT)^{p/2}m^{-\alpha\beta p};$$

and it follows that

$$III_T \leq C_p y_T^{-p}(\log y_T)^p T^{p/2+1}m^{p/2-1-\alpha\beta p}. \tag{55}$$

For the second term, let $r_{s-t,k} = \mathbb{E}(X_{s,k}X_{t,k})$ for $k = 1, 2$; we have

$$\begin{aligned}
\sum_{b=1}^{b_T^*}\mathbb{E}(R_{T,b}^2|\xi_{b-2}) &\leq 2\sum_{b=1}^{b_T^*}\left[\sum_{s,t\in\mathcal{B}_b^*}(r_{s-t,1}Y_{s,1}Y_{t,1} + r_{s-t,2}Y_{s,2}Y_{t,2})\right] \\
&= \sum_{1\leq s\leq t\leq T}a_{s,t,1}X_{s,1}X_{t,1} + \sum_{1\leq s\leq t\leq T}a_{s,t,2}X_{s,2}X_{t,2}.
\end{aligned} \tag{56}$$

By (39) and (40), we know $\sum_{l\in\mathbb{Z}}|r_{l,k}| < \infty$ for $k = 1, 2$, and hence $|a_{s,t,k}| \leq CT$. It follows that the expectations of the two terms in (56) are all less than CT^2, and

$$II_T \leq C_\beta U\left[T, m, \frac{y_T^2}{T(\log y_T)^2}\right] + C_\beta U\left[T, \lfloor m^\beta\rfloor, \frac{y_T^2}{T(\log y_T)^2}\right]. \tag{57}$$

Combining (53), (54), (55) and (57), we have shown that $U(T, m, z_T)$ is bounded from above by

$$\begin{aligned}
&U(T, \lfloor m^\beta\rfloor, z_T - 2y_T) \\
&\quad + C_\beta U\left[T, \lfloor m^\beta\rfloor, \frac{y_T^2}{T(\log y_T)^2}\right] + C_\beta U\left[T, m, \frac{y_T^2}{T(\log y_T)^2}\right] \\
&\quad + C_{p,M,\delta,\beta}[y_T^{-M} + y_T^{-p/2}Tm^{p/2-1-\alpha\beta p/2} \\
&\qquad\qquad + y_T^{-p}(\log y_T)^p T^{p/2+1}m^{p/2-1-\alpha\beta p}].
\end{aligned} \tag{58}$$

Since $\sup_{\{c_{s,t}\}} \|\mathbb{E}_0 R_T\|_{p/2} \le C_p T$ by (41), by applying (58) recursively to deal with the last term on the first line of (58) for q times such that $(y_T/T)^{-2^q p} = O[y_T^{-(M+1)}]$, we have

$$U(T, m, z_T) \le C_{p,M,\delta,\beta}\big[U(T, \lfloor m^\beta \rfloor, z_T - 2y_T) + y_T^{-p/2} T m^{p/2-1-\alpha\beta p/2} + y_T^{-p}(\log y_T)^p T^{p/2+1} m^{p/2-1-\alpha\beta p} + y_T^{-M}\big]. \tag{59}$$

Using the preceding arguments similarly, we can show that when $1 \le m \le 3$

$$U[T, m, z_T/(2j_T)] \le C_{M,p,\delta}[z_T^{-p/2}(\log T)T + z_T^{-p}(\log z_T)^{p+1} T^{p/2+1} + z_T^{-M}].$$

The details of the derivation are omitted. Applying (59) recursively for at most $j_T - 1$ times, we have the first bound for $U(T, m, z_T)$,

$$\begin{aligned} U(T, m, z_T) &\le C_{p,M,\delta,\beta}^{j_T}\{U[T, 3, z_T/(2j_T)] + z_T^{-p/2}(\log z_T) T (m^{p/2-1-\alpha\beta p/2} + 1) \\ &\quad + z_T^{-p}(\log z_T)^{p+1} T^{p/2+1}(m^{p/2-1-\alpha\beta p} + 1) + z_T^{-M}\} \\ &\le C_{p,\delta,\beta}^{j_T}(\log z_T)^{p+1}(z_T^{-p/2} T + z_T^{-p} T^{p/2+1})(m^{p/2-1-\alpha\beta p/2} + 1). \end{aligned} \tag{60}$$

Now plugging (60) back into (58) for the last two terms on the first line and using the condition $T^{1+\delta/2} = o(y_T)$, we have

$$U(T, m, z_T) \le U(T, \lfloor m^\beta \rfloor, z_T - 2y_T) + C_{p,\delta,\beta}[y_T^{-p/2} T (m^{p/2-1-\alpha\beta p/2} + 1)]. \tag{61}$$

Again by applying (61) for at most $j_T - 1$ times, we obtain the second bound for $U(T, m, z_T)$:

$$U(T, m, z_T) \le C_{p,\delta,\beta} z_T^{-p/2}(T \log T)(m^{p/2-1-\alpha\beta p/2} + 1).$$

The proof of the claim (52) is complete. □

8. Conclusion. In this paper we use Toeplitz's connection of eigenvalues of matrices and Fourier transforms of their entries, and obtain optimal bounds for tapered covariance matrix estimates by applying asymptotic results of spectral density estimates. Many problems are still unsolved; for example, can we improve the convergence rate of the thresholded estimate in Theorem 5? What is the asymptotic distribution of the maximum eigenvalues of the estimated covariance matrices? We hope that the approach and results developed in this paper can be useful for other high-dimensional covariance matrix estimation problems in time series. Such problems are relatively less studied compared to the well-known theory of random matrices which requires i.i.d. entries or multiple i.i.d. copies.

Acknowledgments. We are grateful to an Associate Editor and the referees for their many helpful comments.

SUPPLEMENTARY MATERIAL

Additional technical proofs (DOI: 10.1214/11-AOS967SUPP; .pdf). We give the proofs of Remark 5 and Lemma 9, as well as a few remarks on Lemma 9.

REFERENCES

ADENSTEDT, R. K. (1974). On large-sample estimation for the mean of a stationary random sequence. *Ann. Statist.* **2** 1095–1107. MR0368354

AN, H. Z., CHEN, Z. G. and HANNAN, E. J. (1983). The maximum of the periodogram. *J. Multivariate Anal.* **13** 383–400. MR0716931

ANDREWS, D. W. K. (1984). Nonstrong mixing autoregressive processes. *J. Appl. Probab.* **21** 930–934. MR0766830

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

BAI, Z. D. and YIN, Y. Q. (1993). Limit of the smallest eigenvalue of a large-dimensional sample covariance matrix. *Ann. Probab.* **21** 1275–1294. MR1235416

BENTKUS, R. and RUDZKIS, R. (1976). Large deviations for estimates of the spectrum of a stationary Gaussian sequence. *Litovsk. Mat. Sb.* **16** 63–77, 253. MR0436510

BERCU, B., GAMBOA, F. and ROUAULT, A. (1997). Large deviations for quadratic forms of stationary Gaussian processes. *Stochastic Process. Appl.* **71** 75–90. MR1480640

BERCU, B., GAMBOA, F. and LAVIELLE, M. (2000). Sharp large deviations for Gaussian quadratic forms with applications. *ESAIM Probab. Stat.* **4** 1–24 (electronic). MR1749403

BICKEL, P. J. and LEVINA, E. (2008a). Covariance regularization by thresholding. *Ann. Statist.* **36** 2577–2604. MR2485008

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

BRYC, W. and DEMBO, A. (1997). Large deviations for quadratic functionals of Gaussian processes. *J. Theoret. Probab.* **10** 307–332. Dedicated to Murray Rosenblatt. MR1455147

BRYC, W., DEMBO, A. and JIANG, T. (2006). Spectral measure of large random Hankel, Markov and Toeplitz matrices. *Ann. Probab.* **34** 1–38. MR2206341

BÜHLMANN, P. and KÜNSCH, H. R. (1999). Block length selection in the bootstrap for time series. *Comput. Statist. Data Anal.* **31** 295–310.

BURKHOLDER, D. L. (1988). Sharp inequalities for martingales and stochastic integrals. *Colloque Paul Lévy sur les Processus Stochastiques* (*Palaiseau*, 1987). *Astérisque* **157-158** 75–94. MR0976214

CAI, T. T., ZHANG, C.-H. and ZHOU, H. H. (2010). Optimal rates of convergence for covariance matrix estimation. *Ann. Statist.* **38** 2118–2144. MR2676885

DEMBO, A. and ZEITOUNI, O. (1998). *Large Deviations Techniques and Applications*, 2nd ed. *Applications of Mathematics* (*New York*) **38**. Springer, New York. MR1619036

DJELLOUT, H., GUILLIN, A. and WU, L. (2006). Moderate deviations of empirical periodogram and non-linear functionals of moving average processes. *Ann. Inst. Henri Poincaré Probab. Stat.* **42** 393–416. MR2242954

EL KAROUI, N. (2005). Recent results about the largest eigenvalue of random covariance matrices and statistical application. *Acta Phys. Polon. B* **36** 2681–2697. MR2188088

EL KAROUI, N. (2008). Operator norm consistent estimation of large-dimensional sparse covariance matrices. *Ann. Statist.* **36** 2717–2756. MR2485011

FREEDMAN, D. A. (1975). On tail probabilities for martingales. *Ann. Probab.* **3** 100–118. MR0380971

FURRER, R. and BENGTSSON, T. (2007). Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants. *J. Multivariate Anal.* **98** 227–255. MR2301751

GAMBOA, F., ROUAULT, A. and ZANI, M. (1999). A functional large deviations principle for quadratic forms of Gaussian stationary processes. *Statist. Probab. Lett.* **43** 299–308. MR1708097

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

GRENANDER, U. and SZEGÖ, G. (1958). *Toeplitz Forms and Their Applications*. Univ. California Press, Berkeley. MR0094840

HAEUSLER, E. (1984). An exact rate of convergence in the functional central limit theorem for special martingale difference arrays. *Z. Wahrsch. Verw. Gebiete* **65** 523–534. MR0736144

HORN, R. A. and JOHNSON, C. R. (1990). *Matrix Analysis*. Cambridge Univ. Press, Cambridge. Corrected reprint of the 1985 original. MR1084815

JOHANSSON, K. (2000). Shape fluctuations and random matrices. *Comm. Math. Phys.* **209** 437–476. MR1737991

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

KAKIZAWA, Y. (2007). Moderate deviations for quadratic forms in Gaussian stationary processes. *J. Multivariate Anal.* **98** 992–1017. MR2325456

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

LIN, Z. and LIU, W. (2009). On maxima of periodograms of stationary processes. *Ann. Statist.* **37** 2676–2695. MR2541443

LIU, W. and SHAO, Q.-M. (2010). Cramér-type moderate deviation for the maximum of the periodogram with application to simultaneous tests in gene expression time series. *Ann. Statist.* **38** 1913–1935. MR2662363

LIU, W. and WU, W. B. (2010). Asymptotics of spectral density estimates. *Econometric Theory* **26** 1218–1245. MR2660298

MARČENKO, V. A. and PASTUR, L. A. (1967). Distribution of eigenvalues in certain sets of random matrices. *Mat. Sb.* (*N.S.*) **72** 507–536. MR0208649

MCMURRY, T. L. and POLITIS, D. N. (2010). Banded and tapered estimates for autocovariance matrices and the linear process bootstrap. *J. Time Series Anal.* **31** 471–482. MR2732601

NAGAEV, S. V. (1979). Large deviations of sums of independent random variables. *Ann. Probab.* **7** 745–789. MR0542129

PELIGRAD, M. and WU, W. B. (2010). Central limit theorem for Fourier transforms of stationary processes. *Ann. Probab.* **38** 2009–2022. MR2722793

POLITIS, D. N. (2003). Adaptive bandwidth choice. *J. Nonparametr. Stat.* **15** 517–533. MR2017485

POLITIS, D. N., ROMANO, J. P. and WOLF, M. (1999). *Subsampling*. Springer, New York. MR1707286

RIO, E. (2009). Moment inequalities for sums of dependent random variables under projective conditions. *J. Theoret. Probab.* **22** 146–163. MR2472010

RUDZKIS, R. (1978). Large deviations for estimates of the spectrum of a stationary sequence. *Litovsk. Mat. Sb.* **18** 81–98, 217. MR0519099

SAULIS, L. and STATULEVIČIUS, V. A. (1991). *Limit Theorems for Large Deviations*. *Mathematics and Its Applications* (*Soviet Series*) **73**. Kluwer, Dordrecht. Translated and revised from the 1989 Russian original. MR1171883

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

SOLO, V. (2010). On random matrix theory for stationary processes. In *IEEE International Conference on Acoustics Speech and Signal Processing* (*ICASSP*) 3758–3761. IEEE, Piscataway, NJ.

TOEPLITZ, O. (1911). Zur Theorie der quadratischen und bilinearen Formen von unendlichvielen Veränderlichen. *Math. Ann.* **70** 351–376. MR1511625

TONG, H. (1990). *Nonlinear Time Series*. *Oxford Statistical Science Series* **6**. Oxford Univ. Press, New York. MR1079320

TRACY, C. A. and WIDOM, H. (1994). Level-spacing distributions and the Airy kernel. *Comm. Math. Phys.* **159** 151–174. MR1257246

TURKMAN, K. F. and WALKER, A. M. (1984). On the asymptotic distributions of maxima of trigonometric polynomials with random coefficients. *Adv. in Appl. Probab.* **16** 819–842. MR0766781

TURKMAN, K. F. and WALKER, A. M. (1990). A stability result for the periodogram. *Ann. Probab.* **18** 1765–1783. MR1071824

WIENER, N. (1949). *Extrapolation, Interpolation, and Smoothing of Stationary Time Series. With Engineering Applications*. MIT Press, Cambridge, MA. MR0031213

WOODROOFE, M. B. and VAN NESS, J. W. (1967). The maximum deviation of sample spectral densities. *Ann. Math. Statist.* **38** 1558–1569. MR0216717

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

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

XIAO, H. and WU, W. B. (2011). Asymptotic inference of autocovariances of stationary processes. Available at arXiv:1105.3423.

XIAO, H. and WU, W. B. (2012). Supplement to "Covariance matrix estimation for stationary time series." DOI:10.1214/11-AOS967SUPP.

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

ZANI, M. (2002). Large deviations for quadratic forms of locally stationary processes. *J. Multivariate Anal.* **81** 205–228. MR1906377

ZYGMUND, A. (2002). *Trigonometric Series*. *Vols I*, *II*, 3rd ed. Cambridge Univ. Press, Cambridge. MR1963498

DEPARTMENT OF STATISTICS
UNIVERSITY OF CHICAGO
5734 S. UNIVERSITY AVE
CHICAGO, ILLINOIS 60637
USA
E-MAIL: xiao@galton.uchicago.edu
wbwu@galton.uchicago.edu