SPECTRAL INFERENCE UNDER COMPLEX TEMPORAL DYNAMICS

JUN YANG AND ZHOU ZHOU

Abstract. We develop unified theory and methodology for the inference of evolutionary Fourier power spectra for a general class of locally stationary and possibly nonlinear processes. In particular, simultaneous confidence regions (SCR) with asymptotically correct coverage rates are constructed for the evolutionary spectral densities on a nearly optimally dense grid of the joint time-frequency domain. A simulation based bootstrap method is proposed to implement the SCR. The SCR enables researchers and practitioners to visually evaluate the magnitude and pattern of the evolutionary power spectra with asymptotically accurate statistical guarantee. The SCR also serves as a unified tool for a wide range of statistical inference problems in time-frequency analysis ranging from tests for white noise, stationarity and time-frequency separability to the validation for non-stationary linear models.

Contents

Department of Statistical Sciences, University of Toronto, Canada
E-mail addresses: jun@utstat.toronto.edu, zhou@utstat.toronto.edu.

1. Introduction

It is well known that the frequency content of many real-world stochastic processes evolves over time. Motivated by the limitations of the traditional spectral methods in analyzing non-stationary signals, time-frequency analysis has become one of the major research areas in applied mathematics and signal processing [Coh95; Grö01; Dau90]. Based on various models or representations of the non-stationary signal and its time-varying spectra, time-frequency analysis aims at depicting temporal and spectral information simultaneously and jointly. Roughly speaking, there are three major classes of algorithms in time-frequency analysis: linear algorithms such as short time Fourier transforms (STFT) and wavelet transforms [All77; Mey92; Dau92]; bilinear time-frequency representations such as the Wigner–Ville distribution and more generally the Cohen's class of bilinear time–frequency distributions [Coh95; HBB92] and nonlinear algorithms such as the empirical mode decomposition method [HSLW+98] and the synchrosqueezing transform [DLW11]. Though there exists a vast literature on defining and estimating the time-varying frequency content, statistical inference such as confidence region construction and hypothesis testing has been paid little attention to in time-frequency analysis.

It is clear that the focus as well as the goals of time-frequency analysis are well contained in those of time series analysis, particularly non-stationary time series analysis. Unfortunately it seems that the non-stationary spectral domain theory and methodology in the time series literature have been developed largely independently from time-frequency analysis. One major effort in non-stationary time series analysis lies in forming general classes of non-stationary time series models through their evolutionary spectral representation. Among others, Priestley [Pri65] proposed the notion of evolutionary spectra in a seminar paper. In another seminal work, Dahlhaus [Dah97] defined a general and theoretically tractable class of locally stationary time series models based on their time-varying spectral representation. Nason, Sachs, and Kroisandt [NSK00] studied a class of locally stationary time series from an evolutionary wavelet spectrum perspective and investigated the estimation of the latter spectrum. A second line of research in the non-stationary spectral domain literature involves adaptive estimation of the evolutionary spectra. See for instance [Ada98] for a binary segmentation based method, [ORSM01] for an automatic estimation procedure based on the smooth localized complex exponential (SLEX) transform and [FN06] for a Haar–Fisz technique for the estimation of the evolutionary wavelet spectra. On the statistical inference side, there exists a small number of papers utilizing the notion of evolutionary spectra to test some properties, especially second order stationarity, of a time series. See for instance [Pap10; DPV11; DSR11; JSR15] for tests of stationarity based on properties of the Fourier periodogram or spectral density. See also [Nas13] for a test of stationarity based on the evolutionary wavelet spectra. On the other hand, however, to date there have been no results on the joint and simultaneous inference of the evolutionary spectrum itself for general classes of non-stationary and possibly nonlinear time series to the best of our knowledge.

The purpose of the paper is to develop unified theory and methodology for the joint and simultaneous inference of the evolutionary spectral densities for a general class of locally stationary and possibly nonlinear processes. From a time-frequency analysis perspective, the purpose of the paper is to provide a unified and asymptotically correct method for the simultaneous statistical inference of the STFT-based evolutionary power spectra, one of the most classic and fundamental algorithms in time-frequency analysis. Let $\{X_i^{(N)}\}_{i=1}^N$ be the observed time series or signal. One major contribution of the paper is that we establish a maximum deviation theory for the STFT-based spectral density estimates over a nearly optimally dense grid $\mathcal{G}_N$ in the joint time-frequency domain. Here the optimality of the grid refers to the best balance between computational burden and (asymptotic) correctness in depicting the overall time-frequency stochastic variation of the estimates. We refer the readers to Section 5.1 for a detailed definition and discussion of the optimality. The theory is established for a very general class of possibly nonlinear locally stationary processes which admit a time-varying physical representation in the sense of [ZW09] and serves as a foundation for the joint and simultaneous time-frequency inference of evolutionary spectral densities. Specifically, we are able to prove that the spectral density estimates on $\mathcal{G}_N$ are asymptotically independent quadratic forms of $\{X_i^{(N)}\}_{i=1}^N$. And consequently the maximum deviation of the spectral density estimates on $\mathcal{G}_N$ behaves asymptotically like a Gumbel law. The key technique used in the proofs is a joint time-frequency Gaussian approximation to a class of diverging dimensional quadratic forms of non-stationary time series, which may have wider applicability in evolutionary power spectrum analysis.

A second main contribution of the paper is that we propose a simulation based bootstrap method to implement simultaneous statistical inferences to a wide range of problems in time-frequency analysis. The motivation of the bootstrap is to alleviate the slow convergence of the maximum deviation to its Gumbel limit. The bootstrap simply generates independent normal pseudo samples of length N and approximate the distribution of the target maximum deviation with that of the normalized empirical maximum deviations of the spectral density estimates from the pseudo samples. The similar idea was used in, for example [WZ07; ZW10], for some different problems. The bootstrap is proved to be asymptotically correct and performs reasonably well in the simulations. One important application of the bootstrap is to construct simultaneous confidence regions (SCR) for the evolutionary spectral density, which enables researchers and practitioners to visually evaluate the magnitude and pattern of the evolutionary power spectra with asymptotically accurate statistical guarantee. In particular, the SCR helps one to visually identify which variations in time and/or frequency are genuine and which variations are likely to be produced by random fluctuations. See Section 7.4 for two detailed applications in earthquake and explosion signal processing and finance. On the other hand, the SCR can be applied to a wide range of tests on the structure of the evolutionary spectra or the time series itself. Observe that typically under some specific structural assumptions, the time-varying spectra can be estimated with a faster convergence rate than those estimated by STFT without any prior information. Therefore a generic testing procedure is to estimate the evolutionary spectra under the null hypothesis and check whether the latter estimated spectra can be fully embedded into the SCR. This is a very general procedure and it is asymptotically correct as long as the evolutionary spectra estimated under the null converges faster than the SCR. And the test achieves asymptotic power 1 for local alternatives whose evolutionary spectra deviate from the null with a rate larger than the order of the width of the SCR. Specific examples include tests for non-stationary white noise, weak stationarity and time-frequency separability as well as model validation for locally stationary ARMA models and so on. See Section 5.2 for a detailed discussion and Section 7.4 for detailed implementations of the tests in real data.

Finally, we would like to mention that, under the stationarity assumption, the inference of the spectral density is a classic topic in time series analysis. There is a huge literature on the topic and we will only list a very small number of representative works. Early works on this topic include [Par57; WN67; Bri69; And71; Ros84] among others where asymptotic properties of the spectral density estimates were established under various linearity, strong mixing and joint cumulant conditions. For recent developments see [LW10; PP12; WZ18] among others.

The rest of the paper is organized as follows. We first formulate the problem in Section 2. In Section 3, we study STFT and show that STFTs are asymptotically independent normals under very mild conditions. In Section 4, we study the asymptotic properties of STFT-based spectral density estimates, including consistency and asymptotic normality. In Section 5, we establish a maximum deviation theory for the STFT-based spectral density estimates over a nearly optimally dense grid in the joint time-frequency domain. In Section 6, we discuss tuning parameter selection and propose a simulation-based bootstrap method to implement the simultaneous statistical inference. Simulations and real data analysis are given in Section 7. Proofs of the main results are delayed to Section 8 and many details of the proofs have been put in Appendix A.

2. Problem Formulation

We first define locally stationary time series and its instantaneous covariance and spectral density.

Definition 2.1. (Locally stationary time series [ZW09]) We say $\{X_i^{(N)}\}_{i=1}^N$ is locally stationary time series if there exist nonlinear filter G and

$$X_i^{(N)} = G\left(i/N, \mathcal{F}_i\right), \quad i = 0, \ldots, N, \tag{1}$$

where $\mathcal{F}_i = (\ldots, \epsilon_0, \ldots, \epsilon_{i-1}, \epsilon_i)$ and ϵ_i's are i.i.d. random variables.

Intuitively, if $G(u, \cdot)$ is a smooth function of u where $u \in [0, 1]$, then a time series is locally stationary in the sense that any short segment of the time series is approximately stationary. Rigorous definition of the smoothness of $G(u, \cdot)$ is defined using the notion of stochastic Lipschitz continuity (SLC) which is defined in Definition 2.2 below. Throughout the article, we assume $G(u, \mathcal{F}_i)$ is SLC and the time series $\{X_i^{(N)}\}_{i=1}^N$ is centered, i.e. $\mathbb{E}[X_i^{(N)}] = 0$.

Definition 2.2. (Stochastic Lipschitz continuity) There exists $C, q > 0$, such that for all i and $u, s \in (0, 1)$, the time series satisfies SLC(q) if we have

$$\|G(u, \mathcal{F}_i) - G(s, \mathcal{F}_i)\|_q \leq C|u - s|. \tag{2}$$

Example 2.3. (Locally stationary linear time series) Let ϵ_i be i.i.d. random variables and

$$G(u, \mathcal{F}_i) = \sum_{j=0}^{\infty} a_j(u)\epsilon_{i-j}, \tag{3}$$

where $a_j(u) \in \mathcal{C}^1[0,1]$ for $j = 0, 1, \ldots$. This model was considered in [Dah97]. Verifying the SLC assumption is discussed in [ZW09, Propositions 2 and 3]. $\triangleleft$

Example 2.4. (Time varying threshold AR models) Let $\epsilon_i \in \mathcal{L}^q, q > 0$ be i.i.d. random variables with distribution function F_ϵ and density f_ϵ. Consider the model

$$G(u, \mathcal{F}_i) = a(u)[G(u, \mathcal{F}_{i-1})]^+ + b(u)[-G(u, \mathcal{F}_{i-1})]^+ + \epsilon_i, \quad 0 \le u \le 1, \tag{4}$$

where $a(\cdot), b(\cdot) \in \mathcal{C}^1[0,1]$. Then if $\sup_u[|a(u)| + |b(u)|] < 1$, the SLC($q$) assumption holds. See also [ZW09, Section 4] for more discussions on checking SLC assumption for locally stationary nonlinear time series. $\triangleleft$

For simplicity, we will use X_i to denote $X_i^{(N)}$ in this paper. Without loss of generality, we assume $X_i = 0$ for any $i > N$. We adopt the physical dependence measure [ZW09] to describe the dependence structure of the time series.

Definition 2.5. (Physical dependence measure) Let $\{\epsilon_i'\}$ be an i.i.d. copy of $\{\epsilon_i\}$. Consider the locally stationary time series $\{X_i\}_{i=1}^N$. Assume $\max_{1 \le i \le N} \|X_i\|_p < \infty$ where $\|\cdot\|_p = [\mathbb{E}|\cdot|^p]^{1/p}$ is the $\mathcal{L}_p$ norm of a random variable. For $k \ge 0$, define the k-th physical dependence measure by

$$\delta_p(k) := \sup_{0 \le u \le 1} \|G(u, \mathcal{F}_k) - G(u, (\mathcal{F}_{-1}, \epsilon_0', \epsilon_1, \ldots, \epsilon_k))\|_p. \tag{5}$$

Next, we extend the geometric-moment contraction (GMC) condition [SW07] to the non-stationary setting.

Definition 2.6. (Geometric-moment contraction) We say that the locally stationary time series $\{X_i\}_{i=1}^N$ is GMC(p) if for any k we have $\delta_p(k) = \mathcal{O}(\rho^k)$ form some $\rho \in (0,1)$.

Remark 2.7. Denote $X_{u,i} := G(i/N, \mathcal{F}_{u,i})$ where $\mathcal{F}_{u,i} = (\ldots, \epsilon_{\lfloor uN \rfloor}, \epsilon_{\lfloor uN \rfloor+1}, \ldots, \epsilon_{\lfloor uN \rfloor+i})$. Let ϵ_k' be an i.i.d. copy of ϵ_k and $X_{u,i}' := G(i/N, \mathcal{F}_{u,i}')$ where $\mathcal{F}_{u,i}' = (\ldots, \epsilon_0', \ldots, \epsilon_{\lfloor uN \rfloor}', \epsilon_{\lfloor uN \rfloor+1}, \ldots, \epsilon_{\lfloor uN \rfloor+i})$ is a coupled version of $X_{u,i}$. Then under GMC(p), $p > 0$, there exist $C > 0$ and $0 < \rho = \rho(p) < 1$ that do not depend on u, such that for any u and i, we have

$$\sup_u \mathbb{E}(|X_{u,i}' - X_{u,i}|^p) \le C\rho^i. \tag{6}$$

This is because, when GMC(p) holds, we have $\sup_u \mathbb{E}(|X_{u,i}' - X_{u,i}|^p) \le \sum_{k=i}^\infty \delta_p(k) \le \mathcal{O}(\sum_{k=i}^\infty \rho^k) = \mathcal{O}(\rho^i)$. $\triangleleft$

Example 2.8. (Nonstationary nonlinear time series) Many stationary nonlinear time series models are of the form

$$X_i = R(X_{i-1}, \epsilon_i), \tag{7}$$

where ϵ_i are i.i.d. and R is a measurable function. A natural extension to locally stationary setting is to incorporate the time index u via

$$X_i(u) = R(u, X_{i-1}(u), \epsilon_i), \quad 0 \le u \le 1. \tag{8}$$

Zhou and Wu [ZW09, Theorem 6] showed that one can have a nonstationary process $X_i = X_i^{(N)} = G(i/N, \mathcal{F}_i)$ and the GMC(α) condition holds, if $\sup_u \|R(u, x_0, \epsilon_i)\|_\alpha < \infty$ for some x_0, and

$$\sup_{u\in[0,1]} \sup_{x\neq y} \frac{\|R(u, x, \epsilon_0) - R(u, y, \epsilon_0)\|_\alpha}{|x - y|} < 1. \tag{9}$$

See [ZW09, Section 4.2] for more details. ◁

Definition 2.9. (Instantaneous covariance) Let $u \in [0, 1]$, the instantaneous covariance at u is defined by

$$r(u, k) := \mathrm{Cov}\left(G(u, \mathcal{F}_i), G(u, \mathcal{F}_{i+k})\right). \tag{10}$$

Remark 2.10. Note that the definition of $r(u, k)$ does not depend on the index i for $\mathcal{F}_i$ since $(\cdots, \epsilon_0, \dots, \epsilon_{i-1}, \epsilon_i)$ is a stationary sequence. ◁

Remark 2.11. The assumption of SLC(q) together with $\sup_i \mathbb{E}|X_i|^p < \infty$, where $1/p+1/q = 1$, implies the instantaneous covariance $r(u, k)$ is Lipschitz continuous. That is, for all k and for all $u, s \in [0, 1], u \neq s$, we have

$$|r(u, k) - r(s, k)|/|u - s| \le C, \tag{11}$$

for some finite constant C. The proof is given in Appendix A.15. Therefore, uniformly on u, for any positive integer $n \le N$, we have

$$r(u + \delta_u, k) - r(u, k) = \mathcal{O}(n/N), \quad \forall - n/N \le \delta_u \le n/N. \tag{12}$$

Particularly, if we choose $n = o(\sqrt{N})$ then $r(u + \delta_u, k) - r(u, k) = o(1/n), \forall - n/N \le \delta_u \le n/N$. ◁

Remark 2.12. It can be easily shown that if GMC(2) holds, then $\sup_u |r(u,k)| = \mathcal{O}(\rho^k)$ for some $\rho \in (0,1)$. Also, if $\sup_i \|X_i\|_p < \infty$ and GMC(α) holds with any given $\alpha > 0$, then X_i is GMC(α) with any $\alpha \in (0,p)$. In particular, if GMC(α) holds with some $\alpha \geq 2$, then we must have $\sup_u \sum_{k=-\infty}^{\infty} |r(u,k)| < \infty$ since $\sup_u |r(u,k)| = \mathcal{O}(\rho^k) = o(k^{-2})$. Also, if GMC(2) holds as well as $\sup_i \mathbb{E}(|X_i|^{4+\delta}) < \infty$ for some $\delta > 0$, then GMC(4) holds. ◁

Next, we define spectral density using the instantaneous covariance.

Definition 2.13. (Instantaneous spectral density) Let $u \in [0,1]$, the spectral density at u is defined by

$$f(u,\theta) := \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} r(u,k) \exp(\sqrt{-1}k\theta). \tag{13}$$

In this paper, we always assume $f_* := \inf_{u,\theta} f(u,\theta) > 0$, which is a natural assumption in the time series literature (see e.g. [SW07; LW10]). Finally, we define the STFT, the local periodogram, and the STFT-based spectral density estimates.

Definition 2.14. (Short-time Fourier transform) Let $\tau(\cdot) \leq \tau_* < \infty$ be a kernel with support $[-1/2, 1/2]$ such that $\tau \in \mathcal{C}^1([-1/2,1/2])$ and $\int \tau^2(x)\mathrm{d}x = 1$. Let n be the number of data in a local window and $\theta \in [0, 2\pi)$, the STFT is defined by

$$J_n(u,\theta) := \sum_{i=1}^{N} \tau\left(\frac{i - \lfloor uN \rfloor}{n}\right) X_i \exp(\sqrt{-1}\theta i). \tag{14}$$

Definition 2.15. (Local periodogram)

$$I_n(u,\theta) := \frac{1}{2\pi n} |J_n(u,\theta)|^2. \tag{15}$$

Remark 2.16. Note that defining

$$\hat{r}(u,k) := \frac{1}{n} \sum_{i=1}^{N} \tau\left(\frac{i - \lfloor uN \rfloor}{n}\right) \tau\left(\frac{i + k - \lfloor uN \rfloor}{n}\right) X_i X_{i+k}, \tag{16}$$

then we can write $I_n(u,\theta)$ as

$$I_n(u,\theta) = \frac{1}{2\pi} \sum_{k=-n}^{n} \hat{r}(u,k) \exp(\sqrt{-1}\theta k). \tag{17}$$

◁

Definition 2.17. (STFT-based spectral density estimator) Let $a(\cdot)$ be an even, Lipschitz continuous function with support $[-1, 1]$ and $a(0) = 1$; let B_n be a sequence of positive integers with $B_n \to \infty$ and $B_n/n \to 0$, then the STFT-based spectral density estimator is defined by

$$\hat{f}_n(u, \theta) := \frac{1}{2\pi} \sum_{k=-B_n}^{B_n} \hat{r}(u, k) a(k/B_n) \exp(\sqrt{-1}k\theta). \tag{18}$$

3. Fourier Transforms

In this section, we study the STFT and show that the STFTs are asymptotically independent normals under mild conditions. More specifically, we consider frequencies $\{2\pi j/n : j = 1, \ldots, n\}$, we show that uniformly over a grid of u and j, $\{J_n(u, 2\pi j/n)\}$ are asymptotically independent normal random variables.

Denote the real and imaginary parts of $\{J_n(u, 2\pi j/n)/\sqrt{\pi n f(u, 2\pi j/n)}\}$ by

$$\begin{aligned} Z_{u,j}^{(n)} &= \frac{\sum_{k=1}^{N} \tau\left(\frac{k-\lfloor uN \rfloor}{n}\right) X_k \cos(k2\pi j/n)}{\sqrt{\pi n f(u, 2\pi j/n)}}, \\ Z_{u,j+m}^{(n)} &= \frac{\sum_{k=1}^{N} \tau\left(\frac{k-\lfloor uN \rfloor}{n}\right) X_k \sin(k2\pi j/n)}{\sqrt{\pi n f(u, 2\pi j/n)}}, \quad j = 1, \ldots, m, \end{aligned} \tag{19}$$

where $m := \lfloor (n-1)/2 \rfloor$. Then, we have the following result.

Theorem 3.1. *Let $\Omega_{p,q} = \{c \in \mathbb{R}^{pq} : |c| = 1\}$, where $|\cdot|$ denotes Euclidean norm, and*

$$Z_{U,J} = (Z_{u_1,j_1}^{(n)}, \ldots, Z_{u_1,j_p}^{(n)}, \ldots, Z_{u_q,j_1}^{(n)}, \ldots, Z_{u_q,j_p}^{(n)})^T$$

for $J = (j_1, \ldots, j_p)$ satisfies $1 \le j_1, \ldots, j_p \le 2m$ and $U = (u_1, \ldots, u_q)$ satisfies $0 < u_1 < \cdots < u_q < 1$. Then for any fixed $p, q \in \mathbb{N}$, as $n \to \infty$, we have

$$\sup_J \sup_{c \in \Omega_p} \sup_x |P(c^T Z_{U,J} \le x) - \Phi(x)| = o(1). \tag{20}$$

Proof. See Section 8.1. □

The above theorem shows that if we select any p elements from the canonical frequencies $\{2\pi j/n, j = 1, \ldots, n\}$ and q well-separated points from the re-scaled time, the STFTs are asymptotically independent on the latter time-frequency grid. Moreover, the vector formed by these STFTs is asymptotically jointly normal.

4. Consistency and Asymptotic Normality

In this section, we study the asymptotic properties of the smoothed periodogram estimate $\hat{f}_n(u,\theta)$.

4.1. **Consistency.**

The consistency result for the local spectral density estimate $\hat{f}_n(u,\theta)$ is as follows.

Theorem 4.1. *Assume* GMC(2) *and there exists* $\delta \in (0,4]$ *such that* $\sup_i \mathbb{E}(|X_i|^{4+\delta}) < \infty$. *Let* $B_n \to \infty$, $B_n = \mathcal{O}(n^{\eta})$, $0 < \eta < \delta/(4+\delta)$. *Then*

$$\sup_u \max_{\theta\in[0,\pi]} \sqrt{n/B_n}|\hat{f}_n(u,\theta) - \mathbb{E}(\hat{f}_n(u,\theta))| = \mathcal{O}_{\mathbb{P}}(\sqrt{\log n}). \tag{21}$$

Proof. See Section 8.2. □

Later we will see from Theorem 5.3 that the order $\mathcal{O}_{\mathbb{P}}(\sqrt{\log n})$ on the right hand side of Eq. (21) is indeed optimal.

Remark 4.2. Assume $\sup_i \mathbb{E}|X_i|^p < \infty$ with $p > 4$ and SLC(q) with $1/p + 1/q = 1$. If we further assume the kernel $\tau(\cdot)$ is an even function and $r(u,k)$ is twice continuously differentiable with respect to u, then under the assumptions of Theorem 4.1, whenever $n = o(N^{2/3})$, $B_n = o(\min\{n, N^{1/3}\})$, and $\sup_u \sum_{k\in\mathbb{Z}} k^2|r(u,k)| < \infty$, if $a(\cdot)$ is locally quadratic at 0, i.e.

$$\lim_{u\to 0} u^{-2}[1 - a(u)] = C, \tag{22}$$

where C is a nonzero constant, then we have

$$\sup_u \sup_\theta \left[\mathbb{E}\hat{f}_n(u,\theta) - f(u,\theta) - \frac{C}{B_n^2} f''(u,\theta)\right] = o(1/B_n^2). \tag{23}$$

where $f''(u,\theta) := -\frac{1}{2\pi}\sum_{k\in\mathbb{Z}} k^2 r(u,k)\exp(\sqrt{-1}k\theta)$. The proof is given in Appendix A.13. ◁

4.2. **Asymptotic Normality.**

Developing an asymptotic distribution for the local spectral density estimate is an important problem in spectral analysis of non-stationary time series. This allows one to perform statistical inference such as constructing point-wise confidence intervals and hypothesis testing. In the following, we derive a central limit theorem for $\hat{f}_n(u,\theta)$. The main result is as follows.

Theorem 4.3. *Assume* GMC(2) *and* $\sup_i \mathbb{E}(|X_i|^{4+\delta}) < \infty$ *for some* $\delta > 0$, $B_n \to \infty$ *and* $B_n = o(n/(\log n)^{2+8/\delta})$. *Then*

$$\sqrt{n/B_n}\{\hat{f}_n(u,\theta) - \mathbb{E}(\hat{f}_n(u,\theta))\} \Rightarrow \mathcal{N}(0, \sigma_u^2(\theta)), \tag{24}$$

where $\Rightarrow$ *denotes weak convergence,* $\sigma_u^2(\theta) = [1+\eta(2\theta)]f^2(u,\theta)\int_{-1}^{1} a^2(t)\mathrm{d}t$ *and* $\eta(\theta) = 1$ *if* $\theta = 2k\pi$ *for some integer* k *and* $\eta(\theta) = 0$ *otherwise.*

Proof. See Section 8.3. □

5. Maximum Deviations

The asymptotic normality for $\hat{f}_n(u,\theta)$ derived in the last section cannot be used to construct simultaneous confidence region (SCR) over u and θ. For simultaneous spectral inference under complex temporal dynamics, one needs to know the asymptotic distribution of the maximum deviation, which is an extremely difficult problem. In this section, we derive the asymptotic distribution of maximum deviations of the spectral density estimates. Such results can be used to construct SCR for spectral inference over u and θ. The established maximum deviation theory for the STFT-based spectral density estimates is over a dense grid in the joint time-frequency domain.

- Condition (a): Define $\mathcal{U} = \{u_1, \ldots, u_{C_n}\}$ where $C_n = |\mathcal{U}|$ and $\frac{n}{2N} < u_i < 1 - \frac{n}{2N}$, $i = 1, \ldots, C_n$. For any $u_{i_1}, u_{i_2} \in \mathcal{U}$ with $i_1 \neq i_2$, we have $|u_{i_1} - u_{i_2}| \geq \frac{n}{N}(1 - 1/(\log B_n)^2)$.
- Condition (b): Assume $\sup_k \mathbb{E}|X_k|^p < \infty$ where $p > 4$, and SLC(q) where $1/p + 1/q = 1$. Let α be a constant such that $\frac{3}{4(p-1)} < \alpha < \frac{1}{4}$. Then assume $C_n = o[\min\{(nB_n)^{2\alpha(p-1)-1}, B_n^{1+2\alpha(p-2)} n^{-2-2\gamma}\}]$ for some $\gamma > 0$.
- Condition (c): If $a(\cdot)$ is an even bounded function with bounded support $[-1,1]$, $\lim_{x\to 0} a(x) = a(0) = 1$, $\int_{-1}^{1} a^2(x)\mathrm{d}x < \infty$, and $\sum_{j\in\mathbb{Z}} \sup_{|s-j|\leq 1} |a(jx) - a(sx)| = \mathcal{O}(1)$ as $x \to 0$.
- Condition (d): There exists $0 < \delta_1 < \delta_2 < 1$ and $c_1, c_2 > 0$ such that for all large n, $c_1 n^{\delta_1} \leq B_n \leq c_2 n^{\delta_2}$.

Note that Conditions (c) and (d) are very mild. Condition (a) implies the time interval between any two time points on the grid $\mathcal{U}$ cannot be too close. Condition (b) implies that the total number of selected time points is not too large.

Remark 5.1. Condition (a) implies that $C_n \leq \frac{N}{n}(1-\frac{n}{N})(1-\frac{1}{(\log B_n)^2}) = \mathcal{O}(N/n)$. Although we do not assume $\{u_i\}$ to be equally spaced, However, we suggest in practice choosing $\{u_i\}$ equally spaced and $C_n = \frac{N}{n}(1-\frac{n}{N})(1-\frac{1}{(\log B_n)^2})$ to avoid the tricky problem on how to choose the u_i's and the C_n. ◁

Definition 5.2. (Dense Grid $\mathcal{G}_N$) Let $\mathcal{G}_N$ be a collection of time-frequency pairs such that $(u,\theta) \in \mathcal{G}_N$ if $u \in \mathcal{U}$ and $\theta \in \{\frac{i\pi}{B_n}, i = 0, \ldots, B_n\}$.

The following theorem states that the maximum deviation of the spectral density estimates behaves asymptotically like a Gumbel distribution.

Theorem 5.3. *Under* GMC(2) *and Conditions (a)–(d), we have*

$$\begin{aligned} \mathbb{P}\Bigg[& \max_{(u,\theta)\in\mathcal{G}_N} \frac{n}{B_n} \frac{|\hat{f}_n(u,\theta) - \mathbb{E}(\hat{f}_n(u,\theta))|^2}{f^2(u,\theta)\int_{-1}^{1} a^2(t)\mathrm{d}t} \\ & -2\log B_n - 2\log C_n + \log(\pi \log B_n + \pi \log C_n) \leq x\Bigg] \to e^{-e^{-x/2}}. \end{aligned} \tag{25}$$

Proof. See Section 8.4. □

Theorem 5.3 states the spectral density estimates $\{\hat{f}_n(u,\theta) : (u,\theta) \in \mathcal{G}_N\}$ on a dense grid $\mathcal{G}_N$ consisting of $C_n \times B_n$ total number of pairs of (u,θ) are asymptotically independent quadratic forms of $\{X_i\}_{i=1}^N$. Furthermore, the maximum deviation of the spectral density estimates on $\mathcal{G}_N$ converges to a Gumbel law. This result can be used to construct SCR for the evolutionary spectral densities. Note that Theorem 5.3 is established for a very general class of possibly nonlinear locally stationary processes for the joint and simultaneous time-frequency inference of the evolutionary spectral densities.

5.1. **Near optimality of the grid selection.** Note that there is a trade-off on how dense the grid should be chosen. On the one hand, we hope the grid is dense enough to asymptotically correctly depicting the whole time-frequency stochastic variation of the estimates. On the other hand, making the grid too dense is a waste of computational resources since it does not reveal any extra useful information on the overall variability of the estimates. In the following, we define the notion of asymptotically uniform variation matching of a sequence of dense grids. The purpose of the latter notion is to mathematically determine how dense a sequence of grids should be such that it will adequately capture the overall stochastic variation of the spectral density estimates on the joint time-frequency domain.

Definition 5.4. (Asymptotically uniform variation matching of grids) Consider for a given sequence of bandwidths (n, B_n). Let $\{\tilde{\mathcal{G}}_N\}$ be a sequence of grids of time-frequency pairs $\{(u_i, \theta_j)\}$ with time and frequencies equally spaced i.e. $|u_{i+1} - u_i| = \delta_{\theta,n}$ and $|\theta_{j+1} - \theta_j| = \delta_{u,n}$, respectively. Then the sequence $\{\tilde{\mathcal{G}}_N\}$ is said to be asymptotically uniform variation matching if

(26)
$$\max_{\{u_i,\theta_j\}\in\tilde{\mathcal{G}}_N} \sup_{\{u:|u-u_i|\le\delta_{u,n},\theta:|\theta-\theta_j|\le\delta_{\theta,n}\}} \sqrt{n/B_n}\left|\left[\hat{f}_n(u,\theta) - \mathbb{E}(\hat{f}_n(u,\theta))\right] - \left[\hat{f}_n(u_i,\theta_j) - \mathbb{E}(\hat{f}_n(u_i,\theta_j))\right]\right|$$
$$= o_{\mathbb{P}}(\sqrt{\log n}).$$

Note that we have previously shown in Theorem 4.1 that the uniform stochastic variation of $\hat{f}_n(u,\theta)$ on $(u,\theta) \in (0,1)\times[0,\pi)$ has the order of $\mathcal{O}_{\mathbb{P}}(\sqrt{\log n})$. Combing with Theorem 5.3, we can see the order $\mathcal{O}_{\mathbb{P}}(\sqrt{\log n})$ cannot be improved. Therefore, by a simple chaining argument, we can show if a sequence of grids $\{\tilde{\mathcal{G}}_N\}$ is an asymptotically uniform variation matching, then

(27)
$$\sqrt{n/B_n}\left|\sup_{(u,\theta)\in(0,1)\times[0,\pi)}\left|\hat{f}_n(u,\theta) - \mathbb{E}(\hat{f}_n(u,\theta))\right| - \max_{\{u_i,\theta_j\}\in\tilde{\mathcal{G}}_N}\left|\hat{f}_n(u_i,\theta_j) - \mathbb{E}(\hat{f}_n(u_i,\theta_j))\right|\right| = o_{\mathbb{P}}(\sqrt{\log n}).$$

Hence, the uniform stochastic variation of $\hat{f}_n(u,\theta)$ on $(u,\theta) \in \tilde{\mathcal{G}}_N$ is asymptotically equal to the uniform stochastic variation of $\hat{f}_n(u,\theta)$ on $(u,\theta) \in (0,1)\times[0,\pi)$.

However, a grid that is asymptotically uniform variation matching may be unnecessarily dense which causes a waste of computational resources without depicting any additional useful information. The optimal grid should balance between computational burden and asymptotic correctness in depicting the overall time-frequency stochastic variation of the estimates. Therefore, we hope to choose a sequence of grids as sparse as possible provided it is (nearly) asymptotically uniform variation matching.

Next, we show the sequence of grids used in Theorem 5.3 is indeed nearly optimal in this sense. Recall that in Theorem 5.3, the interval between adjacent frequencies is of order $\delta_{\theta,n} = \Omega(1/B_n)$ and the averaged interval between two adjacent time index is of order $\delta_{u,n} = \Omega(n/N)$, where we define $a_n = \Omega(b_n)$ if $1/a_n = \mathcal{O}(1/b_n)$. In the following, we show that if we choose slightly denser grid by $\delta_{\theta,n} = \mathcal{O}\left(\frac{1}{B_n(\log n)^\alpha}\right)$ and $\delta_{u,n} = \mathcal{O}\left(\frac{n}{N(\log n)^\alpha}\right)$ with arbitrary $\alpha > 0$, then the sequence of grids is asymptotically uniform variation matching.

Since α can be chosen arbitrarily close to zero, the dense grids in Theorem 5.3 is nearly optimal.

Theorem 5.5. *For the time series in Theorem 5.3, a sequence of grids with equally spaced time and frequency intervals $\delta_{u,n}$ and $\delta_{\theta,n}$ is asymptotically uniform variation matching if $\delta_{u,n} = \mathcal{O}\left(\frac{n}{N(\log n)^\alpha}\right)$ and $\delta_{\theta,n} = \mathcal{O}\left(\frac{1}{B_n(\log n)^\alpha}\right)$ for some $\alpha > 0$.*

Proof. See Section 8.5. □

5.2. **Applications of the Simultaneous Confidence Regions.** In this subsection, we illustrate several applications of the proposed SCR for spectral inference. These examples include testing time-varying white noise (Example 5.6), testing stationarity (Example 5.7), testing time-frequency separability or correlation stationarity (Example 5.8), and validating time-varying ARMA models (Example 5.9).

These examples demonstrate that our maximum deviation theory can serve as a foundation for the joint and simultaneous time-frequency inference. In particular, as far as we know, there is no existing methodology in the literature for testing time-frequency separability of locally stationary time series, nor model validation for time-varying ARMA models, although they are certainly very important problems. On the other hand, our proposed SCR serves as an asymptotically valid and visually friendly tool for such tests (see Examples 5.8 and 5.9).

Next, we demonstrate the SCR can be applied to a wide range of tests on the structure of the evolutionary spectra or the time series itself. The key observation is that typically under some specific structural assumptions, the time-varying spectra can be estimated with a faster convergence rate than those estimated by the STFT. Therefore, to test the structure of the evolutionary spectra under the null hypothesis, a generic procedure is to check whether the estimated spectra under the null can be fully embedded into the SCR. Note that this very general procedure is asymptotically correct as long as the evolutionary spectra estimated under the null converges faster than the SCR. The test achieves asymptotic power 1 for local alternatives whose evolutionary spectra deviate from the null with a rate larger than the order of the width of the SCR.

Example 5.6. (Testing time-varying white noise) White noise is a collection of uncorrelated random variables with mean 0 and time-varying variance $\sigma^2(u)$. It can be verified that testing time-varying white noise is equivalent to testing the following null hypothesis:

$$H_0: \quad \forall\theta, \quad f(u,\theta) = g(u), \quad u \in [0,1] \tag{28}$$

for some time-varying function $g(\cdot)$.

Therefore, under null hypothesis we can estimate the function g by

$$\hat{g}(u) := \frac{1}{\pi}\int_0^\pi \hat{f}_n(u,\theta)\mathrm{d}\theta \approx \frac{1}{\pi}\int_0^\pi f(u,\theta)\mathrm{d}\theta = g(u). \tag{29}$$

It can be shown that under null hypothesis the convergence rate is $\mathcal{O}(\sqrt{\log n}/\sqrt{n})$ for $\hat{g}(u) \to g(u)$ uniformly over u, which is faster than the rate of SCR which is $\mathcal{O}(\sqrt{\log n}/\sqrt{n/B_n})$. Therefore, we can apply the proposed SCR to test time-varying white noise. ◁

Example 5.7. (Testing stationarity) Under the null hypothesis that the time series is stationary, we can test

$$H_0: \quad \forall u, \quad f(u,\theta) = h(\theta), \quad \theta \in [0,\pi] \tag{30}$$

for some function $h(\cdot)$.

Under the null hypothesis, we can estimate the function h by

$$\hat{h}(\theta) := \int \hat{f}_n(u,\theta)\mathrm{d}u \approx \int f(u,\theta)\mathrm{d}u = h(\theta). \tag{31}$$

It can be shown the convergence rate for $\hat{h}(\theta) \to h(\theta)$ uniformly over θ is $\mathcal{O}(\sqrt{\log n}/\sqrt{N/B_n})$, which is faster than the rate $\mathcal{O}(\sqrt{\log n}/\sqrt{n/B_n})$ of the SCR. Therefore, we can apply the proposed SCR to test stationarity. ◁

Example 5.8. (Testing time-frequency separability or correlation stationarity) We call a non-stationary time series is time-frequency separable if the spectral satisfies $f(u,\theta) = g(u)h(\theta)$ for some functions $g(\cdot)$ and $h(\cdot)$. It can be verified that testing time-frequency separability is equivalent to testing correlation stationarity, i.e. $\mathrm{corr}(X_i, X_{i+k}) = l(k)$ for some function $l(\cdot)$. Without loss of generality, we can formulate the null hypothesis as

$$H_0: \quad f(u,\theta) = C_0 g(u)h(\theta), \tag{32}$$

for some constant C_0 and $\int_0^1 g(u)\mathrm{d}u = 1$ and $\int_0^\pi h(\theta) = 1$.

Under the null hypothesis, we can estimate C_0, $g(u)$ and $h(\theta)$ by

$$\hat{C}_0 := \int\int \hat{f}_n(u_i, \theta_j)\mathrm{d}u\mathrm{d}\theta \approx \int\int f(u,\theta)\mathrm{d}u\mathrm{d}\theta = C_0, \tag{33}$$

$$\hat{g}(u) := \frac{1}{\hat{C}_0}\int \hat{f}_n(u, \theta_j)\mathrm{d}\theta \approx \frac{1}{C_0}\int f(u,\theta)\mathrm{d}\theta = g(u), \tag{34}$$

$$\hat{h}(\theta) := \frac{1}{\hat{C}_0}\int \hat{f}_n(u_i, \theta)\mathrm{d}u \approx \frac{1}{C_0}\int f(u,\theta)\mathrm{d}u = h(\theta). \tag{35}$$

It can be shown the convergence rates for $\hat{C}_0$, $\hat{g}(u)$, and $\hat{h}(\theta)$ are $\mathcal{O}(1/\sqrt{N})$, $\mathcal{O}(\sqrt{\log n}/\sqrt{n})$, and $\mathcal{O}(\sqrt{\log n}/\sqrt{N/B_n})$, respectively. All of them are faster than the rate of the SCR which is $\mathcal{O}(\sqrt{\log n}/\sqrt{n/B_n})$. Therefore, we can apply the proposed SCR to test the null hypothesis. ◁

Example 5.9. (Validating time-varying ARMA models) The proposed SCR can be used for testing existing parametric/non-parametric models of non-stationary time series. Consider the following null hypothesis of a time-varying ARMA model

$$H_0 : \sum_{i=0}^{p} a_i(t/N)X_{t-i} = \sum_{j=0}^{q} b_j(t/N)\epsilon_{t-j} \tag{36}$$

for $a_0(u) = 1$, smooth functions $a_i(\cdot)$, $b_i(\cdot)$, and ϵ_i are uncorrelated random variables with mean 0 and variance 1. This non-parametric time-varying ARMA model is non-parametric in time domain and parametric in frequency domain. Under the null hypothesis, $\{X_i\}$ are locally stationary time series with spectral density

$$f(u,\theta) = \frac{1}{2\pi}\frac{\left|\sum_{j=0}^{q} b_j(u)\exp(\sqrt{-1}2\pi\theta j)\right|^2}{\left|\sum_{i=0}^{p} a_i(u)\exp(\sqrt{-1}2\pi\theta i)\right|^2}. \tag{37}$$

The spectral density can be fitted using generalized Whittle's method [Dah97], where $a_i(t/N)$ and $b_i(t/N)$ are estimated by minimizing a generalized Whittle function and p and q are selected, for example, by AIC. Note that under the null hypothesis, the spectral density estimated using Whittle's method has a convergence rate $\mathcal{O}(\sqrt{\log n}/\sqrt{n})$ faster than the rate $\mathcal{O}(\sqrt{\log n}/\sqrt{n/B_n})$ by STFT-based methods without prior information. Therefore, to test the fitted non-parametric time-varying ARMA model, we can plot the non-parametric spectral density using the estimated time-varying parameters $a_i(\cdot)$ and $b_i(\cdot)$. Under the null hypothesis, the non-parametric spectral density should fall within our SCR with prescribed probability asymptotically. $\triangleleft$

6. Bootstrap and Tuning Parameter Selection

In Section 6.1, we propose a simulation based bootstrap method to implement simultaneous statistical inferences. The motivation of the bootstrap procedure is to alleviate the slow convergence of the maximum deviation to its Gumbel limit in Theorem 5.3. We discuss methods for tuning parameter selection in Section 6.2.

6.1. The Bootstrap Procedure.

Although Theorem 5.3 shows that SCR can be constructed using the Gumbel distribution, the convergence rate for Theorem 5.3 to hold usually is very slow. We propose a bootstrap procedure to alleviate the slow convergence of the maximum deviations. One important application of the bootstrap is to construct SCR in moderate sample cases.

Let $\{\epsilon_1, \ldots, \epsilon_N\}$ be i.i.d. standard normal random variables. Defining

$$\hat{r}^{\epsilon}(u,k) := \frac{1}{n}\sum_{i=1}^{N} \tau\left(\frac{i-\lfloor uN\rfloor}{n}\right)\tau\left(\frac{i+k-\lfloor uN\rfloor}{n}\right)\epsilon_i\epsilon_{i+k} \tag{38}$$

and

$$\hat{f}_n^{\epsilon}(u,\theta) := \frac{1}{2\pi}\sum_{k=-B_n}^{B_n} \hat{r}^{\epsilon}(u,k)a(k/B_n)\exp(\sqrt{-1}k\theta), \tag{39}$$

it can be easily verified that the following analogy of Theorem 5.3 holds.

$$\begin{aligned}\mathbb{P}\Bigg[&\max_{u\in\mathcal{U}}\max_{0\le i\le B_n} \frac{n}{B_n}\frac{|\hat{f}_n^{\epsilon}(u,\theta_i)-\mathbb{E}(\hat{f}_n^{\epsilon}(u,\theta_i))|^2}{[f^{\epsilon}(u,\theta_i)]^2\int_{-1}^{1}a^2(t)\mathrm{d}t}\\ &-2\log B_n - 2\log C_n + \log(\pi\log B_n+\pi\log C_n)\le x\Bigg] \to e^{-e^{-x/2}}.\end{aligned} \tag{40}$$

Therefore, we propose to construct SCR for $\{\hat{f}_n(u,\theta)\}$ using the empirical distribution of $\hat{f}_n^{\epsilon}(u,\theta)$. More specifically, we generate $\{\epsilon_i\}_{i=1}^N$ independently for N_{MC} times. Let $\bar{f}_n^{\epsilon}(u,\theta_i)$ denotes the sample mean of $\{\hat{f}_{n,m}^{\epsilon}(u,\theta_i), m=1,\ldots,N_{\mathrm{MC}}\}$ from the N_{MC} Monte Carlo experiments. Then we compute the empirical distribution of

$$\max_{u\in\mathcal{U}}\max_{0\le i\le B_n}\frac{|\hat{f}_{n,m}^{\epsilon}(u,\theta_i)-\bar{f}_n^{\epsilon}(u,\theta_i)|^2}{[\bar{f}_n^{\epsilon}(u,\theta_i)]^2}, \quad m=1,\ldots,N_{\mathrm{MC}} \tag{41}$$

to approximate the distribution of

$$\max_{u\in\mathcal{U}} \max_{0\le i\le B_n} \frac{|f(u,\theta_i)-\hat{f}_n(u,\theta_i)|^2}{[\hat{f}_n(u,\theta_i)]^2}, \tag{42}$$

which can be employed to construct the SCR. For example, we estimate the critical value $\gamma_{1-\alpha}^2$ for level $\alpha \in (0,1)$ from the bootstrapped distribution using $\hat{f}_n^{\epsilon}(u,\theta)$, which also approximately satisfies

$$\mathbb{P}\left(\max_{u\in\mathcal{U}} \max_{0\le i\le B_n} \frac{|f(u,\theta_i)-\hat{f}_n(u,\theta_i)|^2}{[\hat{f}_n(u,\theta_i)]^2} \le \gamma_{1-\alpha}^2\right) = 1-\alpha. \tag{43}$$

Therefore, the constructed confidence region is

$$\max\{0,(1-\gamma_{1-\alpha})\hat{f}(u,\theta)\} \le f(u,\theta) \le (1+\gamma_{1-\alpha})\hat{f}(u,\theta). \tag{44}$$

Note that in small sample cases, the lower band for the confidence region can be 0 if the estimated $\gamma_{1-\alpha}$ is larger than 1. This happens when N is not large enough and large B_n and C_n are selected. For large sample sizes, the estimated $\gamma_{1-\alpha}^2$ is typically much smaller than 1, in that case we can further use the following approximation

$$\frac{|f(u,\theta_i)-\hat{f}_n(u,\theta_i)|^2}{[\hat{f}_n(u,\theta_i)]^2} \approx [\log(f(u,\theta_i)/\hat{f}_n(u,\theta_i))]^2. \tag{45}$$

Then the constructed confidence region can be written as

$$\exp(-\gamma_{1-\alpha})\hat{f}_n(u,\theta) \le f(u,\theta) \le \exp(+\gamma_{1-\alpha})\hat{f}_n(u,\theta). \tag{46}$$

Overall, the practical implementation is given as follows

(1) Select B_n and n using the tuning parameter selection method described in Section 6.2;
(2) Compute the critical value using bootstrap described in Section 6.1;
(3) Compute the spectral density estimates by Eq. (18);
(4) Compute the SCR defined in Section 6.1 using the spectral density estimates and the critical value obtained by bootstrap.

6.2. **Tuning parameter selection.** Choosing B_n and n in practice is a non-trivial problem. First of all, we hope to select a dense grid so B_n and C_n are large enough. However, large C_n implies small n. Our theory requires $B_n < n$ and $B_n = o(n)$. Furthermore, Theorem 5.3 requires n at least has an order of $B_n \log(B_n)$ to make the confidence region not too wide.

Overall, when N is small, we hope to choose B_n and C_n relatively large under the constraint $B_n < n$ in order to have an accurate bootstrap procedure. When N is large, we hope to choose B_n and C_n relatively large and n has the order of $B_n \log(B_n)$ in order to have a narrow confidence region.

In our Monte Carlo experiments and real data analysis, we find that the minimum volatility (MV) method [PRW99; Zho13] performs reasonably well. Specifically, the MV method uses the fact that the estimator $\hat{f}_n(u, \theta)$ becomes stable when the block size n and the bandwidth B_n are in an appropriate range. More specifically, we first set a proper interval for n as $[n_l, n_r]$. In our simulations and data analysis, we choose $n_l = 2N^\eta$ and $n_r = 3N^\eta$ if $N \leq 1000$, and $n_l = 3N^\eta$ and $n_r = 4N^\eta$ if $N > 1000$, where $\eta = 0.47 < 0.5$. In order to use the MV method, we first form a two-dimensional grid of all candidate pairs of (n, B_n) such that $n \in [n_l, n_r]$ and $B_n < n/\log(n)$. Then, for each candidate pair (n, B_n), we estimate $\hat{f}_n(u, \theta)$ using the candidate pair for a fixed time-frequency grid of (u, θ). Next, we compute the average variance of the spectral density estimates $\hat{f}_n(u, \theta)$ over the neighborhood of each candidate pair on the two-dimensional grid of all candidate pairs of (n, B_n). Finally, we choose the pair of (n, B_n) which gives the lowest average variance.

7. Simulations and Data Analysis

In this section, we study the performance of the proposed SCR via simulations and real data analysis. In Section 7.1, the accuracy of the proposed bootstrap procedure is studied; The accuracy of tuning parameter selection is considered in Section 7.2; The accuracy and power for hypothesis testing is studied in Section 7.3; Finally, we study real data analysis in Section 7.4. Throughout this section, the kernel $\tau(\cdot)$ is chosen to be a re-scaled Epanechnikov kernel such that $\int \tau^2(x)\mathrm{d}x = 1$, and the kernel $a(\cdot)$ is a re-scaled tri-cube kernel such that $a(0) = 1$. The two kernel functions are defined as follows.

$$\tau(x) := \begin{cases} \frac{\sqrt{30}}{4}(1 - 4x^2), & \text{if } |x| < 1/2, \\ 0, & \text{otherwise,} \end{cases} \qquad a(x) := \begin{cases} (1 - |x|^3)^3, & \text{if } |x| < 1, \\ 0, & \text{otherwise.} \end{cases} \tag{47}$$

7.1. **Accuracy of Bootstrap.**

In this subsection, we study the accuracy of the proposed bootstrap procedure for moderate finite samples (e.g. $N = 400$ or $N = 800$). We consider different examples of locally stationary time series models described in the following Examples 7.1 to 7.5.

Table 1. Simulated Coverage Probability for Example 7.1

		$N = 400$		$N = 800$	
n	B_n	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.1$
72	36	0.03	0.06	0.03	0.06
72	32	0.04	0.07	0.04	0.08
72	28	0.04	0.08	0.05	0.10
54	36	0.03	0.06	0.03	0.06
54	32	0.04	0.08	0.04	0.08
54	28	0.04	0.09	0.05	0.09
36	32	0.05	0.11	0.05	0.11
36	28	0.07	0.13	0.07	0.14

Example 7.1. (Time-varying AR model) We have $X_i = G_0(t_i, \mathcal{F}_i)$, where

$$X_i = a(i/N)X_{i-1} + \epsilon_i, \tag{48}$$

where ϵ_i's are i.i.d. $\mathcal{N}(0,1)$. In this example, we choose $a(u) = 0.3\cos(2\pi u)$, then the model is locally stationary in the sense that the AR(1) coefficient $a(u) = 0.3\cos(2\pi u)$ changes smoothly on the interval $[0,1]$. We ran Monte Carlo experiments for $N_{\mathrm{MC}} = 10000$. The simulated coverage probabilities of the SCR are shown in Table 1. ◁

Example 7.2. (Time-varying ARCH model) generating time-varying ARCH(1) model:

$$X_i = \epsilon_i\sqrt{a_0(i/N) + a_1(i/N)X_{i-1}^2}, \tag{49}$$

where ϵ_i i.i.d. $\mathcal{N}(0,1)$, $a_0(u) > 0, a_1(u) > 0$ and $a_0(u) + a_1(u) < 1$. Note that $\{X_i\}$ is white noise. In this example, we choose $a_0(u) = 0.7$ and $a_1(u) = 0.3\sin(\pi u)$. The simulated coverage probabilities of the SCR from Monte Carlo experiments using $N_{\mathrm{MC}} = 10000$ are shown in Table 2. ◁

Example 7.3. (Time-varying Markov switching model) Suppose $\{S_i\}$ is a Markov chain on state space $\{0,1\}$ with transition matrix P, consider the following time-varying Markov switching model

$$X_i = \begin{cases} a_0(i/N) + b(i/N)X_{i-1} + \epsilon_i, & \text{if } S_i = 0, \\ a_0(i/N) + a_1(i/N) + b(i/N)X_{i-1} + \epsilon_i, & \text{if } S_i = 1. \end{cases} \tag{50}$$

Table 2. Simulated Coverage Probability for Example 7.2

		$N = 400$		$N = 800$	
n	B_n	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.1$
72	36	0.03	0.06	0.03	0.06
72	32	0.04	0.08	0.04	0.08
72	28	0.05	0.10	0.04	0.09
54	36	0.03	0.07	0.03	0.06
54	32	0.05	0.10	0.04	0.09
54	28	0.05	0.12	0.04	0.10
36	32	0.06	0.12	0.07	0.14
36	28	0.07	0.14	0.08	0.14

Table 3. Simulated Coverage Probability for Example 7.3

		$N = 400$		$N = 800$	
n	B_n	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.1$
72	36	0.03	0.07	0.03	0.08
72	32	0.05	0.09	0.04	0.09
72	28	0.06	0.10	0.05	0.10
54	36	0.03	0.06	0.02	0.06
54	32	0.04	0.09	0.03	0.08
54	28	0.05	0.11	0.04	0.10
36	32	0.05	0.11	0.06	0.12
36	28	0.07	0.13	0.08	0.15

where $\{\epsilon_i\}$ are i.i.d. standard Gaussian, $|b| < 1$. In this example, we choose $P = \begin{bmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{bmatrix}$ and $a_0(u) = 0, a_1(u) = 0.3u, b(u) = 0.3\cos(2\pi u)$. The simulated coverage probabilities of the SCR from Monte Carlo experiments with $N_{\text{MC}} = 10000$ are shown in Table 3. $\triangleleft$

Example 7.4. (Time-varying threshold AR model) Suppose $\{\epsilon_i\}$ are i.i.d. standard Gaussian random variables, consider the following threshold AR model

$$X_i = a(i/N)\max(0, X_{i-1}) + b(i/N)\max(0, -X_{i-1}) + \epsilon_i, \tag{51}$$

where $\sup_{u\in[0,1]}[|a(u)| + |b(u)|] < 1$. In this example, we choose $a(u) = 0.3\cos(2\pi u)$ and $b(u) = 0.3\sin(2\pi u)$. The simulated coverage probabilities of the SCR from Monte Carlo experiments with $N_{\text{MC}} = 10000$ are shown in Table 4. $\triangleleft$

Table 4. Simulated Coverage Probability for Example 7.4

		$N = 400$		$N = 800$	
n	B_n	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.1$
72	36	0.05	0.11	0.06	0.12
72	32	0.06	0.12	0.06	0.13
72	28	0.07	0.14	0.08	0.15
54	36	0.05	0.10	0.06	0.12
54	32	0.05	0.11	0.06	0.12
54	28	0.06	0.12	0.07	0.13
36	32	0.08	0.14	0.08	0.14
36	28	0.08	0.14	0.09	0.17

Table 5. Simulated Coverage Probability for Example 7.5

		$N = 400$		$N = 800$	
n	B_n	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.1$
72	36	0.04	0.08	0.05	0.08
72	32	0.05	0.11	0.05	0.10
72	28	0.06	0.13	0.05	0.11
54	36	0.03	0.07	0.04	0.09
54	32	0.05	0.10	0.06	0.11
54	28	0.06	0.13	0.07	0.13
36	32	0.06	0.11	0.08	0.15
36	28	0.09	0.16	0.08	0.15

Example 7.5. (Time-varying bilinear process) Let $\{\epsilon_i\}$ be i.i.d. standard Gaussian, consider the following model

$$X_i = b(i/N)X_{i-1} + \epsilon_i + c(i/N)X_{i-1}\epsilon_{i-1}, \tag{52}$$

where $b^2(u) + c^2(u) < 1$. In this example, we choose $b(u) = 0.3\cos(2\pi u)$ and $c(u) = 0.1\sin(2\pi u)$. The simulated coverage probabilities of the SCR from Monte Carlo experiments using $N_{\mathrm{MC}} = 10000$ are shown in Table 5. $\triangleleft$

The simulated coverage probabilities of the SCR for Examples 7.1 to 7.5 are shown in Tables 1 to 5. According to the results, one can see that the proposed bootstrap works well when B_n and n are chosen in an appropriate range, but slightly sensitive to the choice of B_n. In all examples, if B_n is chosen in the interval $[30, 35]$, the bootstrap is relatively accurate.

TABLE 6. Window Size Selection

Length	$N = 400$			$N = 800$		
Size and Level	(n, B_n)	$\alpha = 0.05$	$\alpha = 0.1$	(n, B_n)	$\alpha = 0.05$	$\alpha = 0.1$
Example 7.1	(54,30)	0.06	0.10	(72,30)	0.05	0.09
Example 7.2	(52,32)	0.04	0.09	(68,30)	0.04	0.08
Example 7.3	(50,32)	0.05	0.09	(72,30)	0.04	0.08
Example 7.4	(50,32)	0.06	0.12	(70,32)	0.06	0.12
Example 7.5	(52,32)	0.04	0.09	(69,31)	0.06	0.11

From the bootstrap results of the examples, the accuracy does not change too much for a relative wide range of n. In the next subsection, we discuss the MV method for selecting B_n and n in practice.

7.2. **Accuracy of Tuning Parameter Selection.** We apply the MV method for Examples 7.1 to 7.5. For all examples, the MV method selects (n, B_n) when $N = 400$ and $N = 800$. The bootstrap accuracy is shown in Table 6. We can see that the MV selected reasonable number B_n, at the same time the selected n increases as N. Note that we have constrained $B_n < n$ in the range of (n, B_n) for tuning parameter selection since $N = 400$ or $N = 800$ is moderate large.

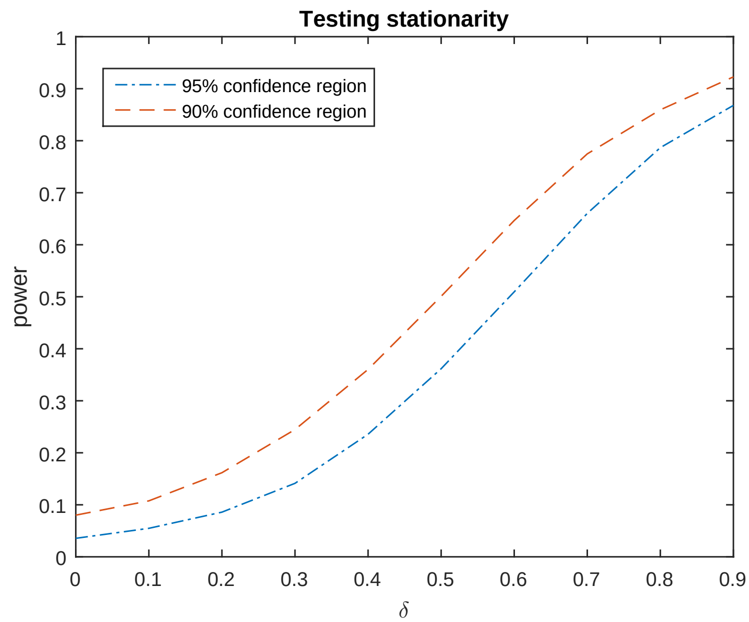
7.3. **Accuracy and Power of Hypothesis Testing.** In this subsection, we study the accuracy and power of hypothesis testing using the proposed SCR. We consider Example 7.6 for testing stationarity and Example 7.7 for testing time-varying white noise. Furthermore, we also consider another example of testing non-parametric model for non-stationary time series, which is given in Example 7.8. In Table 7, we show the accuracy of hypothesis under the null hypothesis for the three examples at level $\alpha = 0.05$ and $\alpha = 0.1$, respectively.

Example 7.6. (Time-varying ARCH model) Consider the following model

$$X_i = \sigma_i \epsilon_i, \quad \sigma_i^2 = a_0(i/N) + a_1(i/N)X_{i-1}^2, \tag{53}$$

where $a_0(u) = 0.3$ and $a_1(u) = 0.2 + \delta u$. Observe that when $\delta = 0$, the model is stationary. When $\delta = 0$, the accuracy of the hypothesis testing for stationarity is studied for two cases, one with $N = 400$, $n = 56$, $B_n = 32$ and the other with $N = 800$, $n = 72$ and $B_n = 32$, all chosen by the MV method. We have shown the simulated coverage probabilities of the SCR in Table 7. Next, we study the power of the hypothesis testing for stationarity using the

FIGURE 1. Simulated Power for testing stationarity for Example 7.6



proposed SCR by increasing δ. We study both 0.05 and 0.1 level confidence regions. The simulated power of the SCR is for the second example ($N = 800$) shown in Fig. 1. $\triangleleft$

Example 7.7. (Time-varying MA model) Consider the following model:

$$X_i = a_0(i/N)\epsilon_i + a_1(i/N)\epsilon_{i-1} \tag{54}$$

where we let $a_0(u) = 0.7 + 0.9\cos(2\pi u)$ and $a_1(u) = \delta a_0(u)$. Clearly, when $\delta = 0$, the model generates time-varying white noise. When $\delta = 0$, we study the accuracy of the hypothesis testing for time-varying white noise using the proposed SCR. The accuracy by the SCR is shown in Table 7, one with $N = 800$, $n = 300$, $B_n = 32$, and the other with $N = 1200$, $n = 420$, $B_n = 32$. We note that in order to test time-varying white noise, relatively larger N is required. We then test time-varying white noise using our proposed SCR by increasing δ for the first example ($N = 800$). The simulated power of the SCR is shown in Fig. 2. $\triangleleft$

According to the simulated power of the SCR shown in Fig. 1 and Fig. 2 for two examples Example 7.6 and Example 7.7, respectively, we can see the the proposed SCR can result in decent power in moderate sample cases.

FIGURE 2. Simulated Power for testing TV white noise for Example 7.7

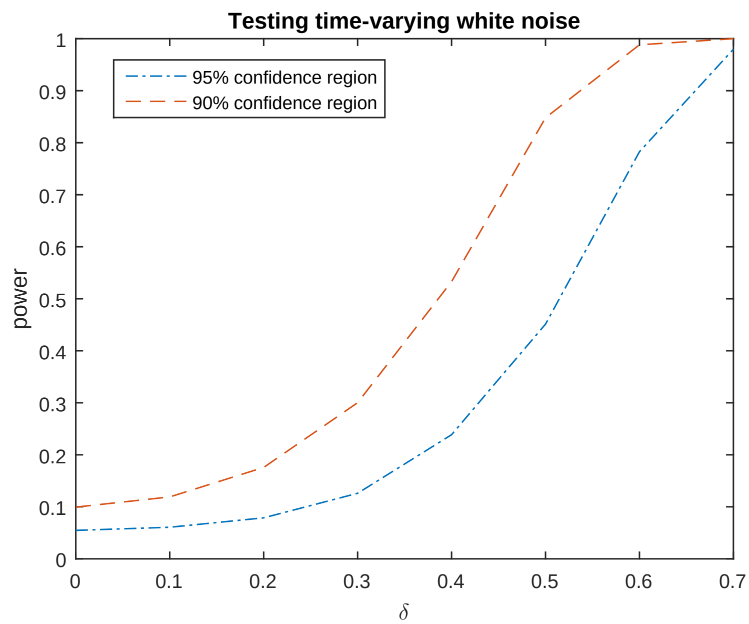


TABLE 7. Accuracy of Hypothesis Testing

Nominal Level		$\alpha = 0.05$	$\alpha = 0.1$		$\alpha = 0.05$	$\alpha = 0.1$
Example 7.6	$N = 400$	0.06	0.12	$N = 800$	0.04	0.09
Example 7.7	$N = 800$	0.05	0.10	$N = 1200$	0.04	0.09
Example 7.8	$N = 400$	0.04	0.09	$N = 800$	0.06	0.12

Example 7.8. (Validating time-varying AR model) Consider the following time-varying AR model

$$\sum_{j=0}^{p} a_j(i/N)X_{i-j} = \sigma(i/N)\epsilon_i, \tag{55}$$

where $a_0(u) = 1$, $a_j(\cdot)$ and $\sigma(\cdot)$ are smooth functions, ϵ_i are i.i.d. with mean 0 and variance 1. Then $\{X_i\}$ is a locally stationary time series with spectral density

$$f(u,\theta) = \frac{\sigma^2(u)}{2\pi}\left|\sum_{j=0}^{p} a_j(u)\exp(\sqrt{-1}2\pi\theta j)\right|^{-2}. \tag{56}$$

Table 8. Real Data: p-values for testing (a) stationarity, (b) time-varying white noise (TV White), (c) time-frequency separability (correlation stationarity).

H_0	Stationarity	TV White Noise	Separability
Earthquake	0.0011^{**}	0.012^{*}	0.064^{+}
Explosion	0.0005^{***}	0.033^{*}	0.61
SP500	0.0001^{***}	0.99	0.99
SP500 (Abs)	0.0004^{***}	0.037^{*}	0.048^{*}

Signif. codes: $(***) < 0.001 \le (**) < 0.01 \le (*) < 0.05 \le (+) < 0.1$.

In this example, we generating time series by $p = 1$ and $a_1(u) = 0.3+0.2u$, $\sigma(u) = 1+0.3u+0.2u^2$, with length $N = 400$ or $N = 800$.

For each generated time series, we fit time-varying AR model with $p = 1$ by minimizing local Whittle likelihood [Dah97], which lead to a non-parametric model, i.e. time-varying AR(1) model. We can then test if the spectral density of the fitted non-parametric time-varying AR model falls into the proposed SCR. The simulated coverage probabilities of the SCR are shown in Table 7 for two cases of $N = 400$ with $(n, B_n) = (56, 32)$ and $N = 800$ with $(n, B_n) = (72, 36)$, respectively. We can see that, under the null hypothesis, the non-parametric time-varying AR model is validated since the coverage probabilities obtained by the non-parametric time-varying AR model match quite well with the prescribed converge probabilities of the proposed SCR. ◁

7.4. **Real Data Analysis.** In this subsection, we present some real data analysis . We study an earthquake and explosion data set from seismology in Example 7.9 and then daily SP500 return from finance in Example 7.10. In all the real data sets considered in this subsection, we have relatively large sample $N > 2000$. For tuning parameter selection, we use the MV method to search (n, B_n) within the region $B_n < n/\log(n)$ which results in a relatively narrow confidence region. Hypothesis tests are performed, including testing stationarity, time-varying white noise, time-frequency separability, on all the data sets.

Example 7.9. (Earthquakes and explosions [SS17]) In this example, we study earthquake data and explosion data from [SS17]. The two time series (see Fig. 3 and Fig. 6) each has length $N = 2048$ representing two phases or arrivals along the surface, denote by phase P: $\{X_i : i = 1, \dots, 1024\}$ and phase S: $\{X_i : i = 1025, \dots, 2048\}$, at a seismic recording station. The recording instruments in Scandinavia are observing earthquakes (in Figs. 3 to 5) and mining explosions (in Figs. 6 to 8) with one of each shown in the figures. The general problem of interest is in distinguishing or discriminating between waveforms generated by earthquakes and those generated by explosions.

From the time domain (see Figs. 3 and 6), one can observe that rough amplitude ratios of the first phase P to the second phase S are different for the two data sets, which tend to be smaller for earthquakes than for explosions. From the spectral density estimates and their confidence regions, the S component for the earthquake (see Fig. 3) shows power at the low frequencies only, and the power remains strong for a long time. In contrast, the explosion (see Fig. 6) shows power at higher frequencies than the earthquake, and the power of the P and S waves does not last as long as in the case of the earthquake.

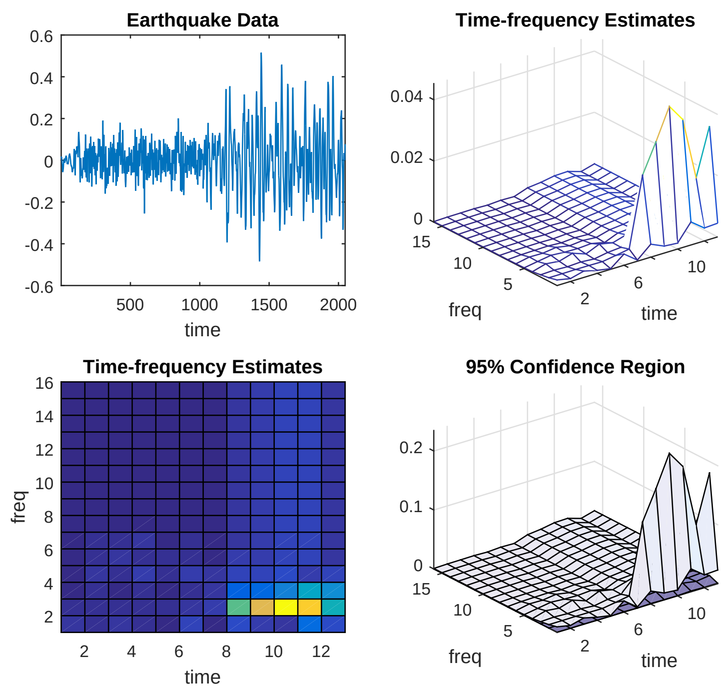
Moreover, we notice from the confidence region at selected time and frequencies that the spectral density of explosion has the similar shape at different time, as well as at different frequencies (see Figs. 7 and 8), however spectral density of earthquake does not seem to have this property (see Figs. 4 and 5). This may suggest that the explosion data are correlation stationary or time-frequency separable. We further perform hypothesis tests on both data sets to confirm our observation (see Table 8). The p-values for testing stationarity and time-varying white noise for both earthquake and explosion are quite small, which implies earthquake and explosion time series are not stationary and not time-varying white noise. However, the p-values for the hypothesis of time-frequency separability (i.e., correlation stationary) is 0.61 for explosion, but 0.064 for earthquake. This interesting result discovers a potential important difference between earthquake and explosion: at least from the analyzed data, explosion tends to be time-frequency separable (correlation stationary) but earthquake does not.

◁

Example 7.10. (SP500 daily returns)

In this example, we analyze daily returns of SP500 from September 23rd, 1991 to August 17th, 2018. We plot the original time series, the spectral density estimates and their confidence regions in Fig. 9. Observing that the SCR in Fig. 9 appears to be quite flat over frequencies, it is reasonable to ask if the time series may be modeled as time-varying white noise. Actually, in the finance literature, it is commonly believed that stock daily returns behave like time-varying white noise. We further confirm this observation by performing hypothesis tests. The results (see Table 8) show that the SP500 time series is not stationary

FIGURE 3. Analysis of Earthquake Data



but it is likely to be a time-varying white noise since the p-value for testing time-varying white noise is 0.99. Furthermore, the p-value for testing time-frequency separability is also quite large which is 0.99.

Next, we turn our focus to the absolute value of SP500 daily returns. Volatility forecasting, i.e. forecasting future absolute values or squared values of the return, is a key problem in finance. The celebrated ARCH/GARCH models are equivalent to exponential smoothings of the absolute or squared returns. The optimal weights in the smoothing are determined fully by the evolutionary spectral density. Hence, to optimally forecast the evolutionary volatility, one needs to fit the absolute returns by an appropriate non-stationary linear model, then apply the fitted model to forecast the future volatility. To date, to our knowledge, there is no methodology for validating non-stationary linear models. In the following, we demonstrate that the proposed SCR is a useful tool for validating non-stationary models for fitting absolute SP500 daily returns.

FIGURE 4. Earthquake Data: selected time

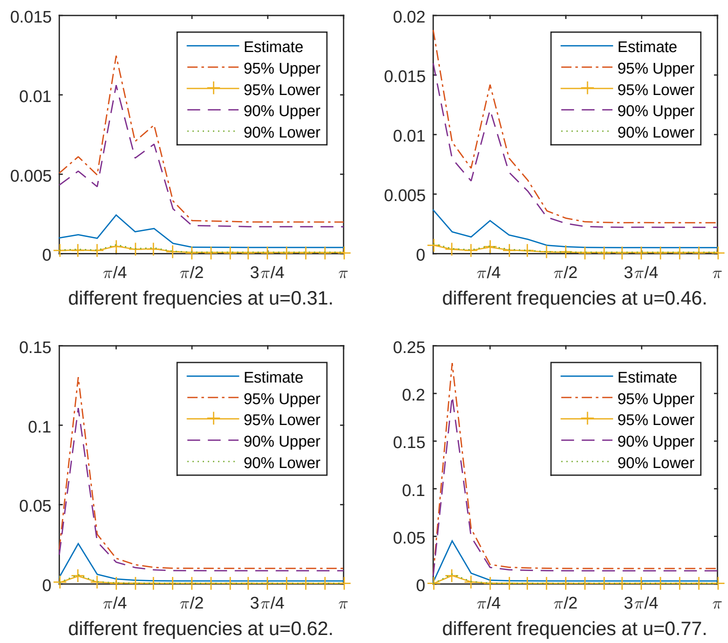


FIGURE 5. Earthquake Data: selected frequencies

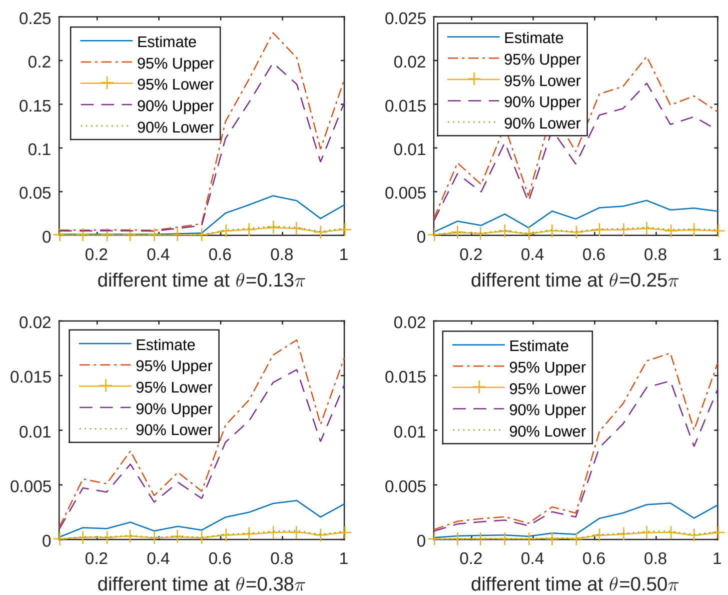
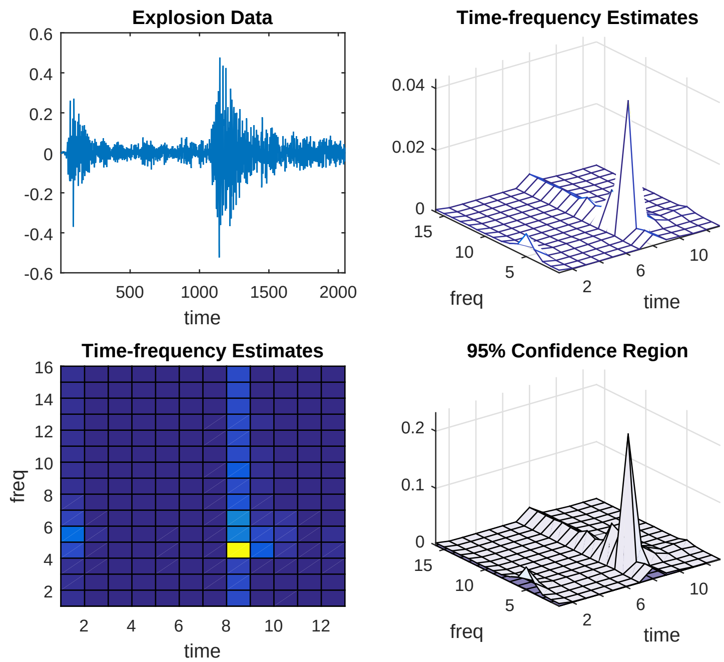


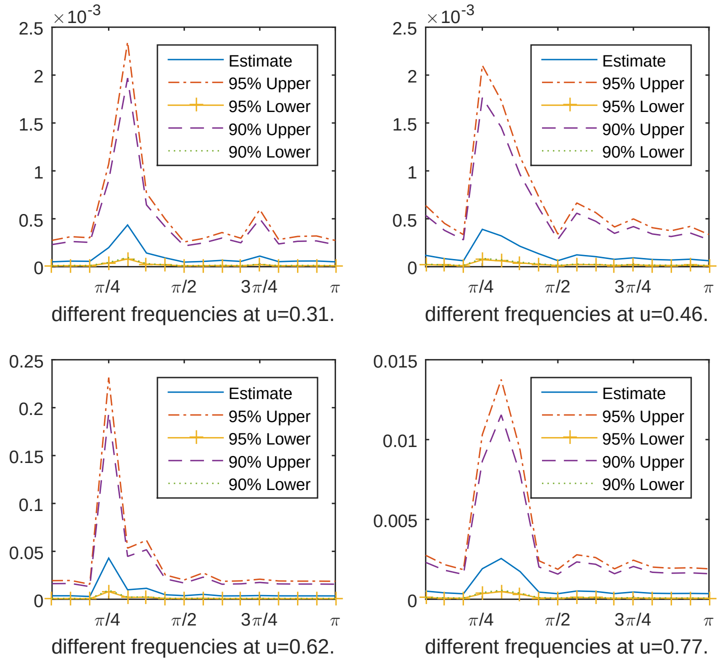
FIGURE 6. Analysis of Explosion Data



We first remove the local mean of the original SP500 time series by kernel smoothing (see Fig. 10). The spectral density estimates and the SCRs are shown in Figs. 11 to 13. We observe from the plots that the spectral density of the absolute SP500 returns behaves quite different from the original SP500 time series. For example, unlike the case for original SP500 time series, the SCR for absolute SP500 in Fig. 12 is not flat over frequencies anymore. We perform the same hypothesis tests again to the absolute SP500 time series. The results (see Table 8) show that the p-values for testing time-varying white noise is 0.037, which is significantly reduced comparing with the case for original SP500 time series (with a p-value of 0.99). Furthermore, the p-value for testing time-frequency separability is 0.048 which is also much smaller than the one for the original SP500 data (with a p-value of 0.99).

Finally, we fit time-varying non-stationary linear models for the absolute SP500 daily returns with mean removed by kernel smoothing. We first fit time-varying AR or ARMA

FIGURE 7. Explosion Data: selected time



models

$$\sum_{i=0}^{p} a_i(t/N)X_{t-i} = \sum_{j=0}^{q} b_j(t/N)\epsilon_{t-j} \tag{57}$$

by minimizing local Whittle likelihood [Dah97]. We then validate if the spectral density from the local parametric AR model falls into the proposed SCR. The p-values for validating time-varying AR/ARMA models are shown in Table 9. One can see that, the p-values for the tv-AR models are quite small, which implies that no tv-AR models up to order 5 is appropriate for fitting absolute SP500 daily returns. For tv-ARMA models, the p-value for tv-ARMA$(1,1)$ equals to 0.019. This suggests that this tv-ARMA model is not appropriate for fitting the absolute SP500 daily returns either. In contrast, the corresponding p-value for validating tv-ARMA$(2,1)$ is 0.79. This interesting observation suggests that tv-ARMA$(2,1)$ may be an appropriate non-parametric model to fit the absolute returns. We further plot the theoretical spectral densities of the fitted time-varying AR(1), AR(4), AR(5), ARMA$(1,1)$, ARMA$(2,1)$, and ARMA$(3,1)$ models in Fig. 14. From Fig. 14, one can see that the spectral

FIGURE 8. Explosion Data: selected frequencies

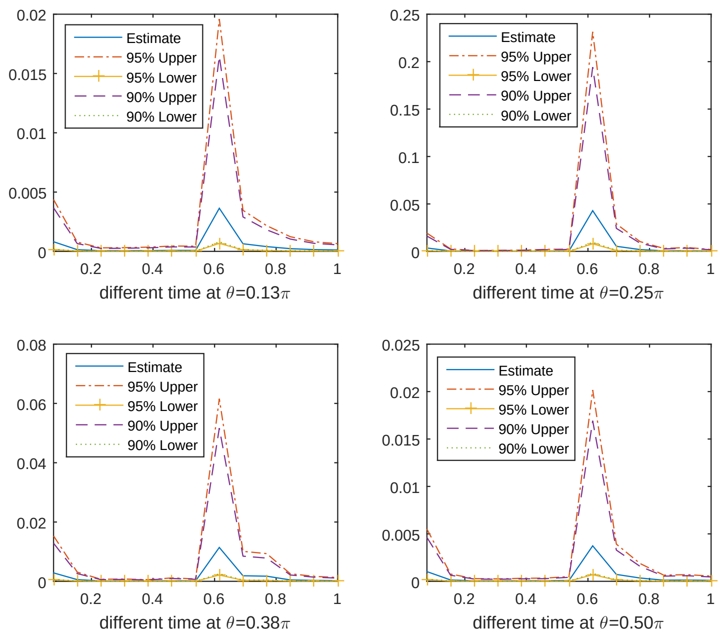


TABLE 9. p-values for fitting time-varying parametric models to absolute SP500

Model	p-value	Model	p-value
tv-AR(1)	0.0066**	tv-ARMA(1, 1)	0.019*
tv-AR(2)	0.0015**	tv-ARMA(2, 1)	0.79
tv-AR(3)	0.0015**	tv-ARMA(3, 1)	0.77
tv-AR(4)	0.0012**	tv-ARMA(4, 1)	0.78
tv-AR(5)	0.0012**	tv-ARMA(5, 1)	0.84

Signif. codes: $(***) < 0.001 \leq (**) < 0.01 \leq (*) < 0.05 \leq (+) < 0.1$.

densities by the tv-AR models are quite different from the STFT-based spectral density estimates. For tv-ARMA models, the spectral density estimates by tv-ARMA$(1, 1)$ are not close to the STFT-based spectral density estimates either. Therefore, based on the proposed SCR, we can conclude the tv-ARMA$(2, 1)$ is an appropriate non-parametric model for the analyzed data and is suggested for short-term future volatility forecasting.

◁

FIGURE 9. Analysis of Daily Returns of SP500

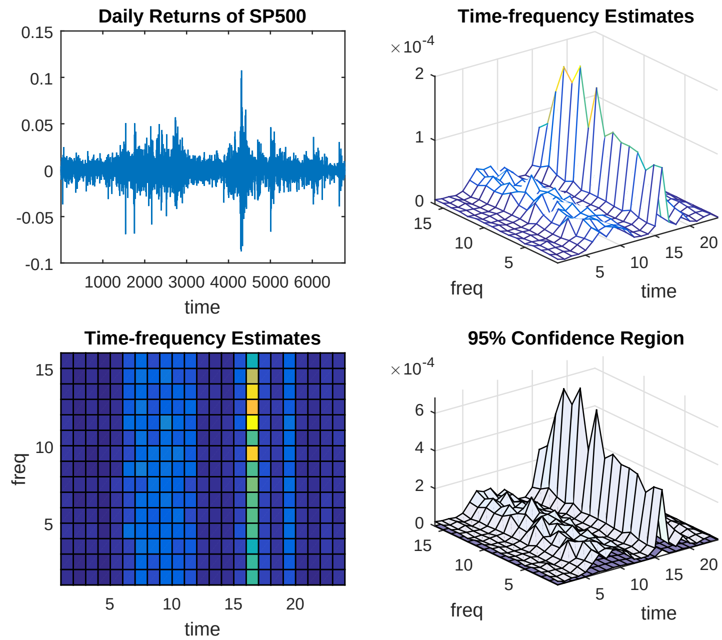


FIGURE 10. Kernel Smoothing for SP500 Absolute Daily Returns

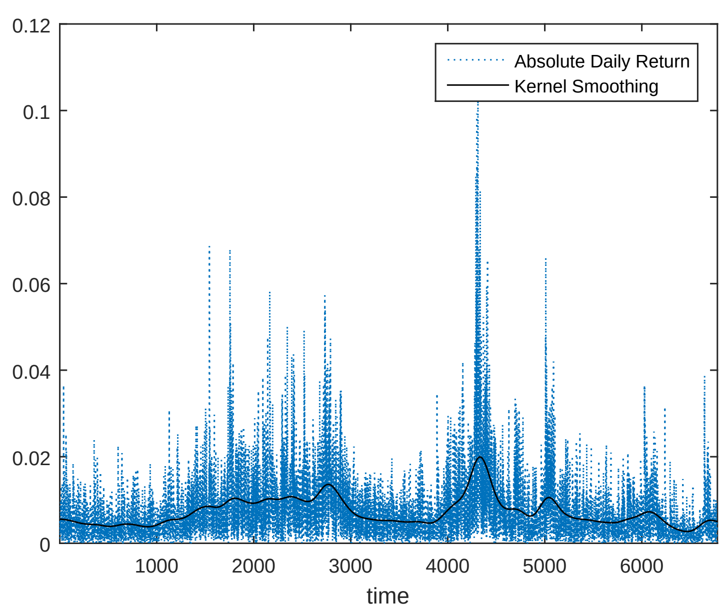
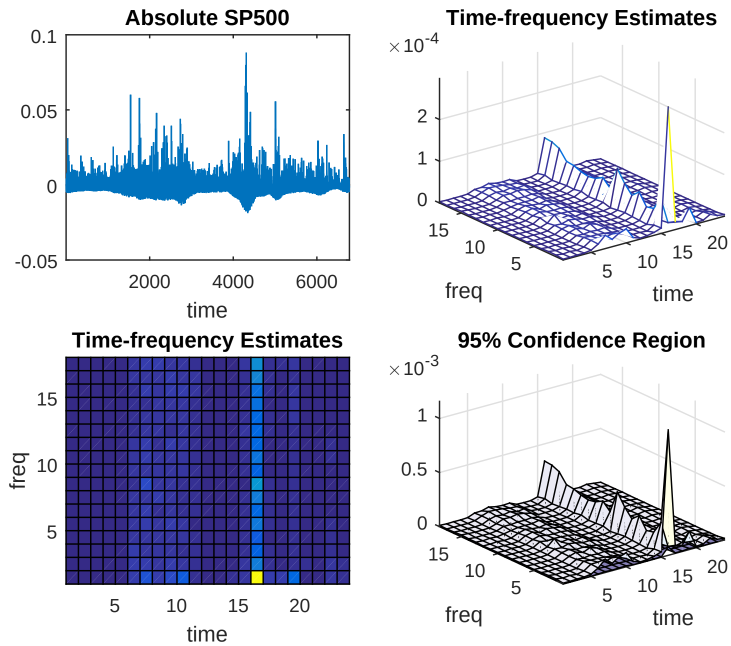


FIGURE 11. Analysis of Absolute SP500 return



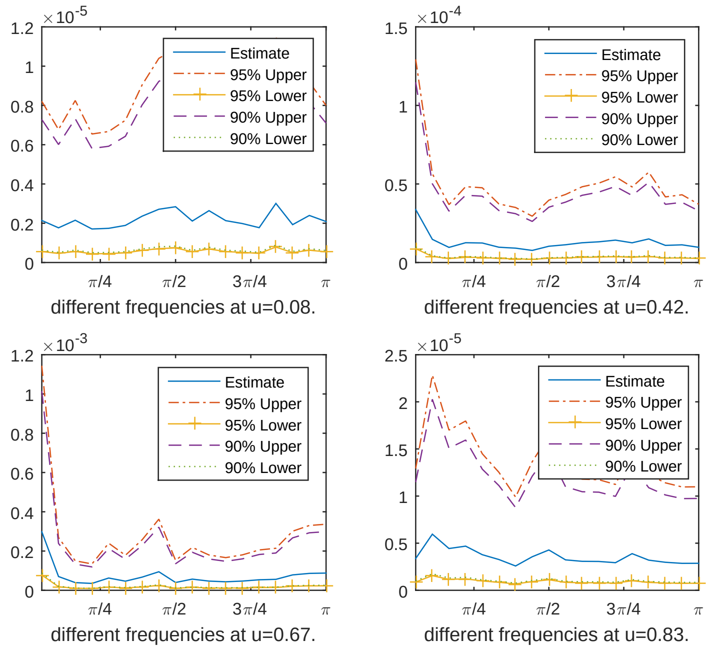
8. Proofs of Main Results

8.1. **Proof of Theorem 3.1.** We prove Theorem 3.1 in two steps. In the first step, we show in Section 8.1.1 that Theorem 3.1 is true for $q = 1$. In this case, we let $\Omega_p = \{c \in \mathbb{R}^p : |c| = 1\}$, $Z_{u,J} = (Z^{(n)}_{u,j_1}, \ldots, Z^{(n)}_{u,j_p})^T$ for $J = (j_1, \ldots, j_p)$ satisfies $1 \le j_1, \ldots, j_p \le 2m$ (recall that $m = \lfloor (n-1)/2 \rfloor$). We prove for any fixed $p \in \mathbb{N}$, as $n \to \infty$, we have

$$\sup_u \sup_J \sup_{c \in \Omega_p} \sup_x |P(c^T Z_{u,J} \le x) - \Phi(x)| = o(1). \tag{58}$$

In the second step of the proof, we show in Section 8.1.2 that for fixed $q \in \mathbb{N}$, for any given $0 < u_1 < \cdots < u_q < 1$, we have $\{(c^{(i)})^T Z_{u_i,J}, i = 1, \ldots, q\}$ are asymptotically independent uniformly over $\{c^{(i)} \in \mathbb{R}^p : |c^{(i)}| = 1\}$ for $i = 1, \ldots, q$. Finally, Theorem 3.1 is proved by combining the two parts.

FIGURE 12. Absolute SP500 return: selected time



8.1.1. *Proof of Eq.* (58). We denote $2\pi j/n$ by θ_j in this proof. Define $\mathcal{P}_k$ to be the projection operator

$$\mathcal{P}_k(X) := \mathbb{E}(X \mid \mathcal{F}_k) - \mathbb{E}(X \mid \mathcal{F}_{k-1}), \tag{59}$$

and $\tilde{X}_k = \mathbb{E}(X_k \mid \epsilon_{k-\ell+1}, \ldots, \epsilon_k)$ are ℓ-dependent conditional expectations of X_k.

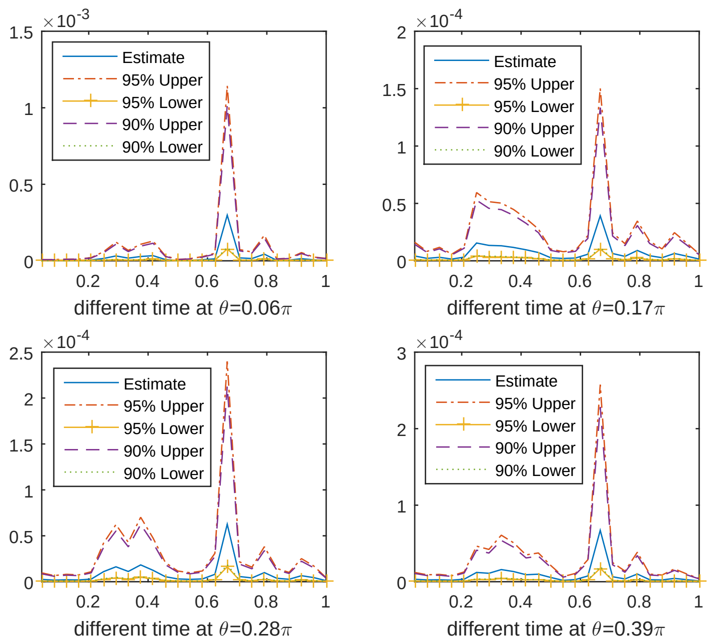
Throughout the proof, we use $\|\cdot\|$ to denote $\|\cdot\|_2$ for simplicity. From the GMC condition, one can easily verify that

$$\sup_k \sum_{j=-\infty}^{k} \|\mathcal{P}_j X_k\| < \infty, \quad \lim_{\ell\to\infty} \sup_k \|X_k - \tilde{X}_k\| = 0. \tag{60}$$

With out loss of generality, we restrict $J = \{j_1, \ldots, j_p\} \in \{1, \ldots, m\}$. Let $c = (c_1, \ldots, c_p)$, define $\mu_{u,k} := \sum_{\ell=1}^{p} \frac{c_\ell \cos(k\theta_{j_\ell})}{\sqrt{\pi f(u, \theta_{j_l})}}$. Then

$$\mu_{u,k} \leq \sum_{\ell=1}^{p} \frac{|c_\ell|}{\sqrt{\pi f_*}} \leq \frac{p}{\sqrt{\pi f_*}} =: \mu_*, \quad \forall c \in \Omega_p, \quad \forall J. \tag{61}$$

FIGURE 13. Absolute SP500 return: selected frequencies



Furthermore, defining

$$T_{u,n} := \sum_{k=1}^{N} \mu_{u,k} \tau \left(\frac{k - \lfloor uN \rfloor}{n} \right) X_k, \quad \tilde{T}_{u,n} := \sum_{k=1}^{N} \mu_{u,k} \tau \left(\frac{k - \lfloor uN \rfloor}{n} \right) \tilde{X}_k, \tag{62}$$

and $\eta := \left(\frac{\|T_{u,n} - \tilde{T}_{u,n}\|}{\sqrt{n}} \right)^{1/2}$, we have the following key lemmas.

Lemma 8.1.

$$\lim_{n \to \infty} \sup_J \sup_c \sup_u \left| \frac{\|T_{u,n}\|^2}{n} - 1 \right|^2 = 0. \tag{63}$$

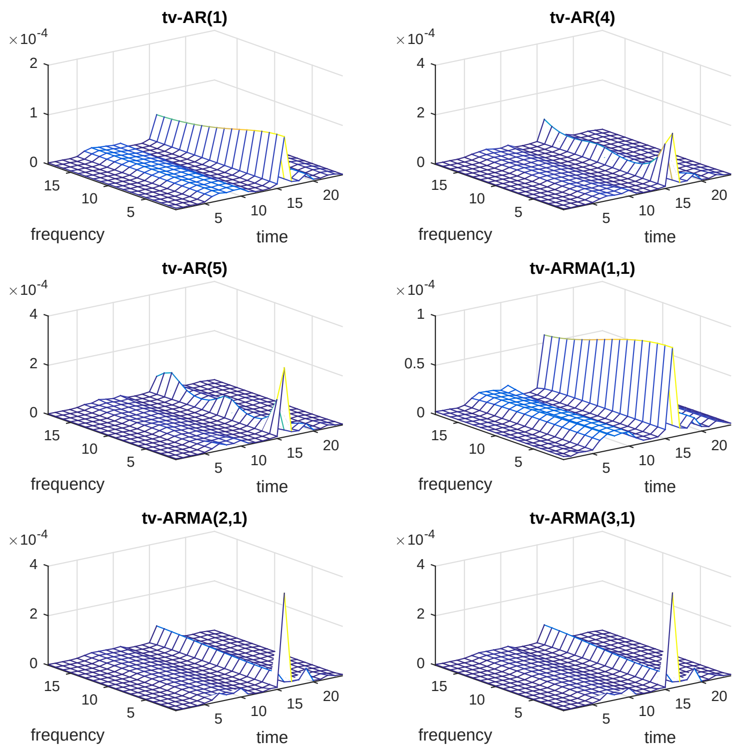
Proof. See Appendix A.1. □

Lemma 8.2.

$$\lim_{\ell \to \infty} \sup_J \sup_c \sup_u \frac{\|T_{u,n} - \tilde{T}_{u,n}\|}{\sqrt{n}} = 0. \tag{64}$$

Proof. See Appendix A.2. □

FIGURE 14. Fitting absolute SP500 daily returns to time-varying ARMA model



Lemma 8.3.

$$\sup_x \left| \mathbb{P}\left(\frac{T_{u,n}}{\sqrt{n}} \le x\right) - \Phi\left(\frac{x}{\|T_{u,n}\|/\sqrt{n}}\right) \right| = \mathcal{O}\left(\mathbb{P}\left(\left|\frac{T_{u,n}-\tilde{T}_{u,n}}{\sqrt{n}}\right| \ge \eta\right) + \delta_n + \eta^2\right), \tag{65}$$

where $\delta_n \to 0$ as $n \to \infty$ uniformly over J, c and u.

Proof. See Appendix A.3. □

Using the above results, we can then prove Eq. (58) as follows. First, by Lemma 8.2 and Chebyshev inequality, we have

$$\mathbb{P}\left(\left|\frac{T_{u,n}-\tilde{T}_{u,n}}{\sqrt{n}}\right| \ge \eta\right) \le \frac{\mathbb{E}(T_{u,n}-\tilde{T}_{u,n})^2/n}{\eta^2} = \eta^2. \tag{66}$$

Next, according to Lemma 8.1, uniformly over J, c and u, for any fixed ℓ, as $n \to \infty$, we have

$$\sup_x \left| \mathbb{P}\left(\frac{T_{u,n}}{\sqrt{n}} \le x \right) - \Phi\left(\frac{x}{\|T_{u,n}\|/n} \right) \right| \to \sup_x \left| \mathbb{P}\left(\frac{T_{u,n}}{\sqrt{n}} \le x \right) - \Phi\left(x \right) \right|. \tag{67}$$

By Lemma 8.3, we have

$$\sup_x \left| \mathbb{P}\left(\frac{T_{u,n}}{\sqrt{n}} \le x \right) - \Phi\left(\frac{x}{\|T_{u,n}\|/n} \right) \right| = \mathcal{O}(2\eta^2 + \delta_n). \tag{68}$$

Note that $\delta_n \to 0$ as $n \to \infty$. Also, uniformly over J, c, u, n we have $\eta \to 0$ as $\ell \to \infty$. Finally, letting $n \to \infty$ then $\ell \to \infty$, we have

$$\sup_x \left| \mathbb{P}\left(\frac{T_{u,n}}{\sqrt{n}} \le x \right) - \Phi\left(x \right) \right| \to 0, \tag{69}$$

uniformly over J, c, and u.

8.1.2. *Proof of asymptotically independence of* $\{(c^{(i)})^T Z_{u_i,J}, i = 1, \ldots, q\}$. We can write $T_{u_i,n}$ and $\tilde{T}_{u_i,n}$ defined in Eq. (62) as $T_{u_i,n,c^{(i)}}$ and $\tilde{T}_{u_i,n,c^{(i)}}$. Then by Lemma 8.2, it suffices to show $\{\tilde{T}_{u_i,n,c^{(i)}}, i = 1, \ldots, q\}$ are asymptotically independent uniformly over $\{c^{(i)} \in \mathbb{R}^p : |c^{(i)}| = 1\}$. Note that in the definition of $\tilde{T}_{u_i,n,c^{(i)}}$, $\tilde{X}_k$ is ℓ-dependent, therefore, $\tilde{T}_{u_1,n,c^{(i)}}$ and $\tilde{T}_{u_2,n,c^{(i)}}$ with $u_2 > u_1$ are independent if $\lfloor (u_2 - u_1)N \rfloor > \ell + 2n$. Since $0 < u_1 < \cdots < u_q < 1$ are fixed, $\min_{i \neq j} |u_i - u_j| > 0$ is bounded away from zero. Therefore, $\{\tilde{T}_{u_i,n,c^{(i)}}, i = 1, \ldots, q\}$ are independent if $\ell < \lfloor (\min_{i \neq j} |u_i - u_j|)N \rfloor - 2n$. Choosing $\ell = o(n)$ and $n = o(N)$, we have $\{\tilde{T}_{u_i,n,c^{(i)}}, i = 1, \ldots, q\}$ are asymptotically independent.

8.2. **Proof of Theorem 4.1.** Throughout the proof, we use $\|\cdot\|$ to denote $\|\cdot\|_2$ for simplicity. We define $X_{u,i,n} := \tau\left(\frac{i - \lfloor n/2 \rfloor}{n}\right) X_{\lfloor uN \rfloor + i - \lfloor n/2 \rfloor}$. For simplicity we will omit the index n and use $X_{u,i}$ for $X_{u,i,n}$. Define

$$Y_{u,i} := Y_{u,i}(\theta) = \frac{1}{2\pi} \sum_{k=-B_n}^{B_n} X_{u,i} X_{u,i+k} a(k/B_n) \cos(k\theta). \tag{70}$$

Then defining

$$g_n(u,\theta) := \sum_{i=1}^{n} Y_{u,i}(\theta), \quad h_n(u,\theta) := \frac{1}{\sqrt{nB_n}} g_n(u,\theta) - \sqrt{nb_n} \hat{f}_n(u,\theta), \tag{71}$$

we have

$$\begin{aligned}&\sqrt{n/B_n}\{\hat{f}_n(u,\theta)-\mathbb{E}(\hat{f}_n(u,\theta))\}\\ &=\frac{g_n(u,\theta)-\mathbb{E}(g_n(u,\theta))}{\sqrt{nB_n}}-h_n(u,\theta)+\mathbb{E}(h_n(u,\theta)).\end{aligned}\tag{72}$$

Next, denote $\tilde{X}_k$ as the ℓ-dependent conditional expectation of X_k, $\tilde{X}_{u,i}$ as the ℓ-dependent conditional expectation of $X_{u,i}$, and $\tilde{Y}_{u,i}$ as the correspondence of sum using $\tilde{X}_{u,i}$ instead of $X_{u,i}$, and $\tilde{g}_n$ as the correspondence of g_n using $\tilde{Y}_{u,i}$ instead of $Y_{u,i}$. Note that under GMC(2) and $\sup_i \mathbb{E}|X_i|^{4+\delta}<\infty$, we know GMC(4) holds. Then we have the following results.

Lemma 8.4. *Under the assumptions of Theorem 4.1,* GMC(4) *holds with* $0<\rho<1$*, we have*

$$\sup_u \|h_n(u,\theta)\| = (nB_n)^{-1/2}\mathcal{O}(B_n),\tag{73}$$

$$\sup_u \sup_i \|Y_{u,i}-\tilde{Y}_{u,i}\| = \mathcal{O}(B_n\rho^{\ell/4}),\tag{74}$$

$$\sup_u \|g_n(u,\theta)-\tilde{g}_n(u,\theta)\| = o(1).\tag{75}$$

Proof. See Appendix A.4. □

Next, we apply the block method to $\{\tilde{Y}_{u,i}(\theta)\}$. Define

$$U_{u,r}(\theta) := \sum_{i=(r-1)(p_n+q_n)+1}^{(r-1)(p_n+q_n)+p_n}\tilde{Y}_{u,i}(\theta),\quad V_{u,r}(\theta) := \sum_{i=(r-1)(p_n+q_n)+p_n+1}^{r(p_n+q_n)}\tilde{Y}_{u,i}(\theta),\quad 1,\dots,k_n,\tag{76}$$

where $k_n := \lfloor n/(p_n+q_n)\rfloor$. Let $p_n = q_n = \lfloor n^{1-4\eta/\delta}(\log n)^{-8/\delta-4}\rfloor$ (i.e. same block length) and $\ell = \ell_n = \lfloor -9\log n/\log\rho\rfloor$ (Note $B_n = o(p_n)$ since $\eta < \delta/(4+\delta)$). Then $U_{u,r}(\theta), r = 1,\dots,k_n$ are independent (not identical distributed) block sums with block length p_n, and $V_{u,r}(\theta), r=1,\dots,k_n$ are independent block sums with block length q_n. Define $U'_{u,r}(\theta) := U_{u,r}(\theta)\mathbf{1}(|U_{u,r}(\theta)|\le d_n)$ where $d_n = \lfloor\sqrt{nB_n}(\log n)^{-1/2}\rfloor$. Then we have the following results.

Lemma 8.5. *Under the assumptions of Theorem 4.1, we have*

$$\sup_u \mathbb{E}(\max_\theta |V_{u,k_n}(\theta)|) = \mathcal{O}(\sqrt{p_n\ell_n}B_n),\tag{77}$$

$$\sup_u \mathbb{E}(\max_\theta |h_n(u,\theta)|) = o(1), \tag{78}$$

$$\sup_u \max_r \max_\theta \operatorname{var}(U_{u,r}(\theta)) = \mathcal{O}(p_n B_n), \tag{79}$$

$$\operatorname{var}(U'_{u,r}(\theta)) = \operatorname{var}(U_{u,r}(\theta))[1+o(1)], \tag{80}$$

where the $o(1)$ in the last equation holds uniformly over θ, r and u.

Proof. See Appendix A.5. □

Lemma 8.6. *Let $U_{u,i}(\theta)$ be one of the block sums with block length p_n. Then we have*

$$\sup_u \sup_i \sup_\theta \|U_{u,i}(\theta)\|_{2+\delta/2} = \mathcal{O}(\ell_n \sqrt{p_n B_n}) \tag{81}$$

Proof. See Appendix A.6. □

Using the previous results Eqs. (73), (75) and (78), we have

$$\begin{aligned}
&\sup_u \max_\theta \sqrt{n/B_n} |\hat{f}_n(u,\theta) - \mathbb{E}(\hat{f}_n(u,\theta))| \\
&\le \frac{\sup_u \max_\theta |\tilde{g}_n(u,\theta) - \mathbb{E}(\tilde{g}_n(u,\theta))| + o(1)}{\sqrt{nB_n}} + \mathcal{O}_{\mathbb{P}}(\sqrt{B_n/n}) + o_{\mathbb{P}}(1) \\
&\le \frac{\sup_u \max_\theta |\sum_{r=1}^{k_n} U_{u,r}(\theta) - \mathbb{E}(\sum_{r=1}^{k_n} U_{u,r}(\theta))|}{\sqrt{nB_n}} \\
&\quad + \frac{\sup_u \max_\theta |\sum_{r=1}^{k_n-1} V_{u,r}(\theta) - \mathbb{E}(\sum_{r=1}^{k_n-1} V_{u,r}(\theta))|}{\sqrt{nB_n}} \\
&\quad + \frac{\sup_u \max_\theta |V_{u,k_n}(\theta)) - \mathbb{E}(V_{u,k_n}(\theta)))|}{\sqrt{nB_n}} + \mathcal{O}_{\mathbb{P}}(\sqrt{B_n/n}) + o_{\mathbb{P}}(1).
\end{aligned} \tag{82}$$

Next, we show the first two terms have a order of $\mathcal{O}_{\mathbb{P}}(\sqrt{\log n})$. We can show the third term is $o_{\mathbb{P}}(\sqrt{\log n})$ using Eq. (77) with similar techniques.

Let $H_{u,n}(\theta) = \sum_{r=1}^{k_n}[U_{u,r}(\theta) - \mathbb{E}(U_{u,r}(\theta))]$ and $H'_{u,n}(\theta) = \sum_{r=1}^{k_n}[U'_{u,r}(\theta) - \mathbb{E}(U'_{u,r}(\theta))]$. Let $\theta_j = \pi j/t_n, j = 0, \ldots, t_n$ where $t_n = \lfloor B_n \log(B_n) \rfloor$. Then, since both $H_{u,n}$ and $H'_{u,n}$ have trigonometric polynomial forms, we can apply the following result from [WN67, Corollary 2.1].

Lemma 8.7. *Let* $p(\lambda) = \sum_{v=-k}^{k} \alpha_v \exp(iv\lambda)$ *be a trigonometric polynomial. Let* $\lambda_i = \pi(i/rk), |i| \le rk$*. then*

$$\max_{|\lambda|\le\pi} |p(\lambda)| \le \max_{|i|\le rk} |p(\lambda_i)/(1 - 3\pi r^{-1})|. \tag{83}$$

Proof. See [WN67, Corollary 2.1]. □

By setting $k = B_n$ and $r = \log(B_n)$ in Lemma 8.7, we get

$$\max_{\theta} |H_{u,n}(\theta)| \le \frac{1}{1 - 3\pi/\log(B_n)} \max_{j\le t_n} |H_{u,n}(\theta_j)|. \tag{84}$$

By Eqs. (79) and (80), there exists a constant C_1 such that

$$\sup_u \max_r \max_\theta \operatorname{var}(U'_{u,r}(\theta)) \le C_1 p_n B_n.$$

Let $\alpha_n := (C_1 n B_n \log n)^{1/2}$, by the union upper bound,

$$\mathbb{P}(\max_{0\le j\le t_n} |H'_{u,n}(\theta_j)| \ge 4\alpha_n) \le \sum_{j=0}^{t_n} \mathbb{P}(|H'_{u,n}(\theta_j)| \ge 4\alpha_n). \tag{85}$$

Then we apply Bernstein's inequality (see Lemma A.2) to $\mathbb{P}(|H'_{u,n}(\theta_j)| \ge 4\alpha_n)$. This leads to, uniformly over u and θ_j,

$$\begin{aligned} \mathbb{P}(|H'_{u,n}(\theta_j)| \ge 4\alpha_n) &\le \exp\left(\frac{-16\alpha_n^2}{2k_n C_1 p_n B_n + \frac{8}{3} d_n \alpha_n}\right) \\ &\le C \exp\left(-\frac{nB_n \log n}{nB_n}\right). \end{aligned} \tag{86}$$

Therefore, uniformly over u, we have

$$\mathbb{P}(\max_{0\le j\le t_n} |H'_{u,n}(\theta_j)| \ge 4\alpha_n) \le \mathcal{O}(t_n)\mathcal{O}(1/n) = o(1). \tag{87}$$

Let $U^*_{u,n}(\theta) = U_{u,n}(\theta) - U'_{u,n}(\theta)$ and $H^*_{u,n}(\theta) = H_{u,n}(\theta) - H'_{u,n}(\theta)$. By the union upper bound and Chebyshev's inequality

$$\begin{aligned} \mathbb{P}(\max_{0\le j\le t_n} |H^*_{u,n}(\theta_j)| \ge 4\alpha_n) &\le \sum_{j=0}^{t_n} \mathbb{P}(|H^*_{u,n}(\theta_j)| \ge 4\alpha_n) \\ &\le \sum_{j=0}^{t_n} \frac{\sum_{i=1}^{k_n} \operatorname{var}(U^*_{u,i}(\theta_j))}{16\alpha_n^2}. \end{aligned} \tag{88}$$

Using Lemma 8.6, $\sup_u \max_i \sup_\theta \|U_{u,i}(\theta)\|_{2+\delta/2} = \mathcal{O}(\ell_n\sqrt{p_nB_n})$, and the following inequality

$$\operatorname{var}(U^*_{u,i}\mathbf{1}_{|U^*_{u,i}|>d_n}) = d_n^2\operatorname{var}\left(\frac{U^*_{u,i}}{d_n}\mathbf{1}_{|U^*_{u,i}|>d_n}\right) \le d_n^2\mathbb{E}\left[\left(\frac{U^*_{u,i}}{d_n}\right)^{2+\delta/2}\right], \tag{89}$$

we have

$$\begin{aligned}
\sum_{j=0}^{t_n}\frac{\sum_{i=1}^{k_n}\operatorname{var}(U^*_{u,i}(\theta_j))}{16\alpha_n^2} &= \mathcal{O}\left(\frac{t_nk_n(\sqrt{p_nB_n}\ell_n)^{2+\delta/2}}{\alpha_n^2d_n^{\delta/2}}\right)\\
&= \mathcal{O}\left(\frac{(B_n\log B_n)(n/p_n)(\sqrt{p_nB_n}\log n)^{2+\delta/2}}{(nB_n\log n)(nB_n)^{\delta/4}(\log n)^{-\delta/4}}\right)\\
&= \mathcal{O}\left(\frac{(p_nB_n)^{1+\delta/4}(\log n)^{2+\delta/2}}{p_n(nB_n)^{\delta/4}(\log n)^{-\delta/4}}\right)\\
&= \mathcal{O}(p_n^{\delta/4}(B_n/n)^{\delta/4}(\log n)^{2+\delta/2+\delta/4}).
\end{aligned} \tag{90}$$

Using $p_n = n^{1-4\eta/\delta}(\log n)^{-8/\delta-4}$ we have $p_n^{\delta/4} = (n^{\delta/4-\eta})(\log n)^{-2-\delta}$. Therefore,

$$\begin{aligned}
\sum_{j=0}^{t_n}\frac{\sum_{i=1}^{k_n}\operatorname{var}(U^*_{u,i}(\theta_j))}{16\alpha_n^2} &= \mathcal{O}\left(\frac{t_nk_n(\sqrt{p_nB_n}\ell_n)^{2+\delta/2}}{\alpha_n^2d_n^{\delta/2}}\right)\\
&= \mathcal{O}(n^{-\eta}B_n^{\delta/4}(\log n)^{-\delta/4}).
\end{aligned} \tag{91}$$

Finally, $B_n = \mathcal{O}(n^\eta), \delta \le 4$ implies $B_n^{\delta/4} = \mathcal{O}(n^\eta)$, so we have

$$\sum_{j=0}^{t_n}\frac{\sum_{i=1}^{k_n}\operatorname{var}(U^*_{u,i}(\theta_j))}{16\alpha_n^2} = o(1). \tag{92}$$

Therefore, uniformly over u, we have $\max_\theta |H'_{u,n}(\theta)| = \mathcal{O}_{\mathbb{P}}(\alpha_n)$ and $\max_\theta |H^*_{u,n}(\theta)| = \mathcal{O}_{\mathbb{P}}(\alpha_n)$. Then

$$\max_\theta |H_{u,n}(\theta)| = \max_\theta |H'_{u,n}(\theta) + H^*_{u,n}(\theta)| = \mathcal{O}_{\mathbb{P}}(\alpha_n) = \mathcal{O}_{\mathbb{P}}(\sqrt{nB_n\log n}). \tag{93}$$

So Eq. (82) has the order of $\mathcal{O}_{\mathbb{P}}(\sqrt{\log n})$.

8.3. **Proof of Theorem 4.3.** Throughout the proof, we use $\|\cdot\|$ to denote $\|\cdot\|_2$ for simplicity. We define $Y_{u,i}$, g_n, h_n, $\tilde{X}_k$, $\tilde{Y}_{u,i}$, $\tilde{g}_n$ the same as in Section 8.2. Therefore, Lemma 8.4 holds.

Next, we apply the block method to $\{\tilde{Y}_{u,i}(\theta)\}$. Define

$$U_{u,r}(\theta) := \sum_{i=(r-1)(p_n+q_n)+1}^{(r-1)(p_n+q_n)+p_n} \tilde{Y}_{u,i}(\theta), \quad V_{u,r}(\theta) := \sum_{i=(r-1)(p_n+q_n)+p_n+1}^{r(p_n+q_n)} \tilde{Y}_{u,i}(\theta), \quad 1, \ldots, k_n, \tag{94}$$

where $k_n := \lfloor n/(p_n+q_n) \rfloor$. Let $\psi_n = n/(\log n)^{2+8/\delta}$, $p_n = \lfloor \psi_n^{2/3} B_n^{1/3} \rfloor$, and $q_n = \lfloor \psi_n^{1/3} B_n^{2/3} \rfloor$. Then we have $p_n, q_n \to \infty$ and $q_n = o(p_n)$. Since $\ell_n = \mathcal{O}(\log n)$, we have $2B_n + \ell_n = o(q_n)$ and $k_n = \lfloor n/(p_n+q_n) \rfloor \to \infty$. Note that $U_{u,r}(\theta), r = 1, \ldots, k_n$ are independent (not identical distributed) block sums with block length p_n, and $V_{u,r}(\theta), r = 1, \ldots, k_n$ are independent block sums with block length q_n. Now the proof of Lemma 8.5 still follows.

Defining $a_n/b_n \to 1$ by $a_n \sim b_n$, we have the following result.

Lemma 8.8. *Let the sequence $s_n \in \mathbb{N}$ satisfy $s_n \leq n$, $s_n = o(n)$ and $B_n = o(s_n)$. Under GMC(4) we have*

$$\left\| \sum_{i=-s_n/2}^{s_n/2} \{Y_{u,i}(\theta) - \mathbb{E}(Y_{u,i}(\theta))\} \right\|^2 \sim s_n B_n \sigma_u^2(\theta), \tag{95}$$

where $\sigma_u^2(\theta) = [1+\eta(2\theta)] f^2(u,\theta) \int_{-1}^{1} a^2(t)\mathrm{d}t$ and $\eta(\theta) = 1$ if $\theta = 2k\pi$ for some integer k and $\eta(\theta) = 0$ otherwise.

Proof. See Appendix A.7. □

According to Lemmas 8.4 and 8.8, for each block $U_{u,r}, r = 1, \ldots, k_n$, we have

$$\begin{aligned} \|U_{u,r} - \mathbb{E}(U_{u,r})\| &= \left\| \sum_{j \in \mathcal{L}_r} \{\tilde{Y}_{u,j} - \mathbb{E}(\tilde{Y}_{u,j})\} \right\| \\ &= \left\| \sum_{j \in \mathcal{L}_r} \{Y_{u,j} - \mathbb{E}(Y_{u,j})\} \right\| + \mathcal{O}\left(\sum_{j \in \mathcal{L}_r} \|Y_{u,j} - \tilde{Y}_{u,j}\| \right) \\ &\sim (p_n B_n \sigma_u^2)^{1/2} + \mathcal{O}(p_n B_n \rho^{\ell_n/4}) \sim (p_n B_n \sigma_u^2)^{1/2}, \end{aligned} \tag{96}$$

where $\mathcal{L}_r = \{j \in \mathbb{N} : (r-1)(p_n+q_n)+1 \leq j \leq r(p_n+q_n) - q_n\}$. Similarly, we can also show

$$\|V_{u,r} - \mathbb{E}(V_{u,r})\| \sim (q_n B_n \sigma_u^2)^{1/2} + \mathcal{O}(q_n B_n \rho^{\ell_n/4}). \tag{97}$$

Then, since $q_n = o(p_n)$, we have

$$\operatorname{var}\left(\sum_{r=1}^{k_n-1} V_{u,r} + V_{u,k_n}\right) = (k_n - 1)\mathcal{O}(q_n B_n \sigma_u^2) + \mathcal{O}((p_n + q_n)B_n) = o(nB_n) \tag{98}$$

which implies that

$$\frac{\sum_r (V_{u,r} - \mathbb{E}(V_{u,r}))}{\sqrt{nB_n}} \Rightarrow 0. \tag{99}$$

Also, by Eq. (73), we have

$$\operatorname{var}(h_n(u,\theta)) = \mathcal{O}(B_n/n) = \mathcal{O}((\log n)^{-2-8/\delta}), \tag{100}$$

which implies

$$h_n(u,\theta) - \mathbb{E}(h_n(u,\theta)) \Rightarrow 0. \tag{101}$$

Therefore, by

$$\begin{aligned} &\sqrt{n/B_n}\{\hat{f}_n(u,\theta) - \mathbb{E}(\hat{f}_n(u,\theta))\} \\ &= \frac{g_n(u,\theta) - \mathbb{E}(g_n(u,\theta))}{\sqrt{nB_n}} - h_n(u,\theta) + \mathbb{E}(h_n(u,\theta)), \end{aligned} \tag{102}$$

we only need to show

$$\frac{\sum_r (U_{u,r} - \mathbb{E}(U_{u,r}))}{\sqrt{nB_n}} \Rightarrow \mathcal{N}(0, \sigma_u^2). \tag{103}$$

We can check the conditions of Lemma A.1 (the Berry–Esseen lemma) as follows.

$$\mathbb{E}\left(\frac{U_{u,r} - \mathbb{E}(U_{u,r})}{\sqrt{nB_n}}\right) = 0; \tag{104}$$

$$\sum_r \frac{\|U_{u,r} - \mathbb{E}(U_{u,r})\|^2}{nB_n} \sim k_n \frac{p_n B_n \sigma_u^2}{nB_n} \sim \sigma_u^2; \tag{105}$$

$$\sum_r \frac{\|U_{u,r} - \mathbb{E}(U_{u,r})\|^{2+\delta/2}}{(nB_n)^{1+\delta/4}} = \mathcal{O}\left(k_n \frac{(\ell_n \sqrt{p_n B_n})^{2+\delta/2}}{(nB_n)^{1+\delta/4}}\right) = \mathcal{O}(\ell_n k_n^{-\delta/4}). \tag{106}$$

Note that

$$
\begin{aligned}
k_n &= \lfloor n/(p_n+q_n)\rfloor = \mathcal{O}(n\psi^{-2/3}B_n^{-1/3}) \\
&= \mathcal{O}(n^{1/3}(\log n)^{(4/3+16/3\delta)}B_n^{-1/3}) = \mathcal{O}((\log n)^{4/3+16/3\delta}),
\end{aligned}
\tag{107}
$$

which implies

$$
\ell_n k_n^{-\delta/4} = \mathcal{O}((\log n)(\log n)^{(-\delta/3-4/3)}) = \mathcal{O}((\log n)^{(-\delta/3-1/3)}) \to 0. \tag{108}
$$

Therefore, the result holds by Lemma A.1.

8.4. **Proof of Theorem 5.3.** Define $D_n = C_n B_n$ and $\theta_i = \frac{i\pi}{B_n}, i = 0, \ldots, B_n$. We use the previous definitions of $X_{u,k}$, the ℓ-dependent $\tilde{X}_{u,k}$, and $\alpha_{n,k-s}$ in Section 8.2. Let $g_n(u,\theta) := [2\pi n\hat{f}_n(u,\theta) - \sum_{k=1}^n X_{u,k}^2] - \mathbb{E}[2\pi n\hat{f}_n(u,\theta) - \sum_{k=1}^n X_{u,k}^2]$, where $\ell = \lfloor n^\gamma \rfloor$ for fixed $\gamma > 0$ which is close to zero. Note that

$$
\begin{aligned}
\hat{f}_n(u,\theta) - \mathbb{E}(\hat{f}_n(u,\theta)) &= \frac{1}{2\pi n}\sum_{1\le k,k'\le n} \alpha_{n,k-k'}[X_{u,k}X_{u,k'} - \mathbb{E}(X_{u,k}X_{u,k'})] \\
&= \frac{1}{2\pi n}\left(g_n(u,\theta) + \sum_{k=1}^n (X_{u,k}^2 - \mathbb{E}X_{u,k}^2)\right).
\end{aligned}
\tag{109}
$$

Therefore,

$$
g_n(u,\theta) = \sum_{1\le k,k'\le n, k\neq k'} \alpha_{n,k-k'}[X_{u,k}X_{u,k'} - \mathbb{E}(X_{u,k}X_{u,k'})]. \tag{110}
$$

Then let $\tilde{g}_n(u,\theta)$ be the corresponding version of $g_n(u,\theta)$ using ℓ-dependent $\{\tilde{X}_{u,k}\}$ instead of $\{X_{u,k}\}$. Define $X'_{u,k} = \tilde{X}_{u,k}\mathbf{1}_{|\tilde{X}_{u,k}|\le (nB_n)^\alpha}$ where $\alpha < \frac{1}{4}$. Next, let $\bar{X}_{u,k} := X'_{u,k} - \mathbb{E}X'_{u,k}$ and define

$$
\begin{aligned}
\bar{g}_n &= 2\sum_{1\le s<k\le n} \alpha_{n,k-s}[\bar{X}_{u,k}\bar{X}_{u,s} - \mathbb{E}(\bar{X}_{u,k}\bar{X}_{u,s})] \\
&= 2\sum_{k=2}^n \bar{X}_{u,k}\sum_{s=1}^{k-1}\alpha_{n,k-s}\bar{X}_{u,s} - 2\mathbb{E}\sum_{k=2}^n \bar{X}_{u,k}\sum_{s=1}^{k-1}\alpha_{n,k-s}\bar{X}_{u,s}.
\end{aligned}
\tag{111}
$$

In the following, we show $g_n(u,\theta)$ can be approximated by $\tilde{g}_n(u,\theta)$.

Lemma 8.9. *Under* GMC(2), *we have* $\max_{u\in\mathcal{U}}\max_{0\le i\le B_n}\mathbb{E}|g_n(u,\theta_i)-\tilde{g}_n(u,\theta_i)| = o(n^{1+\gamma}\rho^{\lfloor n^\gamma\rfloor})$ *and*

$$\max_{u\in\mathcal{U}}\max_{0\le i\le B_n}\frac{|g_n(u,\theta_i)-\tilde{g}_n(u,\theta_i)|}{\sqrt{nB_n}} = o_{\mathbb{P}}(1). \tag{112}$$

Proof. See Appendix A.8. □

Next, we show that $\tilde{g}_n(u,\theta)$ can be approximated by $\bar{g}_n(u,\theta)$.

Lemma 8.10.

$$\mathbb{E}\left(\max_{u\in\mathcal{U}}\max_{\theta}\frac{|\tilde{g}_n(u,\theta)-\bar{g}_n(u,\theta)|}{\sqrt{nB_n}}\right) = o(1). \tag{113}$$

Proof. See Appendix A.9. □

According to Lemma 8.9 and Lemma 8.10, together with $\max_i|\tilde{g}_n(u,\theta_i)-\bar{g}_n(u,\theta_i)|\le \max_\theta|\tilde{g}_n(u,\theta)-\bar{g}_n(u,\theta)|$, we have

$$\max_{u\in\mathcal{U}}\max_{0\le i\le B_n}\frac{|g_n(u,\theta)-\tilde{g}_n(u,\theta)|^2}{nB_n} = o_{\mathbb{P}}(1), \tag{114}$$

and

$$\begin{aligned}&\mathbb{P}\left(\max_{u\in\mathcal{U}}\max_{0\le i\le B_n}\frac{|\tilde{g}_n(u,\theta_i)-\bar{g}_n(u,\theta_i)|^2}{nB_n}\ge y\right)\\ &\le \frac{\mathbb{E}\left(\max_{u\in\mathcal{U}}\max_\theta\frac{|\tilde{g}_n(u,\theta)-\bar{g}_n(u,\theta)|^2}{nB_n}\right)}{y} = o(1).\end{aligned} \tag{115}$$

Since $\max_u\max_i|\mathbb{E}\tilde{g}_n(u,\theta_i)-\mathbb{E}\bar{g}_n(u,\theta_i)|\le\mathbb{E}(\max_u\max_i|\tilde{g}_n(u,\theta_i)-\bar{g}_n(u,\theta_i)|)$, it suffices to show

$$\mathbb{P}\left[\max_{0\le i\le B_n,u\in\mathcal{U}}\frac{|\bar{g}_n(u,\theta_i)-\mathbb{E}(\bar{g}_n(u,\theta_i))|^2}{4\pi^2nB_nf_n^2(u,\theta_i)\int_{-1}^1 a(t)\mathrm{d}t}-2\log D_n+\log(\pi\log D_n)\le x\right]$$

converges to $e^{-e^{-x/2}}$ where $D_n=B_nC_n$.

Let $p_n=\lfloor B_n^{1+\beta}\rfloor$, $q_n=B_n+\ell$, $\ell=\lfloor n^\gamma\rfloor$ and $k_n=\lfloor n/(p_n+q_n)\rfloor$, where γ is small enough and $\beta>0$ is sufficiently close to zero. Split the interval $[1,n]$ into alternating big and small blocks H_j and I_j by

$$\begin{aligned} H_j &= [(j-1)(p_n+q_n)+1, jp_n+(j-1)q_n], \quad 1 \le j \le k_n, \\ I_j &= [jp_n+(j-1)q_n+1, j(p_n+q_n)], \quad 1 \le j \le k_n, \\ I_{k_n+1} &= [k_n(p_n+q_n)+1, n]. \end{aligned} \tag{116}$$

Define $\bar{Y}_{u,k} := \bar{X}_{u,k}\sum_{s=1}^{k-1}\alpha_{n,k-s}\bar{X}_{u,s}$. Then $\bar{g}_n = \sum_{k=1}^{n}(\bar{Y}_{u,k} - \mathbb{E}\bar{Y}_{u,k})$. For $1 \le j \le k_n+1$, let

$$U_j(u,\theta) := \sum_{k\in H_j}(\bar{Y}_{u,k} - \mathbb{E}\bar{Y}_{u,k}), \quad V_j(u,\theta) := \sum_{k\in I_j}(\bar{Y}_{u,k} - \mathbb{E}\bar{Y}_{u,k}). \tag{117}$$

Then $\bar{g}_n = \sum_{j=1}^{k_n} U_j + \sum_{j=1}^{k_n+1} V_j$. Next, define a truncated and normalized version of U_j as

$$\bar{U}_j(u,\theta) := U_j(u,\theta)\mathbf{1}\left(\frac{|U_j(u,\theta)|}{\sqrt{nB_n}} \le \frac{1}{(\log B_n)^4}\right) - \mathbb{E}U_j(u,\theta)\mathbf{1}\left(\frac{|U_j(u,\theta)|}{\sqrt{nB_n}} \le \frac{1}{(\log B_n)^4}\right). \tag{118}$$

In the following, we show that $\bar{g}_n(u,\theta_i) - \mathbb{E}(\bar{g}_n(u,\theta_i))$ can be approximated by $\sum_{j=1}^{k_n}\bar{U}_j(u,\theta_i)$.

Lemma 8.11.

$$\max_{u\in\mathcal{U}}\max_{0\le i\le B_n}\frac{\left|\bar{g}_n(u,\theta_i) - \mathbb{E}(\bar{g}_n(u,\theta_i)) - \sum_{j=1}^{k_n}\bar{U}_j(u,\theta_i)\right|}{\sqrt{nB_n}} = o_{\mathbb{P}}(1). \tag{119}$$

Proof. See Appendix A.10. ☐

Furthermore, we show in the following that $\sum_{j=1}^{k_n}\bar{U}_j(u,\theta_i)$ can be ignored if $i \notin [(\log B_n)^2, B_n - (\log B_n)^2]$.

Lemma 8.12.

$$\mathbb{P}\left(\max_{u\in\mathcal{U}}\max_{i\notin[(\log B_n)^2, B_n-(\log B_n)^2]}\frac{\left|\sum_{j=1}^{k_n}\bar{U}_j(u,\theta_i)\right|}{\sqrt{nB_n}} \ge x\sqrt{\log(B_nC_n)}\right) = o(1). \tag{120}$$

Proof. See Appendix A.11. ☐

Finally, we complete the proof of Eq. (25) by the following result.

Lemma 8.13.

$$\mathbb{P}\left[\max_{u\in\mathcal{U}}\max_{(\log B_n)^2\le i\le B_n-(\log B_n)^2}\frac{\left|\sum_{j=1}^{k_n}\bar{U}_j(u,\theta_i)\right|^2}{4\pi^2 nB_n f_n^2(u,\theta_i)\int_{-1}^{1}a(t)\mathrm{d}t}\right.$$
$$\left.-2\log D_n+\log(\pi\log D_n)\le x\right]\to e^{-e^{-x/2}}. \tag{121}$$

Proof. See Appendix A.12. □

8.5. **Proof of Theorem 5.5.** For simplicity, we denote $\delta_{u,n}$ as δ_u and $\delta_{\theta,n}$ as δ_θ. First, we write

$$\begin{aligned}\hat{f}_n(u,\theta)-\hat{f}_n(u_i,\theta_j)=&\hat{f}_n(u,\theta)-\hat{f}_n(u_i,\theta_j)\\&-\mathbb{E}[\hat{f}_n(u,\theta)-\hat{f}_n(u_i,\theta_j)]+\mathbb{E}[\hat{f}_n(u,\theta)-\hat{f}_n(u_i,\theta_j)].\end{aligned} \tag{122}$$

Then by continuity we have

$$\max_{\{u_i,\theta_j\}}\sup_{\{u:|u-u_i|\le\delta_u,\theta:|\theta-\theta_j|\le\delta_\theta\}}|\mathbb{E}\hat{f}_n(u,\theta)-\mathbb{E}\hat{f}_n(u_i,\theta_j)|=o_{\mathbb{P}}(\sqrt{\log n}). \tag{123}$$

Letting $\hat{g}_n(u,u_i,\theta,\theta_j):=\hat{f}_n(u,\theta)-\hat{f}_n(u_i,\theta_j)$, it suffices to show that

$$\max_{\{u_i,\theta_j\}}\sup_{\{u:|u-u_i|\le\delta_u,\theta:|\theta-\theta_j|\le\delta_\theta\}}|\hat{g}_n(u,u_i,\theta,\theta_j)-\mathbb{E}\hat{g}_n(u,u_i,\theta,\theta_j)|=o_{\mathbb{P}}(\sqrt{\log n}). \tag{124}$$

Note that

$$\begin{aligned}\hat{g}_n(u,u_i,\theta,\theta_j)=&[\hat{f}_n(u,\theta)-\mathbb{E}\hat{f}_n(u,\theta)]\left[1-\frac{\hat{f}_n(u_i,\theta_j)}{\hat{f}_n(u,\theta)}\right]\\&+\mathbb{E}\hat{f}_n(u,\theta)\left[1-\frac{\hat{f}_n(u_i,\theta_j)}{\hat{f}_n(u,\theta)}\right].\end{aligned} \tag{125}$$

Then we can write

$$\begin{aligned}&\sup_{\{u,\theta\}}\hat{g}_n(u,u_i,\theta,\theta_j)\\&\le\sup_{\{u,\theta\}}\left[\hat{f}_n(u,\theta)-\mathbb{E}\hat{f}_n(u,\theta)\right]\sup_{\{u,\theta\}}\left|\frac{\hat{f}_n(u_i,\theta_j)}{\hat{f}_n(u,\theta)}-1\right|\\&+\sup_{\{u,\theta\}}\mathbb{E}\hat{f}_n(u,\theta)\sup_{\{u,\theta\}}\left|\frac{\hat{f}_n(u_i,\theta_j)}{\hat{f}_n(u,\theta)}-1\right|.\end{aligned} \tag{126}$$

Since by Theorem 4.1, we have

$$\sup_{\{u,\theta\}}[\hat{f}_n(u,\theta) - \mathbb{E}\hat{f}_n(u,\theta)] = \mathcal{O}_{\mathbb{P}}(\sqrt{\log n}). \tag{127}$$

Therefore, the following result completes the proof.

Lemma 8.14. *If* $\delta_u = \mathcal{O}(\frac{n}{N(\log n)^\alpha})$ *and* $\delta_\theta = \mathcal{O}(\frac{1}{B_n(\log n)^\alpha})$ *for some* $\alpha > 0$*, then*

$$\max_{\{u_i,\theta_j\}} \sup_{\{u:|u-u_i|\le\delta_u,\theta:|\theta-\theta_j|\le\delta_\theta\}} \left|\frac{\hat{f}_n(u_i,\theta_j)}{\hat{f}_n(u,\theta)} - 1\right| = o_{\mathbb{P}}(1). \tag{128}$$

Proof. See Appendix A.14. □

References

[Ada98] S. Adak. "Time-dependent spectral analysis of nonstationary time series". *Journal of the American Statistical Association* 93.444 (1998), pp. 1488–1501. ISSN: 0162-1459.

[All77] J. Allen. "Short term spectral analysis, synthesis, and modification by discrete Fourier transform". *and Signal Processing IEEE Transactions on Acoustics, Speech* 25.3 (June 1977), pp. 235–238. ISSN: 0096-3518.

[And71] T. W. Anderson. *The statistical analysis of time series*. John Wiley & Sons, Inc., New York-London-Sydney, 1971, pp. xiv+704.

[Ave85] T. Aven. "Upper (lower) bounds on the mean of the maximum (minimum) of a number of random variables". *Journal of Applied Probability* (1985), pp. 723–728.

[Ber62] S. M. Berman. "A law of large numbers for the maximum in a stationary Gaussian sequence". *The Annals of Mathematical Statistics* 33.1 (1962), pp. 93–97.

[Bri69] D. R. Brillinger. "Asymptotic properties of spectral estimates of second order". *Biometrika* 56 (1969), pp. 375–390. ISSN: 0006-3444.

[Coh95] L. Cohen. *Time-Frequency Analysis: Theory and Applications*. Prentice Hall Signal Processing Series. Prentice Hall, 1995.

[CT88] Y. S. Chow and H. Teicher. *Probability Theory: Independence, Interchangeability, Martingales*. Springer US, 1988.

[Dah97] R. Dahlhaus. “Fitting time series models to nonstationary processes”. *The Annals of Statistics* 25.1 (1997), pp. 1–37.

[Dau90] I. Daubechies. “The wavelet transform, time-frequency localization and signal analysis”. *Information Theory, IEEE Transactions on* 36.5 (1990), pp. 961–1005.

[Dau92] I. Daubechies. *Ten lectures on wavelets*. Vol. 61. CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992, pp. xx+357. ISBN: 0-89871-274-2.

[DLW11] I. Daubechies, J. Lu, and H.-T. Wu. “Synchrosqueezed wavelet transforms: an empirical mode decomposition-like tool”. *Applied and Computational Harmonic Analysis. Time-Frequency and Time-Scale Analysis, Wavelets, Numerical Algorithms, and Applications* 30.2 (2011), pp. 243–261. ISSN: 1063-5203.

[DPV11] H. Dette, P. Preuss, and M. Vetter. “A measure of stationarity in locally stationary processes with applications to testing”. *Journal of the American Statistical Association* 106.495 (2011), pp. 1113–1124. ISSN: 0162-1459.

[DSR11] Y. Dwivedi and S. Subba Rao. “A test for second-order stationarity of a time series based on the discrete Fourier transform”. *Journal of Time Series Analysis* 32.1 (2011), pp. 68–91. ISSN: 0143-9782.

[EM97] U. Einmahl and D. M. Mason. “Gaussian approximation of local empirical processes indexed by functions”. *Probability Theory and Related Fields* 107.3 (1997), pp. 283–311.

[FN06] P. Fryzlewicz and G. P. Nason. “Haar-Fisz estimation of evolutionary wavelet spectra”. *Journal of the Royal Statistical Society. Series B. Statistical Methodology* 68.4 (2006), pp. 611–634. ISSN: 1369-7412.

[Grö01] K. Gröchenig. *Foundations of time-frequency analysis*. Springer, 2001.

[HBB92] F. Hlawatsch and G. F. Boudreaux-Bartels. “Linear and quadratic time-frequency signal representations”. *IEEE Signal Processing Magazine* 9.2 (Apr. 1992), pp. 21–67. ISSN: 1053-5888.

[HSLW+98] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu. “The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis”. *Proceedings* of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 454.1971 (1998), pp. 903–995. ISSN: 1364-5021.

[JSR15] C. Jentsch and S. Subba Rao. "A test for second order stationarity of a multivariate time series". *Journal of Econometrics* 185.1 (2015), pp. 124–161. ISSN: 0304-4076.

[LW10] W. Liu and W. B. Wu. "Asymptotics of spectral density estimates". *Econometric Theory* 26.4 (2010), pp. 1218–1245.

[Mey92] Y. Meyer. *Wavelets and operators*. Vol. 37. Cambridge Studies in Advanced Mathematics. Translated from the 1990 French original by D. H. Salinger. Cambridge University Press, Cambridge, 1992, pp. xvi+224. ISBN: 0-521-42000-8; 0-521-45869-2.

[Nas13] G. Nason. "A test for second-order stationarity and approximate confidence intervals for localized autocovariances for locally stationary time series". *Journal of the Royal Statistical Society. Series B. Statistical Methodology* 75.5 (2013), pp. 879–904. ISSN: 1369-7412.

[NSK00] G. P. Nason, R. von Sachs, and G. Kroisandt. "Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum". *Journal of the Royal Statistical Society. Series B. Statistical Methodology* 62.2 (2000), pp. 271–292. ISSN: 1369-7412.

[ORSM01] H. C. Ombao, J. A. Raz, R. von Sachs, and B. A. Malow. "Automatic statistical analysis of bivariate nonstationary time series". *Journal of the American Statistical Association* 96.454 (2001), pp. 543–560. ISSN: 0162-1459.

[Pap10] E. Paparoditis. "Validating stationarity assumptions in time series analysis by rolling local periodograms". *Journal of the American Statistical Association* 105.490 (2010), pp. 839–851. ISSN: 0162-1459.

[Par57] E. Parzen. "On consistent estimates of the spectrum of a stationary time series". *Annals of Mathematical Statistics* 28 (1957), pp. 329–348. ISSN: 0003-4851.

[PP12] E. Paparoditis and D. N. Politis. "Nonlinear spectral density estimation: thresholding the correlogram". *Journal of Time Series Analysis* 33.3 (2012), pp. 386–397. ISSN: 0143-9782.

[Pri65] M. B. Priestley. "Evolutionary spectra and non-stationary processes.(With discussion)". *Journal of the Royal Statistical Society. Series B. Methodological* 27 (1965), pp. 204–237. ISSN: 0035-9246.

[PRW99] D. N. Politis, J. P. Romano, and M. Wolf. *Subsampling.* Springer, 1999.

[Ros84] M. Rosenblatt. "Asymptotic Normality, Strong Mixing and Spectral Density Estimates". *The Annals of Probability* 12.4 (1984), pp. 1167–1180.

[Ros85] M. Rosenblatt. *Stationary Sequences and Random Fields.* Springer, 1985.

[SS17] R. Shumway and D. Stoffer. *Time Series Analysis and Its Applications, With R Examples.* Fourth Edition. Springer, 2017.

[SW07] X. Shao and W. B. Wu. "Asymptotic spectral theory for nonlinear time series". *The Annals of Statistics* 35.4 (2007), pp. 1773–1801.

[Wat54] G. Watson. "Extreme values in samples from m-dependent stationary stochastic processes". *The Annals of Mathematical Statistics* (1954), pp. 798–800.

[WN67] M. B. Woodroofe and J. W. V. Ness. "The Maximum Deviation of Sample Spectral Densities". *The Annals of Mathematical Statistics* 38.5 (1967), pp. 1558–1569.

[WS04] W. B. Wu and X. Shao. "Limit theorems for iterated random functions". *Journal of Applied Probability* 41.2 (2004), pp. 425–436.

[WZ07] W. B. Wu and Z. Zhao. "Inference of trends in time series". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 69.3 (2007), pp. 391–410.

[WZ18] W. B. Wu and P. Zaffaroni. "Asymptotic theory for spectral density estimates of general multivariate time series". *Econometric Theory* 34.1 (2018), pp. 1–22. ISSN: 0266-4666.

[Zho13] Z. Zhou. "Heteroscedasticity and Autocorrelation Robust Structural Change Detection". *Journal of the American Statistical Association* 108.502 (2013), pp. 726–740.

[ZW09] Z. Zhou and W. B. Wu. "Local linear quantile estimation for nonstationary time series". *The Annals of Statistics* (2009), pp. 2696–2729.

[ZW10] Z. Zhou and W. B. Wu. "Simultaneous inference of linear models with time varying coefficients". *Journal of the Royal Statistical Society. Series B. Statistical Methodology* 72.4 (2010), pp. 513–531. ISSN: 1369-7412.

A. Supplemental Material

Lemma A.1. *(Berry-Esseen) If $\{X_i, i \geq 1\}$ are independent random variables $\mathbb{E}(X_i) = 0$, $s_n^2 = \sum_{i=1}^n \mathbb{E}(X_i^2) > 0$, $\sum_{i=1}^n \mathbb{E}|X_i|^{2+\delta} < \infty$, for some $\delta \in (0,1]$ and $S_n = \sum_{i=1}^n X_i$, there exists a universal constant C_δ such that*

$$\sup_{-\infty < x < \infty} |\mathbb{P}(S_n < x s_n) - \Phi(x)| \leq C_\delta \left(\frac{\sum_{i=1}^n \mathbb{E}|X_i|^{2+\delta}}{s_n^{2+\delta}} \right). \tag{129}$$

Proof. See [CT88, pp. 304]. □

A.1. **Proof of Lemma 8.1.** Define $d_{u,n}(h) = \frac{1}{n}\sum_{k=1+h}^n \mu_{u,k}\mu_{u,k-h}$ for $0 \leq h \leq n-1$ and $d_{u,n}(h) = 0$ if $h \geq n$. Since

$$\sum_{k=1}^n \cos(k\theta_{j_\ell}) \cos((k+h)\theta_{j_{\ell'}}) = \frac{n}{2}\cos(h\theta_{j_\ell}) \mathbf{1}_{\{j_\ell = j_{\ell'}\}}, \tag{130}$$

using

$$\begin{aligned} d_{u,n}(h) &= \frac{1}{n}\sum_{k=1+h}^{n+h} \mu_{u,k}\mu_{u,k-h} - \frac{1}{n}\sum_{k=n+1}^{n+h} \mu_{u,k}\mu_{u,k-h} \\ &= \sum_{\ell=1}^p c_\ell^2 \frac{\cos(h\theta_{j_\ell})}{2\pi f(u,\theta_{j_\ell})} - \frac{1}{n}\sum_{k=n+1}^{n+h} \mu_{u,k}\mu_{u,k-h}, \end{aligned} \tag{131}$$

we get that uniformly over J, c and u, there exists K_0 such that

$$\left| d_{u,n}(h) - \sum_{\ell=1}^p c_\ell^2 \frac{\cos(h\theta_{j_\ell})}{2\pi f(u,\theta_{j_\ell})} \right| \leq K_0 \min\left\{ \frac{h}{n}, 1 \right\}. \tag{132}$$

Next, we can write $\|T_{u,n}\|^2/n$ as

$$\begin{aligned} &\frac{1}{n}\mathbb{E}\left(\sum_{k=1}^N \mu_{u,k} \tau\left(\frac{k - \lfloor uN \rfloor}{n} \right) X_k \right)^2 \\ &= d_{u,n}(0) r(u,0) \left[\frac{1}{n}\sum_k \tau\left(\frac{k - \lfloor uN \rfloor}{n} \right)^2 \right] \\ &+ 2\sum_{h=1}^\infty d_{u,n}(h) r(u,h) \left[\frac{1}{n}\sum_k \tau\left(\frac{k - \lfloor uN \rfloor}{n} \right) \tau\left(\frac{k + h - \lfloor uN \rfloor}{n} \right) \right] + o(1). \end{aligned} \tag{133}$$

Furthermore, defining

$$f_n(u,\theta) := \frac{1}{2\pi}\sum_{h=0}^{\infty} r(u,h)\cos(h\theta)\left[\frac{1}{n}\sum_k \tau\left(\frac{k-\lfloor uN\rfloor}{n}\right)\tau\left(\frac{k+h-\lfloor uN\rfloor}{n}\right)\right], \tag{134}$$

we have

$$\begin{aligned}
&\sum_h \left\{\left[\frac{1}{n}\sum_k \tau\left(\frac{k-\lfloor uN\rfloor}{n}\right)\tau\left(\frac{k+h-\lfloor uN\rfloor}{n}\right)\right] r(u,h)\sum_{\ell=1}^{p} c_\ell^2 \frac{\cos(h\theta_{j_\ell})}{2\pi f(u,\theta_{j_\ell})}\right\} \\
&= \sum_{\ell=1}^{p} \frac{c_\ell^2}{2\pi f(u,\theta_{j_\ell})}\sum_h \left\{ r(u,h)\cos(h_{\theta_{j_\ell}})\left[\frac{1}{n}\sum_k \tau\left(\frac{k-\lfloor uN\rfloor}{n}\right)\tau\left(\frac{k+h-\lfloor uN\rfloor}{n}\right)\right]\right\} \\
&= \sum_{\ell=1}^{p} c_\ell^2 \frac{f_n(u,\theta_{j_l})}{f(u,\theta_{j_l})}.
\end{aligned} \tag{135}$$

By the assumptions that $\tau \in \mathcal{C}^1([-1/2,1/2])$, $\int \tau^2(x)\mathrm{d}x = 1$, together with $\sup_u |r(u,h)| = o(h^{-2})$, and $\sum_{h=1}^{\infty} |r(u,h)| < \infty$, we have $f_n(u,\theta) = f(u,\theta) + o(1)$, uniformly over u and θ. This implies that

$$\sum_{\ell=1}^{p} c_\ell^2 \frac{f_n(u,\theta_{j_l})}{f(u,\theta_{j_l})} = \sum_{\ell=1}^{p} c_\ell^2 + o(1) = 1 + o(1). \tag{136}$$

Therefore, uniformly over J and c, we have

$$\begin{aligned}
&\left|\frac{\|T_{u,n}\|^2}{n} - 1\right| - o(1) \\
&\le 2\sum_{h=0}^{\infty}\left|d_{u,n}(h) - \sum_{\ell=1}^{p} c_\ell^2 \frac{\cos(h\theta_{j_\ell})}{2\pi f_n(u,\theta_{j_\ell})}\right| r(u,h)\left[\frac{1}{n}\sum_k \tau\left(\frac{k-\lfloor uN\rfloor}{n}\right)\tau\left(\frac{k+h-\lfloor uN\rfloor}{n}\right)\right] \\
&\le 2\sum_{h=0}^{\infty} K_0 \min\left\{\frac{h}{n},1\right\} r(u,h)\left[\frac{1}{n}\sum_k \tau\left(\frac{k-\lfloor uN\rfloor}{n}\right)\tau\left(\frac{k+h-\lfloor uN\rfloor}{n}\right)\right].
\end{aligned} \tag{137}$$

Finally, since $\sup_u \sum_h |r(u,h)| < \infty$, we have $\sup_u \sum_{h>n} |r(u,h)| \to 0$. Also, as $n \to \infty$,

$$
\begin{aligned}
\sup_u \sum_{h<n} (h/n) r(u,h) &\leq \sup_u \sum_{h<\sqrt{n}} (h/n) r(u,h) + \sup_u \sum_{\sqrt{n} \leq h < n} (h/n) r(u,h) \\
&\leq \sup_u \sum_{h<\sqrt{n}} r(u,h)/\sqrt{n} + \sup_u \sum_{h>\sqrt{n}} r(u,h) \qquad (138)\\
&\to 0.
\end{aligned}
$$

Therefore, $\left|\frac{\|T_{u,n}\|^2}{n} - 1\right| \to 0$.

A.2. **Proof of Lemma 8.2.** Note that $\int \tau(x)\tau(x+h)\mathrm{d}x \leq \frac{1}{2}\int \left[\tau(x)^2 + \tau(x+h)^2\right] \mathrm{d}x = 1$. For simplicity of the proof, we can assume there exists some finite τ_* such that

$$
\frac{1}{n} \sum_k \tau\left(\frac{k - \lfloor uN \rfloor}{n}\right) \tau\left(\frac{k+h-\lfloor uN \rfloor}{n}\right) \leq \tau_*^2. \qquad (139)
$$

Then we have

(140)

$$
\begin{aligned}
&\frac{\|T_{u,n} - \tilde{T}_{u,n}\|}{\sqrt{n}} = \left[\frac{1}{n} \sum_{j=-\infty}^{\lfloor uN + n/2 \rfloor} \|\mathcal{P}_j(T_{u,n} - \tilde{T}_{u,n})\|^2\right]^{1/2} \\
&\leq \mu_* \tau_* \left[\frac{1}{n} \sum_{k=1}^{n} \sum_{j=-\infty}^{\lfloor uN + n/2 \rfloor} \|\mathcal{P}_j(X_{\lfloor uN \rfloor + k - \lfloor n/2 \rfloor} - \tilde{X}_{\lfloor uN \rfloor + k - \lfloor n/2 \rfloor})\|^2\right]^{1/2} \\
&\leq \mu_* \tau_* \max_{k \in \{1,\ldots,n\}} \sum_{j=-\infty}^{\lfloor uN + n/2 \rfloor} \|\mathcal{P}_j(X_{\lfloor uN \rfloor + k - \lfloor n/2 \rfloor} - \tilde{X}_{\lfloor uN \rfloor + k - \lfloor n/2 \rfloor})\| \\
&\leq \mu_* \tau_* \max_{k \in \{1,\ldots,n\}} \sum_{j=-\infty}^{\lfloor uN + n/2 \rfloor} \min\left\{2\|\mathcal{P}_j(X_{\lfloor uN \rfloor + k - \lfloor n/2 \rfloor})\|, \|X_{\lfloor uN \rfloor + k - \lfloor n/2 \rfloor} - \tilde{X}_{\lfloor uN \rfloor + k - \lfloor n/2 \rfloor}\|\right\} \\
&\leq \mu_* \tau_* \sup_k \sum_{j=-\infty}^{k+n} \min\{2\|\mathcal{P}_j(X_k)\|, \|X_k - \tilde{X}_k\|\} \to 0, \quad \text{as } \ell \to \infty.
\end{aligned}
$$

Since the upper bound does not depend on u, the convergence holds uniformly over u.

A.3. **Proof of Lemma 8.3.** In this proof, we omit subscript u for simplicity. Since $\sup_k \mathbb{E}(X_k^2) < \infty$, we have

$$\lim_{t\to\infty} \sup_k \mathbb{E}[X_k^2 \mathbf{1}(|X_k| > t)] = 0. \tag{141}$$

By the property of conditional expectation, we have $\mathbb{E}(\tilde{X}_k^2) < \mathbb{E}(X_k^2)$. Therefore, defining

$$g_n(r) = r^2 \sup_k \mathbb{E}[\tilde{X}_k^2 \mathbf{1}(|\tilde{X}_k| \geq \sqrt{n}/r)], \tag{142}$$

we can get $\lim_{n\to\infty} g_n(r) = 0$ for all given $r > 0$. Also g_n is non-decreasing with r. Then there exists a sequence $\{r_n\}$ such that $r_n \uparrow \infty$ and $g_n(r_n) \to 0$. Note that r_n does not depend on u.

For simplicity, we use $\tilde{X}_{u,k}$ to denote $\tilde{X}_{\lfloor uN \rfloor + k - \lfloor n/2 \rfloor}$. Let $Y_{u,k} = \tilde{X}_{u,k}\mathbf{1}(|\tilde{X}_{u,k}| \leq \sqrt{n}/r_n)$ and $T_{u,n,Y} = \sum_{k=1}^n \mu_{u,k} Y_{u,k}$. Since $\mathbb{E}[\tilde{X}_k^2 \mathbf{1}(|\tilde{X}_k| \geq \sqrt{n}/r)] = o(1/r_n^2)$ by the definition of r_n, we have $\|Y_{u,k} - \tilde{X}_{u,k}\| = o(1/r_n)$. Now since $Y_{u,k} - \tilde{X}_{u,k}$ is ℓ-dependent, we divide each of $\{Y_{u,k}\}$ and $\{\tilde{X}_{u,k}\}$ into ℓ sub-sequences that each sub-sequences has $\lfloor n/\ell \rfloor$ independent elements. Then by triangle inequality we can get

$$\|T_{u,n,Y} - \tilde{T}_{u,n}\| \leq \sum_{a=1}^{\ell} \left\| \sum_{b=a,a+\ell,\ldots}^{n} \mu_{u,b}(Y_{u,b} - \tilde{X}_{u,b}) \right\| = o(\sqrt{n}/r_n). \tag{143}$$

Next, divide the sequence of $\{Y_{u,k}\}$ into pieces of length $p_n + \ell$ where $p_n = \lfloor r_n^{1/4} \rfloor$.

$$U_{u,t} = \sum_{a\in B_t} \mu_{u,a} Y_{u,a} \tag{144}$$

where $B_t = \{a \in \mathbb{N} : 1 + (t-1)(p_n + \ell) \leq a \leq p_n + (t-1)(p_n+\ell)\}$. Note that for given u, $\{U_{u,t}\}$ are independent (but not identically distributed) for different t.

Define $V_{u,t} = \sum_{t=1}^{t_n} U_{u,t}$, then the difference between $V_{u,t}$ and $T_{u,n,Y}$ is the sum of those dropped ℓ terms in each piece. Since ℓ is fixed and there are t_n blocks, we have $\|T_{u,n,Y} - V_{u,t}\| = \mathcal{O}(\sqrt{t_n})$.

Furthermore, since

$$(\sqrt{n}/r_n)^2 \mathbb{P}(|\tilde{X}_k| \geq \sqrt{n}/r)) \geq \mathbb{E}[\tilde{X}_k^2 \mathbf{1}(|\tilde{X}_k| \geq \sqrt{n}/r)] = o(1/r_n^2) \tag{145}$$

we have $P(|\tilde{X}_k| \geq \sqrt{n}/r)) = o(1/n)$. Then, using

$$\begin{aligned}[\mathbb{E}(Y_k)]^2 &= [\mathbb{E}(\tilde{X}_k) - \mathbb{E}(Y_k)]^2 = [\mathbb{E}\tilde{X}_k \mathbf{1}(|\tilde{X}_k| \geq \sqrt{n}/r)]^2 \\ &\leq \mathbb{E}(\tilde{X}_k^2 \mathbf{1}(|\tilde{X}_k| \geq \sqrt{n}/r))\mathbb{P}(|\tilde{X}_k| \geq \sqrt{n}/r)) = o(1/r_n^2)o(1/n)\end{aligned} \tag{146}$$

we have $\mathbb{E}(Y_k) = o(\frac{1}{\sqrt{n}r_n})$, which implies $|\mathbb{E}(V_n)| = O(n)|\mathbb{E}(Y_k)| = o(\sqrt{n}/r_n)$.

Next, defining $W = (V_n - \mathbb{E}(V_n))/\sqrt{n}$ and $\Delta = \tilde{T}_n/\sqrt{n} - W$, we get

$$\begin{aligned}\sqrt{n}\|\Delta\| = \|\tilde{T}_n - V_n + \mathbb{E}(V_n)\| &\leq |\mathbb{E}(V_n)| + \|V_n - \tilde{T}_n\| \\ &\leq |\mathbb{E}(V_n)| + \|V_n - T_{n,Y}\| + \|T_{n,Y} - \tilde{T}_n\| \\ &= o(\sqrt{n}/r_n) + \mathcal{O}(\sqrt{t_n} + \sqrt{n}/r_n) = \mathcal{O}(\sqrt{t_n}).\end{aligned} \tag{147}$$

Next, we apply Lemma A.1 to $\{U_t - \mathbb{E}(U_t), t = 1, \dots, t_n\}$. Recall that $V_n = \sum_{t=1}^{t_n} U_t$ and $W = (V_n - \mathbb{E}(V_n))/\sqrt{n}$, then

$$\begin{aligned}&\sup_x |\mathbb{P}(V_n - \mathbb{E}(V_n) < x\|V_n - \mathbb{E}(V_n)\|) - \Phi(x)| \\ &= \sup_x |\mathbb{P}(W < x\|W\|) - \Phi(x)| \\ &\leq C\sum_{t=1}^{t_n} \mathbb{E}|U_t - \mathbb{E}(U_t)|^3 \|V_n - \mathbb{E}(V_n)\|^{-3} \\ &\leq C\sum_{t=1}^{t_n} \mathbb{E}|U_t|^3 \|V_n - \mathbb{E}(V_n)\|^{-3}.\end{aligned} \tag{148}$$

Next, we get upper bounds of $\mathbb{E}|U_t|^3$ and $\|V_n - \mathbb{E}(V_n)\|^{-3}$. First, by Hölder's inequality $\sum_{a\in B_t} |Y_a| \leq (\sum_{a\in B_t} |Y_a|^3)^{1/3}(\sum_{a\in B_t} 1)^{2/3}$, we have

$$\mathbb{E}|U_t|^3 \leq \mu_*^3 \mathbb{E}\left|\sum_{a\in B_t} Y_a\right|^3 \leq \mu_*^3 p_n^2 \sum_{a\in B_t} \mathbb{E}|Y_a|^3 \leq \mu_*^3 p_n^2 \sum_{a\in B_t} \mathbb{E}(\frac{\sqrt{n}}{r_n}|Y_a|^2) = \mathcal{O}\left(\mu_*^3 p_n^3 \frac{\sqrt{n}}{r_n}\right). \tag{149}$$

For sequences a_n and b_n, we define $a_n = \Theta(b_n)$ if both $a_n = \mathcal{O}(b_n)$ and $b_n = \mathcal{O}(a_n)$. Then, using the definition of $\Theta(\cdot)$, the variance of $\sum_{a\in B_t} \mu_a Y_a$ has the order of $\Theta(p_n)$ because Y_a is ℓ-dependent. Then the variance of V_n has the order of $\Theta(t_n p_n) = \Theta(n)$. Thus, $\|V_n - \mathbb{E}(V_n)\|^{-3}$ has a order of $\Theta(n^{-3/2})$. Overall, we have

$$\sup_x |\mathbb{P}(W < x\|W\|) - \Phi(x)| \leq \mathcal{O}(\mu_*^3 p_n^3(\sqrt{n}/r_n))\Theta(n) = \mathcal{O}(p_n^{-2}). \tag{150}$$

To complete the proof, we first replace $V_n = \sum_t \sum_{a \in B_t} \mu_a Y_a$ by $\tilde{T}_n = \sum_k \mu_k \tilde{X}_k$ then by $T_n = \sum_k X_k$. Since

$$\{W \leq x - \delta, |\Delta| < \delta\} \subseteq \{W + \Delta \leq x\} \subseteq \{W \leq x + \delta\} \cup \{|\Delta| \geq \delta\}, \tag{151}$$

we have

$$\mathbb{P}(W \leq x - \delta) - \mathbb{P}(|\Delta| \geq \delta) \leq \mathbb{P}(W + \Delta \leq x) \leq \mathbb{P}(W \leq x + \delta) + \mathbb{P}(|\Delta| \geq \delta). \tag{152}$$

Furthermore, one can get

$$\begin{aligned} &\sup_x |\mathbb{P}(W < x\|W\|) - \Phi(x)| \\ &= \sup_x |\mathbb{P}(W < x) - \Phi(x/\|W\|)| \\ &= \sup_x \left|\mathbb{P}(\tilde{T}_n/\sqrt{n} - \Delta < x) - \Phi(x/\|W\|)\right|. \end{aligned} \tag{153}$$

Using

$$\mathbb{P}(W < x - \delta) - \mathbb{P}(|\Delta| \geq \delta) \leq \mathbb{P}(\tilde{T}_n/\sqrt{n} < x) \leq \mathbb{P}(W < x + \delta) + \mathbb{P}(|\Delta| \geq \delta), \tag{154}$$

we get

$$\sup_x \left|\mathbb{P}(\tilde{T}_n/\sqrt{n} < x) - \mathbb{P}(W < x)\right| \leq \mathbb{P}(|\Delta| \geq \delta) = \mathcal{O}(\|\Delta\|^2/\delta^2) = \mathcal{O}(p_n^{-1}/\delta^2). \tag{155}$$

Also

$$\begin{aligned} &\sup_x |\Phi(x/\|W\|) - \phi(x/\|W + \Delta\|)| \\ &= \mathcal{O}(\|W + \Delta\|/\|W\| - 1) = \mathcal{O}(\|\Delta\|) = \mathcal{O}(\sqrt{t_n/n}) = \mathcal{O}(p_n^{-1/2}). \end{aligned} \tag{156}$$

Letting $\delta = p_n^{-1/4}$ we have

$$\sup_x \left|\mathbb{P}(\tilde{T}_n/\sqrt{n} < x) - \Phi(x/\|W + \Delta\|)\right| = \mathcal{O}(p_n^{-2}) + \mathcal{O}(p_n^{-1/2}) + \mathcal{O}(p_n^{-1/2}). \tag{157}$$

Finally, use the above technique again with $\Delta_1 = (T_n - \tilde{T}_n)/\sqrt{n}$ and $\delta = \|\Delta_1\|^{1/2}$, we get

$$\sup_x \left|\mathbb{P}(T_n/\sqrt{n} < x) - \Phi(\sqrt{n}x/\|T_n\|)\right| = \mathcal{O}(\mathbb{P}(|\Delta_1| \geq \|\Delta_1\|^{1/2}) + p_n^{-1/2} + \|\Delta_1\|). \tag{158}$$

Lemma A.2. *(Bernstein's inequality) Let $X_1, \dots, X_n$ be independent zero-mean random variables. Suppose $|X_i| \le M$ a.s., for all i. Then for all positive t,*

$$\mathbb{P}\left(\sum_i X_i > t\right) \le \exp\left(\frac{-\frac{1}{2}t^2}{\sum \mathbb{E}(X_i^2) + \frac{1}{3}Mt}\right). \tag{159}$$

Definition A.3. Let $(U_1, \dots, U_k)$ be a random vector. Then the joint cumulant is defined as

$$\operatorname{cum}(U_1, \dots, U_k) = \sum (-1)^p (p-1)! \mathbb{E}\left(\prod_{j \in V_1} U_j\right) \dots \mathbb{E}\left(\prod_{j \in V_p} U_j\right), \tag{160}$$

where $V_1, \dots, V_p$ is a partition of the set $\{1, 2, \dots, k\}$ and the sum is taken over all such partitions.

Lemma A.4. *Assume GMC(α) with $\alpha = k$ for some $k \in \mathbb{N}$, and $\sup_t \mathbb{E}(|X_t|^k) < \infty$ Then there exists a constant $C > 0$ such that for all u and $0 \le m_1 \le \dots \le m_{k-1}$,*

$$|cum(X_{u,0}, X_{u,m_1}, \dots, X_{u,m_{k-1}})| \le C\rho^{m_{k-1}/[k(k-1)]}, \tag{161}$$

where $X_{u,i} := \tau\left(\frac{i - \lfloor n \rfloor/2 \rfloor}{n}\right) X_{\lfloor uN \rfloor + i - \lfloor n/2 \rfloor}$.

Proof. Since $\tau(\cdot)$ is bounded, we have $\sup_u \sup_i \mathbb{E}(|X_{u,i}|^k) < \infty$. We extend [WS04, Proposition 2] to the cases of locally stationary time series.

Given $1 \le l \le k-1$, by multi-linearity of joint cumulants, we replace X_{u,m_i} by independent X'_{u,m_i} for all $i \ge l$ as follows

$$\begin{aligned}
J &:= \operatorname{cum}(X_{u,0}, X_{u,m_1}, \dots, X_{u,m_{k-1}}) \\
&= \operatorname{cum}(X_{u,0}, X_{u,m_1}, \dots, X_{u,m_{l-1}}, X'_{u,m_l}, \dots, X'_{u,m_{k-1}}) \\
&\quad + \operatorname{cum}(X_{u,0}, X_{u,m_1}, \dots, X_{u,m_{l-1}}, X_{u,m_l} - X'_{u,m_l}, \dots, X_{u,m_{k-1}}) \\
&\quad \dots \\
&\quad + \operatorname{cum}(X_{u,0}, X_{u,m_1}, \dots, X_{u,m_{l-1}}, X'_{u,m_l}, \dots, X_{u,m_{k-1}} - X'_{u,m_{k-1}}) \\
&=: B + \sum_{i=l}^{k-1} A_i.
\end{aligned} \tag{162}$$

Note that $(X_{u,0}, X_{u,m_1}, \dots, X_{u,m_{l-1}})$ is independent with $(X'_{u,m_l}, \dots, X'_{u,m_{k-1}})$. By [Ros85, pp.35], we have $B = 0$. Suppose we have

$$|A_i| \le \frac{C}{k}\rho^{(m_i - m_{l-1})/k} \le \frac{C}{k}\rho^{(m_l - m_{l-1})/k} \tag{163}$$

for $l \le i \le k-1$ and some constant C that does not depend on l. Then $|J| \le C\rho^{(m_l - m_{l-1})/k}$ for any $1 \le l \le k-1$. Then we get

$$|J| \le C \min_l \rho^{(m_l - m_{l-1})/k} = C\rho^{\max_l \frac{m_l - m_{l-1}}{k}} \le C\rho^{m_{k-1}/k(k-1)}. \tag{164}$$

Next, we show Eq. (163). In particular, we show the case $i = l$ and the other cases can be proven similarly. Note that $\mathbb{E}(|X_{u,i}|^k)$ is uniformly bounded, by the definition of joint cumulants in Definition A.3, we only need to show that for $V \subset \{0, \dots, k-1\}$ such that $l \notin V$, we have

$$\mathbb{E}\left((X_{u,m_l} - X'_{u,m_l}) \prod_{j \in V} X_{u,m_j}\right) \le C\rho^{(m_l - m_{l-1})/k}. \tag{165}$$

Letting $|V|$ be the cardinality of the set V, then $|V| \le k-1$, and we have

$$\begin{aligned} \left|\mathbb{E}\left[\left(\prod_{j \in V} X_{u,m_j}\right)^{\frac{1+|V|}{|V|}}\right]\right| &\le \mathbb{E}\left[\left(\frac{1}{|V|}\sum_{j \in V} |X_{u,m_j}|^{|V|}\right)^{\frac{1+|V|}{|V|}}\right] \\ &\le \mathbb{E}\left(\frac{1}{|V|}\sum_{j \in V} |X_{u,m_j}|^{1+|V|}\right) \\ &\le \max_{j \in V} \mathbb{E}\left(|X_{u,m_j}|^{1+|V|}\right) \le M. \end{aligned} \tag{166}$$

By Hölder's inequality and Jensen's inequality

$$\begin{aligned} &\left|\mathbb{E}\left((X_{u,m_l} - X'_{u,m_l}) \prod_{j \in V} X_{u,m_j}\right)\right| \\ &\le \left\|X_{u,m_l} - X'_{u,m_l}\right\|_{1+|V|} \left\|\prod_{j \in V} X_{u,m_j}\right\|_{\frac{1+|V|}{|V|}} \\ &\le \left\|X_{u,m_l} - X'_{u,m_l}\right\|_k M^{\frac{|V|}{1+|V|}} \le (C'\rho^{m_l - m_{l-1}})^{1/k} M' \le C\rho^{(m_l - m_{l-1})/k}. \end{aligned} \tag{167}$$

$\square$

A.4. **Proof of Lemma 8.4.** First, letting $\alpha_k = a(k/B_n)\cos(k\theta)$, we have

$$h_n(u,\theta) = \frac{1}{2\pi\sqrt{nB_n}}\left(\sum_{k=0}^{B_n}\sum_{j=n-k+1}^{n} X_{u,j}X_{u,j+k}\alpha_k + \sum_{k=-B_n}^{-1}\sum_{j=n+k+1}^{n} X_{u,j}X_{u,j+k}\alpha_k\right). \tag{168}$$

By the summability of cumulants of orders 2 and 4 [Ros85, page 185], one can get

$$\sup_u \operatorname{var}\left(\sum_{k=0}^{B_n}\sum_{j=n-k+1}^{n} X_{u,j}X_{u,j+k}\alpha_k\right) = \mathcal{O}(B_n^2). \tag{169}$$

Therefore, we have $\sup_u \|h_n(u,\theta)\| = (nB_n)^{-1/2}\mathcal{O}(B_n)$.

Next, note that by the assumption of GMC(4) defined in Eq. (6), we have

$$\sup_u \sup_i \mathbb{E}(|X_{u,i} - \tilde{X}_{u,i}|^4) \le C\rho^{\ell_n}. \tag{170}$$

Then we have

$$\begin{aligned}
\sup_u \sup_i \|Y_{u,i} - \tilde{Y}_{u,i}\| &\le \sup_u \sup_i \frac{1}{2\pi}\sum_{k=-B_n}^{B_n} \|X_{u,i}X_{u,i+k} - \tilde{X}_{u,i}\tilde{X}_{u,i+k}\||\alpha_k| \\
&\le C \sup_u \sup_i \sum_{k=-B_n}^{B_n} \|(X_{u,i} - \tilde{X}_{u,i})X_{u,i+k} + \tilde{X}_{u,i}(X_{u,i+k} - \tilde{X}_{u,i+k})\| \\
&= \mathcal{O}(B_n)\sup_u \sup_i \|X_{u,i} - \tilde{X}_{u,i}\| \\
&= \mathcal{O}(B_n)\sup_u \sup_i (\mathbb{E}(|X_{u,i} - \tilde{X}_{u,i}|)^4)^{1/4} \\
&= \mathcal{O}(B_n\rho^{\ell_n/4}).
\end{aligned} \tag{171}$$

Finally

$$\sup_u \|g_n(u,\theta) - \tilde{g}_n(u,\theta)\| = \mathcal{O}\left(\sup_u \sum_{i=1}^{n} \|Y_{u,i} - \tilde{Y}_{u,i}\|\right) = \mathcal{O}(nB_n\rho^{\ell_n/4}) = o(1). \tag{172}$$

A.5. **Proof of Lemma 8.5.** To show Eq. (77), since α_k is bounded, letting $z_n = k_n(p_n + q_n) + 1 - q_n$, we have

$$\sup_u \mathbb{E}(\max_\theta |V_{u,k_n}(\theta)|) \le C \sum_{j=-B_n}^{B_n} \sup_u \mathbb{E}|\sum_{i=z_n}^{n} \tilde{X}_{u,i}\tilde{X}_{u,i+j}|. \tag{173}$$

Since $\tilde{X}_{u,i}\tilde{X}_{u,i+j}$ is $2\ell_n$-dependent, if $|j| < \ell_n$, we have

$$\sup_u \left\| \sum_{i=z_n}^{n} \tilde{X}_{u,i}\tilde{X}_{u,i+j} \right\| = \mathcal{O}(2\ell_n\sqrt{(n-z_n)/2\ell_n}) = \mathcal{O}(\sqrt{q_n\ell_n}) = \mathcal{O}(\sqrt{p_n\ell_n}). \tag{174}$$

If $|j| \le \ell_n$, since $\mathbb{E}(\tilde{X}_{u,i}\tilde{X}_{u,i+j}\tilde{X}_{u,i'}\tilde{X}_{u,i'+j}) = 0$ if $|i-i'| > \ell_n$, we have

$$\begin{aligned}\sup_u \left\| \sum_{i=z_n}^{n} \tilde{X}_{u,i}\tilde{X}_{u,i+j} \right\|^2 &= \sup_u \sum_{i,i'=z_n}^{n} \mathbb{E}(\tilde{X}_{u,i}\tilde{X}_{u,i+j}\tilde{X}_{u,i'}\tilde{X}_{u,i'+j}) \\ &= \sup_u \sum_{i'=i-\ell_n}^{i+\ell_n}\sum_{i=z_n}^{n} \mathbb{E}(\tilde{X}_{u,i}\tilde{X}_{u,i+j}\tilde{X}_{u,i'}\tilde{X}_{u,i'+j}) \\ &= \mathcal{O}(q_n\ell_n) = \mathcal{O}(p_n\ell_n),\end{aligned} \tag{175}$$

where we have used the assumption $\sup_i \mathbb{E}(|X_i|^{4+\delta}) < M$. Therefore, we get Eq. (77).

To show Eq. (78), we first define $\tilde{h}_n(u,\theta)$ by replacing X_i by $\tilde{X}_i$. Then we can prove similarly to Eq. (171) that

$$\sup_u \mathbb{E}(\max_\theta |h_n(u,\theta) - \tilde{h}_n(u,\theta)|) = o(1). \tag{176}$$

Therefore, it suffices to show $\sup_u \mathbb{E}(\max_\theta |\tilde{h}_n(u,\theta)|) = o(1)$. Using similar technique to Eq. (173) we can show that

$$\sup_u \mathbb{E}(\max_\theta |\tilde{h}_n(u,\theta)|) = \frac{1}{\sqrt{nB_n}}\mathcal{O}(\sqrt{B_n\ell_n}B_n) = \mathcal{O}(\sqrt{\ell_n}B_n/\sqrt{n}) = o(1), \tag{177}$$

where we have used $\eta < \frac{1}{2}$ and $\sqrt{\ell_n}B_n/\sqrt{n} = \mathcal{O}((\log n)^{1/2}n^{\eta-1/2}) = o(1)$.

To show Eq. (79), we note that GMC(4) implies the absolute summability of cumulants up to the fourth order. Also, for zero-mean random variables X, Y, Z, W, the joint cumulants

$$\operatorname{cum}(X,Y,Z,W) = \mathbb{E}(XYZW) - \mathbb{E}(XY)\mathbb{E}(ZW) - \mathbb{E}(XZ)\mathbb{E}(YW) - \mathbb{E}(XW)\mathbb{E}(YZ). \tag{178}$$

Therefore, letting $\mathcal{L}_r$ be the set of the indices i's such that $Y_{u,i}$ belongs to the block corresponding to $U_{u,r}$, we have

$$\begin{aligned}
\mathrm{var}(U_{u,r}(\theta)) &= \left\| \sum_{i\in\mathcal{L}_r} \sum_{k=-B_n}^{B_n} [X_{u,i}X_{u,i+k} - \mathbb{E}(X_{u,i}X_{u,i+k})]\alpha_k \right\|^2 \\
&= \sum_{i,j\in\mathcal{L}_r} \sum_{k,l=-B_n}^{B_n} \mathbb{E}\{[X_{u,i}X_{u,i+k} - \mathbb{E}(X_{u,i}X_{u,i+k})][X_{u,j}X_{u,j+l} - \mathbb{E}(X_{u,j}X_{u,j+l})]\alpha_k\alpha_l\} \\
&= \sum_{i,j\in\mathcal{L}_r} \sum_{k,l=-B_n}^{B_n} \mathrm{cum}(X_{u,i}, X_{u,i+k}, X_{u,j}, X_{u,j+l})\alpha_k\alpha_l \qquad (179)\\
&\quad + \sum_{i,j\in\mathcal{L}_r} \sum_{k,l=-B_n}^{B_n} \mathbb{E}(X_{u,i}X_{u,j})\mathbb{E}(X_{u,i+k}X_{u,j+l})\alpha_k\alpha_l \\
&\quad + \sum_{i,j\in\mathcal{L}_r} \sum_{k,l=-B_n}^{B_n} \mathbb{E}(X_{u,i}X_{u,j+l})\mathbb{E}(X_{u,i+k}X_{u,j})\alpha_k\alpha_l,
\end{aligned}$$

where the first term is finite since the fourth cumulants are summable. For the second term (the last term can also be shown similarly), we use the condition Eq. (12), so that

$$\mathbb{E}(X_{u,i}X_{u,j})\mathbb{E}(X_{u,i+k}X_{u,j+l}) = [r(u,i-j) + o(1/n)][r(u,i-j+k-j) + o(1/n)]. \qquad (180)$$

Then using $p_n = o(n)$, $B_n = o(p_n)$ and $\sup_u \sum_{k=-\infty}^{\infty} |r(u,k)| < \infty$, one can get

$$\begin{aligned}
&\sup_u \max_r \max_\theta \sum_{i,j\in\mathcal{L}_r} \sum_{k,l=-B_n}^{B_n} [r(u,i-j) + o(1/n)][r(u,i-j+k-j) + o(1/n)] \\
&= \sup_u \max_r \max_\theta \sum_{i,j\in\mathcal{L}_r} r(u,i-j) \left[\sum_{k,l=-B_n}^{B_n} r(u,i-j+k-j) + o(B_n/n) \right] \qquad (181)\\
&\le (2p_n+1)(2B_n+1)(\sup_u \sum_{k=-\infty}^{\infty} |r(u,k)|^2) + o(p_nB_n/n) = \mathcal{O}(p_nB_n).
\end{aligned}$$

To show Eq. (80), we note that

$$\mathrm{var}(U'_{u,r}) = \mathrm{var}(U_{u,r}) \left[1 + \frac{2\mathbb{E}(U'_{u,r})\mathbb{E}(U_{u,r} - U'_{u,r}) - 2\mathrm{var}(U_{u,r} - U'_{u,r})}{\mathrm{var}(U_{u,r})} \right]. \qquad (182)$$

From Lemma 8.8, we know that $\mathrm{var}(U_{u,r}(\theta)) \sim p_n B_n \sigma_u^2(\theta)$ and $\sigma_u^2(\theta) = [1+\eta(2\theta)] f^2(u,\theta) \int_{-1}^{1} a^2(t)\mathrm{d}t \geq f_*^2 \int_{-1}^{1} a^2(t)\mathrm{d}t > 0$. Thus, it suffices to show

$$\sup_u \sup_r \sup_\theta \mathbb{E}(U'_{u,r})\mathbb{E}(U_{u,r} - U'_{u,r}) = o(p_n B_n), \quad \sup_u \sup_r \sup_\theta \mathrm{var}(U_{u,r} - U'_{u,r}) = o(p_n B_n). \tag{183}$$

By Lemma 8.6, applying similar inequalities as Eq. (89), we have

$$\begin{aligned}
&\sup_u \sup_i \sup_\theta \mathrm{var}(U_{u,r} - U'_{u,r}) \\
&\leq \sup_u \sup_i \sup_\theta \frac{\|U_{u,r}\|_{2+\delta/2}^{2+\delta/2}}{d_n^{\delta/2}} \\
&= \mathcal{O}((\ell_n \sqrt{p_n B_n})^{2+\delta/2} (\sqrt{nB_n}(\log n)^{-1/2})^{-\delta/2}) \\
&= \mathcal{O}(p_n B_n)\mathcal{O}((\log n)^{2+3\delta/4} (\sqrt{p_n B_n})^{\delta/2} (\sqrt{nB_n})^{-\delta/2}) \\
&= o(p_n B_n).
\end{aligned} \tag{184}$$

Finally, since $\mathbb{E}(U'_{u,r}) \leq \mathbb{E}(|U_{u,r}|) \leq [\mathbb{E}(|U_{u,r}|^{2+\delta/2})]^{\frac{1}{2+\delta/2}}$, using again similar inequalities as Eq. (89), we have

$$\begin{aligned}
&\sup_u \sup_r \sup_\theta \mathbb{E}(U'_{u,r})\mathbb{E}(U_{u,r} - U'_{u,r}) \\
&\leq \sup_u \sup_r \sup_\theta \|U_{u,r}\|_{2+\delta/2} \frac{\|U_{u,r}\|_{2+\delta/2}^{2+\delta/2}}{d_n^{1+\delta/2}} \\
&= \mathcal{O}(\|U_{u,r}\|_{2+\delta/2}/d_n) o(p_n B_n) \\
&= \mathcal{O}(\sqrt{p_n/n}(\log n)^{3/2}) o(p_n B_n) = o(p_n B_n).
\end{aligned} \tag{185}$$

A.6. **Proof of Lemma 8.6.** For simplicity, we first consider that u and i are fixed. Without loss of generality, we consider the first block sum ($i = 1$) so

$$U_{u,1}(\theta) = \sum_{j=1}^{p_n} \tilde{Y}_{u,j}(\theta). \tag{186}$$

We will first show

$$\left\| \sum_{j=1}^{p_n} \sum_{k=-B_n}^{B_n} \tilde{X}_{u,} \tilde{X}_{u,j+k} \alpha_k \right\|_{2+\delta/2} = \mathcal{O}(\ell_n \sqrt{p_n B_n}), \tag{187}$$

where $\alpha_k = a(k/B_n)\cos(k\theta)$, and then we conclude $\mathcal{O}(\ell_n\sqrt{p_nB_n})$ is also uniformly over u and i since the assumption $\sup_u \sup_i \mathbb{E}(|X_{u,i}|^{4+\delta}) < M$. We first write by triangle inequality

$$\begin{aligned}
&\left\|\sum_{j=1}^{p_n}\sum_{k=-B_n}^{B_n}\tilde{X}_{u,j}\tilde{X}_{u,j+k}\alpha_k\right\|_{2+\delta/2} \\
&\leq \left\|\sum_{j=1}^{p_n}\sum_{k=-B_n}^{0}\tilde{X}_{u,j}\tilde{X}_{u,j+k}\alpha_k\right\|_{2+\delta/2} + \left\|\sum_{j=1}^{p_n}\sum_{k=0}^{B_n}\tilde{X}_{u,j}\tilde{X}_{u,j+k}\alpha_k\right\|_{2+\delta/2}.
\end{aligned} \tag{188}$$

Now consider two cases (i) $\ell_n = o(B_n)$, then

$$\sum_{j=1}^{p_n}\sum_{k=-B_n}^{0}\tilde{X}_{u,j}\tilde{X}_{u,j+k}\alpha_k = \sum_{j=1}^{p_n}\left(\tilde{X}_{u,j}\sum_{k=-B_n}^{-\ell_n}\tilde{X}_{u,j+k}\alpha_k\right) + \sum_{j=1}^{p_n}\sum_{k=1-\ell_n}^{0}\tilde{X}_{u,j}\tilde{X}_{u,j+k}\alpha_k, \tag{189}$$

where the first term of the right hand side of Eq. (189) satisfies

$$\begin{aligned}
&\left\|\sum_{j=1}^{p_n}\left(\tilde{X}_{u,j}\sum_{k=-B_n}^{-\ell_n}\tilde{X}_{u,j+k}\alpha_k\right)\right\|_{2+\delta/2} \\
&\leq \sum_{h=1}^{\ell_n}\left\|\sum_{j=1}^{\lfloor (p_n-h)/\ell_n\rfloor}\tilde{X}_{u,h+(j-1)\ell_n}\sum_{k=-B_n}^{-\ell_n}\tilde{X}_{u,h+(j-1)\ell_n+k}\alpha_k\right\|_{2+\delta/2}.
\end{aligned} \tag{190}$$

Continuing to divide the sum of $\sum_{k=-B_n}^{-\ell_n}\tilde{X}_{u,h+(j-1)\ell_n+k}\alpha_k$ into ℓ_n parts, then by $\sup_{u,i}\mathbb{E}(|X_{u,i}|^{4+\delta}) < M$, we have that

$$\begin{aligned}
\left\|\sum_{j=1}^{p_n}\left(\tilde{X}_{u,j}\sum_{k=-B_n}^{-\ell_n}\tilde{X}_{u,j+k}\alpha_k\right)\right\|_{2+\delta/2} &= \mathcal{O}(\ell_n)\mathcal{O}(\sqrt{p_n/\ell_n})\mathcal{O}(\ell_n)\mathcal{O}(\sqrt{B_n/\ell_n}) \\
&= \mathcal{O}(\ell_n\sqrt{p_nB_n}),
\end{aligned} \tag{191}$$

which holds uniformly over u and i. Similarly, for the second term of the right hand side of Eq. (189)

$$\begin{aligned} &\left\| \sum_{j=1}^{p_n} \sum_{k=1-\ell_n}^{0} \tilde{X}_{u,j} \tilde{X}_{u,j+k} \alpha_k \right\|_{2+\delta/2} \leq \sum_{k=1-\ell_n}^{0} \left\| \sum_{j=1}^{p_n} \tilde{X}_{u,j} \tilde{X}_{u,j+k} \alpha_k \right\|_{2+\delta/2} \\ &= \sum_{k=1-\ell_n}^{0} \sum_{h=1}^{3\ell_n} \left\| \sum_{j=1}^{\lfloor (p_n-h)/3\ell_n \rfloor} \tilde{X}_{u,h+3j\ell_n} \tilde{X}_{u,h+3j\ell_n+k} \alpha_k \right\|_{2+\delta/2} = \mathcal{O}(\ell_n^2 \sqrt{p_n/\ell_n}). \end{aligned} \tag{192}$$

Note that the order $\mathcal{O}(\ell_n^2 \sqrt{p_n/\ell_n})$ also holds uniformly over u and i. This is because $\|\tilde{X}_{u,h+3j\ell_n} \tilde{X}_{u,h+3j\ell_n+k}\|_{2+\delta/2}$ is uniformly bounded, which can be shown using Cauchy–Schwarz's inequality and $\sup_u \sup_i \mathbb{E}(|X_{u,i}|^{4+\delta}) < M$. Therefore, we have proven that, for case (i), we have $\sup_u \sup_i \sup_\theta \|U_{u,i}(\theta)\|_{2+\delta/2} = \mathcal{O}(\ell_n \sqrt{p_n B_n})$.

For the second case (ii) $B_n = \mathcal{O}(\ell_n)$, we have

$$\begin{aligned} &\left\| \sum_{j=1}^{p_n} \sum_{k=-B_n}^{0} \tilde{X}_{u,j} \tilde{X}_{u,j+k} \alpha_k \right\|_{2+\delta/2} \leq \sum_{k=-B_n}^{0} \left\| \sum_{j=1}^{p_n} \tilde{X}_{u,j} \tilde{X}_{u,j+k} \alpha_k \right\|_{2+\delta/2} \\ &= \sum_{k=-B_n}^{0} \sum_{h=1}^{3\ell_n} \left\| \sum_{j=1}^{\lfloor (p_n-h)/3\ell_n \rfloor} \tilde{X}_{u,h+3j\ell_n} \tilde{X}_{u,h+3j\ell_n+k} \alpha_k \right\|_{2+\delta/2} \\ &= \mathcal{O}(B_n \ell_n \sqrt{p_n/\ell_n}) = \mathcal{O}(\ell_n \sqrt{p_n B_n}), \end{aligned} \tag{193}$$

which is also uniform over u and i.

A.7. **Proof of Lemma 8.8.** Using the property of cumulants in Eq. (178), similarly to Eqs. (179) and (180), one can get

$$
\begin{aligned}
&\left\|\sum_{i=-s_n/2}^{s_n/2}\{Y_{u,i}(\theta)-\mathbb{E}(Y_{u,i}(\theta))\}\right\|^2\\
&=\sum_{i,j=-s_n}^{s_n}\sum_{k,l=-B_n}^{B_n}\operatorname{cum}(X_{u,i},X_{u,i+k},X_{u,j},X_{u,j+l})\alpha_k\alpha_l\\
&\qquad+\sum_{i,j=-s_n}^{s_n}\sum_{k,l=-B_n}^{B_n}r(u,i-j)r(u,i+k-j-l)\alpha_k\alpha_l+o(s_nB_n/n)\\
&\qquad+\sum_{i,j=-s_n}^{s_n}\sum_{k,l=-B_n}^{B_n}r(u,i-j-l)r(u,i+k-j)\alpha_k\alpha_l+o(s_nB_n/n).
\end{aligned}
\tag{194}
$$

By Lemma A.4, we have

$$
\sum_{m_1,m_2,m_3\in\mathbb{Z}}\operatorname{cum}(X_{u,0},X_{u,m_1},X_{u,m_2},X_{u,m_3})<C\sum_{s=0}^{\infty}\rho^{s/[4(4-1)]}<\infty,
\tag{195}
$$

which implies that the first term of the right hand side of Eq. (194) is finite.

Finally, according to [Ros84, Theorem 2, Eqs. (3.9)–(3.12)], one can show

$$
\begin{aligned}
&\sum_{i,j=-s_n}^{s_n}\sum_{k,l=-B_n}^{B_n}r(u,i-j)r(u,i+k-j-l)\alpha_k\alpha_l\\
&+\sum_{i,j=-s_n}^{s_n}\sum_{k,l=-B_n}^{B_n}r(u,i-j-l)r(u,i+k-j)\alpha_k\alpha_l\sim s_nB_n\sigma_u^2(\theta).
\end{aligned}
\tag{196}
$$

Lemma A.5. *Let $\{X_k\}$ be ℓ-dependent with $\mathbb{E}X_k=0$ and $X_k\in\mathcal{L}^p$ with $p\geq 2$. Let $W_n=\sum_{k=1}^n X_k$. Then for any $Q>0$, there exists $C_1,C_2>0$ only depending on Q such that*

$$
\mathbb{P}(|W_n|\geq x)\leq C_1\left(\frac{\ell}{x^2}\mathbb{E}W_n^2\right)^Q+C_1\min\left[\frac{\ell^{p-1}}{x^p}\sum_{k=1}^n\|X_k\|_p^p,\sum_{k=1}^n\mathbb{P}\left(|X_k|\geq C_2\frac{x}{\ell}\right)\right].
\tag{197}
$$

Proof. See [LW10, Lemma 2]. ☐

Lemma A.6. *Let $\{X_t\}$ be ℓ-dependent with $\mathbb{E}X_t=0$, $|X_t|\leq M$ a.s., $\ell\leq n$, and $M\geq 1$. Let $S_{k,l}=\sum_{t=l+1}^{l+k}X_t\sum_{s=1}^{t-1}\alpha_{n,t-s}X_s$, where $l\geq 0$, $l+k\leq n$ and assume $\max_{1\leq t\leq n}|\alpha_{n,t}|\leq K_0$,*

$\max_{1\le t\le n}\mathbb{E}X_t^2 \le K_0$, $\max_{1\le t\le n}\mathbb{E}X_t^4 \le K_0$ *for some* $K_0 > 0$. *Then for any* $x \ge 1$, $y \ge 1$, *and* $Q > 0$,

$$\mathbb{P}(|S_{k,l} - \mathbb{E}S_{k,l}| \ge x) \le 2e^{-y/4} + C_1 n^3 M^2 \left(x^{-2} y^2 \ell^3 (M^2 + k) \sum_{s=1}^{n} \alpha_{n,s}^2 \right)^Q + C_1 n^3 M^2 \sum_{i=1}^{n} \mathbb{P}\left(|X_i| \ge \frac{C_2 x}{y \ell^2 (M + k^{1/2})} \right), \tag{198}$$

where $C_1, C_2 > 0$ *are constants depending only on* Q *and* K_0.

Proof. See [LW10, Proposition 3]. □

Lemma A.7. *Assume* $X_k \in \mathcal{L}^p$, *with* $p > 1$, *and* $\mathbb{E}X_k = 0$. *Let* $C_p = 18p^{3/2}(p-1)^{-1/2}$ *and* $p' = \min(2, p)$. *Let* $\alpha_1, \dots, \in \mathbb{C}$. *Then under GMC, we have*

$$\left\| \sum_{k=1}^{n} \alpha_k (X_k - \tilde{X}_k) \right\|_p \le C_p \left(\sum_{k=1}^{n} |\alpha_k|^{p'} \right)^{1/p'} o(\rho^\ell), \tag{199}$$

and

$$\| \sum_{k=1}^{n} \alpha_k X_k \|_p \le C \left(\sum_{k=1}^{n} |\alpha_k|^{p'} \right)^{1/p'}, \quad \| \sum_{k=1}^{n} \alpha_k \tilde{X}_k \|_p \le C \left(\sum_{k=1}^{n} |\alpha_k|^{p'} \right)^{1/p'}, \tag{200}$$

for some constant C.

Proof. This lemma follows from [LW10, Lemma 1] with $\Theta_{\ell+1,p} = o(\sum_{j=\ell+1}^{\infty} \rho^j) = o(\rho^\ell)$. □

Lemma A.8. *Assume* $\mathbb{E}X_{u,k} = 0$, $\sup_u \mathbb{E}|X_{u,k}|^{2p} < \infty$, $p \ge 2$. *Let*

$$L_{n,u} = \sum_{1\le j\le j'\le n} \alpha_{j'-j} X_{u,j} X_{u,j'}, \quad \tilde{L}_{n,u} = \sum_{1\le j\le j'\le n} \alpha_{j'-j} \tilde{X}_{u,j} \tilde{X}_{u,j'}, \tag{201}$$

where $\alpha_1, \dots, \in \mathbb{C}$. *Then under GMC, we have*

$$\frac{\sup_u \|L_{n,u} - \mathbb{E}L_{n,u} - (\tilde{L}_{n,u} - \mathbb{E}\tilde{L}_{n,u})\|_p}{n^{1/2} (\sum_{s=1}^{n-1} |\alpha_s|^2)^{1/2}} = o(\ell \rho^\ell). \tag{202}$$

Proof. For fixed u, if $\mathbb{E}|X_{u,k}|^{2p} < \infty$, the result follow from [LW10, Proposition 1] with $\Theta_{0,2p} = o(1)$ and $d_{\ell,2p} = \sum_{t=0}^{\infty} \min\{o(\rho^t), o(\rho^\ell)\} = o(\ell\rho^\ell)$. Since we have $\sup_u \mathbb{E}|X_{u,k}|^{2p} < \infty$ the proof of [LW10, Proposition 1] also hold uniformly over u. □

Lemma A.9. *Assume* $\mathbb{E}X_{u,k} = 0$, $\sup_u \mathbb{E}X_{u,k}^4 < \infty$ *and* GMC(2). *Let* $\alpha_j = \beta_j \exp(ij\theta)$, *where* $i = \sqrt{-1}$, $\theta \in \mathbb{R}$, $\beta_j \in \mathbb{R}$, $1 - n \le j \le -1$, $m \in \mathbb{N}$ *and* $\tilde{L}_{n,u} = \sum_{1 \le j < t \le n} \alpha_{j-t} \tilde{X}_{u,j} \tilde{X}_{u,t}$. *Define*

$$D_k(u,\theta) = A_{u,k} - \mathbb{E}(A_{u,k} \,|\, \mathcal{F}_{u,k-1}), \quad M_n(u,\theta) = \sum_{t=1}^{n} D_t(u,\theta)^* \sum_{j=1}^{t-1} \alpha_{j-t} D_j(u,\theta), \tag{203}$$

where $(\cdot)^*$ *denotes complex conjugate,* $\mathcal{F}_{u,k-1} := \mathcal{F}_{\lfloor uN - n/2 \rfloor + k - 1}$, *and* $A_{u,k} = \sum_{t=0}^{\infty} \mathbb{E}(\tilde{X}_{u,t+k} \,|\, \mathcal{F}_{u,k}) \exp(ij\theta)$. *Then*

$$\sup_u \frac{\|\tilde{L}_{n,u} - \mathbb{E}\tilde{L}_{n,u} - M_n(u,\theta)\|}{m^{3/2} n^{1/2} \sup_k \|X_{u,k}\|_4^2} \le C V_m^{1/2}(\beta), \tag{204}$$

where

$$V_m(\beta) = \max_{1-n \le i \le -1} \beta_i^2 + m \sum_{j=-1}^{-n-1} |\beta_j - \beta_{j-1}|^2. \tag{205}$$

Proof. For fixed u, the result comes from [LW10, Proposition 2]. Since here we have assumed $\sup_u \mathbb{E}X_{u,k}^4 < \infty$, following the proof of [LW10, Proposition 2], the upper bound also holds uniformly over u. □

Lemma A.10. *Suppose that* $\mathbb{E}X_k = 0$, $\sup_u \mathbb{E}X_k^4 < \infty$, *and* GMC(2) *holds, then*

(1) *We have*

$$\left| \frac{\mathbb{E}[(g_n(u_1,\theta_1) - \mathbb{E}g_n(u_1,\theta_1))(g_n(u_2,\theta_2) - \mathbb{E}g_n(u_2,\theta_2))]}{nB_n} \right| = \mathcal{O}(1/(\log B_n)^2), \tag{206}$$

uniformly on $(u_1, u_2, \theta_1, \theta_2)$ *such that either* $(u_1, u_2) \in \mathcal{U}^2$ *or* $(\theta_1, \theta_2) \in \Theta^2$ *where* $\mathcal{U}^2 = \{(u_1, u_2) : \frac{n}{2N} \le u_1 \le u_2 \le 1 - \frac{n}{2N}, |u_1 - u_2| \ge \frac{n}{N}(1 - 1/(\log B_n)^2)\}$ *and* $\Theta^2 = \{(\theta_1, \theta_2) : 0 \le \theta_1 < \theta_2 \le \pi - B_n^{-1}(\log B_n)^2, |\theta_1 - \theta_2| \ge B_n^{-1}(\log B_n)^2\}$.

(2) *For* $\alpha_n > 0$ *with* $\limsup \alpha_n < 1$, *we have*

$$\left| \frac{\mathbb{E}[(g_n(u_1,\theta_1) - \mathbb{E}g_n(u_1,\theta_1))(g_n(u_2,\theta_2) - \mathbb{E}g_n(u_2,\theta_2))]}{4\pi^2 n B_n f(u_1,\theta_1) f(u_2,\theta_2) \int_{t=-1}^{1} a^2(t) \mathrm{d}t} \right| \le \alpha_n, \tag{207}$$

uniformly on $(u_1, u_2, \theta_1, \theta_2)$ *such that either* $(u_1, u_2) \in \mathcal{U}^2$ *or* $(\theta_1, \theta_2) \in \bar{\Theta}^2$ *where* $\mathcal{U}^2 = \{(u_1, u_2) : \frac{n}{2N} \le u_1 \le u_2 \le 1 - \frac{n}{2N}, |u_1 - u_2| \ge \frac{n}{N}(1 - 1/(\log B_n))\}$ *and* $\bar{\Theta}^2 = \{(\theta_1, \theta_2) : B_n^{-1}(\log B_n)^2 \le \theta_1 < \theta_2 \le \pi - B_n^{-1}(\log B_n)^2, |\theta_1 - \theta_2| \ge B_n^{-1}\}$.

(3) *We have*

$$\left| \frac{\mathbb{E}[g_n(u,\theta) - \mathbb{E}g_n(u,\theta)]^2}{4\pi^2 n B_n f^2(u,\theta) \int_{t=-1}^{1} a^2(t)\mathrm{d}t} - 1 \right| = \mathcal{O}(1/(\log B_n)^2), \tag{208}$$

uniformly on $\{(u,\theta) : B_n^{-1}(\log B_n)^2 \le \theta \le \pi - B_n^2(\log B_n)^2, \frac{n}{2N} < u < 1 - \frac{n}{2N}\}$.

Proof. (1) By Lemma A.8 we approximate $g_n - \mathbb{E}g_n$ first by $\tilde{g}_n - \mathbb{E}\tilde{g}_n$. Then by Lemma A.9, we approximate $\tilde{g}_n - \mathbb{E}\tilde{g}_n$ by $M_n(u,\theta)$, where $M_n(u,\theta) = \sum_{t=1}^n D_t(u,\theta)^* \sum_{j=1}^{t-1} \alpha_{n,j-t} D_j(u,\theta)$. Then it is suffices to show $|\mathbb{E}[M_n(u_1,\theta_1) - M_n^*(u_1,\theta_1)][M_n(u_2,\theta_2) - M_n^*(u_2,\theta_2)]| \le C\frac{nB_n}{(\log B_n)^2}$ and $|\mathbb{E}[M_n(u_1,\theta_1) + M_n^*(u_1,\theta_1)][M_n(u_2,\theta_2) + M_n^*(u_2,\theta_2)]| \le C\frac{nB_n}{(\log B_n)^2}$. We only prove the first inequality here, since the other inequality can be proved similarly. Define

$$r_n(u_1,\theta_1,u_2,\theta_2) := |\mathbb{E}[M_n(u_1,\theta_1) + M_n^*(u_1,\theta_1)][M_n(u_2,\theta_2) + M_n^*(u_2,\theta_2)]|. \tag{209}$$

Since the martingale differences $\{D_t(u,\theta)\}$ are uncorrelated but not independent, we further define $N_n(u,\theta) = \sum_{t=1}^n D_t(u,\theta)^* \sum_{j=1}^{t-\ell-1} \alpha_{n,j-t} D_j(u,\theta)$, then $\|M_n(u,\theta) - N_n(u,\theta)\| = \mathcal{O}(\sqrt{n\ell})$ and $|r_n(u_1,\theta_1,u_2,\theta_2)| \le |\tilde{r}_n(u_1,\theta_1,u_2,\theta_2)| + \mathcal{O}(\sqrt{n\ell(nB_n)} + \sqrt{n\ell(B_n^2)})$, where

$$\tilde{r}_n(u_1,\theta_1,u_2,\theta_2) := |\mathbb{E}[N_n(u_1,\theta_1) + N_n^*(u_1,\theta_1)][N_n(u_2,\theta_2) + N_n^*(u_2,\theta_2)]|. \tag{210}$$

Since $\ell = \lfloor n^\gamma \rfloor$ where γ is small enough, it suffices to show $\tilde{r}_n(u_1,\theta_1,u_2,\theta_2) = \mathcal{O}(nB_n/(\log B_n)^2)$. Now substitute $N_n(u,\theta) = \sum_{t=1}^n D_t(u,\theta)^* \sum_{j=1}^{t-\ell-1} \alpha_{n,j-t} D_j(u,\theta)$ to $\tilde{r}_n(u_1,\theta_1,u_2,\theta_2)$.

If $\theta_1 \ne \theta_2$ and $u_1 = u_2$, we have

$$\sum_{t=1}^{n} \sum_{j=1}^{t-\ell-1} 2\mathbb{E}|D_t(u,\theta)D_j(u,\theta)|^2 a^2\left(\frac{t-j}{B_n}\right)[\cos((t-j)(\theta_1+\theta_2)) + \cos((t-j)(\theta_1-\theta_2))]. \tag{211}$$

Now it suffices to show

$$\sum_{t=1}^{n} \sum_{j=1}^{t-\ell-1} a^2\left(\frac{t-j}{B_n}\right)\cos((t-j)(\theta_1 \pm \theta_2)) = \mathcal{O}(nB_n/(\log B_n)^2).$$

Since $|\theta_1-\theta_2| \geq B_n^{-1}(\log B_n)^2$, using $1+2\sum_{k=1}^{n}\cos(k\theta) = \sin((n+1)\theta/2)/\sin(\theta/2) \leq 1/\sin(\theta/2)$, $\sin(x) = \Theta(x)$ when $x \to 0$, and denoting $j = t-s$, we have

$$\sum_{t=1}^{n}\left|\sum_{j=1}^{B_n} a^2(j/B_n)\cos[j(\theta_1 \pm \theta_2)]\right| \leq Cn/(B_n^{-1}(\log B_n)^2) = \mathcal{O}(nB_n/(\log B_n)^2). \tag{212}$$

If $\theta_1 = \theta_2$ but $u_1 \neq u_2$, using Eq. (215) and $n - N|u_1-u_2| \leq n/(\log B_n)^2$, we have

$$\tilde{r}_n(u_1,\theta,u_2,\theta) \leq C\tilde{r}_{n-N|u_1-u_2|}(u,\theta,u,\theta) = \mathcal{O}((n-N|u_1-u_2|)B_n) = \mathcal{O}(nB_n/(\log B_n)^2).$$

(2) When $\theta_1 \neq \theta_2$, using [WN67, Lemma 3.2(ii)] with the assumption on the continuity of $a(\cdot)$ in Theorem 5.3, we have

$$\limsup_n 2(nB_n)^{-1}\sum_{t=1}^{n}\sum_{j=1}^{t-\ell-1} a^2\left(\frac{t-j}{B_n}\right)\cos((t-j)(\theta_1-\theta_2)) < \int a^2(t)\mathrm{d}t. \tag{213}$$

If $\theta_1 = \theta_2$ and $u_1 \neq u_2$ then

$$\begin{aligned}
&\limsup_n 2(nB_n)^{-1}\sum_{t=1}^{n-N|u_1-u_2|}\sum_{j=1}^{t-\ell-1} a^2\left(\frac{t-j}{B_n}\right)\\
&\leq \limsup_n 2(nB_n)^{-1}(n-N|u_1-u_2|)\sum_{j=-B_n}^{B_n} a^2\left(\frac{t-j}{B_n}\right)\\
&\leq \limsup_n 2(nB_n)^{-1}[nB_n/(\log B_n)^2]\int a^2(t)\mathrm{d}t < \int a^2(t)\mathrm{d}t.
\end{aligned} \tag{214}$$

(3) Since $\|D_t(u,\theta)\|^2 = \sum_{j=\ell}^{\ell}\mathbb{E}(\tilde{X}_{u,t}\tilde{X}_{u,t+j})\exp(ij\theta)$. then

$$\begin{aligned}
\tilde{r}_n(u,\theta,u,\theta) &= \mathcal{O}(nB_n/(\log B_n)^2) + \sum_{t=1}^{n}\|D_t(u,\theta)\|^2\sum_{s=-B_n}^{B_n} a^2(s/B_n)\\
&= \mathcal{O}(nB_n/(\log B_n)^2) + 4\pi^2 f^2(u,\theta)nB_n\int a^2(t)\mathrm{d}t.
\end{aligned} \tag{215}$$

□

Lemma A.11. *Let $X_1,\ldots,X_m$ be independent mean zero d-dimensional random vectors that $|X_i| \leq M$. If the underlying probability space is rich enough, one can define independent normally distributed mean zero random vectors $V_1,\ldots,V_m$ such that the covariance matrices* *of V_i and X_i are equal, for all $1 \le i \le m$; furthermore*

$$\mathbb{P}\left(\left|\sum_{i=1}^{m}(X_i - V_i)\right| \ge \delta\right) \le c_1 \exp(-c_2\delta/M). \tag{216}$$

Proof. See [EM97, Fact 2.2]. $\square$

Lemma A.12. *If X and Y have a bi-variate normal distribution with expectations 0, unit variances, and correlation coefficient r, then*

$$\lim_{c\to\infty} \frac{\mathbb{P}(\{X > c\} \cap \{Y > c\})}{[2\pi(1-r)^{\frac{1}{2}}c^2]^{-1}\exp\left(-\frac{c^2}{1+r}\right)(1+r)^{\frac{3}{2}}} = 1, \tag{217}$$

uniformly for all r such that $|r| \le \delta$, for all $0 < \delta < 1$.

Proof. See [Ber62, Lemma 2]. $\square$

A.8. **Proof of Lemma 8.9.** By Markov's inequality, we have

$$\begin{aligned}
&\mathbb{P}\left(\max_{u\in\mathcal{U}}\max_{0\le i\le B_n}\frac{|g_n(u,\theta_i) - \tilde{g}_n(u,\theta_i)|}{\sqrt{nB_n}} \ge 1/\log D_n\right)\\
&\le \sum_{u\in\mathcal{U}}\sum_{0\le i\le B_n}\mathbb{P}\left(\frac{|g_n(u,\theta_i) - \tilde{g}_n(u,\theta_i)|}{\sqrt{nB_n}} \ge 1/\log D_n\right)\\
&\le CB_nC_n\frac{\mathbb{E}[\frac{|g_n(u,\theta_i)-\tilde{g}_n(u,\theta_i)|}{\sqrt{nB_n}}]^{p/2}}{(1/\log D_n)^{p/2}}.
\end{aligned} \tag{218}$$

By Lemma A.8, $\mathbb{E}|g_n(u,\theta_i) - \tilde{g}_n(u,\theta_i)| = o(n^{1+\gamma}\rho^{\lfloor n^\gamma \rfloor})$ uniformly on u and θ_i. Since $D_n = B_nC_n$ is polynomial of n, the GMC assumption guarantees

$$\mathbb{P}\left(\max_{u\in\mathcal{U}}\max_{0\le i\le B_n}\frac{|g_n(u,\theta_i) - \tilde{g}_n(u,\theta_i)|}{\sqrt{nB_n}} \ge 1/\log D_n\right) = o(1). \tag{219}$$

A.9. **Proof of Lemma 8.10.**

Lemma A.13. *Let $X_i, i = 1, \dots, n$ be an arbitrary sequence of real-valued random variables with finite mean and variance. Then*

$$\mathbb{E}(\max_{1\le i\le n} X_i) \le \max_{1\le i\le n}\mathbb{E}X_i + \sqrt{\frac{n-1}{n}\sum_{i=1}^{n}\mathrm{var}(X_i)}. \tag{220}$$

Proof. See [Ave85, Theorem 2.1]. $\square$

First of all, since $a(\cdot)$ has bounded support $[-1,1]$. We only need to consider the case that $|s-k| \le B_n$. Also, let α_* be an upper bound of $\alpha_{n,i}$, and does not depend on u, then

(221)

$$\begin{aligned}
&\mathbb{E}\left(\max_{u\in\mathcal{U}}\max_{\theta}|\tilde{g}_n(u,\theta)-\bar{g}_n(u,\theta)|\right)\\
&\le \alpha_*\,\mathbb{E}\left[\max_{u\in\mathcal{U}}\sum_{2\le k\le n,\max(1,k-B_n)\le s\le k-1}\left|\tilde{X}_{k,u}\tilde{X}_{s,u}-\mathbb{E}(\tilde{X}_{k,u}\tilde{X}_{s,u})-\bar{X}_{k,u}\bar{X}_{s,u}+\mathbb{E}(\bar{X}_{k,u}\bar{X}_{s,u})\right|\right]\\
&\le 2\alpha_*\,\mathbb{E}\left[\max_{u\in\mathcal{U}}\sum_{2\le k\le n,\max(1,k-B_n)\le s\le k-1}\left|\tilde{X}_{k,u}\tilde{X}_{s,u}-\bar{X}_{k,u}\bar{X}_{s,u}\right|\right]\\
&= 2\alpha_*\,\mathbb{E}\left[\max_{u\in\mathcal{U}}\sum_{2\le k\le n,\max(1,k-B_n)\le s\le k-1}\left|\tilde{X}_{k,u}\tilde{X}_{s,u}-\bar{X}_{k,u}\bar{X}_{s,u}-\tilde{X}_{k,u}\bar{X}_{s,u}+\tilde{X}_{k,u}\bar{X}_{s,u}\right|\right]\\
&\le 2\alpha_*\,\mathbb{E}\left[\max_{u\in\mathcal{U}}\left(\sum_{k=2}^{n}|\tilde{X}_{k,u}|\sum_{s=\max\{1,k-B_n\}}^{k-1}|\tilde{X}_{s,u}-\bar{X}_{s,u}|\right)\right]\\
&\quad+2\alpha_*\,\mathbb{E}\left[\max_{u\in\mathcal{U}}\left(\sum_{k=2}^{n}|\tilde{X}_{k,u}-\bar{X}_{k,u}|\sum_{s=\max\{1,k-B_n\}}^{k-1}|\bar{X}_{s,u}|\right)\right].
\end{aligned}$$

Next, we show that the first term of the right hand side of Eq. (221) satisfies

$$\mathbb{E}\left(\max_{u\in\mathcal{U}}\frac{\sum_{k=2}^{n}|\tilde{X}_{k,u}|\sum_{s=\max\{1,k-B_n\}}^{k-1}|\tilde{X}_{s,u}-\bar{X}_{s,u}|}{\sqrt{nB_n}}\right)=o(1). \tag{222}$$

Similar arguments yield the same result for the second term of the right hand side of Eq. (221). Note that

$$\begin{aligned}
&\mathbb{E}\left(\max_{u\in\mathcal{U}}\frac{\sum_{k=2}^{n}|\tilde{X}_{u,k}|\sum_{s=\max\{1,k-B_n\}}^{k-1}|\tilde{X}_{u,s}-\bar{X}_{u,s}|}{\sqrt{nB_n}}\right)\\
&\le \mathbb{E}\left(\max_{u\in\mathcal{U}}\frac{\sum_{k=2}^{n}|\tilde{X}_{u,k}|\sum_{s=\max\{1,k-B_n\}}^{k-\ell}|\tilde{X}_{u,s}-\bar{X}_{u,s}|}{\sqrt{nB_n}}\right)\\
&\quad+\mathbb{E}\left(\max_{u\in\mathcal{U}}\frac{\sum_{k=2}^{n}|\tilde{X}_{u,k}|\sum_{s=\max\{1,k-\ell+1\}}^{k-1}|\tilde{X}_{u,s}-\bar{X}_{u,s}|}{\sqrt{nB_n}}\right).
\end{aligned} \tag{223}$$

Applying Lemma A.13 and using ℓ-independence and Hölder's inequality, we have that uniformly on u

$$
\begin{aligned}
&\mathbb{E}\left(\sum_{k=2}^{n}|\tilde{X}_{u,k}|\sum_{s=\max\{1,k-B_n\}}^{k-\ell}|\tilde{X}_{u,s}-\bar{X}_{u,s}|\right)\\
&=\mathbb{E}\left(\sum_{k=2}^{n}|\tilde{X}_{u,k}|\right)\mathbb{E}\left(\sum_{s=\max\{1,k-B_n\}}^{k-\ell}|\tilde{X}_{u,s}-\bar{X}_{u,s}|\right)\\
&=\mathcal{O}(n)\mathcal{O}(B_n)\mathbb{E}\left|\tilde{X}_{u,k}\mathbf{1}_{|\tilde{X}_{u,k}|>(nB_n)^\alpha}-\mathbb{E}\tilde{X}_{u,k}\mathbf{1}_{|\tilde{X}_{u,k}|>(nB_n)^\alpha}\right|\\
&\le\mathcal{O}(nB_n)(\mathbb{E}\tilde{X}_{u,k}^p)^{1/p}(\mathbb{P}(|\tilde{X}_{u,k}|^p>(nB_n)^{\alpha p}))^{1-1/p}\\
&=\mathcal{O}(nB_n)\mathcal{O}((nB_n)^{-\alpha p})^{1-1/p}=\mathcal{O}((nB_n)^{1-\alpha(p-1)}).
\end{aligned}
\tag{224}
$$

Furthermore, uniformly on u, we also have

$$
\begin{aligned}
&\sqrt{\mathbb{E}\left(\sum_{k=2}^{n}|\tilde{X}_{u,k}|\sum_{s=\max\{1,k-B_n\}}^{k-\ell}|\tilde{X}_{u,s}-\bar{X}_{u,s}|\right)^2}\\
&=\sqrt{\mathbb{E}\left(\sum_{k=2}^{n}|\tilde{X}_{u,k}|\right)^2\mathbb{E}\left(\sum_{s=\max\{1,k-B_n\}}^{k-\ell}|\tilde{X}_{u,s}-\bar{X}_{u,s}|\right)^2}\\
&=\sqrt{\mathcal{O}(n^2)}\sqrt{\mathcal{O}(B_n^2)\mathbb{E}\left|\tilde{X}_{u,k}\mathbf{1}_{|\tilde{X}_{u,k}|>(nB_n)^\alpha}-\mathbb{E}\tilde{X}_{u,k}\mathbf{1}_{|\tilde{X}_{u,k}|>(nB_n)^\alpha}\right|^2}\\
&\le\mathcal{O}(nB_n)(\mathbb{E}\tilde{X}_{u,k}^p)^{1/p}(\mathbb{P}(|\tilde{X}_{u,k}|^p>(nB_n)^{\alpha p}))^{1-1/p}\\
&=\mathcal{O}(nB_n)\mathcal{O}((nB_n)^{-\alpha p})^{1-1/p}=\mathcal{O}((nB_n)^{1-\alpha(p-1)}).
\end{aligned}
\tag{225}
$$

By the assumptions $p>4$ and $(p-1)\alpha>3/4$, we have

$$
\mathbb{E}\left(\max_{u\in\mathcal{U}}\frac{\sum_{k=2}^{n}|\tilde{X}_{u,k}|\sum_{s=\max\{1,k-B_n\}}^{k-\ell}|\tilde{X}_{u,s}-\bar{X}_{u,s}|}{\sqrt{nB_n}}\right)=\mathcal{O}\left(\frac{C_n^{1/2}(nB_n)^{1-\alpha(p-1)}}{(nB_n)^{1/2}}\right)=o(1),
\tag{226}
$$

since we have assumed $C_n^{1/2} = o[(nB_n)^{\alpha(p-1)-\frac{1}{2}}]$. Next, uniformly on u, the second term of the right hand side of Eq. (221) satisfies

$$\begin{aligned}
&\mathbb{E}\left(\sum_{k=2}^{n}|\tilde{X}_{u,k}|\sum_{s=\max\{1,k-\ell+1\}}^{k-1}|\tilde{X}_{u,s}-\bar{X}_{u,s}|\right)\\
&=\mathcal{O}(n\ell)\mathbb{E}\left|\tilde{X}_{u,k}^2\mathbf{1}_{\tilde{X}_{u,k}^2>(nB_n)^{2\alpha}}-\mathbb{E}\tilde{X}_{u,k}^2\mathbf{1}_{\tilde{X}_{u,k}^2>(nB_n)^{2\alpha}}\right|\\
&\le\mathcal{O}(n\ell)\left(\mathbb{E}|\tilde{X}_{u,k}|^p\right)^{2/p}\left(\mathbb{P}(\tilde{X}_{u,k}^p<(nB_n)^{p\alpha})\right)^{1-2/p}\\
&=\mathcal{O}(n\ell)\mathcal{O}((nB_n)^{-\alpha p})^{1-2/p}\\
&=\mathcal{O}(n\ell)\mathcal{O}(nB_n)^{-\alpha(p-2)}.
\end{aligned}\tag{227}$$

Furthermore, uniformly on u, we have

$$\begin{aligned}
&\sqrt{\mathbb{E}\left(\sum_{k=2}^{n}|\tilde{X}_{u,k}|\sum_{s=\max\{1,k-\ell+1\}}^{k-1}|\tilde{X}_{u,s}-\bar{X}_{u,s}|\right)^2}\\
&=\sqrt{\mathcal{O}(n^2\ell^2)\mathbb{E}\left|\tilde{X}_{u,k}^2\mathbf{1}_{\tilde{X}_{u,k}^2>(nB_n)^{2\alpha}}-\mathbb{E}\tilde{X}_{u,k}^2\mathbf{1}_{\tilde{X}_{u,k}^2>(nB_n)^{2\alpha}}\right|^2}\\
&\le\mathcal{O}(n\ell)\left(\mathbb{E}|\tilde{X}_{u,k}|^p\right)^{2/p}\left(\mathbb{P}(\tilde{X}_{u,k}^p<(nB_n)^{p\alpha})\right)^{1-2/p}\\
&=\mathcal{O}(n\ell)\mathcal{O}((nB_n)^{-\alpha p})^{1-2/p}\\
&=\mathcal{O}(n\ell)\mathcal{O}(nB_n)^{-\alpha(p-2)}.
\end{aligned}\tag{228}$$

Overall, we have

$$\mathbb{E}\left(\max_{u\in\mathcal{U}}\frac{\sum_{k=2}^{n}|\tilde{X}_{u,k}|\sum_{s=\max\{1,k-\ell+1\}}^{k-1}|\tilde{X}_{u,s}-\bar{X}_{u,s}|}{\sqrt{nB_n}}\right)=\mathcal{O}\left(\frac{C_n^{1/2}n\ell(nB_n)^{-\alpha(p-2)}}{\sqrt{nB_n}}\right)=o(1),\tag{229}$$

since we have assumed $C_n = o(B_n^{1+2\alpha(p-2)}n^{-2-2\gamma})$.

A.10. **Proof of Lemma 8.11.** We prove this lemma by first showing that

$$\max_{u\in\mathcal{U}}\max_{0\le i\le B_n}\left|\frac{\sum_{j=1}^{k_n+1}V_j(u,\theta_i)}{\sqrt{nB_n}}\right|=o_{\mathbb{P}}(1),\tag{230}$$

and then showing

$$\max_{u\in\mathcal{U}}\max_{0\le i\le B_n}\left|\frac{\sum_{j=1}^{k_n}U_j(u,\theta_i)-\sum_{j=1}^{k_n}\bar{U}_j(u,\theta_i)}{\sqrt{nB_n}}\right|=o_{\mathbb{P}}(1). \tag{231}$$

To show Eq. (230), we note that $\{V_j\}$ are independent. Applying Lemma A.5, we have

$$\mathbb{P}\left(\left|\frac{\sum_{j=1}^{k_n+1}V_j}{\sqrt{nB_n}}\right|\ge\frac{1}{\log B_n}\right)\le C_1\left(\frac{\sum_{j=1}^{k_n+1}\mathbb{E}V_j^2}{nB_n(\log B_n)^{-2}}\right)^Q+C_1\sum_{j=1}^{k_n+1}\mathbb{P}\left(\frac{|V_j|}{\sqrt{nB_n}}\ge\frac{C_2}{\log B_n}\right). \tag{232}$$

Similar to the proof of Lemma 8.12, one can show $\sum_{j=1}^{k_n+1}\mathbb{E}V_j^2=\mathcal{O}(n^{1+\gamma}B_n^{1-\beta})$. Therefore, by choosing γ close to zero and Q large enough, we have

$$\left(\frac{\sum_{j=1}^{k_n+1}\mathbb{E}V_j^2}{nB_n(\log B_n)^{-2}}\right)^Q=\mathcal{O}(n^{-c}), \tag{233}$$

for any $c>0$. For the other term

$$\sum_{j=1}^{k_n+1}\mathbb{P}\left(\frac{|V_j|}{\sqrt{nB_n}}\ge\frac{C_2}{\log B_n}\right), \tag{234}$$

we apply Lemma A.6 with $M=(nB_n)^\alpha$, $k=B_n+\ell$, $\ell=\lfloor n^\gamma\rfloor$ and $y=(\log B_n)^2$, which yields

$$\begin{aligned}
&\mathbb{P}\left(\frac{|V_j|}{\sqrt{nB_n}}\ge\frac{C_2}{\log B_n}\right)\\
&\le 2\exp\left(-\frac{(\log B_n)^2}{4}\right)+\mathcal{O}\left(n^3(nB_n)^{2\alpha}\left(\frac{(\log B_n)^2}{nB_n}(\log B_n)^4\lfloor n^{3\gamma}\rfloor((nB_n)^{2\alpha}+B_n)\right)^Q\right)\\
&\quad+\mathcal{O}\left(n^3(nB_n)^{2\alpha}\sum_{i=1}^n\mathbb{P}\left(|\bar{X}_{i,\ell}|\ge\frac{C_2\frac{\sqrt{nB_n}}{\log B_n}}{(\log B_n)^2\lfloor n^{2\gamma}\rfloor((nB_n)^\alpha+(B_n+\lfloor n^\gamma\rfloor)^{1/2})}\right)\right),
\end{aligned} \tag{235}$$

where the second term of the right hand side is $\mathcal{O}(n^{-c})$ by choosing Q large enough. Since $\alpha<1/4$ and $|\bar{X}_{i,\ell}|<(nB_n)^\alpha$ almost surely, the last term of the right hand side converge to zero almost surely if

$$(nB_n)^\alpha = o\left(\frac{\frac{\sqrt{nB_n}}{\log B_n}}{(\log B_n)^2\lfloor n^{2\gamma}\rfloor((nB_n)^\alpha + (B_n + \lfloor n^\gamma\rfloor)^{1/2})}\right), \tag{236}$$

which can be satisfied by choosing γ close enough to zero. Therefore, by choosing Q large enough so that $\mathcal{O}(C_nB_nn^{-c}) = o(1)$ (Note that this only requires $C_n = o(n^cB_n^{-1})$ for some c, which is always satisfied when C_n is polynomial of n), we have

$$\mathbb{P}\left(\max_{u\in\mathcal{U}}\max_{0\le i\le B_n}\left|\frac{\sum_{j=1}^{k_n+1}V_j}{\sqrt{nB_n}}\right| \ge \frac{1}{\log B_n}\right) \le \mathcal{O}(C_nB_n)\mathbb{P}\left(\left|\frac{\sum_{j=1}^{k_n+1}V_j}{\sqrt{nB_n}}\right| \ge \frac{1}{\log B_n}\right) = o(1), \tag{237}$$

which implies Eq. (230).

To prove Eq. (231), note that

$$U_j - \bar{U}_j(u,\theta) = U_j(u,\theta)\mathbf{1}\left(\frac{|U_j(u,\theta)|}{\sqrt{nB_n}} > \frac{1}{(\log B_n)^4}\right) - \mathbb{E}U_j(u,\theta)\mathbf{1}\left(\frac{|U_j(u,\theta)|}{\sqrt{nB_n}} > \frac{1}{(\log B_n)^4}\right). \tag{238}$$

Therefore, other than using $p_n = B_n^{1+\beta}$ instead of $q_n = B_n + \ell$, the proof of Eq. (231) is essentially the same as the proof of Eq. (230).

A.11. **Proof of Lemma 8.12.** By Lemma A.2, we have

$$\begin{aligned}
&\mathbb{P}\left(\max_{u\in\mathcal{U}}\max_{i\notin[(\log B_n)^2, B_n-(\log B_n)^2]}\left|\frac{\sum_{j=1}^{k_n}\bar{U}_j}{\sqrt{nB_n}}\right| \ge x\sqrt{\log(B_nC_n)}\right)\\
&= \mathcal{O}(C_n)\sum_{i\notin[(\log B_n)^2, B_n-(\log B_n)^2]}\mathbb{P}\left(\left|\frac{\sum_{j=1}^{k_n}\bar{U}_j}{\sqrt{nB_n}}\right| \ge x\sqrt{\log(B_nC_n)}\right)\\
&= \mathcal{O}(C_nB_n + C_n(\log B_n)^2)\mathbb{P}\left(\frac{\sum_{j=1}^{k_n}|\bar{U}_j|}{\sqrt{nB_n}} \ge x\sqrt{\log(B_nC_n)}\right)\\
&= \mathcal{O}(C_nB_n)\exp\left(\frac{-\frac{1}{2}x^2nB_n(\log B_n + \log C_n)}{\sum_{j=1}^{k_n}\mathbb{E}\bar{U}_j^2 + \frac{1}{3}\frac{\sqrt{nB_n}}{(\log B_n)^4}x\sqrt{nB_n(\log B_n + \log C_n)}}\right).
\end{aligned} \tag{239}$$

Note that $U_j = \sum_{k\in H_j}(\bar{Y}_{k,\ell} - \mathbb{E}\bar{Y}_{k,\ell})$, we first divide $\sum_{k\in H_j}(\bar{Y}_{k,\ell} - \mathbb{E}\bar{Y}_{k,\ell})$ into ℓ sums of subsequences. Note that $\bar{Y}_{k,\ell} = \bar{X}_{k,\ell}\sum_{s=1}^{k-1}\alpha_{n,k-s}\bar{X}_{s,\ell} = \bar{X}_{k,\ell}\sum_{s=\max(1,k-B_n)}^{k-1}\alpha_{n,k-s}\bar{X}_{s,\ell}$. Thus, one can get $\|\bar{U}_j\|^2 = \mathcal{O}(\ell B_n^2)$. Then using $\ell = \mathcal{O}(n^\gamma)$ and $k_n = \lfloor n/(p_n + q_n)\rfloor = \mathcal{O}(n/B_n^{1+\beta})$, one can get $\sum_{j=1}^{k_n}\mathbb{E}\bar{U}_j^2 = \mathcal{O}(n^{1+\gamma}B_n^{1-\beta}) = o(nB_n)$ by choosing γ and β such that $n^\gamma B_n^{-\beta} = o(1)$.

Finally, we have

$$\begin{aligned}
&\mathcal{O}(C_nB_n)\exp\left(\frac{-\frac{1}{2}x^2nB_n(\log B_n + \log C_n)}{\sum_{j=1}^{k_n}\mathbb{E}\bar{U}_j^2 + \frac{1}{3}\frac{\sqrt{nB_n}}{(\log B_n)^4}x\sqrt{nB_n(\log B_n+\log C_n)}}\right)\\
&= o\left[C_nB_n\exp\left(\frac{-\frac{1}{2}x^2\log(B_nC_n)}{o(nB_n)/(nB_n)+\frac{1}{3}x\frac{\log(B_nC_n)}{(\log B_n)^4}}\right)\right] \qquad (240)\\
&\to o\left[C_nB_n\exp\left(-\frac{3}{2}x(\log B_n)^4\right)\right] = o(1),
\end{aligned}$$

since $\log C_n + \log B_n = o(\log B_n)^4$ when C_n and B_n are polynomials of n.

A.12. **Proof of Lemma 8.13.**

(1) We first show that for $|i_1 - i_2| \geq (\log B_n)^2/B_n$ or $|u_1 - u_2| \geq \frac{n}{N}(1 - 1/(\log B_n)^2)$, we have

$$\left|\frac{\mathbb{E}\sum_{j=1}^{k_n}\bar{U}_j(u_1,\theta_{i_1})\sum_{j=1}^{k_n}\bar{U}_j(u_2,\theta_{i_2})}{nB_n}\right| = \mathcal{O}(1/(\log B_n)^2). \qquad (241)$$

Note that $\sum_j \bar{U}_j$ can be approximated by $\bar{g}_n$. This is because according to the proof of Lemma 8.11, we have

$$\max_u \max_i \frac{\mathbb{E}|\sum_{j=1}^{k_n}\bar{U}_j(u,\theta_i) - \bar{g}_n(u,\theta_i)|^2}{nB_n} = \mathcal{O}(B_n^{-\epsilon/2}). \qquad (242)$$

Next, we can approximate $\bar{g}_n$ by $\tilde{g}_n$. This is because by Lemma 8.10 we have

$$\max_u \max_\theta \frac{\mathbb{E}|\tilde{g}_n(u,\theta) - \bar{g}_n(u,\theta)|^2}{nB_n} = \mathcal{O}(1/(\log B_n)^2). \qquad (243)$$

Finally, we only need to show

$$\frac{|\mathrm{Cov}(\tilde{g}_n(u_1,\theta_{i_1}), \tilde{g}_n(u_2,\theta_{i_2}))|}{nB_n} = \mathcal{O}(1/(\log B_n)^2), \qquad (244)$$

which has been proved in Lemma A.10(i).

(2) For convenience, we assume $\int a^2(t)\mathrm{d}t = 1$. Select d distinct tuples $(\theta_{i_1}, u_i), i = 1,\dots,d$ that $(\log B_n)^2 \leq i_1 \leq \dots \leq i_d \leq B_n - (\log B_n)^2$ and $u_i \in \mathcal{U}, i = 1,\dots,d$. Let

$\boldsymbol{W}_n = \sum_{j=1}^{k_n} W_j$ where

$$W_j = \left(\frac{\bar{U}_j(u_1, \theta_{i_1})}{f(u_1, \theta_{i_1})}, \dots, \frac{\bar{U}_j(u_d, \theta_{i_d})}{f(u_d, \theta_{i_d})} \right), \quad 1 \le j \le k_n. \tag{245}$$

Note that by Lemma A.10(iii), we have

$$\left| \mathbb{E} \frac{\left(\sum_{j=1}^{k_n} \bar{U}_j(u, \theta) \right)^2}{nB_n} - 4\pi^2 f^2(u, \theta) \right| = \mathcal{O}(1/(\log B_n)^2). \tag{246}$$

Together with Eq. (241), we have

$$\left| \frac{\operatorname{Cov}(\boldsymbol{W}_n)}{nB_n} - 4\pi^2 \boldsymbol{I}_d \right| = \mathcal{O}(1/(\log B_n)^2). \tag{247}$$

Then we approximate $\boldsymbol{W}_n$ by $\boldsymbol{W}_n' = \sum_{j=1}^{k_n} W_j'$ using Lemma A.11, where $\{W_j'\}$ are independent centered normal random vectors. Then by Lemma A.11, we have $\operatorname{Cov}(W_j) = \operatorname{Cov}(W_j')$, for $1 \le j \le k_n$, and

$$\mathbb{P}\left(\frac{|\boldsymbol{W}_n - \boldsymbol{W}_n'|}{\sqrt{nB_n}} \ge 1/\log B_n \right) = \mathcal{O}(e^{-(\log B_n)^3}). \tag{248}$$

Therefore, we have

$$\left| \frac{\operatorname{Cov}(\boldsymbol{W}_n')}{nB_n} - 4\pi^2 \boldsymbol{I}_d \right| = \mathcal{O}(1/(\log B_n)^2). \tag{249}$$

(3) Next, for $z = (z_1, \dots, z_d)$, we define the minimum of $\{z_i\}$ by $|z|_d := \min_{1 \le i \le d}\{z_i\}$. Then we show

$$\mathbb{P}\left(\frac{|\boldsymbol{W}_n|_d}{\sqrt{nB_n}} \ge y_n \right) = (1 + o(1)) \left(\sqrt{8\pi} y_n^{-1} \exp\left(-\frac{y_n^2}{8\pi^2} \right) \right)^d, \tag{250}$$

uniformly on distinct tuples of $\{(u_j, \theta_{i_j}), j = 1, \dots, d : (\log B_n)^2 \le j_1 \le \cdots \le j_d \le B_n - (\log B_n)^2, \frac{n}{2N} < u_j < 1 - \frac{n}{2N}\}$ such that for any two tuples $(u_{j_1}, \theta_{i_{j_1}})$ and $(u_{j_2}, \theta_{i_{j_2}})$, if $u_{j_1} = u_{j_2}$ then $|\theta_{i_{j_1}} - \theta_{i_{j_2}}| \ge (\log B_n)^2/B_n$; if $\theta_{i_{j_1}} = \theta_{i_{j_2}}$ then $|u_{j_1} - u_{j_2}| \ge \frac{n}{N}(1 - 1/(\log B_n)^2)$.

According to Eq. (248), we have

$$\begin{aligned}
&\mathbb{P}\left(\frac{|\boldsymbol{W}_n'|_d}{\sqrt{nB_n}} \geq y_n - \frac{1}{\log B_n}\right) - \mathcal{O}(e^{-(\log B_n)^3}) \\
&\leq \mathbb{P}\left(\frac{|\boldsymbol{W}_n|_d}{\sqrt{nB_n}} \geq y_n\right) \leq \mathbb{P}\left(\frac{|\boldsymbol{W}_n'|_d}{\sqrt{nB_n}} \geq y_n - \frac{1}{\log B_n}\right) + \mathcal{O}(e^{-(\log B_n)^3}).
\end{aligned} \tag{251}$$

From Eq. (249), we have

$$\left|\frac{\mathrm{Cov}^{1/2}(\boldsymbol{W}_n')}{\sqrt{nB_n}} - 2\pi\boldsymbol{I}_d\right| = \mathcal{O}(1/(\log B_n)^2), \tag{252}$$

so that for a standard normal R^d-valued random vector, $\tilde{W}$, the tail probability of $\frac{\mathrm{Cov}^{1/2}(\boldsymbol{W}_n')}{\sqrt{nB_n}}\tilde{W} - 2\pi\boldsymbol{I}_d\tilde{W}$ satisfies

$$\begin{aligned}
&\mathbb{P}\left(\left|\left(\frac{\mathrm{Cov}^{1/2}(\boldsymbol{W}_n')}{\sqrt{nB_n}} - 2\pi\boldsymbol{I}_d\right)\tilde{W}\right| \geq 1/\log B_n\right) \\
&= \mathcal{O}(e^{-(\log B_n)^2/4}).
\end{aligned} \tag{253}$$

Putting together the above results we can use $2\pi|\tilde{W}|_d$ (recall that we defined the minimum of $\{z_i\}$ by $|z|_d := \min_{1\leq i\leq d}\{z_i\}$) instead of $\frac{|\boldsymbol{W}_n'|_d}{\sqrt{nB_n}}$ to bound the tail probability of $\frac{|\boldsymbol{W}_n|_d}{\sqrt{nB_n}}$:

$$\begin{aligned}
&\mathbb{P}(2\pi|\tilde{W}|_d \geq y_n - 2/\log B_n) - \mathcal{O}(e^{-(\log B_n)^2/4}) \\
&\leq \mathbb{P}\left(\frac{|\boldsymbol{W}_n|_d}{\sqrt{nB_n}} \geq y_n\right) \\
&\leq \mathbb{P}(2\pi|\tilde{W}|_d \geq y_n - 2/\log B_n) + \mathcal{O}(e^{-(\log B_n)^2/4}).
\end{aligned} \tag{254}$$

Using the following approximation of tail probability of a standard normal random variable Z,

$$\mathbb{P}(Z > z) = 1 - \Phi(z) \leq \frac{1}{z\sqrt{2\pi}}\exp\left(-\frac{z^2}{2}\right), \tag{255}$$

we can get

$$\mathbb{P}\left(|Z| > \frac{y_n}{2\pi}\right) = 2\,\mathbb{P}\left(Z > \frac{y_n}{2\pi}\right) \leq \sqrt{8\pi}y_n^{-1}\exp\left(-\frac{y_n^2}{8\pi^2}\right). \tag{256}$$

Then we have shown

$$\mathbb{P}\left(\frac{|\boldsymbol{W}_n|_d}{\sqrt{nB_n}} \geq y_n\right) = (1+o(1))\left(\sqrt{8\pi}y_n^{-1}\exp\left(-\frac{y_n^2}{8\pi^2}\right)\right)^d. \tag{257}$$

Similarly, using Lemma A.12 and Lemma A.10(ii), we can also have

$$\begin{aligned} &\mathbb{P}\left(\left|\frac{\sum_{j=1}^{k_n}\bar{U}_j(u_k,\theta_{i_k})}{\sqrt{nB_n}f(u_k,\theta_{i_k})}\right| \geq y_n, k=1,\ldots,d\right) \\ &\leq C\left(\sqrt{8\pi}y_n^{-1}\exp\left(-\frac{y_n^2}{8\pi^2}\right)\right)^{d-2} y_n^{-2}\exp\left(-\frac{y_n^2}{8\pi^2}(1+\delta)\right), \end{aligned} \tag{258}$$

for some $\delta > 0$, uniformly on distinct tuples of $\{(u_j,\theta_{i_j}), j=1,\ldots,d : (\log B_n)^2 \leq j_1 \leq \cdots \leq j_d \leq B_n - (\log B_n)^2, \frac{n}{2N} < u_j < 1 - \frac{n}{2N}\}$ such that for any two tuples $(u_{j_1},\theta_{i_{j_1}})$ and $(u_{j_2},\theta_{i_{j_2}})$with $j_1 \leq j_2$, if $u_{j_1} = u_{j_2}$ then if $\theta_{i_{j_1}} = \min_j \theta_{i_j}$ then $|\theta_{i_{j_1}} - \theta_{i_{j_2}}| \geq B_n^{-1}$; otherwise $|\theta_{i_{j_1}} - \theta_{i_{j_2}}| \geq (\log B_n)^2/B_n$; if $\theta_{i_{j_1}} = \theta_{i_{j_2}}$ then $|u_{j_1} - u_{j_2}| \geq \frac{n}{N}(1 - 1/(\log B_n)^2)$.

(4) Finally, we define

$$A_{u,i} = \left\{\frac{|\sum_{j=1}^{k_n}\bar{U}_j(u,\theta_i)|^2}{4\pi^2 nB_n f^2(u,\theta_i)} \geq 2\log B_n + 2\log C_n - \log(\pi\log B_n + \pi\log C_n) + x\right\} \tag{259}$$

and we show

$$\mathbb{P}\left(\bigcup_{(\log B_n)^2 \leq i \leq B_n - (\log B_n)^2, u\in\mathcal{U}} A_{u,i}\right) \to 1 - e^{-e^{-x/2}}. \tag{260}$$

To this end, we define

$$\tilde{A}_u = \bigcup_{(\log B_n)^2 \leq i \leq B_n - (\log B_n)^2} A_{u,i} \tag{261}$$

and

$$P_{t,u} := \sum_{(\log B_n)^2 \leq i_1 < \cdots < i_t \leq B_n - (\log B_n)^2} \mathbb{P}(A_{u,i_1} \cap \cdots \cap A_{u,i_t}). \tag{262}$$

Then by Bonferroni's inequality, we have for every fixed k and u that

$$\sum_{t=1}^{2k}(-1)^{t-1}P_{t,u} \leq \mathbb{P}(\tilde{A}_u) \leq \sum_{t=1}^{2k-1}(-1)^{t-1}P_{t,u}. \tag{263}$$

Next following the proof of [Wat54, Theorem] and [WN67, Theorem 3.3] based on Eq. (250) and Eq. (258), we can show

$$P_{t,u} \to [B_n \mathbb{P}(A_{u,i})]^t / t! \tag{264}$$

as $n \to \infty$. As shown in [Wat54, pp.799], with Eq. (250) and Eq. (258), when $n \to \infty$, we have

$$P_{t,u} \to [(B_n - 2(\log B_n)^2)^t / t! + \mathcal{O}(B_n - 2(\log B_n)^2)^{t-1}] \mathbb{P}(A_{u,i})^t. \tag{265}$$

Therefore, we have shown

$$\mathbb{P}(\tilde{A}_u) \to 1 - e^{-[B_n \mathbb{P}(A_{u,i})]}. \tag{266}$$

Finally, we use the above techniques again to show

$$\mathbb{P}\left(\bigcup_{u \in \mathcal{U}} \tilde{A}_u\right) \to 1 - e^{-e^{-x/2}}, \tag{267}$$

which means we only need to show

$$C_n \mathbb{P}(\tilde{A}_u) \to \exp(-x/2). \tag{268}$$

Letting $y_n^2/4\pi^2 = 2\log B_n + 2\log C_n - \log(\pi \log B_n + \pi \log C_n) + x$, as in Eq. (250), we have

$$\begin{aligned}
& C_n \mathbb{P}(\tilde{A}_u) \to C_n B_n \mathbb{P}(A_{u,i}) \to C_n B_n \mathbb{P}\left(|N| > \frac{y_n}{2\pi}\right) \\
& \to \frac{C_n B_n}{y_n} \sqrt{8\pi} \exp\left(-\frac{y_n^2}{8\pi^2}\right) \\
& \to C_n B_n \frac{\sqrt{8\pi}}{\sqrt{8\pi^2}\sqrt{\log B_n + \log C_n}} \exp\left(-\frac{x}{2}\right) \frac{\sqrt{\pi \log B_n + \pi \log C_n}}{B_n C_n} \\
& \to \exp\left(-\frac{x}{2}\right).
\end{aligned} \tag{269}$$

A.13. **Proof of Remark 4.2.** First of all, by the assumption GMC(2)

$$\begin{aligned}\mathbb{E}\hat{f}_n(u,\theta)-f(u,\theta)&=\frac{1}{2\pi}\left[\sum_{k=-B_n}^{B_n}\mathbb{E}\hat{r}(u,k)a(k/B_n)-\sum_{k\in\mathbb{Z}}r(u,k)\right]\exp(\sqrt{-1}k\theta)\\&=\frac{1}{2\pi}\sum_{k=-B_n}^{B_n}\left[\mathbb{E}\hat{r}(u,k)a(k/B_n)-r(u,k)\right]\exp(\sqrt{-1}k\theta)+\mathcal{O}(\rho^{B_n}).\end{aligned} \tag{270}$$

Next, by the SLC condition, we know $r(u,k)$ is Lipschitz. Together with the Lipschitz condition of $\tau(\cdot)$, we have

$$\mathbb{E}\hat{r}(u,k)=\frac{1}{n}\sum_{i=1}^{N}\tau\left(\frac{i-\lfloor uN\rfloor}{n}\right)\tau\left(\frac{i+k-\lfloor uN\rfloor}{n}\right)\mathbb{E}(X_iX_{i+k}) \tag{271}$$

$$=\frac{1}{n}\sum_{i=\lfloor uN\rfloor-\frac{n}{2}}^{\lfloor uN\rfloor+\frac{n}{2}}\tau\left(\frac{i-\lfloor uN\rfloor}{n}\right)\tau\left(\frac{i+k-\lfloor uN\rfloor}{n}\right)\left[r(i/N,k)+\mathcal{O}(k/N)\right] \tag{272}$$

$$=\frac{1}{n}\sum_{i=\lfloor uN\rfloor-\frac{n}{2}}^{\lfloor uN\rfloor+\frac{n}{2}}\left[\tau\left(\frac{i-\lfloor uN\rfloor}{n}\right)^2+o(k/n)\right]r(i/N,k)+\mathcal{O}(k/N) \tag{273}$$

Since $r(u,k)$ is twice continuously differentiable with respect to u, we have

$$\mathbb{E}\hat{r}(u,k)=\frac{1}{n}\sum_{i=\lfloor uN\rfloor-\frac{n}{2}}^{\lfloor uN\rfloor+\frac{n}{2}}\tau\left(\frac{i-\lfloor uN\rfloor}{n}\right)^2\left[r(u,k)+\left(\frac{i-\lfloor uN\rfloor}{N}\right)\frac{\partial r(u,k)}{\partial u}+\mathcal{O}(n^2/N^2)\right] \tag{274}$$

$$+o(k/n)r(i/N,k)+\mathcal{O}(k/N) \tag{275}$$

Furthermore, since $\tau(\cdot)$ is an even function

$$\frac{1}{n}\sum_{i=\lfloor uN\rfloor-\frac{n}{2}}^{\lfloor uN\rfloor+\frac{n}{2}}\tau\left(\frac{i-\lfloor uN\rfloor}{n}\right)^2\left(\frac{i-\lfloor uN\rfloor}{N}\right)\frac{\partial r(u,k)}{\partial u}=0. \tag{276}$$

Therefore, we have

$$\mathbb{E}\hat{r}(u,k)=\left[\int\tau^2(x)\mathrm{d}x+o(1/n)\right]r(u,k)+\mathcal{O}(n^2/N^2)+o(k/n)r(u,k)+\mathcal{O}(k/N) \tag{277}$$

$$=r(u,k)+o(k/n+1/n)r(u,k)+\mathcal{O}(k/N+n^2/N^2). \tag{278}$$

Therefore, by the locally quadratic property of $a(\cdot)$ at 0, we have

$$\begin{aligned}
&\mathbb{E}\hat{r}(u,k)a(k/B_n) - r(u,k) \\
&= \mathbb{E}\hat{r}(u,k)\left[a(0) + a'(0)k/B_n + \frac{1}{2}a''(0)k^2/B_n^2 + o(k^2/B_n^2)\right] - r(u,k) \\
&= -C\left(\frac{k^2}{B_n^2} + o(k/n)\right) r(u,k) + \mathcal{O}(k/N + n^2/N^2).
\end{aligned} \tag{279}$$

Then, using the fact that if $\theta \notin \{0, \pi\}$, we know

$$\sum_{k=0}^{B_n} \cos(k\theta) = \frac{1}{2} + \frac{\sin(\frac{2B_n+1}{2}\theta)}{2\sin(\theta/2)}, \quad \sum_{k=1}^{B_n} \sin(k\theta) = \frac{\sin\frac{B_n\theta}{2}\sin\frac{(B_n+1)\theta}{2}}{\sin(\theta/2)}, \tag{280}$$

then for fixed $\theta \notin \{0, \pi\}$, we have

$$\sum_{k=0}^{B_n} \cos(k\theta) = \mathcal{O}(1), \quad \sum_{k=0}^{B_n} k\cos(k\theta) = \mathcal{O}(B_n). \tag{281}$$

If $\sup_u \sum_{k\in\mathbb{Z}} |r(u,k)|k^2 < \infty$ and $B_n = o(n)$, then

$$\mathbb{E}\hat{f}_n(u,\theta) - f(u,\theta) + \frac{C}{2\pi}\sum_{k\in\mathbb{Z}} \frac{k^2 r(u,k)\exp(\sqrt{-1}k\theta)}{B_n^2} = \mathcal{O}(B_n/N + n^2/N^2). \tag{282}$$

Finally, $B_n = o(N^{1/3})$ implies $\mathcal{O}(B_n/N) = o(1/B_n^2)$. Also, $n = o(N^{2/3})$ and $B_n = o(N^{1/3})$ implies $\mathcal{O}(B_n^2 n^2/N^2) = o(1)$.

A.14. **Proof of Lemma 8.14.** First, we pick any (u_0, θ_0) such that $|u_0 - u| \le \delta_u$ and $|\theta_0 - \theta| \le \delta_\theta$, then

$$\hat{f}_n(u_0,\theta_0) - \hat{f}_n(u,\theta) = \frac{1}{2\pi}\sum_{k=-B_n}^{B_n} a(k/B_n)[\hat{r}(u_0,k)\exp(\sqrt{-1}k\theta_0) - \hat{r}(u,k)\exp(\sqrt{-1}k\theta)]. \tag{283}$$

Using $\tau\left(\frac{i-\lfloor u_0 N\rfloor}{n}\right)=\tau\left(\frac{i-\lfloor uN\rfloor}{n}\right)+\mathcal{O}\left(\frac{\delta_u N}{n}\right)$, we have

$$\hat{r}(u_0,k)\exp(\sqrt{-1}k\theta_0)=\frac{1}{n}\sum_{i=1}^{N}\tau\left(\frac{i-\lfloor u_0N\rfloor}{n}\right)\tau\left(\frac{i+k-\lfloor u_0N\rfloor}{n}\right)(X_iX_{i+k})\exp(\sqrt{-1}k\theta_0) \tag{284}$$

$$=\frac{1}{n}\sum_{i=\lfloor uN\rfloor-\frac{n}{2}}^{\lfloor uN\rfloor+\frac{n}{2}}\left[\tau\left(\frac{i-\lfloor uN\rfloor}{n}\right)\tau\left(\frac{i+k-\lfloor uN\rfloor}{n}\right)+\mathcal{O}\left(\frac{\delta_uN}{n}\right)\right](X_iX_{i+k})\exp(\sqrt{-1}k\theta_0). \tag{285}$$

Note that $\exp(\sqrt{-1}k\theta_0)=\exp(\sqrt{-1}k\theta)[\exp(\sqrt{-1}k(\theta_0-\theta))]$ and $\cos(k\theta_0)=\cos(k\theta)\cos(k(\theta_0-\theta))-\sin(k\theta)\sin(k(\theta_0-\theta))$. Therefore, we have

$$\hat{f}_n(u_0,\theta_0)=\frac{1}{2\pi}\sum_{k=-B_n}^{B_n}a(k/B_n)\hat{r}(u_0,k)\exp(\sqrt{-1}k\theta)\exp(\sqrt{-1}k(\theta_0-\theta)) \tag{286}$$

$$=\frac{1}{2\pi}\sum_{k=-B_n}^{B_n}a(k/B_n)\hat{r}(u,k)\cos(k\theta)\left[1+\mathcal{O}\left(\frac{\delta_uN}{n}\right)\right][1+\mathcal{O}(k\delta_\theta)] \tag{287}$$

$$-\frac{1}{2\pi}\sum_{k=-B_n}^{B_n}a(k/B_n)\hat{r}(u_0,k)\sin(k\theta)\mathcal{O}(k\delta_\theta) \tag{288}$$

$$=\hat{f}_n(u,\theta)\left[1+\mathcal{O}\left(\frac{\delta_uN}{n}\right)\right][1+\mathcal{O}(B_n\delta_\theta)]+\mathcal{O}_{\mathbb{P}}(B_n\delta_\theta), \tag{289}$$

where we have used the fact that the GMC condition implies $\sum_{k=0}^{B_n}kr(u,k)=\mathcal{O}(\sum_{k=0}^{B_n}k\rho^k)=\mathcal{O}(B_n)$. Note that we have assumed that $f(u,\theta)>f_*>0$ uniformly over u and θ, so we can write $\mathcal{O}_{\mathbb{P}}(B_n\delta_\theta)=(B_n\delta_\theta)\mathcal{O}_{\mathbb{P}}(\hat{f}_n(u,\theta))$. Therefore, we have

$$\hat{f}_n(u_0,\theta_0)-\hat{f}_n(u,\theta)=\mathcal{O}(\delta_uN/n+B_n\delta_\theta)\mathcal{O}_{\mathbb{P}}(\hat{f}_n(u,\theta)), \tag{290}$$

which implies

$$\left|\frac{\hat{f}_n(u_0,\theta_0)}{\hat{f}_n(u,\theta)}-1\right|=\mathcal{O}_{\mathbb{P}}(\delta_uN/n+B_n\delta_\theta). \tag{291}$$

In order to make it equal to $o_{\mathbb{P}}(1)$, we only need $\delta_u=o(n/N)$ and $\delta_\theta=o(1/B_n)$. Therefore, choosing $\alpha>0$, $\delta_u=\mathcal{O}\left(\frac{n}{N(\log n)^\alpha}\right)$ and $\delta_\theta=\mathcal{O}\left(\frac{1}{B_n(\log n)^\alpha}\right)$ is sufficient.

A.15. **Proof of Remark 2.11.** By triangle inequality and Hölder's inequality, we have

$$|r(u,k) - r(s,k)| \tag{292}$$

$$= |\mathbb{E}\left[G(u,\mathcal{F}_i)G(u,\mathcal{F}_{i+k}) - G(s,\mathcal{F}_i)G(s,\mathcal{F}_{i+k})\right]| \tag{293}$$

$$\leq \| \left[G(u,\mathcal{F}_i) - G(s,\mathcal{F}_i)\right] G(u,\mathcal{F}_{i+k})\|_1 + \| \left[G(u,\mathcal{F}_{i+k}) - G(s,\mathcal{F}_{i+k})\right] G(s,\mathcal{F}_i)\|_1 \tag{294}$$

$$\leq \|G(u,\mathcal{F}_i) - G(s,\mathcal{F}_i)\|_q \|G(u,\mathcal{F}_{i+k})\|_p + \|G(u,\mathcal{F}_{i+k}) - G(s,\mathcal{F}_{i+k})\|_q \|G(s,\mathcal{F}_i)\|_p \tag{295}$$

$$\leq C|u-s|. \tag{296}$$