Spectral Inference for High Dimensional Time Series

Chi Zhang, *Student Member, IEEE,* and Danna Zhang, *Member, IEEE,*

***Abstract*—Spectral analysis plays a fundamental role in the study of time series. While there is a well-developed asymptotic theory for spectral density estimate in low-dimensional cases, a corresponding distributional theory for high-dimensional time series is still lacking. This paper aims to fill this gap by introducing a comprehensive inference theory for the spectral density estimate of high-dimensional time series with possibly nonlinear generating systems and non-Gaussian distributions. Our result is built across different dimensions and frequencies and can serve as a versatile tool for addressing various time series inference challenges. Additionally, we present two distinct resampling methods aimed at practical implementation of high-dimensional spectral inference, each accompanied by a theoretical justification of its validity.**

***Index Terms*—Asymptotic distribution, high dimensional time series, resampling, simultaneous inference, spectral density.**

I. Introduction

THE information era has witnessed an explosion in the collection of high-dimensional time series data across a wide range of areas. Spectral density, as an adequate description of serial auto- and cross-correlation in the frequency domain, plays an essential role in helping understanding the dynamics in serially correlated data without necessarily needing to develop complex parametric models in high dimensions. It has been widely applied in measuring dynamic co-movement of multiple economic time series [1], detecting functional brain network connectivity in neuroscience [2], [3], and investigating multichannel event-related potentials in cognitive science [4]. Besides, it is a powerful tool for waveform analysis in geophysics [5], wave analysis in oceanography [6], [7], signal analysis in weather radar [8], flow analysis in transportation engineering [9], climate oscillation analysis [10] and studies in many other disciplines. Accurate estimation of the spectral density matrix is also a crucial problem in methodological studies. For example, spectral and cross-spectral density estimates can be used in linear prediction of vector-valued stationary processes [11], parameter estimation of generalized least squares type for dynamic models [12], [13], instrumental variable estimation of general linear systems [14], and estimation of generalized dynamic factor models [15], [16]. Also, it permits optimal linear interpolation of multiple missing values in multivariate time series [17].

To fix the idea, for $p \in \mathbb{N}^+$ and stationary process $X_t \in \mathbb{R}^p$, let $\Gamma(k) = (\Gamma_{ij}(k))_{i,j=1}^p = \mathrm{Cov}(X_0, X_k)$ be the autocovariance matrix function. For all $1 \le i, j \le p$, assume that $\sum_{k\in\mathbb{Z}} |\Gamma_{ij}(k)| < \infty$. The spectral density matrix function is defined by

$$F(\theta) = \frac{1}{2\pi} \sum_{k\in\mathbb{Z}} \Gamma(k) e^{-\iota k\theta}, \text{ where } \iota = \sqrt{-1}.$$

Properties of spectral estimates have been extensively explored for one dimensional stationary processes in many classical time series textbooks; see, for example [18], [19], [20], [21], [22] among others. See also [23] and [24] for more references. Without loss of generality, we assume $\mathbb{E}X_t = 0$ throughout the paper. It is well known that the matrix-valued periodogram

$$\widehat{G}_T(\theta) = \frac{1}{2\pi T} \Big(\sum_{t=1}^T X_t e^{-\iota t\theta} \Big) \Big(\sum_{t=1}^T X_t^\top e^{\iota t\theta} \Big)$$

is an asymptotically unbiased but inconsistent estimator of $F(\theta)$. Consistent estimators of the spectral density function can be obtained by tapering the autocovariance estimator or equivalently by smoothing the periodogram. One natural choice is Welch's [25] non-overlapped segment averaging spectral density estimator given by

$$\widehat{F}(\theta) = \frac{1}{2\pi T} \sum_{r=1}^{T/B} \Big(\sum_{t=(r-1)B+1}^{rB} X_t e^{-\iota t\theta} \Big) \Big(\sum_{t=(r-1)B+1}^{rB} X_t^\top e^{\iota t\theta} \Big) \tag{1}$$

where $B = o(T)$ is the window size and for notational convenience it is assumed that T is divisible by B; see [1], [19], [26] for more lag-window type variants using different smoothing functions.

In this work, we shall consider a general framework of high-dimensional stationary processes of a causal form

$$X_t = G(\mathcal{F}_t) = (X_{t1}, \ldots, X_{tp})^\top, \tag{2}$$

where $\mathcal{F}_t = (\varepsilon_t, \varepsilon_{t-1}, \ldots)$, ε_t, $t \in \mathbb{Z}$, are i.i.d. random elements, $G(\cdot) = (g_1(\cdot), \ldots, g_p(\cdot))^\top$ is an $\mathbb{R}^p$-valued measurable function such that $X_t = G(\mathcal{F}_t)$ is a well-defined random vector. In the univariate case where $p = 1$, framework (2) provides a natural paradigm for both linear and nonlinear time series models and represents a huge class of stationary processes which appear frequently in practice; see [27], [28], [29], [30], [31] among many others. Estimation and inference theory on spectral density has been well developed by [23] and [24] in the one dimensional case and [32] in the multivariate case. By allowing the data-generating function to be $\mathbb{R}^p$-valued where p may diverge to infinity, it extends a large number of existing low dimensional stationary processes into their

Chi Zhang is with the Department of Mathematics, University of California, San Diego, CA 92131, USA (email: chz018@ucsd.edu).

Danna Zhang is with the Department of Mathematics, University of California, San Diego, CA 92131, USA (email: daz076@ucsd.edu).

This work was supported by NSF DMS-1916290 and Hellman fellowship.

high-dimensional counterparts in a natural way; see [33], [34] and [35] for examples. In the high-dimensional case, [36], [37] and [38] concerned the convergence of spectral density matrix estimates under different moment and dependence assumptions on the underlying process. [39] established the convergence in distribution of the maximum sample coherence for Gaussian time series to the Gumbel distribution under the high-dimensional regime when $B = O(T^\rho)$ for some $0 < \rho < 1$ and $p/B \to c$ for some constant $c \in (0, \infty)$. It is still an open problem on how to perform simultaneous inference of spectral density matrix for high-dimensional non-Gaussian time series. The main goal of this paper is to develop a comprehensive distributional theory for the high-dimensional spectral density estimate. In particular, we shall establish the asymptotic normality for point-wise inference and a Gaussian approximation result for simultaneous inference. We take into account all of the following features that may arise in real-world applications involving time series data: (i) possibly nonlinear generating system, (ii) temporal dependence and cross-sectional dependence, (iii) high-dimensional data and (iv) non-Gaussian distribution. As another important contribution, we introduce two practical implementation approaches for simultaneous inference of high-dimensional spectral density matrix: one is Gaussian multiplier resampling method and another is subsampling. There has been an extensive study on resampling methods for low dimensional time series including the widely used moving block bootstrap [40], [41] and subsampling [42], [43]. The validity of these methods in high dimensions is still unknown. In the recent decade, Gaussian multiplier boostrap was introduced in [44] for high-dimensional mean vector inference. As an extension, [45] proposed a blockwise multiplier bootstrap. We shall provide rigorous theoretical justification on resampling approaches to infer spectral density for high-dimensional time series.

The paper is organized as follows. In Section II, we introduce the functional dependence measures for high-dimensional time series and establish the convergence and point-wise asymptotic normality of spectral estimates. In Section III, we study simultaneous inference of spectral density by establishing a Gaussian approximation result for the maximum deviation of the spectral density estimate over dimensions and frequencies, which could be used to tackle a variety of time series inference problems. We also introduce resampling methods to perform statistical inference in practice and provide theoretical justification on the validity. In Section IV, we discuss potential extensions of our results. Section V presents a simulation study and real data analysis. The proofs of the main results are provided in Section VI and part of the proofs are relegated to the Appendix.

We first introduce some notation. For a random variable X and $q \geq 1$, we define $\|X\|_q = (\mathbb{E}|X|^q)^{1/q}$. Denote the operator $\mathbb{E}_0$ with $\mathbb{E}_0(X) := X - \mathbb{E}X$. Define the projection operators $\mathcal{P}_t \cdot = \mathbb{E}(\cdot|\mathcal{F}_t) - \mathbb{E}(\cdot|\mathcal{F}_{t-1})$ and $\mathcal{P}^t \cdot = \mathbb{E}(\cdot|\mathcal{F}^t) - \mathbb{E}(\cdot|\mathcal{F}^{t+1})$, where $\mathcal{F}_t = (\varepsilon_t, \varepsilon_{t-1}, \ldots)$ and $\mathcal{F}^t = (\varepsilon_t, \varepsilon_{t+1}, \ldots)$. For a complex number $a \in \mathbb{C}$, denote the real and imaginary part of a as $\mathsf{Re}(a)$ and $\mathsf{Im}(a)$ respectively. For a matrix $A = (a_{ij})_{i,j} \in \mathbb{C}^{m \times n}$, define the element-wise ℓ_∞ norm $|A|_\infty = \max_{i,j} |a_{ij}|$. Denote the transpose of A as $A^\top$ and the conjugate transpose of A as A^H. For two real numbers, set $x \vee y = \max(x, y)$ and $x \wedge y = \min(x, y)$. We use $C, C_1, C_2, \cdots$ to denote positive constants whose values may differ from place to place. A constant with a symbolic subscript is used to emphasize the dependence of the value on the subscript. Throughout the paper, we use r, s, t to denote time indexes and use i, j, h, l to denote dimension indexes.

II. Consistency and Asymptotic Normality of Spectral Estimate

To develop an asymptotic theory for spectral estimates, we need to introduce appropriate dependence measures. Assume that $\max_{1 \leq j \leq p} \|X_{tj}\|_q < \infty$ for some $q \geq 2$. Following the idea in [31], for $t \geq 0$ and $1 \leq j \leq p$, we define functional dependence measure

$$\delta_{t,q,j} = \|g_j(\varepsilon_t, \varepsilon_{t-1}, \ldots) - g_j(\varepsilon_t, \ldots, \varepsilon_1, \varepsilon_0^*, \varepsilon_{-1}, \ldots)\|_q, \quad (3)$$

where ε_0^* is an i.i.d. copy of ε_0. We can observe that $\delta_{t,q,j}$ measures the distance between the original process $X_{tj} = g_j(\varepsilon_t, \varepsilon_{t-1}, \ldots)$ and the decoupled process $g_j(\varepsilon_t, \ldots, \varepsilon_1, \varepsilon_0^*, \varepsilon_{-1}, \ldots)$ with ε_0 replaced by ε_0^* and other innovations kept the same. Thus, it quantifies the effect of the innovation vector ε_0 on the process (X_{tj}), which can be interpreted as a possibly nonlinear impulse response function. In this sense, we call $\delta_{t,q,j}$ the dependence measure for the j-th component process $(X_{\cdot j})$ at lag t and moment order q. By convention, $\delta_{t,q,j} = 0$ for $t < 0$. For high-dimensional time series, it is challenging to depict the dependence structure, because both the temporal and the cross-sectional dependence need to be concerned. A main advantage of the representation (3) is that it lets us define physically meaningful and easily workable dependence measures, even for high-dimensional cases. Apparently, $\delta_{t,q,j}$ measures the temporal dependence at lag t. As each component X_{tj} is dependent on the p-variate vectors $X_{t-1}, X_{t-2}, \ldots$, $\delta_{t,q,j}$ incorporates the cross-sectional dependence as well.

Equipped with the dependence measures (3), we define the q-th dependence adjusted moment: assume that there exists some constant $\rho \in (0, 1)$ such that

$$\|X_{\cdot j}\|_q = \sup_{m \geq 0} \rho^{-m} \Delta_{m,q,j} < \infty, \text{where } \Delta_{m,q,j} = \sum_{t=m}^{\infty} \delta_{t,q,j}. \quad (4)$$

Here a geometric moment contracting (GMC) condition is imposed on the cumulative (tail) dependence measure $\Delta_{m,q,j} = \sum_{t=m}^{\infty} \delta_{t,q,j}$ by noting that $\Delta_{m,q,j} \leq \|X_{\cdot j}\|_q \cdot \rho^m$ for all $m \geq 0$. We can interpret the quantity $\|X_{\cdot j}\|_q$ as the q-th moment of the component process $(X_{\cdot j})$ by taking dependence into account. Elementary calculations show that, if X_{tj}, $t \in \mathbb{Z}$, are i.i.d., then $\|X_{tj}\|_q \leq \|X_{\cdot j}\|_q \leq 2\|X_{tj}\|_q$, suggesting that the dependence adjusted moment is equivalent to the classical $\mathcal{L}^q$ moment for the independent case. For high-dimensional time series, it could happen that the quantity $\|X_{\cdot j}\|_q$ may be expressed as a function of p and may increase with p. In the following examples, we work out a bound for the dependence measure of high-dimensional time series. This is a key step in applying our theorems.

Example 1 (High Dimensional Linear Processes)**.** *Let $\varepsilon_t = (\varepsilon_{t1}, \ldots, \varepsilon_{tp})$, $t \in \mathbb{Z}$, be i.i.d innovation vectors, where ε_{tj} are i.i.d. random variables with mean 0 and $\mu_q = \|\varepsilon_{ti}\|_q < \infty$ for some $q \geq 2$. Let A_k, $k \geq 0$, be $p \times p$ real coefficient matrices such that $\sum_{k=0}^{\infty} tr(A_k A_k^\top) < \infty$. By Kolmogorov's three-series theorem, the p-dimensional linear process*

$$X_t = \sum_{k=0}^{\infty} A_k \varepsilon_{t-k} \tag{5}$$

is well-defined. Many researchers have worked on this model recently; see for example, [46], [47], [48], [49], [50] among others. It includes the vector AR processes [51], [52], [53], [54] and MA processes [55], [56], [57]. Denote the (j, j')-th entry of A_t by $A_{t,jj'}$. We assume that there exist some $0 < \rho < 1$ and $0 < K_p < \infty$ such that $\max_{1 \leq j \leq p}(\sum_{j'=1}^{p} A_{t,jj'}^2)^{1/2} \leq K_p \rho^t$ for all $t \geq 0$. By Theorem 2.1 of [58], we can obtain

$$\delta_{t,q,j} = \| \sum_{j'=1}^{p} A_{t,jj'}(\varepsilon_{0j'} - \varepsilon_{0j'}^*)\|_q \leq \sqrt{q}\mu_q K_p \rho^t.$$

Example 2 (High dimensional autoregressive conditional heteroskedasticity models)**.** *We consider the vector ARCH process $X_t = (X_{t1}, \ldots, X_{tp})^\top$ with*

$$X_{tj} = \varepsilon_{tj}(a_j + X_{t-1}^\top A_j X_{t-1})^{1/2}, \quad j = 1, \ldots, p, \tag{6}$$

where ε_{tj} are i.i.d. random variables with mean 0 and $\mu_q = \|\varepsilon_{tj}\|_q < \infty$ for some $q \geq 2$, a_j are positive constants, and A_j are $p \times p$ nonnegative-definite matrices with the spectral norms denoted by $\lambda_j = \|A_j\|_2$. Denote $\varepsilon_t = (\varepsilon_{t1}, \ldots, \varepsilon_{tp})$ and $X_t = \boldsymbol{f}(X_{t-1}, \varepsilon_t)$. Since for any $X, Y \in \mathbb{R}^p$,

$$|(a_j + X^\top A_j X)^{1/2} - (a_j + Y^\top A_j Y)^{1/2}|^2 \leq \lambda_j |X - Y|_2^2, \tag{7}$$

it follows that

$$|\boldsymbol{f}(X, \varepsilon_t) - \boldsymbol{f}(Y, \varepsilon_t)|_2 \leq (\sum_{j=1}^{p} \lambda_j \varepsilon_{tj}^2)^{1/2} |X - Y|_2.$$

Assume that $\rho := \|(\sum_{j=1}^{p} \lambda_j \varepsilon_{tj}^2)^{1/2}\|_q < 1$. Denote $K_{p,q} = \|(\sum_{j=1}^{p} a_j \varepsilon_{tj}^2)^{1/2}\|_q = \||\boldsymbol{f}(\boldsymbol{0}, \varepsilon_t)|_2\|_q$. By the arguments for Theorem 2 in [59], the recursion (6) *admits a stationary solution $X_t = \boldsymbol{g}(\varepsilon_t, \varepsilon_{t-1}, \ldots)$. By* (6) *and* (7)*, we can compute the dependence measure*

$$\delta_{t,q,j} \leq \mu_q K_{p,q} \sqrt{\lambda_j} \rho^t / (1 - \rho).$$

Before proceeding, we state the main assumption required in studying the asymptotic properties of the spectral density estimator (1).

Assumption 1. *There exists some constant $\alpha \geq 0$ such that*

$$\|X_{\cdot j}\|_{\psi_\alpha} := \sup_{q \geq 2} q^{-\alpha} \|X_{\cdot j}\|_q < \infty.$$

Denote $\|X_{\cdot}\|_{\psi_\alpha} = \max_{1 \leq j \leq p} \|X_{\cdot j}\|_{\psi_\alpha}$.

In Assumption 1, it is assumed that X_{tj} has finite moment with any order. For $\alpha \geq 0$, $\|X_{\cdot j}\|_{\psi_\alpha}$ is interpreted as the dependence adjusted exponential type norm for the j-th component process and $\|X_{\cdot}\|_{\psi_\alpha}$ is interpreted as the uniform dependence adjusted norm by taking the maximum over all the component processes, which accounts for high dimensionality. By this definition, if X_{tj}, $t \in \mathbb{Z}$, are i.i.d., $\|X_{\cdot j}\|_{\psi_\alpha}$ is equivalent to the classical sub-Gaussian norm when $\alpha = 1/2$ or sub-exponential norm when $\alpha = 1$, due to the equivalence of $\|X_{\cdot j}\|_q$ and $\|X_{tj}\|_q$. The parameter α measures how fast $\|X_{\cdot j}\|_q$ increases with q. With this assumption, we are able to deal with tails which are heavier than Gaussian tails: the larger α is, the heavier the tail is. It is worth noting that Assumption 1 imposes some additional restrictions on the underlying process, even when the innovation has finite moments of all orders. These restrictions are characterized by the parameter α, which quantifies the rate at which the q-th dependence-adjusted norm $\|X_{\cdot j}\|_q$ increases with q. Specifically, under the setting of Example 1, it requires $\mu_q \leq q^{\alpha - 1/2}(1 - \rho)K_p^{-1}\|X_{\cdot j}\|_{\psi_\alpha}$ for all $1 \leq j \leq p$. In the case of Example 2, the condition becomes $\mu_q K_{p,q} \leq q^\alpha (1 - \rho)^2 \lambda_j^{-1/2} \|X_{\cdot j}\|_{\psi_\alpha}$ for all $1 \leq j \leq p$.

For notational convenience, we assume B is an even integer and $B \geq 4$. Define the set of discrete Fourier frequencies in $[0, \pi]$

$$\Theta = \{2k\pi/B : k = 0, 1, 2, \ldots, (B/2)\}.$$

Theorem 1 below provides a uniform non-asymptotic tail probability bound for $\max_{\theta \in \Theta} |\widehat{F}(\theta) - \mathbb{E}\widehat{F}(\theta)|_\infty$, the maximum deviation of the spectral estimate from its expectation across different dimensions and frequencies. Since the process can be nonlinear, non-Gaussian and high-dimensional, it can be quite involved to derive a tail probability inequality for $\max_{\theta \in \Theta} |\widehat{F}(\theta) - \mathbb{E}\widehat{F}(\theta)|_\infty$. In our framework, equipped with functional dependence measures, we can have a closed form of the upper bound depending on the sample size T, window size B, dimension p, parameter α and dependence adjusted norm $\|X_{\cdot}\|_{\psi_\alpha}$.

Theorem 1. *Let Assumption 1 be satisfied for some $\alpha \geq 0$. Then there exists some constant C_α such that for any $x > 0$,*

$$\mathbb{P}\Big(\max_{\theta \in \Theta} |\widehat{F}(\theta) - \mathbb{E}\widehat{F}(\theta)|_\infty \geq x\Big) \lesssim Bp^2 \cdot \exp\Big[-C_\alpha \Big(\frac{\sqrt{T}x}{\sqrt{B}\|X_{\cdot}\|_{\psi_\alpha}^2}\Big)^{\frac{1}{2\alpha+1}}\Big]. \tag{8}$$

Remark 1. *It is worth mentioning that the result in Theorem 1 also applies to the uniform deviation for θ over the real line. In fact, for a trigonometric polynomial of order B, we can bound the uniform deviation by taking the maximum over a fine grid, i.e., by Theorem 7.28 in Ch. X of [60], for $B' = 4B$ and $\theta'_k = 2\pi k/B'$,*

$$\max_{\theta \in \mathbb{R}} |\widehat{F}(\theta) - \mathbb{E}\widehat{F}(\theta)|_\infty \leq 2 \max_{0 \leq k \leq B'} |\widehat{F}(\theta'_k) - \mathbb{E}\widehat{F}(\theta'_k)|_\infty.$$

Remark 2. *A closely related result to our Theorem 1 is Theorem 3.1 in [37], which presented a tail probability inequality for smoothed periodogram under finite sub-exponential moment. Our requirement and availability are significantly different from theirs. Firstly, we are able to deal with a wide class of exponential type tails characterized by α. Moreover, based on (8), we can obtain*

$$\max_{\theta \in \Theta} |\widehat{F}(\theta) - \mathbb{E}\widehat{F}(\theta)|_\infty = O_{\mathbb{P}}\Big(\|X_{\cdot}\|_{\psi_\alpha}^2 \sqrt{\frac{B}{T}} \log^{2\alpha+1}(Bp)\Big). \tag{9}$$

If $\|X.\|_{\psi_\alpha} \asymp 1$ and $B \asymp T^c$ for some $0 < c < 1$, as a natural requirement of consistency, we can allow the dimension p to increase exponentially with T, i.e.,

$$\log p = o\big(T^{\frac{1-c}{4\alpha+2}}\big),$$

while [37] allows a narrower range $p = O(n^C)$ for some $C > 0$. Proposition 4.3 of [38] delivered a probability inequality for lag-window spectral estimates of Gaussian processes:

$$\mathbb{P}\big(\max_{\theta\in\Theta} |\widehat{F}(\theta) - \mathbb{E}\widehat{F}(\theta)|_\infty \geq x\big) \lesssim Bp^2 \cdot \exp\Big[-C\min\Big(\frac{Tx^2}{B\Phi_2^4}, \frac{Tx}{B\Phi_2^2}\Big)\Big], \quad (10)$$

where $\Phi_2 = \max_{1\leq j\leq p} \|X_{\cdot j}\|_2$. (10) suggests two types of bounds for the tail probability: a sub-Gaussian type tail $\exp(-CTx^2/B\Phi_2^4)$ and a sub-exponential type tail $\exp(-CTx/B\Phi_2^2)$. Our result concerns the tails heavier than Gaussian tails and the probability bound is thus less sharper than the Gaussian case. In view of (9), the error rate is non-decreasing with α, p and the dependence adjusted norm $\|X.\|_{\psi_\alpha}$. In other words, the rate is less sharper when the process has a heavier tail, higher dimension and stronger dependence.

We also note that Theorem 4.1 in [38] established a concentration bound on spectral density estimators for high-dimensional and locally stationary process, assuming finite polynomial moment. Specifically, we revise the component-wise dependence measure in (3) and define the L_∞ dependence measure as

$$w_{t,q} = \||G(\varepsilon_t, \varepsilon_{t-1}, \ldots) - G(\varepsilon_t, \ldots, \varepsilon_1, \varepsilon_0^*, \varepsilon_{-1}, \ldots)|_\infty\|_q. \quad (11)$$

We further assume that $\||X.|_\infty\|_q = \sup_{m\geq 0} \rho^{-m} \sum_{t=m}^{\infty} w_{t,q} < \infty$ for some $0 < \rho < 1$. Under the stationarity assumption and with $q > 4$, Theorem 6.1 in [38] yields

$$\mathbb{P}\big(\max_{\theta\in\Theta} |\widehat{F}(\theta) - \mathbb{E}\widehat{F}(\theta)|_\infty \geq x\big) \lesssim \frac{T^{1-q/2}B^{q/2}(1\vee \log p)^{5q/4}\||X.|_\infty\|_q^q}{x^{q/2}} + Bp^2 \cdot \exp\Big(-\frac{CTx^2}{B\Phi_4^4}\Big). \quad (12)$$

This result, a Nagaev-type inequality, provides two types of bounds on the tail probability: a polynomial tail bound and a sub-Gaussian type tail bound. By comparison, our result involves only one term when assuming a general exponential-type tail quantified by $\|X.\|_{\psi_\alpha}$. In Section IV-A, we discuss an extension to the case with polynomial decayed dependence measures, as explored in [38].

Remark 3. *An alternative estimator of $F(\theta)$, which is applicable when the mean $\mathbb{E}(X_t)$ is unknown, is the sample-mean-adjusted version given by*

$$\tilde{F}(\theta) = \frac{1}{2\pi T}\sum_{r=1}^{T/B}\Big(\sum_{t=(r-1)B+1}^{rB}(X_t - \bar{X})e^{-\iota t\theta}\Big) \times \Big(\sum_{t=(r-1)B+1}^{rB}(X_t - \bar{X})^\top e^{\iota t\theta}\Big), \quad (13)$$

where $\bar{X} = T^{-1}\sum_{t=1}^T X_t$. Obviously $\tilde{F}(\theta) \equiv \widehat{F}(\theta)$ for $\theta \in \Theta$ but $\theta \neq 0$, since $\sum_{t=(r-1)B+1}^{rB} e^{-\iota t\theta} = 0$. If $\theta = 0$, we can obtain $|\tilde{F}(\theta) - \widehat{F}(\theta)|_\infty \lesssim B|\bar{X}|_\infty^2$, while for $\theta \notin \Theta$, $|\tilde{F}(\theta) - \widehat{F}(\theta)|_\infty \lesssim B|\bar{X}|_\infty|\bar{X}(\theta)|_\infty$, where $\bar{X}(\theta) = T^{-1}\sum_{t=1}^T X_t e^{-\iota t\theta}$. Applying Theorem 3 in [34] to $\bar{X}$ and $\bar{X}(\theta)$, and using the Bonferroni technique, we can conclude that the difference $|\tilde{F}(\theta) - \widehat{F}(\theta)|_\infty$ is negligible and Theorem 1 also holds for $\tilde{F}(\theta)$.

In Proposition 2, we concern the bias of the spectral estimate.

Proposition 2. *Denote $\Phi_2 = \max_{1\leq j\leq p}\|X_{\cdot j}\|_2$. Then we have*

$$\max_{\theta\in\Theta} |\mathbb{E}\widehat{F}(\theta) - F(\theta)|_\infty \lesssim B^{-1}\Phi_2^2. \quad (14)$$

We can select the window size B by considering a trade-off between the stochastic deviation bound in (9) and the bias order provided in Proposition 2. For example, if Φ_2 and $\|X.\|_{\psi_\alpha}$ are both of a constant order, to minimize the rate $\sqrt{B/T}\log^{2\alpha+1}(Bp) + B^{-1}$, it is suggested that

$$B \asymp \Big(\frac{T}{\log^{4\alpha+2}(p\vee T)}\Big)^{1/3}.$$

The order of the best B is analytically expressed in terms of the sample size T, the dimension p and the parameter α.

To perform statistical inference of spectral density, one shall establish the distributional theory of the spectral estimate. Theorem 3 below focuses on the joint convergence of two arbitrary entries in $\widehat{F}(\theta)$.

Theorem 3. *Denote $\Phi_4 = \max_{1\leq j\leq p}\|X_{\cdot j}\|_4$. Assume $B/T \to 0$ and $\Phi_4^8(\log B)/B \to 0$. Take any $\theta, \theta' \in \Theta$ and any $1 \leq i, j, h, l \leq p$. Then there exists some bivariate (complex) normal random variable $\boldsymbol{\xi} = (\xi_{ij}(\theta), \xi_{hl}(\theta'))^\top$ such that*

$$\sqrt{\frac{T}{B}}\begin{pmatrix}\widehat{F}_{ij}(\theta) - \mathbb{E}\widehat{F}_{ij}(\theta)\\ \widehat{F}_{hl}(\theta') - \mathbb{E}\widehat{F}_{hl}(\theta')\end{pmatrix} - \boldsymbol{\xi} \xrightarrow{d} \mathbf{0}.$$

(i) If $\theta = \theta' \in (0, \pi)$, $\boldsymbol{\xi} = (\xi_{ij}(\theta), \xi_{hl}(\theta))^\top$ is a bivariate complex normal distribution with mean 0, covariance matrix

$$\mathbb{E}(\boldsymbol{\xi}\boldsymbol{\xi}^H) = \begin{pmatrix}F_{ii}(\theta)F_{jj}(\theta) & F_{ih}(\theta)F_{lj}(\theta)\\ F_{hi}(\theta)F_{jl}(\theta) & F_{hh}(\theta)F_{ll}(\theta)\end{pmatrix} \triangleq \Sigma_1$$

and relation matrix (a.k.a. pseudo-covariance matrix)

$$\mathbb{E}(\boldsymbol{\xi}\boldsymbol{\xi}^\top) = \begin{pmatrix}F_{ij}^2(\theta) & F_{il}(\theta)F_{hj}(\theta)\\ F_{il}(\theta)F_{hj}(\theta) & F_{hl}^2(\theta)\end{pmatrix} \triangleq \Sigma_2.$$

(ii) If $\theta = \theta' = 0$ or π, $\boldsymbol{\xi}$ is a bivariate normal distribution with mean 0 and covariance matrix

$$E(\boldsymbol{\xi}\boldsymbol{\xi}^\top) = \Sigma_1 + \Sigma_2.$$

(iii) If $\theta \neq \theta'$, $\xi_{ij}(\theta)$ and $\xi_{hl}(\theta')$ are independent (complex) normal distributions. In particular, the off-diagonals in matrices $\mathbb{E}(\boldsymbol{\xi}\boldsymbol{\xi}^H)$ and $\mathbb{E}(\boldsymbol{\xi}\boldsymbol{\xi}^\top)$ are all 0.

Theorem 3 covers all possible cases for the asymptotic joint behavior of two (cross-)spectral density estimates $\widehat{F}_{ij}(\theta)$ and $\widehat{F}_{hl}(\theta')$. By (i) and (ii), the covariance matrix (as well as the pseudo-covariance matrix for complex-valued spectral estimates) of the two at the same frequency is fully characterized by the entries of $F(\theta)$, which will further be used in performing simultaneous inference of spectral density in Section III. Result (iii) suggests the asymptotic independence of spectral density estimates at different frequencies. By Proposition 2, with the bias concerned, we can replace $\mathbb{E}\widehat{F}_{ij}(\theta)$ with $F_{ij}(\theta)$ in the statement of Theorem 3 under the additional condition $T\Phi_2^4 = o(B^3)$,

By considering the marginal case of the joint convergence, we can study the limiting distribution of $\widehat{F}_{ij}(\theta)$, the estimator of each cross spectral density. We represent the limiting distribution of $\sqrt{T/B}(\widehat{F}_{ij}(\theta) - \mathbb{E}\widehat{F}_{ij}(\theta))$ as $\xi_{ij}(\theta) = \mathsf{Re}(\xi_{ij}(\theta)) + \iota\mathsf{Im}(\xi_{ij}(\theta))$. By Theorem 3 (i), for $\theta \in (0, \pi)$, we can conclude $(\mathsf{Re}(\xi_{ij}(\theta)), \mathsf{Im}(\xi_{ij}(\theta)))^\top$ is a bivariate normal distribution with mean 0. Looking into the diagonals of the covariance matrices, we can get the variance of the real part and imaginary part as follows

$$\mathsf{Var}(\mathsf{Re}(\xi_{ij}(\theta)) = \frac{1}{2}F_{ii}(\theta)F_{jj}(\theta) + \frac{1}{2}\mathsf{Re}(F_{ij}^2(\theta)), \quad (15)$$

$$\mathsf{Var}(\mathsf{Im}(\xi_{ij}(\theta)) = \frac{1}{2}F_{ii}(\theta)F_{jj}(\theta) - \frac{1}{2}\mathsf{Re}(F_{ij}^2(\theta)), \quad (16)$$

$$\mathsf{Cov}(\mathsf{Re}(\xi_{ij}(\theta), \mathsf{Im}(\xi_{ij}(\theta)) = \frac{1}{2}\mathsf{Im}(F_{ij}^2(\theta)), \quad (17)$$

while Theorem 3 (ii) implies $\xi_{ij}(\theta) = N(0, F_{ij}^2(\theta) + F_{ii}(\theta)F_{jj}(\theta))$ if $\theta = 0$ or π. This result agrees with Theorem 3 in [32] for the single cross spectral estimate and agrees with Theorem 2 in [24] for the real-valued spectrum estimate by letting $i = j$. The central limit theorem of the spectral estimate is highly non-trivial even for the special case of one dimensional time series. Many previous results require restrictive conditions on the underlying process such as linear processes [18], strong mixing conditions [61] and Gaussian processes [62]. Our framework covers a wide class of possibly nonlinear stationary processes and our conditions involving functional dependence measures can be easily verified. In addition, we impose a mild moment condition to deal with non-Gaussian data.

III. Simultaneous Inference of Spectral Density

In this section, we shall consider the simultaneous inference of the spectral density matrix for high-dimensional time series. In the one dimensional case, the distribution of $\max_{\theta\in[0,\pi]} |\widehat{F}(\theta) - F(\theta)|$ or the discretized version $\max_{\theta\in\Theta} |\widehat{F}(\theta) - F(\theta)|$ allows one to construct the confidence band of $F(\theta)$ or conduct a parametric specification test for $F(\theta)$. Gumbel convergence of this maximum deviation has been established by earlier work. For example, in the one dimensional case, [63] dealt with linear processes and [64] worked on Gaussian processes. Under the framework of (2) with $p = 1$, [24] derived the asymptotic distribution of the maximum deviation for the spectral density estimate and it was generalized to the multivariate case by [32] concerning the cross-spectral density. [39] considered the asymptotic distribution of the maximum sample coherence for Gaussian time series under the high-dimensional regime when $B = O(T^\rho)$ for some $0 < \rho < 1$ and $p/B \to c$ for some constant $c \in (0, \infty)$. There is little investigation on the simultaneous inference of spectral density for non-Gaussian processes with a possibly ultra-high dimension. In this section, we shall develop an inference theory on high-dimensional spectral density.

A. Gaussian Approximation for High Dimensional Spectral Estimates

Performing simultaneous inference of the high-dimensional spectral density matrix is a challenging task, especially when the dimension p is comparable to or even greater than the sample size T. If p can also diverge to infinity, [65] showed that the central limit theorem will fail for i.i.d. random vectors if $\sqrt{T} = o(p)$. Classical probability tools are inadequate for formulating a distribution theory for high-dimensional cases. We shall consider an alternative form: Gaussian approximation for the largest magnitude of the entries of the spectral density matrix. The idea of Gaussian approximation for high-dimensional random vectors was previously adopted in mean vector inference; see, for example [35], [44] and [66]. Addressing the spectral density matrix still poses a new challenge. Firstly, the spectral density matrix exhibits complex-valued off-diagonals (cross-spectral densities) when $p > 1$. Secondly, the spectral estimate manifests as a quadratic form in the time-dependent samples, further complicating the analytical landscape. In this section, we will address both aspects and formulate a new distribution theory to facilitate the inference of the spectral density matrix. In particular, we consider two types of deviation:

$$\boldsymbol{F} = \sqrt{\frac{T}{B}} \max_{\theta\in\Theta} \max\big\{|\mathsf{Re}(\widehat{F}(\theta) - \mathbb{E}\widehat{F}(\theta))|_\infty, |\mathsf{Im}(\widehat{F}(\theta) - \mathbb{E}\widehat{F}(\theta))|_\infty\big\};$$

$$\tilde{\boldsymbol{F}} = \sqrt{\frac{T}{B}} \max_{\theta\in\Theta} |\widehat{F}(\theta) - \mathbb{E}\widehat{F}(\theta)|_\infty.$$

The former separates the real and imaginary parts of each element, while the latter concerns the maximum modulus of these entries. While it is quite natural to consider the largest modulus of those complex-valued elements, it is worth mentioning that investigation on the real and imaginary parts separately of cross-spectral density can be of important use in many applications. For example, the imaginary part has gained increasing popularity to identify brain interaction in EEG and MEG connectivity studies [3], [67], [68], [69], [70], etc.

For the convenience of simultaneous inference, we exclude the two end frequencies $0, \pi$ and consider $\Theta = \{\theta_k : \theta_k = 2k\pi/B, k = 1, 2, \ldots, (B/2 - 1)\}$. Let $\mathbf{vec}_0(\widehat{F}) \in \mathbb{R}^{(B-2)p^2}$ be the vectorization of $(\mathsf{Re}(\widehat{F}(\theta)), \mathsf{Im}(\widehat{F}(\theta)))_{\theta\in\Theta}$. We aim to establish a distributional theory for $\boldsymbol{F}$ by the idea of Gaussian approximation. In particular, denote

$$\boldsymbol{\Sigma} = \lim_{B,T\to\infty} \frac{T}{B}\mathsf{Cov}(\mathbf{vec}_0(\widehat{F})) \text{ and } \boldsymbol{Z} \sim N(0,\boldsymbol{\Sigma}). \quad (18)$$

By Theorem 3, we can work out the expression of each entry in $\boldsymbol{\Sigma}$. For $\theta \in (0,\pi)$ and arbitrary combination of $i,j,h,l \in \{1,2,\ldots,p\}$, it can be computed that

$$\begin{aligned}
&\lim_{B,T\to\infty} \frac{T}{B}\mathsf{Cov}(\mathsf{Re}(\widehat{F}_{ij}(\theta)), \mathsf{Re}(\widehat{F}_{hl}(\theta)) \\
&\qquad = \frac{1}{2}\mathsf{Re}(F_{ih}(\theta)F_{lj}(\theta)) + \frac{1}{2}\mathsf{Re}(F_{il}(\theta)F_{hj}(\theta)), \\
&\lim_{B,T\to\infty} \frac{T}{B}\mathsf{Cov}(\mathsf{Im}(\widehat{F}_{ij}(\theta)), \mathsf{Im}(\widehat{F}_{hl}(\theta)) \\
&\qquad = \frac{1}{2}\mathsf{Re}(F_{ih}(\theta)F_{lj}(\theta)) - \frac{1}{2}\mathsf{Re}(F_{il}(\theta)F_{hj}(\theta)), \\
&\lim_{B,T\to\infty} \frac{T}{B}\mathsf{Cov}(\mathsf{Im}(\widehat{F}_{ij}(\theta)), \mathsf{Re}(\widehat{F}_{hl}(\theta)) \\
&\qquad = \frac{1}{2}\mathsf{Im}(F_{il}(\theta)F_{hj}(\theta)) + \frac{1}{2}\mathsf{Im}(F_{ih}(\theta)F_{lj}(\theta)), \\
&\lim_{B,T\to\infty} \frac{T}{B}\mathsf{Cov}(\mathsf{Re}(\widehat{F}_{ij}(\theta)), \mathsf{Im}(\widehat{F}_{hl}(\theta)) \\
&\qquad = \frac{1}{2}\mathsf{Im}(F_{il}(\theta)F_{hj}(\theta)) - \frac{1}{2}\mathsf{Im}(F_{ih}(\theta)F_{lj}(\theta)).
\end{aligned}$$

All the other entries of $\boldsymbol{\Sigma}$ are set to be 0 by the asymptotic independence of two spectral density estimates at different frequencies.

Before proceeding, we state the main assumptions required in studying the distribution of $\boldsymbol{F}$.

Assumption 2. *There exists some constant $c_1 > 0$ such that*

$$\min_{\theta\in\Theta} F_{ii}(\theta) \geq c_1 \text{ for all } 1 \leq i \leq p.$$

Assumption 3. *There exists some constant $c_2 > 0$ such that*

$$\min_{\theta\in\Theta}\{F_{ii}(\theta)F_{jj}(\theta) - |\mathsf{Re}(F_{ij}^2(\theta))|\} \geq c_2 \text{ for all } 1 \leq i \neq j \leq p.$$

In view of (15) and (16), Assumption 2 and Assumption 3 ensure all the variances of non-constantly-zero entries in $\boldsymbol{Z}$ are bounded away from 0, where the former requires non-vanishing spectrum for each component process and the latter accounts for a lower bound for the asymptotic variances of $\mathsf{Re}(\widehat{F}_{ij}(\theta))$ and $\mathsf{Im}(\widehat{F}_{ij}(\theta))$ when $i \neq j$.

In Theorem 4, we shall present the main result for the approximate distribution on spectral estimate. That is, the distribution of $\boldsymbol{F}$ or $\tilde{\boldsymbol{F}}$ can be approximated by its Gaussian analogue.

Theorem 4. *Let Assumption 1, Assumption 2 and Assumption 3 be satisfied. For $\kappa_1 = \max\{7, 2\alpha+6\}$ and $\kappa_2 = 4\alpha+4$, let*

$$\mathcal{E}_{T,B,p} = \Big(\frac{B\log^{\kappa_1}(Tp)}{T}\Big)^{\frac{1}{6}} + \Big(\frac{\log^{\kappa_2}(Bp)}{B}\Big)^{\frac{1}{2}} + \Big(\frac{\log^5(Bp)}{B}\Big)^{\frac{1}{6}}. \quad (19)$$

Assume that $\|X_\cdot\|_{\psi_\alpha}^2 \mathcal{E}_{T,B,p} \to 0$. Then it holds that

$$\begin{aligned}
d_{T,B,p} &:= \sup_{x\in\mathbb{R}} |\mathbb{P}(\boldsymbol{F} \leq x) - \mathbb{P}(|\boldsymbol{Z}|_\infty \leq x)| \lesssim \|X_\cdot\|_{\psi_\alpha}^2 \mathcal{E}_{T,B,p}, \\
\tilde{d}_{T,B,p} &:= \sup_{x\in\mathbb{R}} |\mathbb{P}(\tilde{\boldsymbol{F}} \leq x) - \mathbb{P}(|\tilde{\boldsymbol{Z}}|_\infty \leq x)| \lesssim \|X_\cdot\|_{\psi_\alpha}^2 \mathcal{E}_{T,B,p},
\end{aligned}$$

where $\boldsymbol{Z} = (Z_1, Z_2, \ldots, Z_{(B-2)p^2})^\top$ is defined in (18) and $\tilde{\boldsymbol{Z}} = \big(\sqrt{Z_{2j-1}^2 + Z_{2j}^2}\big)_{j=1,\ldots,(B/2-1)p^2}$.

In the low dimensional case when p is a fixed number, the conditions of Theorem 4 are simplified to the very mild requirement $B = o(T/\log^{\kappa_1}(T))$. We can also investigate how large the dimension p can be allowed for the validity of the Gaussian approximation result in high dimensions. For example, if $\|X_\cdot\|_{\psi_\alpha} \asymp 1$, we shall require

$$\log p = o\Big(\min\big\{\big(\frac{T}{B}\big)^{1/\kappa_1}, B^{1/\kappa_2}, B^{1/5}\big\}\Big)$$

If we choose $B \asymp T^c$ for some $0 < c < 1$, p can be allowed to increase exponentially with the sample size T, i.e., $\log p = o(T^C)$ where $C = \min\{(1-c)/\kappa_1, c/\kappa_2, c/5\}$.

We now consider the version centered by $F(\theta)$. Define

$$\begin{aligned}
\boldsymbol{F}^\star &= \sqrt{\frac{T}{B}} \max_{\theta\in\Theta} \max\big\{|\mathsf{Re}(\widehat{F}(\theta) - F(\theta))|_\infty, \\
&\qquad\qquad |\mathsf{Im}(\widehat{F}(\theta) - F(\theta))|_\infty\big\}; \\
\tilde{\boldsymbol{F}}^\star &= \sqrt{\frac{T}{B}} \max_{\theta\in\Theta} |\mathsf{Re}(\widehat{F}(\theta) - F(\theta))|_\infty. \quad (20)
\end{aligned}$$

By accounting for the bias of $\widehat{F}(\theta)$, the Gaussian approximation result also holds for $\boldsymbol{F}^\star$ and $\tilde{\boldsymbol{F}}^\star$; see Corollary 5 below.

Corollary 5. *Let the assumptions of Theorem 4 be satisfied. Further assume that*

$$\frac{T\log(Bp)\|X_\cdot\|_{\psi_\alpha}^4}{B^3} \to 0. \quad (21)$$

Then we have

$$\begin{aligned}
d_{T,B,p}^\star &:= \sup_{x\in\mathbb{R}} |\mathbb{P}(\boldsymbol{F}^\star \leq x) - \mathbb{P}(|\boldsymbol{Z}|_\infty \leq x)| \to 0, \\
\tilde{d}_{T,B,p}^\star &:= \sup_{x\in\mathbb{R}} |\mathbb{P}(\tilde{\boldsymbol{F}}^\star \leq x) - \mathbb{P}(|\tilde{\boldsymbol{Z}}|_\infty \leq x)| \to 0.
\end{aligned}$$

Our results in Theorem 4 and Corollary 5 concern the maximum deviation over all the entries in $\widehat{F}(\theta)$ and over all the frequencies in Θ. In various inference scenarios, we only need to select a subset of these elements and the asymptotic covariance matrix $\boldsymbol{\Sigma}$ should also have an adaptively reduced size to account for the variances and pairwise covariances of elements in the chosen subset. In Section III-B we provide some examples as such.

B. Applications

Our result can be applied to address a variety of time series inference problems. In hypothesis testing, we can compute the test statistics defined by (20), where $\widehat{F}(\theta)$ is calculated from the sample data, and the plugged-in $F(\theta)$ is from the null hypothesis. Corollary 5 provides the theoretical guarantee of validity.

An immediate application is to test the spectrum for any component process marginally, i.e., $H: F_{ii}(\theta) = f_i(\theta)$, for $1 \leq i \leq p$, where $f_i(\theta)$ is the hypothesized spectral density for the component process $(X_{ti})_t$. In this case, $\widehat{F}_{ii}(\theta)$ is real-valued and only Assumption 2 is required to ensure a non-vanishing variance of $\widehat{F}_{ii}(\theta)$. The two test statistics are equal

$$\boldsymbol{F}^\star = \tilde{\boldsymbol{F}}^\star = \sqrt{\frac{T}{B}} \max_{\theta\in\Theta} \max_{1\le i\le p} |\widehat{F}_{ii}(\theta) - f_i(\theta)|$$

and $\boldsymbol{\Sigma}$ consists of covariances of $\widehat{F}_{ii}(\theta)$ and $\widehat{F}_{jj}(\theta)$ for $1 \le i, j, \le p$ and $\theta \in \Theta$.

Moreover, we can provide a new solution to high-dimensional white noise test, which is used for serial correlation detection of time series. Classical portmanteau-type tests including [71] and [72] are popular choices in the one dimensional case, and are well generalized to the multivariate case (e.g., [73], [74]). high-dimensional white noise tests are developed for linear processes by [75] concerning the sum of squared singular values of the first few lagged sample autocovariance matrices and for β-mixing processes by [76] using the maximum absolute auto-correlations and cross-correlations (see [77] for a similar test statistic formulation). The forementioned tests in the time domain concern the null hypothesis $H : \Gamma(k) = 0$, for $1 \le k \le L$, to test for serial correlation and present the additional problem of selecting the number of lags to be considered, which does not appear in the frequency domain by stating the null hypothesis in terms of the spectral density matrix function, namely, $H : F(\theta) = F_0$, where F_0 consists of constant-valued entries free of θ. The most common statistics in the one dimensional case are based on standardized cumulative periodogram; see [78], [79], [80], [81], just to name a few. With the given F_0 inserted in our statistic (20), our result can be used to design a new high-dimensional white noise test for a large class of stationary processes under very mild regularity conditions. To practically select an initial F_0, we observe that $F(\theta) = (2\pi)^{-1}\Gamma(0)$ if X_t is white noise, where $\Gamma(0) = \mathrm{Cov}(X_t, X_t)$ denotes the covariance matrix of the stationary process. We can refer to existing methods of consistently estimating the covariance matrix for high-dimensional time series, including the sample covariance matrix studied in [38], the regularized covariance estimator under sparse settings in [33], and the robust covariance estimator for fat-tailed processes proposed in [82].

We can also test for pairwise independence between two component time series. Unlike the tests based on the cross-correlation function in the time domain (e.g., [83], [84], [85]), we test $H : F_{ij}(\theta) \equiv 0$ for fixed $i \ne j$. Suppose Assumption 3 holds for the selected i, j and let $p = 2$ in the theory. We shall use the test statistic

$$\tilde{\boldsymbol{F}}_{ij}^\star = \sqrt{\frac{T}{B}} \max_{\theta\in\Theta} \max\{|\mathsf{Re}(\widehat{F}_{ij}(\theta))|, |\mathsf{Im}(\widehat{F}_{ij}(\theta))|\}$$

or $\tilde{\boldsymbol{F}}_{ij}^\star = \sqrt{T/B}\max_{\theta\in\Theta} |\widehat{F}_{ij}(\theta)|$. Our Corollary 5 provides a new criteria for the cross-spectral density testing.

Spectral analysis has been widely used to detect functional network connectivity; see [3], [86], [87], [88] and [89] for applications involving neuroimaging data. Denote the partition of the index set $\{1, 2, \ldots, p\}$ as K disjoint subsets $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_K$ with $\cup_{k=1}^K \mathcal{A}_k = \{1, 2, \ldots, p\}$. We can identify disconnected brain regions by testing the joint independence of the K subset processes $(X_{ti})_{i\in\mathcal{A}_k}$, $k = 1, \ldots, K$. That is, the problem is to consider $H : F_{ij}(\theta) \equiv 0$ for any combinations of $i \in \mathcal{A}_{k_1}$, $j \in \mathcal{A}_{k_2}$ and $1 \le k_1 \ne k_2 \le K$ with the test statistic

$$\tilde{\boldsymbol{F}}^\star = \sqrt{\frac{T}{B}} \max_{\theta\in\Theta} \max_{1\le k_1\ne k_2\le K} \max_{i\in\mathcal{A}_{k_1}, j\in\mathcal{A}_{k_2}} |\widehat{F}(\theta)|.$$

C. Resampling Methods

To perform statistical inference on high-dimensional spectral density in practice, given the level $\theta_0 \in (0, 1)$, one needs to work out the cut-off value $c_{1-\theta_0}$, the $(1-\theta_0)$-quantile of $|\boldsymbol{Z}|_\infty$ or $|\tilde{\boldsymbol{Z}}|_\infty$ for $\boldsymbol{Z} \sim N(0, \boldsymbol{\Sigma})$. Despite the theoretical advancements on the distribution approximation of the spectral estimate, practical implementation remains challenging due to the involvement of unknown covariance matrix $\boldsymbol{\Sigma}$. In spectral inference, we leverage the unique feature that the asymptotic covariances of the spectral estimates (1) can be entirely expressed through the true spectral densities. Hence, a good estimator of $\widehat{F}(\theta)$ with Theorem 1 concerning the convergence of $\widehat{F}(\theta)$ ensures a good plug-in estimator of $\boldsymbol{\Sigma}$.

In this context, we carry out the implementation of estimating the distribution of $\boldsymbol{F}^\star$. The case of $\tilde{\boldsymbol{F}}^\star$ can be dealt with in a similar way. By Theorem 3 and looking into their real and imaginary parts pairwisely, we have the nice property of expressing the entries in $\boldsymbol{\Sigma}$ in terms of the entries in $F(\theta)$, as justified in Section III-A. We can implement a Gaussian multiplier resampling method for spectral inference as follows:

1) Given the observations $X_1, \ldots, X_t$, estimate $F(\theta)$ by $\widehat{F}(\theta)$.
2) Plug the entries of $\widehat{F}(\theta)$ into the formula of entries of $\boldsymbol{\Sigma}$ to obtain $\widehat{\boldsymbol{\Sigma}}$.
3) Obtain $\widehat{c}_{1-\theta_0}$, the conditional $(1-\theta_0)$-quantile of $|\widehat{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{Z}_0|_\infty$ given $(X_t)_{t=1}^T$, where $\boldsymbol{Z}_0$ is a standard normal random vector, independent of $(X_t)_{t=1}^T$.

The technique of Gaussian multiplier bootstrap was employed in [44] for high-dimensional mean vector inference in the independent setting. As an extension, blockwise multiplier bootstrap was proposed by [45] for M-dependent processes. In particular, to apply the blockwise multiplier bootstrap for M-dependent processes, the method divides the sample into consecutive large and small blocks of size N and M, respectively, with $N \gg M$. The statistic over these blocks is then computed and further multiplied by i.i.d. Gaussian variables to facilitate statistical inference. However, for the general weakly dependent processes addressed in our paper, selecting appropriate values for N and M is challenging and introduces additional uncertainty into the inference. Based on our Gaussian approximation theory for spectral estimates, in which $\boldsymbol{\Sigma}$ is specified by entries from the true $F(\theta)$, directly substituting $\widehat{F}(\theta)$ to estimate $\boldsymbol{\Sigma}$ offers a straightforward approach by avoiding the need for additional block-size selection.

Theorem 6 is introduced to concern the validity of our approach.

Theorem 6. *Let the conditions in Corollary 5 be satisfied. Further assume*

$$\sqrt{\frac{B}{T}} \log^{2\alpha+3}(Bp)\|X_\cdot\|_{\psi_\alpha}^4 \to 0. \tag{22}$$

Then

$$\sup_{\theta_0\in(0,1)} |\mathbb{P}(\boldsymbol{F}^\star \geq \widehat{c}_{1-\theta_0}) - \theta_0| \to 0. \tag{23}$$

If we use $\tilde{\boldsymbol{F}}^\star$, it suffices to slightly modify the third step to obtain $\tilde{c}_{1-\theta_0}$, the conditional $(1-\theta_0)$-quantile of $|\tilde{\widehat{\boldsymbol{Z}}}|_\infty$, where

$$\widehat{\boldsymbol{Z}} = (\widehat{Z}_1, \ldots, \widehat{Z}_{(B-2)p^2})^\top = \widehat{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{Z}_0,$$

$\boldsymbol{Z}_0$ is a standard normal random vector, independent of $(X_t)_{t=1}^T$ and

$$\tilde{\widehat{\boldsymbol{Z}}} = \big(\sqrt{\widehat{Z}_{2j-1}^2 + \widehat{Z}_{2j}^2}\big)_{j=1,\ldots,(B/2-1)p^2}.$$

Another widely employed and effective resampling technique is subsampling, which constructs the resampling distribution by taking the subsamples of the observations. It has been commonly used in inferring low dimensional time series. Some pioneering work includes [42], [43], [90]; see [91] for a comprehensive overview. Despite its intuitive appeal, there has been relatively limited exploration into the theoretical properties of the subsampling method in high dimensions. Making an additional noteworthy contribution, we provide a thorough theoretical justification for the validity of high-dimensional spectral inference through subsampling. Consequently, we obtain a consistent estimator for the distribution function of $\boldsymbol{F}^\star$, enabling the implementation of tests and confidence intervals based on $\boldsymbol{F}^\star$. To the best of our knowledge, our work is the first to provide theoretical assurance for high-dimensional spectral inference utilizing subsampling. It is anticipated that the novel proofs can be adapted to offer theoretical support in other high-dimensional subsampling problems.

Let $M = o(T)$ be the subsample size and $B = o(M)$ be the window size. Over each subsample $(X_\ell, X_{\ell+1}, \ldots, X_{\ell+M-1})$, we apply (1) to estimate the spectral density matrix, denoted by $\widehat{F}_\ell(\theta)$. Recall $\widehat{F}(\theta)$ is the estimator over the whole sample $X_1, \ldots, X_t$. We define

$$\begin{aligned}\boldsymbol{F}_\ell = \sqrt{\frac{M}{B}}\max_{\theta\in\Theta}\max\big\{&|\mathsf{Re}(\widehat{F}_\ell(\theta) - \widehat{F}(\theta))|_\infty, \\ &|\mathsf{Im}(\widehat{F}_\ell(\theta) - \widehat{F}(\theta))|_\infty\big\}.\end{aligned}$$

For $\mathcal{L} = T - M + 1$, define the empirical distribution function

$$\boldsymbol{Q}(x) = \frac{1}{\mathcal{L}}\sum_{\ell=1}^{\mathcal{L}} \mathbf{1}\{\boldsymbol{F}_\ell \leq x\}.$$

The deviation based on the maximum modulus of spectral densities can be dealt with similarly with $\boldsymbol{F}_\ell$ replaced by

$$\tilde{\boldsymbol{F}}_\ell = \sqrt{\frac{M}{B}}\max_{\theta\in\Theta}|\widehat{F}_\ell(\theta) - \widehat{F}(\theta)|_\infty.$$

Theorem 7. *Assume that $M = o(T)$ and $B = o(M)$. Let the conditions of Corollary 5 be satisfied, in which T is replaced by the subsample size M. Further assume the following conditions are met as $B, M, T \to \infty$:*

$$T^{-1}M\log^{4\alpha+5}(Bp)\|X.\|_{\psi_\alpha}^8 \to 0 \tag{24}$$

and

$$B^{-3}M\log^3(Bp)\|X.\|_{\psi_\alpha}^8 \to 0. \tag{25}$$

Then we have

$$\sup_{x\in\mathbb{R}}\big|\boldsymbol{Q}(x) - \mathbb{P}(|\boldsymbol{Z}|_\infty \leq x)\big| \xrightarrow{\mathbb{P}} 0.$$

Remark 4. *The subsampling approach is also valid for consistent spectral estimates by other smoothing methods when the asymptotic covariance matrix $\boldsymbol{\Sigma}$ may not be fully characterized. For example, with a general lag-window estimator, or estimating the entries of $\widehat{F}(\theta)$ element-wisely using different smoothing functions, the Gaussian approximation result still holds if the asymptotic variances of the real and imaginary parts of each element are non-vanishing, but the variances may rely on the smoothing functions in a subtle way and working out a closed-form of $\boldsymbol{\Sigma}$ could be highly non-trivial. In such cases, we can choose to implement subsampling to avoid the estimation of $\boldsymbol{\Sigma}$.*

IV. Extension

A. Relaxing the Moment and Dependence Conditions

In this section, we shall discuss the generalization of the results in Section II and Section III by relaxing the moment and dependence conditions. Specifically, we assume the existence of the q-th moment for some $q > 4$, as considered by [35] and [38]. This assumption is weaker than Assumption 1, which requires moments to exist at all orders. Using the L_∞ dependence measure defined in (11) and the associated q-th dependence-adjusted norm $\||X.|_\infty\|_q$, we can follow the proof of Theorem 6.1 in [38] to obtain the tail probability bound for $q > 4$:

$$\begin{aligned}&\mathbb{P}\big(\max_{\theta\in\Theta}|\widehat{F}(\theta) - \mathbb{E}\widehat{F}(\theta)|_\infty \geq x\big) \\ &\lesssim \frac{T^{1-q/2}B^{q/2}(1\vee\log p)^{5q/4}\||X.|_\infty\|_q^q}{x^{q/2}} \\ &\qquad + Bp^2\cdot\exp\Big(-\frac{CTx^2}{B\Phi_4^4}\Big).\end{aligned} \tag{26}$$

Using the concentration inequality in (26) as a replacement for Theorem 1 and applying Gaussian approximation with finite polynomial moments (cf. Proposition 2.1, case (E.2) in [92]) as a substitute for Lemma 14, we can employ similar arguments to derive the Gaussian approximation result and demonstrate the validity of the two resampling methods under polynomial moment conditions.

While the geometric moment contracting (GMC) condition considered in (4) has been widely applied in spectral inference (e.g., [23] and [32]), a more relaxed condition is polynomial decay, which assumes

$$\|X_{\cdot j}\|_{q,\alpha} = \sup_{m\geq 0}(m\vee 1)^{-\alpha}\sum_{t=m}^\infty \delta_{t,q,j} < \infty$$

for some $\alpha > 0$. In Example 1, for the linear process $X_t = \sum_{k=0}^\infty A_k\varepsilon_{t-k}$, we assume that there exist some $\beta > 1$ and $0 < K_p < \infty$ such that $\max_{1\leq j\leq p}(\sum_{j'=1}^p A_{t,jj'}^2)^{1/2} \leq K_p(t\vee 1)^{-\beta}$ for any $t \geq 0$, where $A_{t,jj'}$ denotes the (j,j')-th entry of A_t. This allows us to bound $\delta_{t,q,j} = \|\sum_{j'=1}^p A_{t,jj'}(\varepsilon_{0j'} - \varepsilon_{0j'}^*)\|_q \leq C_q\mu_q K_p(t\vee 1)^{-\beta}$. Thus, the upper bound of the dependence measure exhibits polynomial decay with $\alpha = \beta - 1$. Essentially, replacing the original tail cumulative dependence measure of the order ρ^m with the order $m^{-\alpha}$ allows all proofs to extend to the case where the dependence measure decays polynomially. Notably, an additional assumption, $\max_{1 \le j \le p} \sum_{t=0}^{\infty} t^2 \delta_{t,4,j} < \infty$, is required to ensure the summability of fourth-order joint cumulants, while this condition is automatically satisfied under the geometric decay as stated in Lemma 9.

B. Lag-window Estimator of Spectral Density

Next, we discuss the possibility of extending the theory to the general lag-window estimator, defined as

$$\check{F}(\theta) = \frac{1}{2\pi T} \sum_{k=-B}^{B} \mathcal{K}(k/B) \widehat{\Gamma}(k) e^{-\iota k \theta}, \tag{27}$$

where $\mathcal{K}(\cdot)$ is a continuous, even and bounded kernel function on $[-1, 1]$ satisfying $\mathcal{K}(0) = 1$, $B = o(T)$ is the window size, and $\widehat{\Gamma}(k)$ are the sample autocovariance matrices with $\widehat{\Gamma}(k) = T^{-1} \sum_{t=k+1}^{T} X_{t-k} X_t^\top$ for $k \ge 0$ and $\widehat{\Gamma}(k) = \widehat{\Gamma}(-k)^\top$ for $k < 0$. By an inspection of the proof for Theorem 1, applying equation (41) of [93] to derive the key step (32), the same concentration bound in Theorem 1 applies to the lag-window estimator $\check{F}(\theta)$.

For the distributional theory, we observe that the separation needs to be larger than the order B^{-1} to ensure asymptotic uncorrelation of spectral estimates at neighboring frequencies. This behavior for the lag-window estimator is also supported by Section 6.2.4 in [19], Lemma A.5 in [24] and Lemma A.6 in [32]. Therefore, the covariance matrix $\mathbf{\Sigma}$ contains many non-zero entries, reflecting the strong correlations at close frequencies, making it difficult to compute. After verification, our Gaussian approximation results still hold with $\mathbf{\Sigma} = \lim_{B,T\to\infty} \frac{T}{B} \mathrm{Cov}(\mathbf{vec}_0(\widehat{F}))$ and $\boldsymbol{Z} \sim N(0, \mathbf{\Sigma})$; however, obtaining a closed form for all entries of $\mathbf{\Sigma}$ is highly non-trivial. Consequently, while the Gaussian multiplier resampling method, which relies on explicit computation of $\mathbf{\Sigma}$, is not applicable, subsampling remains valid.

V. Numerical Results

A. Simulation Study

In this section, we conduct a simulation study to evaluate the accuracy of the Gaussian approximation results presented in Section III-A, followed by an assessment of the performance of the Gaussian multiplier resampling method and the subsampling method proposed in Section III-B. We consider the linear process

$$X_t = \sum_{k=0}^{\infty} A_k \varepsilon_{t-k}, \tag{28}$$

where the coefficient matrices A_k are constructed with a Toeplitz structure, parameterized by a correlation factor ρ. Specifically, $A_k = \rho^k M$ where $M = (m_{ij})_{i,j=1}^p$ and $m_{ij} = \rho^{|i-j|+1}$, with $\rho \in (0, 1)$. To ensure the process exhibits sub-exponential decay, we consider errors $\varepsilon_i = (\varepsilon_{i1}, \cdots, \varepsilon_{ip})$, where ε_{ij} are i.i.d. random variables following a Gamma

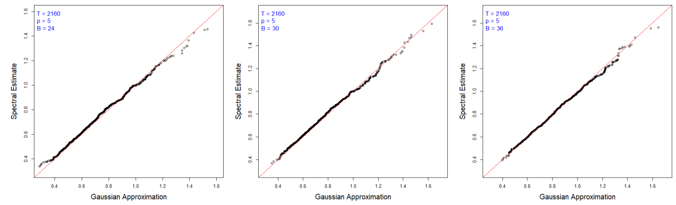

Fig. 1. Q-Q plot comparing maximum deviation F^* and its Gaussian approximation $|Z|_\infty$ when $p = 5$.

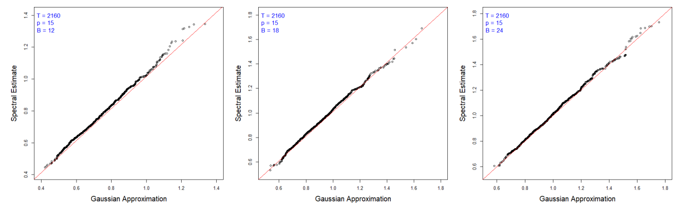

Fig. 2. Q-Q plot comparing maximum deviation F^* and its Gaussian approximation $|Z|_\infty$ when $p = 15$.

distribution $\Gamma(2, 1)$. We adopt the following numerical configurations for our analysis: decay rate $\rho = 0.5$, sample size $T = 2160$, dimension $p = 5, 15$, with the sum in (28) truncated to $X_t = \sum_{k=0}^{2000} A_k \varepsilon_{t-k}$.

It is worth noting that the window size B and subsample size M play important roles in the inference procedure. In high-dimensional settings, where both the sample size T and the vector dimension $(B-2)p^2$ in Gaussian approximation can be large, selecting these tuning parameters is challenging and sensitive to constant scaling. In practice, we recommend $B = T^{1/2}/\log p$, which minimizes the error bound in Theorem 4 sharpest in terms of the polynomial orders of T and $\log p$. Given the complicated error bounds in the subsampling procedure, determining an ideal order for the subsample size M is highly non-trivial. We suggest using $M = T^{1/3}B$, which meets the additional conditions (24) and (25). With the suggested orders $B = T^{1/2}/\log p$ and $M = T^{1/3}B$, we adjust B to the nearest even divisor of T and M to the closest integer divisible by B for computational convenience. For example, for $p = 5$, it can be computed that $T^{1/2}/\log p \approx 28.877$, so we set $B = 30$, an even divisor of T and choose $M = 360$. For comparison, we also examine a smaller bandwidth $B = 24$ with $M = 288$ and a larger bandwidth $B = 36$ with $M = 432$. Similarly, for $p = 15$, we use $B = 18$ and $M = 216$ as the best choices, and include comparisons with $B = 12$, $M = 144$, and $B = 24$, $M = 288$.

Figure 1 and Figure 2 display Q-Q plots comparing $\boldsymbol{F}^*$, as defined in (20), and its Gaussian approximation $|\boldsymbol{Z}|_\infty$, as defined in (18), based on 1000 repetitions. In each figure, a Q-Q plot is generated for each of the 3 values of B considered. As shown, the fit aligns closely with the reference line when $p = 5$ and show slightly less accuracy at the tail but remains satisfactory when $p = 15$. These results validate our Gaussian approximation through simulations. We also observe that the suggested B provides the best approximation across each set of three plots.

We further employ the Gaussian multiplier resampling

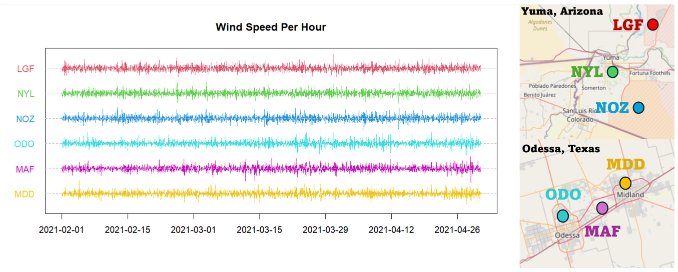

Fig. 3. Standardized first-order differences of hourly wind speed series between 02/01/2021 to 04/30/2021 from selected stations in Arizona and Texas.

method and subsampling to evaluate the validity of these two approaches. Specifically, we compare the values of F^* with the cut-off values $\widehat{c}_{1-\theta_0}$ obtained from the resampling and subsampling methods discussed in Section III-B using nominal levels $\theta_0 = 0.1$ and 0.05, corresponding to the 90% and 95% empirical quantiles respectively. Table I shows the coverage probability for each case, representing the proportion of instances where the cut-off value $\widehat{c}_{1-\theta_0}$ exceeds F^* over 1000 repetitions. The results indicate that our simultaneous confidence bands achieve the desired coverage levels. Additionally, the procedure becomes more sensitive to the choice of B as p increases, particularly for the Gaussian multiplier resampling method.

TABLE I
EMPIRICAL COVERAGES OF SIMULTANEOUS CONFIDENCE BANDS.

Dimension	Bandwidth	Resampling		Subsampling	
		90%	95%	90%	95%
p = 5	B = 24	87.8	93.4	87.6	92.9
	B = 30	89.8	94.3	89.4	95.1
	B = 36	90.6	94.7	88.1	93.0
p = 15	B = 12	84.7	91.2	89.3	94.7
	B = 18	87.5	93.7	90.3	95.2
	B = 24	86.4	92.7	87.4	94.6

B. Data Analysis

In this section, we test out method on wind data obtained from the Iowa State University Environmental Mesonet (IEM) Automated Surface Observing System (ASOS) Network, a publicly accessible database ([94], [95]). ASOS stations, primarily located at airports, provide weather observations and general reports for the National Weather Service (NWS), the Federal Aviation Administration (FAA), and the Department of Defense (DOD). For this study, we analyze data collected from six airports, three located in Arizona: LGF (Laguna Army Airfield), NYL (Yuma International Airport), and NOZ (Rolle Airfield), and three in Texas: ODO (Odessa-Schlemeyer Field), MAF (Midland International Air and Space Port), and MDD (Midland Airpark). By examining the cross-spectra across these 15 pairs, we aim to test for pairwise independence between each pair of wind speed time series, based on the theory presented in Section III-B.

We consider the hourly wind speed from 02/01/2021 to 04/30/2021 for each location. Since our method relies on the assumption of stationarity, we select this time period to avoid extreme values and non-stationarity associated with rainfall and storms, which are prevalent in other months. We note that there is virtually no rainfall during this period at these locations. Previous studies on wind behavior have shown that certain locations exhibit quasi-periodic patterns influenced by geographic and climatic factors, such as jet streams in summer and winter ([96]), which we aim to avoid in our analysis. To eliminate any local trends, we conduct a joint analysis of the 6-dimensional data by taking the first-order differences of each time series. Each differenced series is further standardized by subtracting its mean and dividing by its standard deviation to improve computational stability. Figure 3 presents the resulting 6-dimensional time series data, consisting of observations at 2136 time points, along with the 6 airport locations.

We shall consider the hypothesis test: $H : F_{ij}(\theta) = 0$ for each pairs among the six airports, where $1 \le i < j \le 6$ with significance level $\theta_0 = 0.05$. The test statistic $\boldsymbol{F}^{\star}_{ij}$, defined in Section III-B, is employed. Unlike in the simulation study, the true $\boldsymbol{\Sigma}$ is not available in real data analysis. To validate our test, we apply resampling and subsampling methods as discussed in Section III-C, obtaining (i) the conditional $(1-\alpha)$-quantile $\widehat{c}_{1-\alpha}$ via resampling, and (ii) the empirical quantile from the empirical distribution $Q(x)$ by subsampling. Following the recommendations discussed in Section V-A, we set $B = 24$ and $M = 288$.

Table II presents the test statistics and cut-off values obtained through resampling and subsampling. Results indicate that for pairs within the groups (LGF, NYL, NOZ) and (ODO, MAF, MDD), the test statistics significantly exceed the corresponding cut-off values, leading us to reject the null hypothesis of independence within these pairs. Additionally, for tests between pairs from different groups, the test statistics are smaller than the cut-off values, so we fail to reject the null hypothesis, supporting independence between these pairs. This finding aligns with geographic proximity, as locations close to each other tend to share similar wind profile patterns, whereas locations farther apart do not exhibit this similarity.

VI. Proofs

In this section, we provide the proofs of the results introduced in the main body.

A. Proofs of Results in Section II

Proof of Theorem 1. Denote $n = T/B$. For $1 \le r \le n$ and $1 \le i \le p$, define

$$Y_{ri}(\theta) = \sum_{t=(r-1)B+1}^{rB} X_{ti} e^{-\iota t\theta}.$$

To prove the result (8), we shall first consider the quadratic form

$$R_{ij}(\theta) = 2\pi T \widehat{F}_{ij}(\theta) = \sum_{r=1}^{n} Y_{ri}(\theta) Y_{rj}(\theta)$$

and apply the Markov inequality: for $C_0 > 0$ and $\eta > 0$ which are to be determined later,

$$\begin{aligned} &\mathbb{P}\big(|R_{ij}(\theta) - \mathbb{E}R_{ij}(\theta)| \ge x\big) \\ &\le e^{-\eta x^{C_0}} \mathbb{E}[\exp(\eta |R_{ij}(\theta) - \mathbb{E}R_{ij}(\theta)|^{C_0})]. \end{aligned} \tag{29}$$

TABLE II
PAIRWISE CROSS-SPECTRAL TEST RESULTS

Location	Station pairs	Test statistic	Rsampling cut-off value	Subsampling cut-off value
Yuma	LGF & NYL	0.648	0.418	0.391
	LGF & NOZ	0.593	0.420	0.507
	NYL & NOZ	0.545	0.431	0.493
Odessa	ODO & MAF	0.905	0.399	0.496
	ODO & MDD	0.805	0.408	0.431
	MAF & MDD	0.889	0.417	0.433
Yuma vs. Odessa	LGF & ODO	0.321	0.400	0.445
	LGF & MAF	0.311	0.416	0.410
	LGF & MDD	0.333	0.412	0.496
	NYL & ODO	0.299	0.412	0.446
	NYL & MAF	0.225	0.413	0.402
	NYL & MDD	0.264	0.421	0.392
	NOZ & ODO	0.263	0.410	0.478
	NOZ & MAF	0.286	0.422	0.400
	NOZ & MDD	0.372	0.418	0.409

Take the martingale decomposition $R_{ij}(\theta) - \mathbb{E}R_{ij}(\theta) = \sum_{l=-\infty}^{T} \mathcal{P}_l R_{ij}(\theta)$. By Theorem 2.1 of [58], we have for $q \geq 4$,

$$\begin{aligned}&\|R_{ij}(\theta) - \mathbb{E}R_{ij}(\theta)\|_{q/2}^2 \\ \leq\ & (q/2-1)\sum_{l=-\infty}^{T}\|\mathcal{P}_l\sum_{r=1}^{n}Y_{ri}(\theta)Y_{rj}(\theta)\|_{q/2}^2 \\ \leq\ & (q/2-1)\sum_{l=-\infty}^{T}\Big(\sum_{r=1}^{n}\|\mathcal{P}_lY_{ri}(\theta)Y_{ij}(\theta)\|_{q/2}\Big)^2. \quad (30)\end{aligned}$$

By Equation (33) in Proposition 8 of [97], we can get

$$\|Y_{ri}(\theta)\|_q \leq (q-1)^{1/2}\sqrt{B}\Delta_{0,q,i} \leq (q-1)^{1/2}\sqrt{B}\|X_{\cdot i}\|_q.$$

By Equation (31) in Proposition 8 of [97], we have

$$\|Y_{ri}(\theta) - Y_{ri,\{l\}}(\theta)\|_q \leq \sum_{t=(r-1)B+1}^{rB}\delta_{t-l,q,i}.$$

By Jensen's inequality, the triangle inequality and Hölder's inequality, it then follows that

$$\begin{aligned}&\sum_{r=1}^{n}\|\mathcal{P}_lY_{ri}(\theta)Y_{rj}(\theta)\|_{q/2} \\ \leq\ & \sum_{r=1}^{n}\|Y_{ri}(\theta)Y_{rj}(\theta) - Y_{rj,\{l\}}(\theta)Y_{rj,\{l\}}(\theta)\|_{q/2} \\ \leq\ & \sum_{r=1}^{n}(\|Y_{ri}(\theta)\|_q\|Y_{rj}(\theta) - Y_{rj,\{l\}}(\theta)\|_q \\ & \quad + \|Y_{ri} - Y_{ri,\{l\}}(\theta)\|_q\|Y_{rj,\{l\}}(\theta)\|_q) \\ \leq\ & (q-1)^{1/2}\sqrt{B}\Big(\|X_{\cdot i}\|_q\sum_{t=1}^{T}\delta_{t-l,q,j} + \|X_{\cdot j}\|_q\sum_{t=1}^{T}\delta_{t-l,q,i}\Big). \quad (31)\end{aligned}$$

By (30) and (31), elementary manipulation implies

$$\begin{aligned}&\|R_{ij}(\theta) - \mathbb{E}R_{ij}(\theta)\|_{q/2} \\ &\quad \leq \sqrt{(q-2)(q-1)}\sqrt{TB}\|X_{\cdot i}\|_q\|X_{\cdot j}\|_q. \quad (32)\end{aligned}$$

Under Assumption 1 and noting that $\|X_{\cdot}\|_{\psi_\alpha} = \max_{1\leq j\leq p}\|X_{\cdot j}\|_{\psi_\alpha}$, (32) suggests

$$\|R_{ij}(\theta) - \mathbb{E}R_{ij}(\theta)\|_{q/2} \leq (q-1)q^{2\alpha}\sqrt{TB}\|X_{\cdot}\|_{\psi_\alpha}^2. \quad (33)$$

Write the negative binomial expansion $(1-s)^{-1/2} = 1 + \sum_{h=1}^{\infty}c_hs^h$ with $c_h = (2h)!/(2^{2h}(h!)^2)$ for $|s| < 1$. By Stirling's formula, we have $c_h \sim (h\pi)^{-1/2}$ as $h \to \infty$. Hence, there exists absolute constants $C_1, C_2 > 0$ such that for all $h \geq 1$,

$$C_1(h/e)^hc_h^{-1} \leq h! \leq C_2(h/e)^hc_h^{-1}. \quad (34)$$

Let $C_0 = 1/(2\alpha+1)$. For any positive integer h such that $C_0h \geq 2$, (33) leads to

$$\|R_{ij}(\theta) - \mathbb{E}R_{ij}(\theta)\|_{C_0h} \leq (2C_0h-1)(2C_0h)^{2\alpha}\sqrt{TB}\|X_{\cdot}\|_{\psi_\alpha}^2.$$

Let $\eta_0 = (2eC_0T^{C_0/2}B^{C_0/2}\|X_{\cdot}\|_{\psi_\alpha}^{2C_0})^{-1}$. Notice that $2\alpha+1 = 1/C_0$. Then

$$\begin{aligned}&\frac{\eta^h\mathbb{E}|R_{ij}(\theta) - \mathbb{E}R_{ij}(\theta)|^{C_0h}}{h!} \\ \leq\ & \frac{\eta^h(2C_0h-1)^{C_0h}(2C_0h)^{2\alpha C_0h}(TB)^{C_0h/2}\|X_{\cdot}\|_{\psi_\alpha}^{2C_0h}}{C_1(h/e)^hc_h^{-1}} \\ \leq\ & \frac{c_h\eta^h(2C_0h-1)^{C_0h}}{C_1\eta_0^h(2C_0h)^{C_0h}} \leq \frac{c_h\eta^h}{C_1\sqrt{e}\eta_0^h}.\end{aligned}$$

If $C_0h < 2$, then $\|R_{ij}(\theta) - \mathbb{E}R_{ij}(\theta)\|_{C_0h} \leq \|R_{ij}(\theta) - \mathbb{E}R_{ij}(\theta)\|_2 \leq \sqrt{6}\cdot 4^{2\alpha}\sqrt{TB}\|X_{\cdot}\|_{\psi_\alpha}^2$. So we have

$$\begin{aligned}&\mathbb{E}[\exp(\eta|R_{jk}(\theta) - \mathbb{E}R_{ij}(\theta)|^{C_0})] \\ \leq\ & 1 + \sum_{1\leq h<2/C_0}\frac{\eta^h(\sqrt{6}\cdot 4^{2\alpha}\sqrt{TB}\|X_{\cdot}\|_{\psi_\alpha}^2)^{C_0h}}{h!} \\ & + \sum_{h\geq 2/C_0}\frac{c_h\eta^h}{C_1\sqrt{e}\eta_0^h} \\ \leq\ & 1 + C_{C_0}\sum_{h=1}^{\infty}c_h\frac{\eta^h}{\eta_0^h} \leq 1 + C_{C_0}\frac{\eta/\eta_0}{(1-\eta/\eta_0)^{1/2}}.\end{aligned}$$

By choosing $\eta = \eta_0/2$ in (29) and recalling $C_0 = 1/(2\alpha+1)$, we have

$$\begin{aligned}&\mathbb{P}\big(|R_{ij}(\theta) - \mathbb{E}R_{ij}(\theta)| \geq x\big) \\ &\quad \lesssim \exp\Big\{-C_\alpha\Big(\frac{x}{\sqrt{TB}\|X_{\cdot}\|_{\psi_\alpha}^2}\Big)^{\frac{1}{2\alpha+1}}\Big\}. \quad (35)\end{aligned}$$

As a result, (8) follows in view of (35), the Bonferroni inequality and the simple scaling $\widehat{F}_{ij}(\theta) = (2\pi T)^{-1}R_{ij}(\theta)$. $\square$

Proof of Proposition 2. By the stationarity of the process, we have

$$\mathbb{E}\widehat{F}_{ij}(\theta) = \frac{1}{2\pi}\sum_{k=-B}^{B}\big(1-\frac{|k|}{B}\big)\Gamma_{ij}(k)e^{-\iota k\theta}.$$

It follows that

$$\begin{aligned}2\pi|\mathbb{E}\widehat{F}_{ij}(\theta) - F_{ij}(\theta)| &\le \sum_{k:|k|\ge B}|\Gamma_{ij}(k)| + \sum_{k=-B+1}^{B-1}\frac{|k|}{B}|\Gamma_{ij}(k)| \\ &:= A_1 + A_2.\end{aligned} \tag{36}$$

Since $X_{ti} = \sum_{r=0}^{\infty}\mathcal{P}_{t-r}(X_{ti})$, for $k\ge 0$, we have

$$\begin{aligned}|\Gamma_{ij}(k)| &= \Big|\sum_{r=0}^{\infty}\mathbb{E}[\mathcal{P}_{t-r}(X_{ti})\mathcal{P}_{t-r}(X_{(t+k)j})]\Big| \\ &\le \sum_{r=0}^{\infty}\delta_{r,2,i}\delta_{r+k,2,j},\end{aligned} \tag{37}$$

which implies $A_1 \lesssim \Delta_{0,2,i}\Delta_{B,2,j}$. In view of $\Delta_{0,2,i}\le \|X_{\cdot i}\|_2$ and $\Delta_{B,2,j}\le \rho^B\|X_{\cdot j}\|_2$, we have $A_1 \lesssim \rho^B\|X_{\cdot i}\|_2\|X_{\cdot j}\|_2$. Furthermore, we can get

$$\begin{aligned}A_2 &\le 2B^{-1}\sum_{k=1}^{B-1}\sum_{r=0}^{\infty}k\delta_{r,2,i}\delta_{r+k,2,j} \\ &\le 2B^{-1}\|X_{\cdot j}\|_2\sum_{k=1}^{B-1}k\rho^k\sum_{r=0}^{\infty}\delta_{r,2,i} \\ &\lesssim B^{-1}\|X_{\cdot i}\|_2\|X_{\cdot j}\|_2.\end{aligned} \tag{38}$$

By combining the bounds of A_1 and A_2, we have for any $1\le i,j\le p$ and θ,

$$|\mathbb{E}\widehat{F}_{ij}(\theta) - F_{ij}(\theta)| \lesssim B^{-1}\|X_{\cdot i}\|_2\|X_{\cdot j}\|_2. \tag{39}$$

□

Proof of Theorem 3. Let $n=T/B$. Define $\mathcal{B}_r = \{(r-1)B+1,\dots,rB\}$ and $\mathcal{D}_r = \{(r-1)B+1,\dots,(r-1)B+B-m\}$ for $r=1,\dots,n$. Then we can write

$$\widehat{F}(\theta) = \big(\widehat{F}_{ij}(\theta)\big)_{i,j=1}^{p} = \frac{1}{2\pi T}\sum_{r=1}^{n}\sum_{t,s\in\mathcal{B}_r}X_tX_s^\top e^{-\iota(t-s)\theta}.$$

For $m\ge 0$, define $X_{t,m} = \mathbb{E}(X_t|\varepsilon_{t-m},\varepsilon_{t-m+1},\dots,\varepsilon_t)$. Let

$$\begin{aligned}\widehat{F}_m(\theta) &= \big(\widehat{F}_{ij,m}(\theta)\big)_{i,j=1}^{p} \\ &= \frac{1}{2\pi T}\sum_{r=1}^{n}\sum_{t,s\in\mathcal{B}_r}X_{t,m}X_{s,m}^\top e^{-\iota(t-s)\theta}.\end{aligned}$$

Further define

$$\begin{aligned}\widehat{H}_{r,m}(\theta) &= \big(\widehat{H}_{ij,r,m}(\theta)\big)_{i,j=1}^{p} \\ &= \frac{1}{2\pi B}\sum_{t,s\in\mathcal{D}_r}X_{t,m}X_{s,m}^\top e^{-\iota(t-s)\theta},\end{aligned}$$

$$\widehat{S}_m(\theta) = \big(\widehat{S}_{ij,m}(\theta)\big)_{i,j=1}^{p} = \frac{1}{n}\sum_{r=1}^{n}\widehat{H}_{r,m}(\theta).$$

By Lemma 9 in [93],

$$\|\mathbb{E}_0(\widehat{F}_{ij}(\theta)) - \mathbb{E}_0(\widehat{F}_{ij,m}(\theta))\|_2 \lesssim \sqrt{\frac{B}{T}}\rho^m\Phi_4^2. \tag{40}$$

Similarly as (71), we can obtain

$$\|\mathbb{E}_0(\widehat{F}_{ij,m}(\theta)) - \mathbb{E}_0(\widehat{S}_{ij,m}(\theta))\|_2 \lesssim \sqrt{\frac{m}{T}}\Phi_4^2. \tag{41}$$

Choosing $m\asymp(\log B)/\log(\rho^{-1})$, by (40) and (41),

$$\sqrt{\frac{T}{B}}\|\mathbb{E}_0(\widehat{F}_{ij}(\theta)) - \mathbb{E}_0(\widehat{S}_{ij,m}(\theta))\|_2 \lesssim \sqrt{\frac{\log B}{B}}\Phi_4^2 \to 0. \tag{42}$$

As a result, the distribution of $\sqrt{T/B}(\widehat{F}_{ij}(\theta) - \mathbb{E}\widehat{F}_{ij}(\theta))$ is dominated by that of $\sqrt{T/B}(\widehat{S}_{ij}(\theta) - \mathbb{E}\widehat{S}_{ij}(\theta))$. Note that $\widehat{H}_{r,m}(\theta)$, $r=1,\dots,n$, are independent and identically distributed. It remains to concern the covariance structure of $\widehat{H}_{1,m}(\theta)$. Take $m\asymp(\log B)/\log(\rho^{-1})$ in (70). Under the condition $\Phi_4^8(\log B)/B\to 0$, we have $\mathcal{E}_{B,m}^{(2)}\to 0$. Note that $(\widehat{H}_{r,m}(\theta))_{r=1}^{n}$ is an i.i.d. sequence. Theorem 3 follows in view of Lemma 10, Lemma 11, the classical central limit theorem, the Cramér-Wold device and the construction of bivariate complex random variable. □

B. Proofs of Results in Section III

Proof of Theorem 4. We shall mainly focus on the case of $\boldsymbol{F}$. $\tilde{\boldsymbol{F}}$ can be dealt with by similar proofs. Recall the definitions of $\widehat{F}_m(\theta)$ and $\widehat{S}_m(\theta)$ in (73) and (75), respectively. Denote

$$\begin{aligned}\boldsymbol{S}_m = &(\mathsf{Re}(\widehat{S}_m(\theta_0)), \mathsf{Im}(\widehat{S}_m(\theta_0)),\dots, \\ &\mathsf{Re}(\widehat{S}_m(\theta_{B/2})), \mathsf{Im}(\widehat{S}_m(\theta_{B/2}))).\end{aligned}$$

Choose $m\asymp\log B/\log\rho^{-1}$ and denote $n=T/B$. We start with

$$\begin{aligned}d_{T,B,p} \le &\sup_{x\in\mathbb{R}}\big|\mathbb{P}(\boldsymbol{F}\le x) - \mathbb{P}(\sqrt{n}\max_{\theta\in\Theta}|\mathbb{E}_0\boldsymbol{S}_m|_\infty\le x)\big| \\ &+ \sup_{x\in\mathbb{R}}\big|\mathbb{P}(\sqrt{n}\max_{\theta\in\Theta}|\mathbb{E}_0\boldsymbol{S}_m|_\infty\le x) - \mathbb{P}(|\boldsymbol{Z}|_\infty\le x)\big| \\ =:& d_{T,B,p}^{\circ} + d_{T,B,p}^{\dagger}.\end{aligned} \tag{43}$$

By the triangle inequality, for every $y>0$, we have

$$\begin{aligned}d_{T,B,p}^{\circ} =& \sup_{x\in\mathbb{R}}\big|\mathbb{P}(\boldsymbol{F}\ge x) - \mathbb{P}(\sqrt{n}\max_{\theta\in\Theta}|\mathbb{E}_0\boldsymbol{S}_m|_\infty\ge x)| \\ \le& \mathbb{P}(\sqrt{n}\max_{\theta\in\Theta}|\mathbb{E}_0(\widehat{F}(\theta)) - \mathbb{E}_0(\widehat{S}_m(\theta))|_\infty\ge y) \\ &+ \sup_{x\in\mathbb{R}}\mathbb{P}(|\sqrt{n}\max_{\theta\in\Theta}|\mathbb{E}_0\boldsymbol{S}_m|_\infty - x|\le y) \\ =:& d_{T,B,p}^{\circ(1)} + d_{T,B,p}^{\circ(2)}.\end{aligned}$$

By Lemma 12 and the Bonferroni inequality, for every $y>0$,

$$\begin{aligned}&\mathbb{P}(\sqrt{n}\max_{\theta\in\Theta}|\mathbb{E}_0(\widehat{F}(\theta)) - \mathbb{E}_0(\widehat{F}_m(\theta))|_\infty\ge y) \\ \lesssim& Bp^2\exp\Big\{-C_\alpha\Big(\frac{y}{\rho^m\|X_{\cdot}\|_{\psi_\alpha}^2}\Big)^{\frac{1}{2\alpha+1}}\Big\}.\end{aligned} \tag{44}$$

And by Lemma 13, we have

$$\mathbb{P}(\sqrt{n}\max_{\theta\in\Theta}|\mathbb{E}_0(\widehat{F}_m(\theta))-\mathbb{E}_0(\widehat{S}_m(\theta))|_\infty\geq y)$$
$$\lesssim Bp^2\exp\Big\{-C_\alpha\Big(\frac{\sqrt{B}y}{\sqrt{m}\|X.\|_{\psi_\alpha}^2}\Big)^{\frac{1}{2\alpha+1}}\Big\}. \quad (45)$$

With the choice of m, it follows that

$$d_{T,B,p}^{\circ(1)}\lesssim Bp^2\exp\Big\{-C_\alpha\Big(\frac{\sqrt{B}y}{\sqrt{\log B}\|X.\|_{\psi_\alpha}^2}\Big)^{\frac{1}{2\alpha+1}}\Big\}. \quad (46)$$

We next consider

$$d_{T,B,p}^{\dagger(1)}=\big|\mathbb{P}(\sqrt{n}\max_{\theta\in\Theta}|\mathbb{E}_0\boldsymbol{S}_m|_\infty\leq x)-\mathbb{P}(|\boldsymbol{Z}_m|_\infty\leq x)\big|,$$
$$d_{T,B,p}^{\dagger(2)}=\big|\mathbb{P}(|\boldsymbol{Z}_m|_\infty\leq x)-\mathbb{P}(|\boldsymbol{Z}|_\infty\leq x)\big|$$

Then we have

$$d_{T,B,p}^{\dagger}\leq d_{T,B,p}^{\dagger(1)}+d_{T,B,p}^{\dagger(2)},$$
$$d_{T,B,p}^{\circ(2)}\leq 2d_{T,B,p}^{\dagger(1)}+2d_{T,B,p}^{\dagger(2)}+\sup_{x\in\mathbb{R}}\mathbb{P}(||\boldsymbol{Z}|_\infty-x|\leq y).$$

As $\mathcal{E}_{B,m}^{(2)}=(\rho^m+\sqrt{m/B})\Phi_4^4\lesssim\sqrt{m/B}\Phi_4^4=o(1)$, Assumption 2, Assumption 3 and Lemma 11 ensure that the variance of each non-zero entry of $\boldsymbol{S}_m$ is lower bounded. Applying in Lemma 14, we have

$$d_{T,B,p}^{\dagger(1)}\lesssim\Big(\frac{\|X.\|_{\psi_\alpha}^{12}\log^{\kappa_1}((Bp)\vee n)}{n}\Big)^{\frac{1}{6}}, \quad (47)$$

for $\kappa_1=\max\{7,2\alpha+6\}$. By Lemma 11 again, it follows that

$$|\boldsymbol{\Sigma}_m-\boldsymbol{\Sigma}|_\infty\lesssim\sqrt{\frac{m}{B}}\Phi_4^4. \quad (48)$$

Lemma 3.1 in [44] implies

$$d_{T,B,p}^{\dagger(2)}\lesssim\big(\sqrt{\frac{m}{B}}\Phi_4^4\big)^{1/3}\log^{2/3}(Bp), \quad (49)$$

By Lemma 15, for every $y>0$.

$$\sup_{x\in\mathbb{R}}\mathbb{P}\big(||\boldsymbol{Z}|_\infty-x|\leq y\big)\lesssim y\sqrt{\log(Bp)}. \quad (50)$$

Choosing

$$y=c_\alpha\frac{\|X.\|_{\psi_\alpha}^2\log^{2\alpha+3/2}(Bp)}{B^{1/2}}$$

with a sufficiently large c_α, the proof is completed by combining the results (43) – (50).

To deal with $\tilde{\boldsymbol{F}}$, the Gaussian approximation result in Lemma 14 concerning the class of all hyperrectangles also holds for $\tilde{\boldsymbol{F}}$ approximated by the distribution of $|\tilde{\boldsymbol{Z}}|_\infty$. The case of $\tilde{\boldsymbol{F}}$ corresponds to the class of s-sparsely convex sets with $s=2$ in the framework of [92]. We can apply the same arguments of Proposition 4.2 in their papar to obtain the Gaussian approximation result. And the anti-concentration inequality still holds for $|\tilde{\boldsymbol{Z}}|_\infty$ in view of (50) and $|\boldsymbol{Z}|_\infty\leq|\tilde{\boldsymbol{Z}}|_\infty\leq\sqrt{2}|\boldsymbol{Z}|_\infty$. □

Proof of Corollary 5. Note that

$$d_{T,B,p}^{\star}\leq\sup_{x\in\mathbb{R}}|\mathbb{P}(\boldsymbol{F}^\star\leq x)-\mathbb{P}(\boldsymbol{F}\leq x)|+d_{T,B,p}.$$

By the triangle inequality, for every $y>0$, we have

$$\sup_{x\in\mathbb{R}}|\mathbb{P}(\boldsymbol{F}^\star\leq x)-\mathbb{P}(\boldsymbol{F}\leq x)|$$
$$\leq\mathbb{P}\Big(\sqrt{\frac{T}{B}}\max_{\theta\in\Theta}|\mathbb{E}\widehat{F}(\theta)-F(\theta)|_\infty\geq y\Big)+\sup_{x\in\mathbb{R}}\mathbb{P}\big(|\boldsymbol{F}-x|\leq y\big).$$

By (50) and the definition of $d_{T,B,p}$, we have

$$\sup_{x\in\mathbb{R}}\mathbb{P}\big(|\boldsymbol{F}-x|\leq y\big)\lesssim d_{T,B,p}+y\sqrt{\log(Bp)}.$$

By Proposition 2, choosing $y=C(T/B)^{1/2}B^{-1}\Phi_2^2$ for a sufficiently large C, condition (21) guarantees $y\sqrt{\log(Bp)}\to 0$, which together with Theorem 4 implies $d_{T,B,p}^\star\to 0$. The case of $\tilde{\boldsymbol{F}}^\star$ can be verified in a similar way. □

Proof of Theorem 6. By the formula of entries in $\boldsymbol{\Sigma}$ as given in Section III-A, the triangle inequality and (67), we have

$$|\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}|_\infty\lesssim\Phi_2^2\max_{\theta\in\Theta}|\widehat{F}_{ij}(\theta)-F_{ij}(\theta)|_\infty. \quad (51)$$

Theorem 1 and Proposition 2 imply

$$\max_{\theta\in\Theta}|\widehat{F}_{ij}(\theta)-F_{ij}(\theta)|_\infty\lesssim\sqrt{\frac{B}{T}}\log^{2\alpha+1}(Bp)\|X.\|_{\psi_\alpha}^2+\frac{\|X.\|_{\psi_\alpha}^2}{B}.$$

Following the arguments of Theorem 3.1 in [44], for every $x>0$,

$$\sup_{\theta_0\in(0,1)}|\mathbb{P}(\boldsymbol{F}^\star\geq\widehat{c}_{1-\theta_0})-\theta_0|$$
$$\lesssim d_{T,B}^\star+\mathcal{W}(x)+\mathbb{P}(|\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}|_\infty\geq x),$$

where $\mathcal{W}(x)=x^{1/3}\log^{2/3}(Bp)$ for $x>0$. Then (23) follows under the additional condition (22) and the conditions of Corollary 5. □

Proof of Theorem 7. For $y\in\mathbb{R}$, let

$$w(y)=\big[1-\min\{0,\max(y,-1)\}^2\big]^2$$

. For any $a>1$ and $b\in\mathbb{R}$, let $w_{a,b}(y)=w(a(y-b))$. It is easy to observe that

$$w_{a,b+a^{-1}}(y)\leq\mathbf{1}\{y>b\}\leq w_{a,b}(y)\leq\mathbf{1}\{y>b-a^{-1}\}.$$

Then we have for any $x>0$,

$$\frac{1}{\mathcal{L}}\sum_{\ell=1}^{\mathcal{L}}w_{a,x+a^{-1}}(\boldsymbol{F}_\ell)-\mathbb{P}(|\boldsymbol{Z}|_\infty>x)$$
$$\leq\frac{1}{\mathcal{L}}\sum_{\ell=1}^{\mathcal{L}}\mathbf{1}\{\boldsymbol{F}_\ell>x\}-\mathbb{P}(|\boldsymbol{Z}|_\infty>x)$$
$$\leq\frac{1}{\mathcal{L}}\sum_{\ell=1}^{\mathcal{L}}w_{a,x}(\boldsymbol{F}_\ell)-\mathbb{P}(|\boldsymbol{Z}|_\infty>x).$$

It leads to

$$\sup_x|\boldsymbol{Q}(x)-\mathbb{P}(|\boldsymbol{Z}|_\infty\leq x)|\leq\sup_{x>0}\mathbb{P}(x<|\boldsymbol{Z}|_\infty\leq x+a^{-1})+\Delta,$$

where

$$\Delta=\sup_{x>0}\Big|\frac{1}{\mathcal{L}}\sum_{\ell=1}^{\mathcal{L}}w_{a,x}(\boldsymbol{F}_\ell)-\mathbb{P}(|\boldsymbol{Z}|_\infty>x)\Big|.$$

By Lemma 15, we have

$$\sup_{x>0}\mathbb{P}(x<|\boldsymbol{Z}|_\infty\le x+a^{-1})\le Ca^{-1}\sqrt{\log(Bp^2)}\lesssim a^{-1}\sqrt{\log(Bp)}.$$

We choose a such that $a^{-1}\sqrt{\log(Bp)}=o(1)$. It suffices to prove $\Delta\xrightarrow{\mathbb{P}}0$. Denote $\boldsymbol{\Sigma}_0=\mathrm{diag}(\boldsymbol{\Sigma})$. Let $x_0=4\sqrt{|\boldsymbol{\Sigma}_0|_\infty\log(Bp^2)}$ and $x_1=x_0+1$. Consider

$$\Delta_1=\sup_{x>x_1}\Big|\frac{1}{\mathcal{L}}\sum_{\ell=1}^{\mathcal{L}}w_{a,x}(\boldsymbol{F}_\ell)-\mathbb{P}(|\boldsymbol{Z}|_\infty>x)\Big|,$$
$$\Delta_2=\sup_{0<x\le x_1}\Big|\frac{1}{\mathcal{L}}\sum_{\ell=1}^{\mathcal{L}}w_{a,x}(\boldsymbol{F}_\ell)-\mathbb{P}(|\boldsymbol{Z}|_\infty>x)\Big|.$$

We firstly deal with Δ_1. By the Bonferroni inequality,

$$\mathbb{P}(|\boldsymbol{Z}|_\infty>x_1)\le Bp^2\cdot\max_j\mathbb{P}(|Z_j|>x_1)\le\frac{1}{Bp^2}\to 0. \tag{52}$$

Also, we can obtain

$$\mathbb{E}\sup_{x>x_1}\frac{1}{\mathcal{L}}\sum_{\ell=1}^{\mathcal{L}}w_{a,x}(\boldsymbol{F}_\ell)\le\frac{1}{\mathcal{L}}\sum_{\ell=1}^{\mathcal{L}}\mathbb{E}w_{a,x_1}(\boldsymbol{F}_\ell)\le\frac{1}{\mathcal{L}}\sum_{\ell=1}^{\mathcal{L}}\mathbb{P}(\boldsymbol{F}_\ell>x_1-a^{-1}). \tag{53}$$

For each $1\le\ell\le\mathcal{L}$, define

$$\boldsymbol{F}_\ell^\star=\sqrt{\frac{M}{B}}\max_{\theta\in\Theta}\max\big\{|\mathsf{Re}(\widehat{F}_\ell(\theta)-F(\theta))|_\infty,\ |\mathsf{Im}(\widehat{F}_\ell(\theta)-F(\theta))|_\infty\big\}.$$

Recall that

$$\boldsymbol{F}^\star=\sqrt{\frac{T}{B}}\max_{\theta\in\Theta}\max\big\{|\mathsf{Re}(\widehat{F}(\theta)-F(\theta))|_\infty,\ |\mathsf{Im}(\widehat{F}(\theta)-F(\theta))|_\infty\big\}. \tag{54}$$

Since $a>1$, by the triangle inequality and Corollary 5, it follows that

$$\begin{aligned}&\mathbb{P}(\boldsymbol{F}_\ell>x_1-a^{-1})\\ &\le\mathbb{P}(\boldsymbol{F}_\ell^\star>x_0/2)+\mathbb{P}(\boldsymbol{F}^\star>\sqrt{T/M}\cdot x_0/2)\\ &\le 2\mathbb{P}(|\boldsymbol{Z}|_\infty>x_0/2)+d_{M,B,p}+d_{T,B,p}\\ &\le\frac{2}{Bp^2}+d_{M,B,p}+d_{T,B,p}\to 0.\end{aligned} \tag{55}$$

By (52), (53) and (55), it follows that $\Delta_1\xrightarrow{\mathbb{P}}0$. To deal with Δ_2, we discretize the interval $[0,x_1]$ by evenly spaced points. Let K be large such that $K^{-1}\Phi_2^2\log(Bp)=o(1)$ and $\tilde{x}_k=k\cdot x_1/K$ for $k=0,1,\ldots,K$. By the triangle inequality,

$$\Delta_2\le\Delta_2^{(1)}+\Delta_2^{(2)}+\Delta_2^{(3)},$$

where

$$\Delta_2^{(1)}=\max_{1\le k\le K}\mathbb{P}(\tilde{x}_{k-1}<|\boldsymbol{Z}|_\infty\le\tilde{x}_k),$$
$$\Delta_2^{(2)}=\max_{0\le k\le K}\Big|\frac{1}{\mathcal{L}}\sum_{\ell=1}^{\mathcal{L}}\big[\mathbb{E}w_{a,\tilde{x}_k}(\boldsymbol{F}_\ell)-\mathbb{P}(|\boldsymbol{Z}|_\infty>\tilde{x}_k)\big]\Big|,$$
$$\Delta_2^{(3)}=\max_{0\le k\le K}\Big|\frac{1}{\mathcal{L}}\sum_{\ell=1}^{\mathcal{L}}\big[w_{a,\tilde{x}_k}(\boldsymbol{F}_\ell)-\mathbb{E}w_{a,\tilde{x}_k}(\boldsymbol{F}_\ell)\big]\Big|.$$

By Theorem 3 and (67), $|\boldsymbol{\Sigma}_0|_\infty\le C\Phi_2^4$. By Lemma 15,

$$\Delta_2^{(1)}\le Cx_1K^{-1}\sqrt{\log(Bp^2)}\lesssim K^{-1}\sqrt{|\boldsymbol{\Sigma}_0|_\infty}\log(Bp)\to 0.$$

Before working on $\Delta_2^{(2)}$, we consider

$$\begin{aligned}d_{M,B}^\ell&:=\sup_{x>0}|\mathbb{P}(\boldsymbol{F}_\ell\le x)-\mathbb{P}(|\boldsymbol{Z}|_\infty\le x)|\\ &=\sup_{x>0}|\mathbb{P}(\boldsymbol{F}_\ell>x)-\mathbb{P}(|\boldsymbol{Z}|_\infty>x)|.\end{aligned}$$

By the triangle inequality, Corollary 5 and Lemma 15, for any $b>0$,

$$\begin{aligned}&\sup_{x>0}|\mathbb{P}(\boldsymbol{F}_\ell\le x)-\mathbb{P}(\boldsymbol{F}_\ell^\star\le x)|\\ &\le\mathbb{P}(\boldsymbol{F}^\star>b\sqrt{T/M})+2d_{M,B,p}+Cb\sqrt{\log(Bp^2)}.\end{aligned}$$

Choosing $b=2\sqrt{M|\boldsymbol{\Sigma}_0|_\infty\log(Bp^2)/T}$, it follows that

$$\mathbb{P}(\boldsymbol{F}^\star>b\sqrt{T/M})\le\mathbb{P}(|\boldsymbol{Z}|_\infty>b\sqrt{T/M})+d_{T,B,p}\le\frac{1}{Bp^2}+d_{T,B,p}.$$

If $M\log^2(Bp)\Phi_2^4/T\to 0$, we have for any $1\le\ell\le\mathcal{L}$,

$$d_{M,B}^\ell\le\sup_{x>0}|\mathbb{P}(\boldsymbol{F}_\ell\le x)-\mathbb{P}(\boldsymbol{F}_\ell^\star\le x)|+d_{M,B,p}\to 0.$$

Now we deal with $\Delta_2^{(2)}$. On the one hand,

$$\mathbb{E}w_{a,\tilde{x}_k}(\boldsymbol{F}_\ell)\ge\mathbb{P}(\boldsymbol{F}_\ell>\tilde{x}_k)\ge\mathbb{P}(|\boldsymbol{Z}|_\infty>\tilde{x}_k)-d_{M,B}^\ell.$$

On the other hand,

$$\mathbb{E}w_{a,\tilde{x}_k}(\boldsymbol{F}_\ell)\le\mathbb{P}(\boldsymbol{F}_\ell>\tilde{x}_k-a^{-1})\le\mathbb{P}(|\boldsymbol{Z}|_\infty>\tilde{x}_k-a^{-1})+d_{M,B}^\ell,$$

where

$$\begin{aligned}&\mathbb{P}(|\boldsymbol{Z}|_\infty>\tilde{x}_k-a^{-1})\\ &\le\mathbb{P}(|\boldsymbol{Z}|_\infty>\tilde{x}_k)+\mathbb{P}(\tilde{x}_k-a^{-1}<|\boldsymbol{Z}|_\infty\le\tilde{x}_k)\\ &\le\mathbb{P}(|\boldsymbol{Z}|_\infty>\tilde{x}_k)+Ca^{-1}\sqrt{\log(Bp^2)}.\end{aligned}$$

Hence it follows that

$$\Delta_2^{(2)}\le\max_\ell d_{M,B,p}^\ell+Ca^{-1}\sqrt{\log(Bp^2)}\to 0.$$

It remains to consider $\Delta_2^{(3)}$. We are to show $\Delta_2^{(3)}\xrightarrow{\mathbb{P}}0$ by verifying $\mathbb{E}(\Delta_2^{(3)})^2\to 0$. Define

$$\mathcal{U}=\max_{0\le k\le K}\max_{1\le\ell,\ell'\le\mathcal{L}}\Big|\mathsf{Cov}\big(w_{a,\tilde{x}_k}(\boldsymbol{F}_\ell),w_{a,\tilde{x}_k}(\boldsymbol{F}_{\ell'})\big)-\mathsf{Cov}\big(w_{a,\tilde{x}_k}(\boldsymbol{F}_\ell^\star),w_{a,\tilde{x}_k}(\boldsymbol{F}_{\ell'}^\star)\big)\Big|.$$

Since $0\le w(y)\le 1$ for all $y\in\mathbb{R}$, by Hölder's inequality,

$$\mathcal{U}\le 4\max_{0\le k\le K}\max_{1\le\ell\le\mathcal{L}}\|w_{a,\tilde{x}_k}(\boldsymbol{F}_\ell)-w_{a,\tilde{x}_k}(\boldsymbol{F}_\ell^\star)\|_2.$$

By elementary calculations, we can obtain $\sup_{y\in\mathbb{R}}|w_{a,b}'(y)|\le 2a$, which implies for any $0\le k\le K$ and $1\le\ell\le\mathcal{L}$,

$$\|w_{a,\tilde{x}_k}(\boldsymbol{F}_\ell)-w_{a,\tilde{x}_k}(\boldsymbol{F}_\ell^\star)\|_2\le 2a\|\boldsymbol{F}_\ell-\boldsymbol{F}_\ell^\star\|_2\le 2a\sqrt{\frac{M}{T}}\|\boldsymbol{F}^\star\|_2.$$

Note that

$$\sqrt{\frac{B}{T}}\|\boldsymbol{F}^\star\|_2 \le \|\max_{\theta\in\Theta}|\widehat{F}(\theta)-\mathbb{E}\widehat{F}(\theta)|_\infty\|_2 + \max_{\theta\in\Theta}|\mathbb{E}\widehat{F}(\theta)-F(\theta)|_\infty.$$

By Proposition 2,

$$\max_{\theta\in\Theta}|\mathbb{E}\widehat{F}(\theta)-F(\theta)|_\infty \lesssim B^{-1}\Phi_2^2.$$

By Theorem 1,

$$\|\max_{\theta\in\Theta}|\widehat{F}(\theta)-\mathbb{E}\widehat{F}(\theta)|_\infty\|_2 \lesssim \sqrt{\frac{B}{T}}\log^{2\alpha+1}(Bp)\cdot\|X_\cdot\|_{\psi_\alpha}^2.$$

It can be obtained that

$$\mathcal{U} \lesssim a\sqrt{\frac{M}{B^3}}\Phi_2^2 + a\sqrt{\frac{M}{T}}\log^{2\alpha+1}(Bp)\cdot\|X_\cdot\|_{\psi_\alpha}^2. \tag{56}$$

Then we have

$$\begin{aligned}&\mathbb{E}(\Delta_2^{(3)})^2\\ \le\ &\sum_{k=0}^K\frac{1}{\mathcal{L}^2}\sum_{\ell=1}^{\mathcal{L}}\sum_{\ell'=1}^{\mathcal{L}}\mathrm{Cov}\big(w_{a,\tilde{x}_k}(\boldsymbol{F}_\ell),w_{a,\tilde{x}_k}(\boldsymbol{F}_{\ell'})\big)\\ \le\ &\sum_{k=0}^K\frac{1}{\mathcal{L}^2}\sum_{\ell=1}^{\mathcal{L}}\sum_{\ell'=1}^{\mathcal{L}}\mathrm{Cov}\big(w_{a,\tilde{x}_k}(\boldsymbol{F}_\ell^\star),w_{a,\tilde{x}_k}(\boldsymbol{F}_{\ell'}^\star)\big)+(K+1)\mathcal{U}.\end{aligned} \tag{57}$$

For any $m\ge 1$, let $\widehat{F}_{\ell,m}(\theta)$ be the m-approximation of $\widehat{F}_\ell(\theta)$ where $\{X_t\}$ is replaced by $\{X_{t,m}\}$. Define

$$\boldsymbol{F}_{\ell,m}^\star = \sqrt{\frac{M}{B}}\max_{\theta\in\Theta}\max\big\{|\mathsf{Re}(\widehat{F}_{\ell,m}(\theta)-F(\theta))|_\infty, |\mathsf{Im}(\widehat{F}_{\ell,m}(\theta)-F(\theta))|_\infty\big\}.$$

Note that $\boldsymbol{F}_0^\star$ is independent of $\boldsymbol{F}_{\ell,m_\ell}^\star$ for $m_\ell=\lfloor(\ell-1)/2\rfloor$ and $\ell\ge 2M-1$. Following (57) and by the stationarity of the process $\{\boldsymbol{F}_\ell^\star\}_{\ell\in\mathcal{Z}}$, we have

$$\begin{aligned}&\mathbb{E}(\Delta_2^{(3)})^2\\ \le\ &\sum_{k=0}^K\frac{2}{\mathcal{L}}\sum_{\ell=0}^{\mathcal{L}-1}\big|\mathrm{Cov}\big(w_{a,\tilde{x}_k}(\boldsymbol{F}_0^\star),w_{a,\tilde{x}_k}(\boldsymbol{F}_\ell^\star)\big)\big|+2K\mathcal{U}\\ \le\ &\sum_{k=0}^K\frac{2}{\mathcal{L}}\sum_{\ell=1}^{\mathcal{L}-1}\big|\mathrm{Cov}\big(w_{a,\tilde{x}_k}(\boldsymbol{F}_0^\star),w_{a,\tilde{x}_k}(\boldsymbol{F}_\ell^\star)-w_{a,\tilde{x}_k}(\boldsymbol{F}_{\ell,m_\ell}^\star)\big)\big|\\ &+\frac{4(K+1)M}{\mathcal{L}}+2K\mathcal{U}\\ \le\ &\frac{2}{\mathcal{L}}\sum_{k=0}^K\sum_{\ell=1}^{\mathcal{L}}\min\big\{1,\|w_{a,\tilde{x}_k}(\boldsymbol{F}_\ell^\star)-w_{a,\tilde{x}_k}(\boldsymbol{F}_{\ell,m_\ell}^\star)\|_2\big\}\\ &+\frac{4(K+1)M}{\mathcal{L}}+2K\mathcal{U}.\end{aligned} \tag{58}$$

By similar arguments in deriving (56), we have

$$\begin{aligned}&\|w_{a,\tilde{x}_k}(\boldsymbol{F}_\ell^\star)-w_{a,\tilde{x}_k}(\boldsymbol{F}_{\ell,m_\ell}^\star)\|_2\\ &\le 2a\sqrt{\frac{M}{B}}\|\max_{\theta\in\Theta}|\widehat{F}_\ell(\theta)-\widehat{F}_{\ell,m_\ell}(\theta)|_\infty\|_2.\end{aligned}$$

By Lemma 12, we have

$$\begin{aligned}&\|\max_{\theta\in\Theta}|\widehat{F}_\ell(\theta)-\widehat{F}_{\ell,m_\ell}(\theta)|_\infty\|_2\\ &\lesssim \rho^{\ell/2}\sqrt{\frac{B}{M}}\log^{2\alpha+1}(Bp)\|X_\cdot\|_{\psi_\alpha}^2.\end{aligned}$$

As a result, for any $0\le k\le K$,

$$\|w_{a,\tilde{x}_k}(\boldsymbol{F}_\ell^\star)-w_{a,\tilde{x}_k}(\boldsymbol{F}_{\ell,m_\ell}^\star)\|_2 \lesssim a\rho^{\ell/2}\log^{2\alpha+1}(Bp)\|X_\cdot\|_{\psi_\alpha}^2.$$

By (58) and $\mathcal{L}\asymp T$, $\mathbb{E}(\Delta_2^{(3)})^2\to 0$ follows if it holds that (i) $\mathcal{K}aT^{-1}\log^{2\alpha+1}(Bp)\|X_\cdot\|_{\psi_\alpha}^2\to 0$, (ii) $KM/T\to 0$, and (iii) $K\mathcal{U}\to 0$, where $\mathcal{U}$ is bounded in (56). We recall that the following conditions are additionally required to control other terms in the previous proof: (iv) $a^{-1}\sqrt{\log(Bp)}\to 0$, (v) $\mathcal{K}^{-1}\Phi_2^2\log(Bp)\to 0$, (vi) $M\log^2(Bp)\Phi_2^4/T\to 0$. In view of $\Phi_2\lesssim\|X_\cdot\|_{\psi_\alpha}$, elementary manipulation indicates that theses conditions can be reduced to (24) and (25), under which the existence of satisfactory a and $\mathcal{K}$ is ensured. □

VII. Conclusion

Spectral density is a pivotal component in the analysis of time series data. While a systematic asymptotic theory has been developed for spectral estimates in low-dimensional cases, the distributional theory for high-dimensional time series remains underexplored. Our work endeavors to fill this gap by formulating an inference theory specifically tailored for the high-dimensional spectral density matrix. Our primary focus is on establishing asymptotic normality for point-wise inference and a Gaussian approximation for simultaneous inference. Our distribution theory serves as a powerful tool for addressing a diverse range of challenges in time series inference. Additionally, we introduce two distinct resampling methods designed to facilitate practical implementation of high-dimensional spectral inference. Importantly, we provide theoretical justification for the validity of these methods, enhancing their reliability in real-world applications. We expect that our framework, distribution theory of spectral estimate and implementation methods will be useful in many high-dimensional inference problems and practical applications that involve time series data.

VIII. Appendix

In the Appendix, we provide the lemmas that are used in Section VI and their proofs .

A. Useful Lemmas for Expectation and Covariance of Spectral Estimates

Lemma 8. *Denote $\Phi_2=\max_{1\le j\le p}\|X_{\cdot j}\|_2$. For $m\ge 0$, define $X_{t,m}=\mathbb{E}(X_t|\varepsilon_{t-m},\varepsilon_{t-m+1},\ldots,\varepsilon_t)$ and*

$$\widehat{G}_{ij,m}(\theta)=\frac{1}{2\pi B}\sum_{t,s=1}^B X_{ti,m}X_{sj,m}e^{-\iota(t-s)\theta}. \tag{59}$$

Then for any $1\le i,j\le p$ and $0\le\theta\le\pi$,

$$\big|\mathbb{E}\widehat{G}_{ij,m}(\theta)-F_{ij}(\theta)\big|\lesssim(\rho^m+B^{-1})\Phi_2^2. \tag{60}$$

Proof of Lemma 8. By the definitions of $\widehat{G}_{ij}(\theta)$ and $\widehat{G}_{ij,m}(\theta)$, we have

$$(2\pi B)\mathbb{E}\widehat{G}_{ij}(\theta) = \mathbb{E}\Big(\sum_{t=1}^{B} X_{ti}e^{-\iota t\theta}\Big)\Big(\sum_{s=1}^{B} X_{sj}e^{\iota s\theta}\Big),$$
$$(2\pi B)\mathbb{E}\widehat{G}_{ij,m}(\theta) = \mathbb{E}\Big(\sum_{t=1}^{B} X_{ti,m}e^{-\iota t\theta}\Big)\Big(\sum_{s=1}^{B} X_{sj,m}e^{\iota s\theta}\Big).$$

By Hölder's inequality,

$$(2\pi B)|\mathbb{E}\widehat{G}_{ij}(\theta) - \mathbb{E}\widehat{G}_{ij,m}(\theta)| \le \mathrm{I} + \mathrm{II}, \tag{61}$$

where

$$\mathrm{I} = \Big\|\sum_{t=1}^{B}(X_{ti} - X_{ti,m})e^{-\iota t\theta}\Big\|_2\Big\|\sum_{s=1}^{B} X_{sj}e^{\iota s\theta}\Big\|_2,$$
$$\mathrm{II} = \Big\|\sum_{t=1}^{B} X_{ti,m}e^{-\iota t\theta}\Big\|_2\Big\|\sum_{s=1}^{B}(X_{sj} - X_{sj,m})e^{\iota s\theta}\Big\|_2.$$

By Lemma 1 in [24], we have $\|\sum_{s=1}^{B} X_{sj}e^{\iota s\theta}\|_2 \lesssim \sqrt{B}\Delta_{0,2,j}$, $\|\sum_{t=1}^{B} X_{ti,m}e^{-\iota t\theta}\|_2 \lesssim \sqrt{B}\Delta_{0,2,i}$, $\|\sum_{t=1}^{B}(X_{ti} - X_{ti,m})e^{-\iota t\theta}\Big\|_2 \lesssim \sqrt{B}\Delta_{m+1,2,i}$ and $\|\sum_{s=1}^{B}(X_{sj} - X_{sj,m})e^{\iota s\theta}\Big\|_2 \lesssim \sqrt{B}\Delta_{m+1,2,j}$. Since $\Delta_{0,2,i} \le \|X_{\cdot i}\|_2$ and $\Delta_{m+1,2,i} \le \Delta_{m,2,i} \le \rho^m\|X_{\cdot i}\|_2$, it follows that

$$\mathrm{I} \lesssim B\rho^m\|X_{\cdot i}\|_2\|X_{\cdot j}\|_2, \quad \mathrm{II} \lesssim B\rho^m\|X_{\cdot i}\|_2\|X_{\cdot j}\|_2. \tag{62}$$

By the proof of Proposition 2, we have

$$|\mathbb{E}\widehat{G}_{ij}(\theta) - F_{ij}(\theta)| \lesssim B^{-1}\|X_{\cdot i}\|_2\|X_{\cdot j}\|_2. \tag{63}$$

By (61), (62), (63) and the triangle inequality, (60) follows by noting $\Phi_2 = \max_{1\le j\le p}\|X_{\cdot j}\|_2$. ☐

Lemma 9. *For $1 \le i,j,h,l \le p$ and $k_1,k_2,k_3 \in \mathbb{Z}$, denote the 4-th joint cumulant as*

$$\gamma_{i,j,h,l}(k_1,k_2,k_3) = \mathit{Cum}(X_{0i}, X_{k_1j}, X_{k_2h}, X_{k_3l}).$$

Denote $\Phi_4 = \max_{1\le j\le p}\|X_{\cdot j}\|_4$. Then for some constant $C_\rho > 0$,

$$\sum_{k_1,k_2,k_3\in\mathbb{Z}} |\gamma_{i,j,h,l}(k_1,k_2,k_3)| \le C_\rho\Phi_4^4.$$

Proof of Lemma 9. We can apply the similar arguments of proving Theorem 21 in [97] without extra difficulty. ☐

Lemma 10. *Denote $\Phi_4 = \max_{1\le j\le p}\|X_{\cdot j}\|_4$. Take any $\theta,\theta' \in \Theta$ and any $1 \le i,j,h,l \le p$. Consider $\widehat{G}_{ij,m}(\theta)$ and $\widehat{G}_{hl,m}(\theta)$ as given by (59). Denote $\mathcal{E}_{B,m}^{(1)} := (\rho^m + B^{-1})\Phi_4^4$.*
(i) If $\theta \in (0,\pi)$,

$$|\mathbb{E}(\widehat{G}_{ij,m}(\theta)\widehat{G}_{hl,m}(\theta)) - \mathbb{E}\widehat{G}_{ij,m}(\theta)\mathbb{E}\widehat{G}_{hl,m}(\theta) - F_{il}(\theta)F_{hj}(\theta)| \lesssim \mathcal{E}_{B,m}^{(1)},$$
$$|\mathbb{E}(\widehat{G}_{ij,m}(\theta)\widehat{G}_{hl,m}^H(\theta)) - \mathbb{E}\widehat{G}_{ij,m}(\theta)\mathbb{E}\widehat{G}_{hl,m}^H(\theta) - F_{ih}(\theta)F_{lj}(\theta)| \lesssim \mathcal{E}_{B,m}^{(1)}.$$

(ii) If $\theta = 0$ or π,

$$|\mathbb{E}(\widehat{G}_{ij,m}(\theta)\widehat{G}_{hl,m}(\theta)) - \mathbb{E}\widehat{G}_{ij,m}(\theta)\mathbb{E}\widehat{G}_{hl,m}(\theta) - F_{ih}(\theta)F_{lj}(\theta) - F_{il}(\theta)F_{hj}(\theta)| \lesssim \mathcal{E}_{B,m}^{(1)}.$$

(iii) If $\theta \ne \theta'$,

$$|\mathbb{E}(\widehat{G}_{ij,m}(\theta)\widehat{G}_{hl,m}(\theta')) - \mathbb{E}\widehat{G}_{ij,m}(\theta)\mathbb{E}\widehat{G}_{hl,m}(\theta')| \lesssim \mathcal{E}_{B,m}^{(1)},$$
$$|\mathbb{E}(\widehat{G}_{ij,m}(\theta)\widehat{G}_{hl,m}^H(\theta')) - \mathbb{E}\widehat{G}_{ij,m}(\theta)\mathbb{E}\widehat{G}_{hl,m}^H(\theta')| \lesssim \mathcal{E}_{B,m}^{(1)},$$

Proof. For any $\theta,\theta' \in \Theta$ and any $1 \le i,j,h,l \le p$, we can write

$$\mathbb{E}\widehat{G}_{ij,m}(\theta)\widehat{G}_{hl,m}(\theta') = \frac{1}{4\pi B^2}\sum_{t,s,u,v=1}^{B}\mathbb{E}(X_{ti,m}X_{sj,m}X_{uh,m}X_{vl,m}) \times e^{-\iota[(t-s)\theta+(u-v)\theta']}. \tag{64}$$

As $\mathbb{E}X_{ti,m} = 0$ for any $t \in \mathbb{Z}$ and $1 \le i \le p$, by the decomposition of the 4-th joint cumulant, we have

$$\begin{aligned}&\mathbb{E}(X_{ti,m}X_{sj,m}X_{uh,m}X_{vl,m})\\ &= \mathbb{E}(X_{ti,m}X_{sj,m})\mathbb{E}(X_{uh,m}X_{vl,m})\\ &\quad+ \mathbb{E}(X_{ti,m}X_{uh,m})\mathbb{E}(X_{sj,m}X_{vl,m})\\ &\quad+ \mathbb{E}(X_{ti,m}X_{vl,m})\mathbb{E}(X_{sj,m}X_{uh,m})\\ &\quad+ \mathrm{Cum}(X_{ti,m}, X_{sj,m}, X_{uh,m}, X_{vl,m}).\end{aligned}$$

Denote

$$\mathrm{I} = \frac{1}{4\pi^2B^2}\sum_{t,s,u,v=1}^{B}\mathbb{E}(X_{ti,m}X_{sj,m})\mathbb{E}(X_{uh,m}X_{vl,m}) \times e^{-\iota[(t-s)\theta+(u-v)\theta']},$$
$$\mathrm{II} = \frac{1}{4\pi^2B^2}\sum_{t,s,u,v=1}^{B}\mathbb{E}(X_{ti,m}X_{uh,m})\mathbb{E}(X_{sj,m}X_{vl,m}) \times e^{-\iota[(t-s)\theta+(u-v)\theta']},$$
$$\mathrm{III} = \frac{1}{4\pi^2B^2}\sum_{t,s,u,v=1}^{B}\mathbb{E}(X_{ti,m}X_{vl,m})\mathbb{E}(X_{sj,m}X_{uh,m}) \times e^{-\iota[(t-s)\theta+(u-v)\theta']},$$
$$\mathrm{IV} = \frac{1}{4\pi^2B^2}\sum_{t,s,u,v=1}^{B}\mathrm{Cum}(X_{ti,m}, X_{sj,m}, X_{uh,m}, X_{vl,m}) \times e^{-\iota[(t-s)\theta+(u-v)\theta']}.$$

It can be easily seen that $\mathrm{I} = (\mathbb{E}\widehat{G}_{ij,m}(\theta))(\mathbb{E}\widehat{G}_{hl,m}(\theta'))$. Applying Lemma 9 to the process after m-approximation, we have $|\mathrm{IV}| \lesssim B^{-1}\Phi_4^4$.

Next we shall consider II and III. For $\theta = \theta' = 0$ or π, $\widehat{G}_{ij,m}(\theta)$ and $\widehat{G}_{hl,m}(\theta)$ are both real. It follows that

$$\begin{aligned}\mathrm{II} &= \frac{1}{4\pi^2B^2}\sum_{t,s,u,v=1}^{B}\mathbb{E}(X_{ti,m}X_{uh,m})\mathbb{E}(X_{sj,m}X_{vl,m}) \times e^{-\iota[(t-s)\theta+(u-v)\theta]}\\ &= \frac{1}{4\pi^2B^2}\sum_{t,s,u,v=1}^{B}\mathbb{E}(X_{ti,m}X_{uh,m})\mathbb{E}(X_{sj,m}X_{vl,m}) \times e^{-\iota[(t-s)\theta-(u-v)\theta]}\\ &= \mathbb{E}\widehat{G}_{ih,m}(\theta)\mathbb{E}\widehat{G}_{lj,m}(\theta).\end{aligned}$$

Observe that

$$\mathrm{II} = \frac{1}{4\pi^2 B^2}\Big(\sum_{t,u=1}^{B} \Gamma_{ih,m}(t-u)e^{-\iota(t\theta+u\theta')}\Big) \times \Big(\sum_{s,v=1}^{B} \Gamma_{jl,m}(s-v)e^{\iota(s\theta+v\theta')}\Big),$$

where $\Gamma_{ij,m}(k) = \mathrm{Cov}(X_{0i,m}, X_{kj,m})$ for any $1 \le i,j \le p$ and $k \in \mathbb{Z}$. We have

$$\sum_{t,u=1}^{B} \Gamma_{ih,m}(t-u)e^{-\iota(t\theta+u\theta')} = \sum_{k=-B+1}^{B-1} \Gamma_{ih,m}(k)e^{-\iota k\theta} \sum_{u=1}^{B-|k|} e^{-\iota u(\theta+\theta')}. \quad (65)$$

If $\theta + \theta' \neq 0$ or 2π, by the fact that $\sum_{u=1}^{B} e^{-\iota u(\theta+\theta')} = 0$, we have $|\sum_{u=1}^{B-|k|} e^{-\iota u(\theta+\theta')}| \le |k|$. By similar arguments in (37) and (38), we have

$$\sum_{k=1}^{B-1} k|\Gamma_{ih,m}(k)| \le \sum_{k=1}^{B-1}\sum_{r=0}^{\infty} k\delta_{r,2,i}\delta_{r+k,2,h} \lesssim \Phi_2^2,$$

Dealing with $\sum_{s,v=1}^{B} \Gamma_{jl,m}(s-v)e^{\iota(s\theta+v\theta')}$ in the same way, we can obtain $|\mathrm{II}| \lesssim B^{-2}\Phi_2^4$ if $\theta + \theta' \neq 0$ or 2π.

As for III, if $\theta = \theta'$, it can be observed that $\mathrm{III} = \mathbb{E}\widehat{G}_{il,m}(\theta))(\mathbb{E}\widehat{G}_{hj,m}(\theta)$. If $\theta \neq \theta'$, we can write

$$\mathrm{III} = \frac{1}{4\pi^2 B^2}\Big(\sum_{t,v=1}^{B} \Gamma_{il,m}(t-v)e^{-\iota(t\theta-v\theta')}\Big) \times \Big(\sum_{s,u=1}^{B} \Gamma_{jh,m}(s-u)e^{\iota(s\theta-v\theta')}\Big),$$

and apply similar arguments as (65) to get the same bound.

Combining $\mathrm{I}, \mathrm{II}, \mathrm{III}$ and IV and noting that $\Phi_2 \le \Phi_4$, we have

$$|\mathbb{E}(\widehat{G}_{ij,m}(\theta)\widehat{G}_{hl,m}(\theta')) - \mathbb{E}\widehat{G}_{ij,m}(\theta)\mathbb{E}\widehat{G}_{hl,m}(\theta') - V_m| \lesssim B^{-1}\Phi_4^4, \quad (66)$$

where $V_m = \mathbb{E}\widehat{G}_{il,m}(\theta)\mathbb{E}\widehat{G}_{hj,m}(\theta)$ if $\theta = \theta' \in (0,\pi)$, $V_m = \mathbb{E}\widehat{G}_{ih,m}(\theta)\mathbb{E}\widehat{G}_{lj,m}(\theta) + \mathbb{E}\widehat{G}_{il,m}(\theta)\mathbb{E}\widehat{G}_{hj,m}(\theta)$ if $\theta = \theta' = 0$ or π, and $V_m = 0$ if $\theta \neq \theta'$. Let $V = F_{il}(\theta)F_{hj}(\theta)$ if $\theta = \theta' \in (0,\pi)$, $V = F_{ih}(\theta)F_{lj,m}(\theta) + F_{il,m}(\theta)F_{hj,m}(\theta)$ if $\theta = \theta' = 0$ or π, and $V = 0$ if $\theta \neq \theta'$. We now account for the bias between V_m and V. By (37), for any $1 \le i,j \le p$ and θ,

$$|F_{ij}(\theta)| \le 2\sum_{k=0}^{\infty}|\Gamma_{ij}(k)| \le 2\sum_{k=0}^{\infty}\sum_{r=0}^{\infty}\delta_{r,2,i}\delta_{r+k,2,j} \le 2\Phi_2^2. \quad (67)$$

By the triangle inequality, Lemma 8 and (67),

$$|V_m - V| \lesssim (\rho^m + B^{-1})\Phi_2^4. \quad (68)$$

In view of (66), (68) and $\Phi_2 \le \Phi_4$, the desired results for $\mathbb{E}(\widehat{G}_{ij,m}(\theta)\widehat{G}_{hl,m}(\theta')) - \mathbb{E}\widehat{G}_{ij,m}(\theta)\mathbb{E}\widehat{G}_{hl,m}(\theta')$ in (i), (ii) and (iii) can be reached.

When we consider $\mathbb{E}(\widehat{G}_{ij,m}(\theta)\widehat{G}_{hl,m}^H(\theta'))$, (64) becomes

$$\mathbb{E}\widehat{G}_{ij,m}(\theta)\widehat{G}_{hl,m}^H(\theta') = \frac{1}{4\pi B^2}\sum_{t,s,u,v=1}^{B}\mathbb{E}(X_{ti,m}X_{sj,m}X_{uh,m}X_{vl,m}) \times e^{-[\iota(t-s)\theta-(u-v)\theta']}.$$

We can apply the same manipulation to deal with $\mathbb{E}(X_{ti,m}X_{sj,m}X_{uh,m}X_{vl,m})$ and define $\mathrm{I}^\diamond$, $\mathrm{II}^\diamond$, $\mathrm{III}^\diamond$ and $\mathrm{IV}^\diamond$ with the multiplier $e^{-\iota[(t-s)\theta-(u-v)\theta']}$ in place of $e^{-\iota[(t-s)\theta+(u-v)\theta']}$. It follows that $\mathrm{I}^\diamond = (\mathbb{E}\widehat{G}_{ij,m}(\theta))(\mathbb{E}\widehat{G}_{hl,m}^H(\theta'))$. Similar to III, we can derive $\mathrm{II}^\diamond = \mathbb{E}\widehat{G}_{ih,m}(\theta)\mathbb{E}\widehat{G}_{lj,m}(\theta)$ if $\theta = \theta'$ and $|\mathrm{II}^\diamond| \lesssim B^{-2}\Phi_2^4$ if $\theta \neq \theta'$. And

$$\mathrm{III}^\diamond = \frac{1}{4\pi^2 B^2}\Big(\sum_{t,u=1}^{B}\Gamma_{il,m}(t-v)e^{-\iota(t\theta+v\theta')}\Big) \times \Big(\sum_{u,s=1}^{B}\Gamma_{jh,m}(s-u)e^{\iota(s\theta+u\theta')}\Big)$$

can be dealt with similarly as II. □

Lemma 11. *For any $1 \le m \le B$, $1 \le i,j \le p$ and $\theta \in \Theta$, define*

$$\widehat{H}_{ij,m}(\theta) = \frac{1}{2\pi B}\sum_{t,s=1}^{B-m} X_{ti,m}X_{sj,m}e^{-\iota(t-s)\theta}. \quad (69)$$

Let

$$\mathcal{E}_{B,m}^{(2)} = \big(\rho^m + \sqrt{\frac{m}{B}}\big)\Phi_4^4. \quad (70)$$

Then Lemma 10 (i), (ii) and (iii) hold for $H_{ij}(\theta)$ and $H_{hl}(\theta')$ with $\mathcal{E}_{B,m}^{(1)}$ replaced by $\mathcal{E}_{B,m}^{(2)}$.

Proof of Lemma 11. Recall that

$$\widehat{G}_{ij,m}(\theta) = \frac{1}{2\pi B}\sum_{t,s=1}^{B} X_{ti,m}X_{sj,m}e^{-\iota(t-s)\theta}.$$

Then we have $2\pi B(\widehat{G}_{ij,m}(\theta) - \widehat{H}_{ij,m}(\theta)) = D_1 + D_2 - D_3$, where

$$D_1 = \sum_{t=B-m+1}^{B}\sum_{s=1}^{B} X_{ti,m}X_{sj,m}e^{-\iota(t-s)\theta},$$
$$D_2 = \sum_{s=B-m+1}^{B}\sum_{t=1}^{B} X_{ti,m}X_{sj,m}e^{-\iota(t-s)\theta},$$
$$D_3 = \sum_{t,s=B-m+1}^{B} X_{ti,m}X_{sj,m}e^{-\iota(t-s)\theta}.$$

Applying similar arguments of deriving equation (41) in Lemma 8 of [93] to the process $(X_{ti,m})_t$ and $(X_{tj,m})_t$, we have

$$\|\mathbb{E}_0 D_1\|_2 \lesssim \sqrt{Bm}\Phi_4^2,\ \|\mathbb{E}_0 D_2\|_2 \lesssim \sqrt{Bm}\Phi_4^2,\ \|\mathbb{E}_0 D_3\|_2 \lesssim m\Phi_4^2.$$

It follows that

$$\|\mathbb{E}_0\widehat{H}_{ij,m}(\theta) - \mathbb{E}_0\widehat{G}_{ij,m}(\theta)\|_2 \lesssim \sqrt{\frac{m}{B}}\Phi_4^2. \quad (71)$$

Applying Lemma 8 of [93] again, it follows that

$$\|\mathbb{E}_0\widehat{H}_{ij,m}(\theta)\|_2 \lesssim \Phi_4^2,\ \ \|\mathbb{E}_0\widehat{G}_{ij,m}(\theta)\|_2 \lesssim \Phi_4^2.$$

By the triangle inequality and Hölder's inequality, we have

$$\begin{aligned}
&\left|\mathbb{E}(\mathbb{E}_0\widehat{H}_{ij,m}(\theta)\mathbb{E}_0\widehat{H}_{hl,m}(\theta')) - \mathbb{E}(\mathbb{E}_0\widehat{G}_{ij,m}(\theta)\mathbb{E}_0\widehat{G}_{hl,m}(\theta'))\right| \\
&\le \|\mathbb{E}_0\widehat{H}_{ij,m}(\theta)\|_2\|\mathbb{E}_0\widehat{G}_{hl,m}(\theta') - \mathbb{E}_0\widehat{G}_{hl,m}(\theta')\|_2 \\
&\quad + \|\mathbb{E}_0\widehat{G}_{hl,m}(\theta')\|_2\|\mathbb{E}_0\widehat{H}_{ij,m}(\theta) - \mathbb{E}_0\widehat{G}_{ij,m}(\theta)\|_2 \\
&\lesssim \sqrt{\frac{m}{B}}\Phi_4^4.
\end{aligned} \tag{72}$$

Similarly, we can obtain

$$\begin{aligned}
&\left|\mathbb{E}(\mathbb{E}_0\widehat{H}_{ij,m}(\theta)\mathbb{E}_0\widehat{H}^*_{hl,m}(\theta')) - \mathbb{E}(\mathbb{E}_0\widehat{G}_{ij,m}(\theta)\mathbb{E}_0\widehat{G}^*_{hl,m}(\theta'))\right| \\
&\lesssim \sqrt{\frac{m}{B}}\Phi_4^4.
\end{aligned}$$

Therefore, Lemma 10 also holds for $H_{ij}(\theta)$ and $H_{hl}(\theta')$ with the additional error bound $\sqrt{m/B}\Phi_4^4$. $\square$

B. Probability Inequalities for High Dimensional Time Series

Let $n = T/B$. Define $\mathcal{B}_r = \{(r-1)B+1, \dots, rB\}$ and $\mathcal{D}_r = \{(r-1)B+1, \dots, (r-1)B+B-m\}$ for $r = 1, \dots, n$. For $m \ge 1$, let $X_{t,m} = \mathbb{E}(X_t|\varepsilon_{t-m}, \varepsilon_{t-m+1}, \dots, \varepsilon_t)$. Define

$$\begin{aligned}
\widehat{F}_m(\theta) &= \big(\widehat{F}_{ij,m}(\theta)\big)_{i,j=1}^p \\
&= \frac{1}{2\pi T}\sum_{r=1}^n \sum_{t,s\in\mathcal{B}_r} X_{t,m}X_{s,m}^\top e^{-\iota(t-s)\theta},
\end{aligned} \tag{73}$$

$$\begin{aligned}
\widehat{H}_{r,m}(\theta) &= \big(\widehat{H}_{ij,r,m}(\theta)\big)_{i,j=1}^p \\
&= \frac{1}{2\pi B}\sum_{t,s\in\mathcal{D}_r} X_{t,m}X_{s,m}^\top e^{-\iota(t-s)\theta},
\end{aligned} \tag{74}$$

$$\widehat{S}_m(\theta) = \big(\widehat{S}_{ij,m}(\theta)\big)_{i,j=1}^p = \frac{1}{n}\sum_{r=1}^n \widehat{H}_{r,m}(\theta). \tag{75}$$

Lemma 12. *Let Assumption 1 be satisfied for some $\alpha \ge 0$. Then there exists constants $C_\alpha > 0$ such that for any $x > 0$,*

$$\begin{aligned}
&\mathbb{P}(|\mathbb{E}_0(\widehat{F}_{ij}(\theta)) - \mathbb{E}_0(\widehat{F}_{ij,m}(\theta))| \ge x\big) \\
&\qquad \lesssim \exp\Big\{-C_\alpha\Big(\frac{\sqrt{T}x}{\sqrt{B}\rho^m\|X_\cdot\|_{\psi_\alpha}^2}\Big)^{\frac{1}{2\alpha+1}}\Big\}.
\end{aligned}$$

Proof of Lemma 12. We follow similar arguments of proving Theorem 1. In particular, we use $\tilde{R}_{ij}(\theta) = 2\pi T(\widehat{F}_{ij}(\theta) - \widehat{F}_{ij,m}(\theta))$ to replace $R_{ij}(\theta)$. By Lemma 9 in [93], we have

$$\begin{aligned}
\|\tilde{R}_{ij}(\theta) - \mathbb{E}\tilde{R}_{ij}(\theta)\|_{q/2} &\le C(q-2)\rho^m\sqrt{TB}\|X_{\cdot i}\|_q\|X_{\cdot j}\|_q \\
&\le C(q-2)q^{2\alpha}\rho^m\sqrt{TB}\|X_\cdot\|_{\psi_\alpha}^2.
\end{aligned} \tag{76}$$

With (76) in place of (33) and by similar arguments of Theorem 1, the result can be proved without extra difficulty. $\square$

Lemma 13. *Let Assumption 1 be satisfied for some $\alpha \ge 0$. Then there exists constants $C_\alpha > 0$ such that for any $x > 0$,*

$$\begin{aligned}
&\mathbb{P}(|\mathbb{E}_0(\widehat{F}_{ij,m}(\theta)) - \mathbb{E}_0(\widehat{S}_{ij,m}(\theta))| \ge x\big) \\
&\qquad \lesssim \exp\Big\{-C_\alpha\Big(\frac{\sqrt{T}x}{\sqrt{m}\|X_\cdot\|_{\psi_\alpha}^2}\Big)^{\frac{1}{2\alpha+1}}\Big\}.
\end{aligned} \tag{77}$$

Proof of Lemma 13. Write $2\pi T(\widehat{F}_{ij,m}(\theta) - \widehat{S}_{ij,m}(\theta)) = \mathrm{I} + \mathrm{II} + \mathrm{III}$, where

$$\begin{aligned}
\mathrm{I} &= \sum_{r=1}^n \sum_{t=rB-m+1}^{rB} \sum_{s=(r-1)B}^{rB} X_{ti,m}X_{sj,m}^\top e^{-\iota(t-s)\theta}, \\
\mathrm{II} &= \sum_{r=1}^n \sum_{s=rB-m+1}^{rB} \sum_{t=(r-1)B}^{rB} X_{ti,m}X_{sj,m}^\top e^{-\iota(t-s)\theta}, \\
\mathrm{III} &= \sum_{r=1}^n \sum_{t,s=rB-m+1}^{rB} X_{ti,m}X_{sj,m}^\top e^{-\iota(t-s)\theta}.
\end{aligned}$$

We can adopt similar arguments of proving (35) to obtain

$$\begin{aligned}
&\mathbb{P}\big(|\mathrm{I} - \mathbb{E}\mathrm{I}| \ge x\big) \lesssim \exp\Big\{-C_\alpha\Big(\frac{x}{\sqrt{Tm}\|X_\cdot\|_{\psi_\alpha}^2}\Big)^{\frac{1}{2\alpha+1}}\Big\}, \\
&\mathbb{P}\big(|\mathrm{II} - \mathbb{E}\mathrm{II}| \ge x\big) \lesssim \exp\Big\{-C_\alpha\Big(\frac{x}{\sqrt{Tm}\|X_\cdot\|_{\psi_\alpha}^2}\Big)^{\frac{1}{2\alpha+1}}\Big\}, \\
&\mathbb{P}\big(|\mathrm{III} - \mathbb{E}\mathrm{III}| \ge x\big) \lesssim \exp\Big\{-C_\alpha\Big(\frac{\sqrt{B}x}{\sqrt{T}m\|X_\cdot\|_{\psi_\alpha}^2}\Big)^{\frac{1}{2\alpha+1}}\Big\}.
\end{aligned}$$

Then (77) can be obtained by combining the three parts. $\square$

Denote $\theta_k = 2\pi k/B$ for $k \in \mathbb{Z}$. Define

$$\begin{aligned}
\boldsymbol{H}_m = (&\mathsf{Re}(\widehat{H}_m(\theta_0)), \mathsf{Im}(\widehat{H}_m(\theta_0)), \dots, \\
&\mathsf{Re}(\widehat{H}_m(\theta_{B/2})), \mathsf{Im}(\widehat{H}_m(\theta_{B/2}))),
\end{aligned}$$

where

$$\begin{aligned}
\widehat{H}_m(\theta) &= (\widehat{H}_{ij,m}(\theta))_{i,j=1}^p \\
&= \frac{1}{2\pi B}\Big(\sum_{t=1}^{B-m} X_{t,m}e^{-\iota t\theta}\Big)\Big(\sum_{t=1}^{B-m} X_{t,m}^\top e^{\iota t\theta}\Big).
\end{aligned}$$

Let $\tilde{p} = Bp^2 - (B/2-1)p$ and let $\tilde{\boldsymbol{W}} = (\tilde{W}_1, \tilde{W}_2, \dots, \tilde{W}_{\tilde{p}})^\top = \mathbf{vec}_0(\boldsymbol{H}_m)$ be the vectorization of $\boldsymbol{H}_m$ where the constantly zero entries, i.e., $\mathsf{Im}(\widehat{H}_{ii,m}(\theta))$ for $1 \le i \le p$ and all the entries in $\mathsf{Im}(\widehat{H}_m(0))$ and $\mathsf{Im}(\widehat{H}_m(\pi))$, are excluded. Denote $\boldsymbol{\Sigma}_m = \mathrm{Cov}(\mathbf{vec}_0(\boldsymbol{H}_m))$ and $\boldsymbol{Z}_m \sim N(0, \boldsymbol{\Sigma}_m)$.

Lemma 14. *Suppose Assumption 1 is satisfied for some for some $\alpha \ge 0$. Assume that there exists some constant $b_0 > 0$ such that $\min_{1\le j\le\tilde{p}}\mathrm{Var}(\tilde{W}_j) \ge b_0$. Let $\boldsymbol{W}_1, \dots, \boldsymbol{W}_n$ be i.i.d. copies of $\tilde{\boldsymbol{W}}$ and $\boldsymbol{Z}_m \sim N(0, \mathrm{Cov}(\mathbf{vec}_0(\boldsymbol{H}_m)))$. Then for $\bar{\boldsymbol{W}}_n = n^{-1}\sum_{i=1}^n \boldsymbol{W}_i$,*

$$\begin{aligned}
&\sup_x |\mathbb{P}(\sqrt{n}|\bar{\boldsymbol{W}}_n - \mathbb{E}\bar{\boldsymbol{W}}_n|_\infty \le x) - \mathbb{P}(|\boldsymbol{Z}_m|_\infty \le x)| \\
&\qquad \lesssim \Big(\frac{\|X_\cdot\|_{\psi_\alpha}^{12}\log^{\kappa_1}(\tilde{p}\vee n)}{n}\Big)^{\frac{1}{6}},
\end{aligned} \tag{78}$$

where $\kappa_1 = \max\{7, 2\alpha+6\}$.

Proof of Lemma 14. Without loss of generality, we may assume that

$$\frac{\|X_\cdot\|_{\psi_\alpha}^{12}\log^{\kappa_1}(\tilde{p}\vee n)}{n} \le C, \tag{79}$$

where C is a sufficiently small constant. Otherwise, (78) trivially holds by taking the constant in $\lesssim$ large enough. We shall apply Theorem 2.1 in [92] to prove this Lemma. Define $L = \max_{1\leq j\leq \tilde{p}} \mathbb{E}|\mathbb{E}_0 \tilde{W}_j|^3$. For $\phi \geq 1$, define

$$M_{n,W}(\phi) = \mathbb{E}\big[|\mathbb{E}_0 \tilde{\boldsymbol{W}}|_\infty^3 \mathbf{1}\{|\mathbb{E}_0 \tilde{\boldsymbol{W}}|_\infty > \sqrt{n}/(4\phi \log \tilde{p})\}\big],$$
$$M_{n,Z}(\phi) = \mathbb{E}\big[|\boldsymbol{Z}_m|_\infty^3 \mathbf{1}\{|\boldsymbol{Z}_m|_\infty > \sqrt{n}/(4\phi \log \tilde{p})\}\big].$$

Denote $M_n(\phi) = M_{n,W}(\phi) + M_{n,Z}(\phi)$. By Theorem 2.1 in [92], there exist constants $K_1, K_2 > 0$ depending only on b_0 such that for every constant $\bar{L} \geq L$,

$$\sup_x |\mathbb{P}(\sqrt{n}|\bar{\boldsymbol{W}}_n - \mathbb{E}\bar{\boldsymbol{W}}_n|_\infty \leq x) - \mathbb{P}(|\boldsymbol{Z}_m|_\infty \leq x)|$$
$$\leq K_1\Big[\Big(\frac{\bar{L}^2 \log^7 \tilde{p}}{n}\Big)^{1/6} + \frac{M_n(\phi_n)}{\bar{L}}\Big] \tag{80}$$

with

$$\phi_n = K_2\Big(\frac{\bar{L}^2 \log^4 \tilde{p}}{n}\Big)^{-1/6}.$$

Applying Equation (41) in Lemma 8 of [93] for the third moment, we have $2\pi\|\mathbb{E}_0 \tilde{W}_j\|_3 \leq 4\sqrt{10}\Phi_6^2$. Since $\Phi_6 \leq 6^\alpha \|X.\|_{\psi_\alpha}$, we can denote

$$L = \max_{1\leq j\leq p} \mathbb{E}|\mathbb{E}_0 \tilde{W}_j|^3 \leq K_\alpha^3 \|X.\|_{\psi_\alpha}^6.$$

Choose

$$\bar{L} = K_\alpha^3 \log^{c_0}(\tilde{p} \vee n)\|X.\|_{\psi_\alpha}^6, \tag{81}$$

where $c_0 \geq 0$ and is to be determined later. For $C \leq (K_2/K_\alpha)^6$, (79) implies $\phi_n := K_2(n^{-1}\bar{L}^2 \log^4 \tilde{p})^{-1/6} \geq 1$. By similar arguments of proving (35), with $C_0 = 1/(1+2\alpha) \in (0, 1]$, we have

$$\mathbb{P}(|\mathbb{E}_0 \tilde{\boldsymbol{W}}|_\infty \geq x) \lesssim \tilde{p} \exp\big(-\frac{x^{C_0}}{C_\alpha \|X.\|_{\psi_\alpha}^{2C_0}}\big). \tag{82}$$

For $\tau = \sqrt{n}/(4\phi_n \log \tilde{p})$, we have

$$M_{n,W}(\phi_n) = \tau^3 \mathbb{P}(|\mathbb{E}_0 \tilde{\boldsymbol{W}}|_\infty > \tau) + 3\int_\tau^\infty x^2 \mathbb{P}(|\mathbb{E}_0 \tilde{\boldsymbol{W}}|_\infty > x)dx. \tag{83}$$

Using integration by parts, it can be obtained that

$$\int_\tau^\infty x^2 \exp\big(-\frac{x^{C_0}}{C_\alpha \|X.\|_{\psi_\alpha}^{2C_0}}\big)dx$$
$$= \frac{C_\alpha \|X.\|_{\psi_\alpha}^{2C_0} \tau^{3-C_0}}{C_0} \exp\big(-\frac{\tau^{C_0}}{C_\alpha \|X.\|_{\psi_\alpha}^{2C_0}}\big)$$
$$+ \frac{C_\alpha (3 - C_0)\|X.\|_{\psi_\alpha}^{2C_0}}{C_0} \int_\tau^\infty x^{2-C_0} \exp\big(-\frac{x^{C_0}}{C_\alpha \|X.\|_{\psi_\alpha}^{2C_0}}\big)dx. \tag{84}$$

Note that $\tau = (4K_2)^{-1}(n\bar{L}/\log \tilde{p})^{1/3}$. Let $c_0 = \max\{0, C_0{}^{-1} - 2\}$ and let C in (79) be small enough such that

$$\frac{\tau^{C_0}}{C_\alpha \|X.\|_{\psi_\alpha}^{2C_0} \log(\tilde{p} \vee n)} \geq \max\{3, (6 - 2C_0)/C_0\}.$$

By (82), (83) and (84), elementary calculation implies $M_{n,W}(\phi_n) \lesssim \|X.\|_{\psi_\alpha}^6/n$. Since $\boldsymbol{Z}_m$ is a Gaussian vector and Lemma 8 in [93] suggests $\mathrm{Var}(\tilde{W}_j) \leq [(2\pi)^{-1}4^{2\alpha+1}\sqrt{3}\|X.\|_{\psi_\alpha}^2]^2$, we have

$$P(|\boldsymbol{Z}_m|_\infty \geq x)$$
$$\leq 2\sum_{j=1}^{\tilde{p}} \int_x^\infty \frac{1}{\sqrt{2\pi \mathrm{Var}(\tilde{W}_j)}} \exp\big(-\frac{y^2}{2\mathrm{Var}(\tilde{W}_j)}\big)dy$$
$$\lesssim \frac{\tilde{p}\|X.\|_{\psi_\alpha}^2}{x} \exp\big(-\frac{x^2}{C'_\alpha \|X.\|_{\psi_\alpha}^2}\big).$$

A similar argument of dealing with $M_{n,W}(\phi_n)$ also implies $M_{n,Z}(\phi_n) \lesssim \|X.\|_{\psi_\alpha}^6/n$. By (80) and (81), (78) follows in view of the fact that $M_n(\phi_n)/\bar{L}$ is negligible compared to the first term. □

Lemma 15 (Nazarov's inequality, [98])**.** *Let $Z = (Z_1, \ldots, Z_p)^\top$ be a centered Gaussian random vector in $\mathbb{R}^p$ such that $\mathbb{E}(Z_j^2) \geq c$ for all $j = 1, \ldots, p$ and some constant $c > 0$. Then for every $z \in \mathbb{R}^p$ and $a > 0$,*

$$\mathbb{P}(Z \leq z + a) - \mathbb{P}(Z \leq z) \leq Ca\sqrt{\log p},$$

where C is a positive constant depending on c only.

Acknowledgement

The authors thank the Associate Editor and three anonymous referees for their many constructive comments.

References

[1] E. J. Hannan. *Multiple time series*. John Wiley & Sons, 1970.

[2] Joachim Gross, Jan Kujala, Matti Hämäläinen, Lars Timmermann, Alfons Schnitzler, and Riitta Salmelin. Dynamic imaging of coherent sources: studying neural interactions in the human brain. *Proceedings of the National Academy of Sciences*, 98(2):694–699, 2001.

[3] A.M. Bastos and J.M. Schoffelen. A tutorial review of functional connectivity analysis methods and their interpretational pitfalls. *Frontiers in systems neuroscience*, 9:175, 2016.

[4] Mingzhou Ding, Steven L Bressler, Weiming Yang, and Hualou Liang. Short-window spectral analysis of cortical event-related potentials by adaptive multivariate autoregressive modeling: data preprocessing, model validation, and variability assessment. *Biological cybernetics*, 83:35–45, 2000.

[5] B.M. Båth. *Spectral Analysis in Geophysics*. ISSN. Elsevier Science, 2012.

[6] Jun He. Coherence and cross-spectral density matrix analysis of random wind and wave in deep water. *Ocean Engineering*, 197:106930, 2020.

[7] Finn-Idar Grøtta Giske, Bernt Johan Leira, and Ole Øiseth. Efficient computation of cross-spectral densities in the stochastic modelling of waves and wave loads. *Applied Ocean Research*, 62:70–88, 2017.

[8] David A Warde and Sebastián M Torres. The autocorrelation spectral density for doppler-weather-radar signal analysis. *IEEE transactions on geoscience and remote sensing*, 52(1):508–518, 2013.

[9] Anthony Stathopoulos and Matthew G Karlaftis. Spectral and cross-spectral analysis of urban traffic flows. In *ITSC 2001. 2001 IEEE Intelligent Transportation Systems. Proceedings (Cat. No. 01TH8585)*, pages 820–825. IEEE, 2001.

[10] Michael Ghil, MR Allen, MD Dettinger, K Ide, D Kondrashov, ME Mann, Andrew W Robertson, A Saunders, Y Tian, Fa Varadi, et al. Advanced spectral methods for climatic time series. *Reviews of geophysics*, 40(1):3–1, 2002.

[11] M. Rosenblatt. The multidimensional prediction problem. *Proceedings of the National Academy of Sciences*, 43(11):989–992, 1957.

[12] E.J. Hannan. Regression for time series. *Time series analysis*, pages 17–37, 1963.

[13] P.M. Robinson. Non-linear regression for multiple time-series. *Journal of Applied Probability*, 9(4):758–768, 1972.

[14] P.M. Robinson. Automatic frequency domain inference on semiparametric and nonparametric models. *Econometrica: Journal of the Econometric Society*, pages 1329–1363, 1991.

[15] M. Forni, M. Hallin, M. Lippi, and L. Reichlin. The generalized dynamic-factor model: Identification and estimation. *Review of Economics and statistics*, 82(4):540–554, 2000.

[16] M. Forni, M. Hallin, M. Lippi, and P. Zaffaroni. Dynamic factor models with infinite-dimensional factor space: Asymptotic analysis. *Journal of Econometrics*, 199(1):74–92, 2017.

[17] T.S. McElroy and D.N. Politis. Optimal linear interpolation of multiple missing values. *Statistical Inference for Stochastic Processes*, 25:1–13, 2022.

[18] T.W. Anderson. *The Statistical Analysis of Time Series*. Wiley, 1971.

[19] M.B. Priestley. *Spectral Analysis and Time Series*. Academic Press, 1981.

[20] M. Rosenblatt. *Stationary Sequences and Random Fields*. Springer, 1985.

[21] P.J. Brockwell and R.A. Davis. *Time series theory and methods*. Springer, 1991.

[22] D. R. Brillinger. *Time series: data analysis and theory*. SIAM, 2001.

[23] X. Shao and W.B. Wu. Asymptotic spectral theory for nonlinear time series. *The Annals of Statistics*, 35(4):1773–1801, 2007.

[24] W. Liu and W.B. Wu. Asymptotics of spectral density estimates. *Econometric Theory*, 26(4):1218–1245, 2010.

[25] P. Welch. The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on audio and electroacoustics*, 15(2):70–73, 1967.

[26] H. R. Neave. A comparison of lag window generators. *Journal of the American Statistical Association*, 67(337):152–158, 1972.

[27] N. Wiener. *Nonlinear Problems in Random Theory*. Wiley, New York, 1958.

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

[29] M.B. Priestley. *Nonlinear and Nonstationary Time Series Analysis*. Academic Press, London, 1988.

[30] H. Tong. *Non-linear Time Series: A Dynamical System Approach*. Oxford University Press, 1990.

[31] W.B. Wu. Nonlinear system theory: another look at dependence. *Proceedings of the National Academy of Sciences of the United States of America*, 102(40):pp. 14150–14154, 2005.

[32] Wei Biao Wu and Paolo Zaffaroni. Asymptotic theory for spectral density estimates of general multivariate time series. *Econometric Theory*, 34(1):1–22, 2018.

[33] X. Chen, M. Xu, and W. B. Wu. Covariance and precision matrix estimation for high-dimensional time series. *The Annals of Statistics*, 41(6):2994–3021, 12 2013.

[34] W.B. Wu and Y.N. Wu. Performance bounds for parameter estimates of high-dimensional linear models with correlated errors. *Electronic Journal of Statistics*, 10(1):352–379, 2016.

[35] D. Zhang and W. B. Wu. Gaussian approximation for high dimensional time series. *The Annals of Statistics*, 45(5):1895–1919, 2017.

[36] Y. Sun, Y. Li, A. Kuceyeski, and S. Basu. Large spectral density matrix estimation by thresholding. *arXiv preprint arXiv:1812.00532*, 2018.

[37] M. Fiecas, C. Leng, W. Liu, and Y. Yu. Spectral analysis of high-dimensional time series. *Electronic Journal of Statistics*, 13(2):4079–4101, 2019.

[38] D. Zhang and W. B. Wu. Convergence of covariance and spectral density estimates for high-dimensional locally stationary processes. *The Annals of Statistics*, 49(1):233–254, 2021.

[39] A. Rosuel, P. Loubaton, and P. Vallet. On the asymptotic distribution of the maximum sample spectral coherence of gaussian time series in the high dimensional regime. *arXiv preprint arXiv:2107.02891*, 2021.

[40] Hans R Künsch. The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, pages 1217–1241, 1989.

[41] R.Y. Liu and K. Singh. Moving block jackknife and bootstrap capture weak dependence. *Exploring the limits of bootstrap, R. Lepage & L. Billard eds*, 225:248, 1992.

[42] D.N. Politis and J.P. Romano. Large sample confidence regions based on subsamples under minimal assumptions. *The Annals of Statistics*, pages 2031–2050, 1994.

[43] P. Hall and B. Jing. On sample reuse methods for dependent data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(4):727–737, 1996.

[44] V. Chernozhukov, D. Chetverikov, and K. Kato. Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *The Annals of Statistics*, 41(6):2786–2819, 2013.

[45] Xianyang Zhang and Guang Cheng. Bootstrapping high dimensional time series. *arXiv preprint arXiv:1406.1037*, 2014.

[46] M. Bhattacharjee and A. Bose. Estimation of autocovariance matrices for infinite dimensional vector linear

process. *Journal of Time Series Analysis*, 35(3):262–281, 2014.

[47] M. Bhattacharjee and A. Bose. Large sample behaviour of high dimensional autocovariance matrices. *The Annals of Statistics*, 44(2):598–628, 2016.

[48] B. Jin, C. Wang, ZD Bai, K K Nair, and M. Harding. Limiting spectral distribution of a symmetrized auto-cross covariance matrix. *The Annals of Applied Probability*, 24(3):1199–1225, 2014.

[49] H. Liu, A. Aue, and D. Paul. On the marčenko–pastur law for linear time series. *The Annals of Statistics*, 43(2):675–712, 2015.

[50] X. Chen, M. Xu, and W.B. Wu. Regularized estimation of linear functionals of precision matrices for high-dimensional time series. *IEEE Transactions on Signal Processing*, 64(24):6459–6470, 2016.

[51] Christopher A Sims. Macroeconomics and reality. *Econometrica*, 48(1):1–48, 1980.

[52] James H Stock and Mark W Watson. Vector autoregressions. *Journal of Economic perspectives*, 15(4):101–115, 2001.

[53] R.S. Tsay. *Analysis of Financial Time Series*, volume 543. John Wiley & Sons, 2005.

[54] J. Fan, J. Lv, and L. Qi. Sparse high dimensional models in economics. *Annual review of economics*, 3:291–317, 2011.

[55] Gregory C Reinsel. *Elements of multivariate time series analysis*. Springer Science & Business Media, 2003.

[56] Helmut Lütkepohl. *New introduction to multiple time series analysis*. Springer Science & Business Media, 2005.

[57] Peter J Brockwell and Richard A Davis. *Time series: theory and methods*. Springer science & business media, 2009.

[58] E. Rio. Moment inequalities for sums of dependent random variables under projective conditions. *Journal of Theoretical Probability*, 22(1):146–163, 2009.

[59] W. B. Wu and X. Shao. Limit theorems for iterated random functions. *Journal of Applied Probability*, pages 425–436, 2004.

[60] A. Zygmund. *Trigonometric Series. Vol I, II*. Third Ed. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 2002.

[61] M. Rosenblatt. Asymptotic normality, strong mixing and spectral density estimates. *The Annals of Probability*, 12(4):1167–1180, 1984.

[62] R. Y. Bentkus and R. Rudzkis. On the distributions of some statistical estimates of spectral density. *Teoriya Veroyatnostei i ee Primeneniya*, 27(4):739–756, 1982.

[63] M. B. Woodroofe and J. W. Van Ness. The maximum deviation of sample spectral densities. *The Annals of Mathematical Statistics*, pages 1558–1569, 1967.

[64] R. Rudzkis. On the distribution of the maximum deviation of the gaussian stationary time series spectral density estimate. *Lithuanian Mathematical Journal*, pages 118–130, 1985.

[65] S. Portnoy. On the central limit theorem in R^p when $p \to \infty$. *Probability Theory and Related Fields*, 73(4):571–583, 1986.

[66] X. Zhang and G. Cheng. Gaussian approximation for high dimensional vector under physical dependence. *Bernoulli*, 24:2640–2675, 2018.

[67] G. Nolte, O. Bai, L. Wheaton, Z. Mari, S. Vorbach, and M. Hallett. Identifying true brain interaction from eeg data using the imaginary part of coherency. *Clinical neurophysiology*, 115(10):2292–2307, 2004.

[68] L. Marzetti, G. Nolte, M. G. Perrucci, G. L. Romani, and C. Del Gratta. The use of standardized infinity reference in eeg coherency studies. *Neuroimage*, 36(1):48–63, 2007.

[69] L. García Domínguez, J. Stieben, J. L. Pérez Velázquez, and S. Shanker. The imaginary part of coherency in autism: differences in cortical functional connectivity in preschool children. *PLoS One*, 8(10):e75941, 2013.

[70] F.U. Hohlefeld, C. Huchzermeyer, J. Huebl, G.H. Schneider, G. Nolte, C. Brücke, T. Schönecker, A.A. Kühn, G. Curio, and V.V. Nikulin. Functional and effective connectivity in subthalamic local field potential recordings of patients with parkinson's disease. *Neuroscience*, 250:320–332, 2013.

[71] G. E. P. Box and David A. Pierce. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *J. Amer. Statist. Assoc.*, 65:1509–1526, 1970.

[72] G. M. Ljung and G. E. P. Box. On a measure of lack of fit in time series models. *Biometrika*, 65(2):297–303, 1978.

[73] Ratnam V. Chitturi. Distribution of residual autocorrelations in multiple autoregressive schemes. *J. Amer. Statist. Assoc.*, 69:928–934, 1974.

[74] J. R. M. Hosking. The multivariate portmanteau statistic. *J. Amer. Statist. Assoc.*, 75(371):602–608, 1980.

[75] Z. Li, C. Lam, Ji. Yao, and Q. Yao. On testing for high-dimensional white noise. *The Annals of Statistics*, 47(6):3382–3412, 2019.

[76] J. Chang, Q. Yao, and W. Zhou. Testing for high-dimensional white noise using maximum cross-correlations. *Biometrika*, 104(1):111–127, 2017.

[77] R.S. Tsay. Testing serial correlations in high-dimensional time series via extreme value theory. *Journal of Econometrics*, 216(1):106–117, 2020.

[78] M.S. Bartlett. An introduction to stochastic processes. *Cambridge University Press, London*, 1955.

[79] U. Grenander and M. Rosenblatt. *Statistical analysis of stationary time series*. Chelsea, New York, NY, 1957.

[80] J. Durbin. Tests of serial independence based on the cumulated periodogram. *Bulletin of the International Statistical Institute*, 42:1039–1048, 1967.

[81] S.N. Durlauf. Spectral based testing of the martingale hypothesis. *Journal of Econometrics*, 50(3):355–376, 1991.

[82] Danna Zhang. Robust estimation of the mean and covariance matrix for high dimensional time series. *Statistica Sinica*, 31(2):797–820, 2021.

[83] L.D. Haugh. Checking the independence of two covariance-stationary time series: a univariate residual

cross-correlation approach. *Journal of the American Statistical Association*, 71(354):378–385, 1976.
[84] P.D. Koch and S. Yang. A method for testing the independence of two time series that accounts for a potential pattern in the cross-correlation function. *Journal of the American Statistical Association*, 81(394):533–544, 1986.
[85] Y. Hong. Testing for independence between two covariance stationary time series. *Biometrika*, 83(3):615–625, 1996.
[86] Eugene Duff, Leigh Johnston, Jinhu Xiong, Peter Fox, Iven Mareels, and Gary Egan. The power of spectral density analysis for mapping endogenous bold signal fluctuations. *Human brain mapping*, 29:778–90, 07 2008.
[87] S.L. Simpson, F.D. Bowman, and P.J. Laurienti. Analyzing complex functional brain networks: fusing statistics and network science to understand the brain. *Statistics surveys*, 7:1, 2013.
[88] M.A. Lindquist, Y. Xu, M.B. Nebel, and B.S. Caffo. Evaluating dynamic bivariate correlations in resting-state fmri: A comparison study and a new approach. *Neuroimage*, 101:531–546, 2014.
[89] P.A. Robinson, J.A. Henderson, N.C. Gabay, K.M. Aquino, T. Babaie-Janvier, and X. Gao. Determination of dynamic brain connectivity via spectral analysis. *Frontiers in Human Neuroscience*, page 360, 2021.
[90] M. Sherman and E. Carlstein. Replicate histograms. *Journal of the American Statistical Association*, 91(434):566–576, 1996.
[91] D.N. Politis, J.P. Romano, and M. Wolf. *Subsampling*. Springer, 1999.
[92] V. Chernozhukov, D. Chetverikov, and K. Kato. Central limit theorems and bootstrap in high dimensions. *The Annals of Probability*, 45(4):2309–2352, 2017.
[93] H. Xiao and W.B. Wu. Covariance matrix estimation for stationary time series. *The Annals of Statistics*, 40(1):466–493, 2012.
[94] ASOS ASOS. Automated surface observing system: Asos user's guide.[washington, dc]: Us dept. of commerce. *National Oceanic and Atmospheric Administration: Federal Aviation Administration: US Navy: US Dept. of the Air Force, URL https://www. weather. gov/media/asos/aum-toc. pdf*, 1998.
[95] Dennis P Todey, DE Herzmann, and ES Takle. The iowa environmental mesonetcombining observing systems into a single network. In *Sixth Symposium on Integrated Observing Systems*. Iowa State University Ames, IA, USA, 2002.
[96] Martin Stendel, Jennifer Francis, Rachel White, Paul D Williams, and Tim Woollings. The jet stream and climate change. In *Climate change*, pages 327–357. Elsevier, 2021.
[97] H. Xiao and W. B. Wu. Asymptotic inference of autocovariances of stationary processes. *arXiv preprint arXiv:1105.3423*, 2011.
[98] F. Nazarov. On the maximal perimeter of a convex set in r^n with respect to gaussian measure, geometric aspects of func. *Anal*, 1807:169–187, 2003.

Chi Zhang received the B.S. degree in Mathematics and Applied Mathematics from Shanghai Jiao Tong University, Shanghai, China, and the M.S. degree in Mathematics from the University of Wisconsin–Madison, WI, USA. He is currently pursuing the Ph.D. degree in Mathematics with a specialization in Statistics with the University of California San Diego, CA, USA. His research interests include high-dimensional time series and data analytics.

Danna Zhang received the Ph.D. degree in statistics from the University of Chicago, Chicago, IL. She is currently an Assistant Professor in the Department of Mathematics at University of California, San Diego, CA. Her research interests include high-dimensional time series, high-dimensional significance testing, nonlinear/nonstationary time series, time-varying network and high-order statistics.