Spectral Inference under Complex Temporal Dynamics

Jun Yang & Zhou Zhou

Taylor & Francis
Taylor & Francis Group

Check for updates

Spectral Inference Under Complex Temporal Dynamics

Jun Yang and Zhou Zhou

Department of Statistical Sciences, University of Toronto, Toronto, ON, Canada

ABSTRACT

We develop a 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 (SCRs) 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 nonstationary linear models. Supplementary materials for this article are available online.

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 nonstationary signals, time-frequency analysis has become one of the major research areas in applied mathematics and signal processing (Daubechies 1990; Cohen 1995; Gröchenig 2001). Based on various models or representations of the nonstationary 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 (STFTs) and wavelet transforms (Allen 1977; Daubechies 1992; Meyer 1992); bilinear time-frequency representations such as the Wigner–Ville distribution and more generally the Cohen's class of bilinear time-frequency distributions (Hlawatsch and Boudreaux-Bartels 1992; Cohen 1995) and nonlinear algorithms such as the empirical mode decomposition method (Huang et al. 1998) and the synchrosqueezing transform (Daubechies, Lu, and Wu 2011). 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 subject and the goals of time-frequency analysis and nonstationary time series analysis are highly overlapped. Unfortunately, it seems that the nonstationary spectral domain theory and methodology in the time series literature have been developed largely independently from time-frequency analysis. One major effort in nonstationary time series analysis lies in forming general classes of nonstationary time series models through their evolutionary spectral representation. Among others, Priestley (1965) proposed the notion of evolutionary spectra in a seminar paper. In another seminal work, Dahlhaus (1997) defined a general and theoretically tractable class of locally stationary time series models based on their time-varying spectral representation. Nason, von Sachs, and Kroisandt (2000) 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 nonstationary spectral domain literature involves adaptive estimation of the evolutionary spectra. See, for instance, Adak (1998) for a binary segmentation based method, Ombao et al. (2001) for an automatic estimation procedure based on the smooth localized complex exponential (SLEX) transform, and Fryzlewicz and Nason (2006) 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, Paparoditis (2010), Dette, Preuss, and Vetter (2011), Dwivedi and Subba Rao (2011), and Jentsch and Subba Rao (2015) for tests of stationarity based on properties of the Fourier periodogram or spectral density. See also Nason (2013) 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 nonstationary and possibly nonlinear time series to the best of our knowledge.

The purpose of the article is to develop a 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

---

article 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 article 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 Zhou and Wu (2009) 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 $\{X_i^{(N)}\}_{i=1}^N$. 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 nonstationary time series, which may have wider applicability in evolutionary power spectrum analysis.

A second main contribution of the article 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 normally distributed 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, Wu and Zhao (2007) and Zhou and Wu (2010), for 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 (SCRs) 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 hypothesis converges faster than the SCR. Furthermore, the test achieves asymptotically the power 1 for local alternatives whose evolutionary spectra deviate from the null hypothesis with a rate larger than the order of the width of the SCR. Specific examples include tests for nonstationary 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 vast literature on the topic and we will only list a very small number of representative works. Early works on this topic include Parzen (1957), Woodroofe and Ness (1967), Brillinger (1969), Anderson (1971), and Rosenblatt (1984), 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 Liu and Wu (2010), Paparoditis and Politis (2012), and Wu and Zaffaroni (2018), among others.

The rest of the article is organized as follows. We first formulate the problem in Section 2. In Section 3, we study the STFT and show that the STFTs are asymptotically independent Gaussian random variables under very mild conditions. In Section 4, we study the asymptotic properties of the 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 deferred to Section 8 and many details of the proofs have been put in the online supplementary materials.

2. Problem Formulation

We first define locally stationary time series and their instantaneous covariance and spectral density. Throughout the article, we assume the time series $\{X_i^{(N)}\}_{i=1}^N$ is centered, that is, $\mathbb{E}[X_i^{(N)}] = 0$. Furthermore, for a random variable X, define $\|X\|_q := [\mathbb{E}|X|^q]^{1/q}$ and use $\|\cdot\|$ to denote $\|\cdot\|_2$ for simplicity.

Definition 2.1 (Locally stationary time series (Zhou and Wu 2009)). We say $\{X_i^{(N)}\}_{i=1}^N$ is a locally stationary time series if there exists a nonlinear filter G such that

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

where $\mathcal{F}_i = (\dots, \epsilon_0, \dots, \epsilon_{i-1}, \epsilon_i)$ and ϵ_i's are iid random variables. Furthermore, the nonlinear filter G satisfies the stochastic Lipschitz continuity condition, SLC(q), for some $q > 0$; that is, there exists $C > 0$ such that for all i and $u, s \in (0, 1)$, we have

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

Remark 2.1. For time series $X_1, X_2, \dots, X_N$, we rescale the time index as $t_i = i/N$, $i = 1, \dots, N$. Then $\{t_i\}$ forms a dense grid in $[0, 1]$. The rescaled time $u \in [0, 1]$ is a natural

extension of $\{t_i\}_{i=1}^N$ to be continuum. This rescaling provides an asymptotic device for studying locally stationary time series, which was first introduced by Dahlhaus (1997). In particular, the rescaling together with the stochastic Lipschitz continuity assumption ensure that for each X_i, there is a diverging number of data points in its neighborhood with similar distributional properties.

Example 2.1 (Locally stationary linear time series). Let ϵ_i be iid 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 Dahlhaus (1997). Verification of the SLC assumption is discussed in Zhou and Wu (2009, Propositions 2 and 3).

Example 2.2 (Time varying threshold AR models). Let $\epsilon_i \in \mathcal{L}^q, q > 0$ be iid random variables with distribution function F_ϵ and density f_ϵ. Consider the model

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

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 Zhou and Wu (2009, sec. 4) for more discussions on checking the SLC assumption for locally stationary nonlinear time series.

For simplicity, we will use X_i to denote $X_i^{(N)}$ in this article. Without loss of generality, we assume $X_i = 0$ for any $i > N$. We adopt the physical dependence measure Zhou and Wu (2009) to describe the dependence structure of the time series.

Definition 2.2 (Physical dependence measure). Let $\{\epsilon_i'\}$ be an iid copy of $\{\epsilon_i\}$. Consider the locally stationary time series $\{X_i\}_{i=1}^N$. Assume $\max_{1 \leq i \leq N} \|X_i\|_p < \infty$. For $k \geq 0$, define the kth physical dependence measure by

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

Next, we extend the geometric-moment contraction (GMC) condition (Shao and Wu 2007) to the nonstationary setting.

Definition 2.3 (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 $\delta_p(k) = \mathcal{O}(\rho^k)$ for some $\rho \in (0, 1)$.

Let $\mathcal{P}_k(X) := \mathbb{E}(X \mid \mathcal{F}_k) - \mathbb{E}(X \mid \mathcal{F}_{k-1})$ and $\tilde{X}_k^{[\ell]} := \mathbb{E}(X_k \mid \epsilon_{k-\ell+1}, \ldots, \epsilon_k)$ be the ℓ-dependent conditional expectations of X_k. From the GMC(2) condition and $\sup_k \|X_k\| < \infty$, one can easily verify that $\sup_k \sum_{j=-\infty}^{k} \|\mathcal{P}_j X_k\| < \infty$ and $\lim_{\ell \to \infty} \sup_k \|X_k - \tilde{X}_k^{[\ell]}\| = 0$. We refer to Remark A.1 of the online supplementary materials and Shao and Wu (2007) for more discussions on the GMC condition.

Example 2.3 (Nonstationary nonlinear time series). Many stationary nonlinear time series models are of the form

$$X_i = R(X_{i-1}, \epsilon_i), \tag{6}$$

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

$$X_i(u) = R(u, X_{i-1}(u), \epsilon_i), \quad 0 \leq u \leq 1. \tag{7}$$

Zhou and Wu (2009, 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

$$\sup_{u \in [0,1]} \sup_{x \neq y} \frac{\|R(u, x, \epsilon_0) - R(u, y, \epsilon_0)\|_\alpha}{|x-y|} < 1. \tag{8}$$

See Zhou and Wu (2009, sec. 4.2) for more details.

Definition 2.4 (Instantaneous covariance). Let $u \in [0,1]$. The instantaneous covariance at u is defined by

$$r(u, k) := \mathrm{cov}\left(G(u, \mathcal{F}_0), G(u, \mathcal{F}_k)\right). \tag{9}$$

Remark 2.2. 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

$$|r(u, k) - r(s, k)|/|u - s| \leq C, \tag{10}$$

for some finite constant C. The proof is given in Appendix A.16 of the online supplementary materials. Therefore, uniformly on u, for any positive integer $n \leq N$, we have

$$r(u + \delta_u, k) - r(u, k) = \mathcal{O}(n/N), \quad \forall -n/N \leq \delta_u \leq n/N. \tag{11}$$

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

Next, we define the evolutionary spectral density using the instantaneous covariance.

Definition 2.5 (Instantaneous spectral density). Let $u \in [0,1]$. The spectral density at u is defined by

$$f(u, \theta) := \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} r(u, k) \exp(\sqrt{-1}k\theta). \tag{12}$$

Remark 2.3. In the definition of instantaneous spectral density, $u \in [0,1]$ represents the rescaled time (see Remark 2.1 for more discussions) and $\theta \in [0, 2\pi)$ represents the frequency. Different from the usual spectral density for stationary process, the instantaneous spectral density is a two dimensional function of u and θ, which captures the spectral density variation in both time and frequency. The usual spectral density for stationary process is a one-dimensional function of θ and is static over time. The notion of instantaneous spectral density is useful for capturing the dynamics of the spectral evolution over time.

Remark 2.4. Note that, for any fixed time point u, $r(u, k)$ is a nonnegative definite function on the integers. Hence, Bochner's theorem (or Herglotz representation theorem) implies that the covariance function $r(u, k)$ and the spectral density function $f(u, \theta)$ has a one-to-one correspondence at each rescaled time

point u under the GMC condition. Therefore, $r(u, k)$ defined in Definition 2.4 has a one-one-one correspondence to the spectral density $f(u, \theta)$ defined in Definition 2.5 for short range dependent locally stationary time series defined in our article.

In this article, we always assume $f_* := \inf_{u,\theta} f(u, \theta) > 0$, which is a natural assumption in the time series literature (see, e.g., Shao and Wu 2007; Liu and Wu 2010). Finally, we define the STFT, the local periodogram, and the STFT-based spectral density estimates.

Definition 2.6 (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 $\theta \in [0, 2\pi)$. Then the STFT is defined by

$$J_n(u, \theta) := \sum_{i=1}^{N} \tau\left(\frac{i - \lfloor uN \rfloor}{n}\right) X_i \exp(\sqrt{-1}\theta i). \qquad (13)$$

Definition 2.7 (Local periodogram).

$$I_n(u, \theta) := \frac{1}{2\pi n}|J_n(u, \theta)|^2. \qquad (14)$$

Remark 2.5. Note that defining

$$\hat{r}(u, k) := \frac{1}{n}\sum_{i=1}^{N} \tau\left(\frac{i - \lfloor uN \rfloor}{n}\right)\tau\left(\frac{i + k - \lfloor uN \rfloor}{n}\right) X_i X_{i+k}, \qquad (15)$$

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

$$I_n(u, \theta) = \frac{1}{2\pi}\sum_{k=-n}^{n} \hat{r}(u, k)\exp(\sqrt{-1}\theta k). \qquad (16)$$

It is well known that $I_n(u, \theta)$ is an inconsistent estimator of $f(u, \theta)$ due to the fact that $\hat{r}(u, k)$ are inconsistent when k is large. A natural and classic way to overcome this difficulty is to restrict the above summation to relatively small k's only. This leads to the following.

Definition 2.8 (STFT-based spectral density estimator). Let $a(\cdot)$ be an even, Lipschitz continuous kernel 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

$$\hat{f}_n(u, \theta) := \frac{1}{2\pi}\sum_{k=-B_n}^{B_n} \hat{r}(u, k)a(k/B_n)\exp(\sqrt{-1}k\theta). \qquad (17)$$

Remark 2.6. The modified $\hat{f}_n(u, \theta)$ in Equation (17) is not always nonnegative as it depends on the property of the kernel function $a(\cdot)$. According to Andrews (1991, p. 822), if the kernel function further satisfies $\frac{1}{2\pi}\int_{-\infty}^{\infty} a(x)\exp(-\sqrt{-1}\theta x)\mathrm{d}x \geq 0$ for any $\theta \in [0, 2\pi)$, then the modified $\hat{f}_n(u, \theta)$ in Equation (17) is always nonnegative. For example, the Bartlett kernel, $a(x) = (1 - |x|)\mathbf{1}_{\{|x| \leq 1\}}$, and the Parzen kernel, $a(x) = (1 - 6x^2 + 6|x|^3)\mathbf{1}_{\{0 \leq |x| \leq 1/2\}} + 2(1 - |x|)^3\mathbf{1}_{\{1/2 < |x| \leq 1\}}$.

3. Fourier Transforms

In this section, we study the STFT and show that the STFTs are asymptotically independent and normally distributed under mild conditions. More specifically, when 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 and normally distributed 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

$$Z_{u,j}^{(n)} = \frac{\sum_{k=1}^{N} \tau\left(\frac{k - \lfloor uN \rfloor}{n}\right) X_k \cos(k2\pi j/n)}{\sqrt{\pi n f(u, 2\pi j/n)}},$$

$$Z_{u,j+m}^{(n)} = \frac{\sum_{k=1}^{N} \tau\left(\frac{k - \lfloor uN \rfloor}{n}\right) X_k \sin(k2\pi j/n)}{\sqrt{\pi n f(u, 2\pi j/n)}}, \quad j = 1, \ldots, m, \qquad (18)$$

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

Theorem 3.1. Assume GMC(2), SLC(2), and $\sup_k \mathbb{E}(X_k^2) < \infty$. 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 $1 \leq j_1, \ldots, j_p \leq 2m$ 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 that

$$\sup_J \sup_{c \in \Omega_{p,q}} \sup_x |P(c^T Z_{U,J} \leq x) - \Phi(x)| = o(1), \qquad (19)$$

where $\Phi(x)$ is the cumulative distribution function of the standard normal distribution.

Proof. See Section 8.1. $\qquad \square$

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 rescaled time, the STFTs are asymptotically independent on the latter time-frequency grid. Moreover, the vector formed by these STFTs is asymptotically jointly normally distributed.

4. Consistency and Asymptotic Normality

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

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), SLC(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

$$\sup_u \max_{\theta \in [0,\pi]} \sqrt{n/B_n}|\hat{f}_n(u, \theta) - \mathbb{E}(\hat{f}_n(u, \theta))| = \mathcal{O}_{\mathbb{P}}(\sqrt{\log n}). \qquad (20)$$

Proof. See Section 8.2. □

Later we will see from Theorem 5.1 that the order $\mathcal{O}_{\mathbb{P}}(\sqrt{\log n})$ on the right-hand side of Equation (20) is indeed optimal.

Remark 4.1. 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 GMC(2), 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, that is,

$$\lim_{u \to 0} u^{-2}[1 - a(u)] = C, \quad (21)$$

where C is a nonzero constant, then we have

$$\sup_u \sup_\theta \left[\mathbb{E}\hat{f}_n(u, \theta) - f(u, \theta) - \frac{C}{B_n^2} f''(u, \theta) \right] = o(1/B_n^2), \quad (22)$$

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 of the supplementary materials. Therefore, the consistency of $\hat{f}_n(u, \theta)$ is implied by combining Theorem 4.1 and Equation (22).

4.2. Asymptotic Normality

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

Theorem 4.2. Assume GMC(2), SLC(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

$$\sqrt{n/B_n}\{\hat{f}_n(u, \theta) - \mathbb{E}(\hat{f}_n(u, \theta))\} \Rightarrow \mathcal{N}(0, \sigma_u^2(\theta)), \quad (23)$$

where $\Rightarrow$ denotes weak convergence, $\sigma_u^2(\theta) = [1 + \eta(2\theta)]f^2(u, \theta) \int_{-1}^1 a^2(t)dt$ 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 SCR over u and θ. For simultaneous spectral inference under complex temporal dynamics, one needs to know the asymptotic behavior of the maximum deviation of $\hat{f}_n(u, \theta)$ from $f(u, \theta)$ on the joint time-frequency domain, which is an extremely difficult problem. In this section, we establish a maximum deviation theory for the STFT-based spectral density estimates over a dense grid in the joint time-frequency domain. Such results serve as a theoretical foundation for the joint time-frequency inference of the evolutionary spectral densities.

- 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 assume that $|u_{i_1} - u_{i_2}| \geq \frac{n}{N}(1 - 1/(\log B_n)^2)$.

- 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): Assume that $a(\cdot)$ is an even and bounded function with bounded support $[-1, 1]$, $\lim_{x \to 0} a(x) = a(0) = 1$, $\int_{-1}^1 a^2(x)dx < \infty$, 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 that 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 the 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, 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 5.1 (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 5.1. Under GMC(2) and Conditions (a)–(d), we have that, for any $x \in \mathbb{R}$,

$$\mathbb{P}\left[\max_{(u,\theta) \in \mathcal{G}_N} \frac{n}{B_n} \frac{|\hat{f}_n(u, \theta) - \mathbb{E}(\hat{f}_n(u, \theta))|^2}{f^2(u, \theta) \int_{-1}^1 a^2(t)dt} \right.$$
$$\left. - 2\log B_n - 2\log C_n + \log(\pi \log B_n + \pi \log C_n) \leq x \right]$$
$$\to e^{-e^{-x/2}}.$$
$$(24)$$

Proof. See Section 8.4. □

Theorem 5.1 states that the spectral density estimates $\hat{f}_n(u, \theta)$ 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 5.1 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 depict 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 5.2 (Asymptotically uniform variation matching of grids). Consider a given sequence of bandwidths (n, B_n), and let $\{\tilde{\mathcal{G}}_N\}$ be a sequence of grids of time-frequency pairs $\{(u_i, \theta_j)\}$ with time and frequencies equally spaced, that is, $|u_{i+1} - u_i| = \delta_{\theta,n}$ 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

$$
\max_{\{u_i, \theta_j\} \in \tilde{\mathcal{G}}_N} \sup_{\{u:|u-u_i| \le \delta_{u,n}, \theta:|\theta-\theta_j| \le \delta_{\theta,n}\}} \sqrt{n/B_n}
$$
$$
\times \left| \left[\hat{f}_n(u, \theta) - \mathbb{E}(\hat{f}_n(u, \theta)) \right] - \left[\hat{f}_n(u_i, \theta_j) - \mathbb{E}(\hat{f}_n(u_i, \theta_j)) \right] \right|
$$
$$
= o_{\mathbb{P}}(\sqrt{\log n}). \tag{25}
$$

Note that we have previously shown in Theorem 4.1 that the uniform stochastic variation of $\sqrt{n/B_n}\hat{f}_n(u, \theta)$ on $(u, \theta) \in (0, 1) \times [0, \pi)$ has the order $\mathcal{O}_{\mathbb{P}}(\sqrt{\log n})$. In combination with Theorem 5.1, 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

$$
\sqrt{n/B_n} \left| \sup_{(u,\theta) \in (0,1) \times [0,\pi)} \left| \hat{f}_n(u, \theta) - \mathbb{E}(\hat{f}_n(u, \theta)) \right| \right.
$$
$$
\left. - \max_{\{u_i, \theta_j\} \in \tilde{\mathcal{G}}_N} \left| \hat{f}_n(u_i, \theta_j) - \mathbb{E}(\hat{f}_n(u_i, \theta_j)) \right| \right| = o_{\mathbb{P}}(\sqrt{\log n}). \tag{26}
$$

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 $(u, \theta) \in (0, 1) \times [0, \pi)$. In other words, $\max_{\{u_i, \theta_j\} \in \tilde{\mathcal{G}}_N} \left| \hat{f}_n(u_i, \theta_j) - \mathbb{E}(\hat{f}_n(u_i, \theta_j)) \right|$ and $\sup_{(u,\theta) \in (0,1) \times [0,\pi)} \left| \hat{f}_n(u, \theta) - \mathbb{E}(\hat{f}_n(u, \theta)) \right|$ have the same limiting distribution.

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. Furthermore, if the grid is too dense, the limiting distribution is different from our main result and is unknown to the best of our knowledge. 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 5.1 is indeed nearly optimal in this sense. Recall that in Theorem 5.1,

the interval between adjacent frequencies is of order $\delta_{\theta,n} = \Omega(1/B_n)$ and the averaged interval between two adjacent time indices 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 a sequence of slightly denser grids with $\delta_{\theta,n} = \mathcal{O}\left(\frac{1}{B_n(\log n)^\alpha}\right)$ and $\delta_{u,n} = \mathcal{O}\left(\frac{n}{N(\log n)^\alpha}\right)$ where α is any fixed positive constant, then the latter sequence of grids is asymptotically uniform variation matching. Since α can be chosen arbitrarily close to zero, the dense grids in Theorem 5.1 are nearly optimal.

Theorem 5.2. Under the assumptions of Theorem 5.1, 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 Appendix A.14 of the supplementary materials. □

5.2. Applications of the Simultaneous Confidence Regions

In this subsection, we illustrate several applications of the proposed SCR for joint time-frequency inference. These examples include testing time-varying white noise (Example 5.1), testing stationarity (Example 5.2), testing time-frequency separability or correlation stationarity (Example 5.3), and validating time-varying ARMA models (Example 5.4).

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 the above purposes (see Examples 5.3 and 5.4).

To implement the tests, observe 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 hypothesis 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 hypothesis converges faster than the SCR. The test achieves asymptotic power 1 for local alternatives whose evolutionary spectra deviate from the null hypothesis with a rate larger than the order of the width of the SCR.

Example 5.1 (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:

$$
H_0: \quad \forall \theta, \quad f(u, \theta) = g(u), \quad u \in [0, 1] \tag{27}
$$

for some time-varying function $g(\cdot)$. Consider the following optimization problem:

$$g_0(u) := \arg\min_{\tilde{g}} \frac{1}{\pi} \int_0^\pi |f(u,\theta) - \tilde{g}(u)|^2 \mathrm{d}\theta. \qquad (28)$$

That is, we would like to find a function of u which is closest to $f(u,\theta)$ in L_2 distance. Direct calculations show that $g_0(u) = \frac{1}{\pi}\int_0^\pi f(u,\theta)\mathrm{d}\theta$. Therefore, under the null hypothesis we can estimate the function g in Equation (27) by

$$\hat{g}(u) := \frac{1}{\pi} \int_0^\pi \hat{f}_n(u,\theta)\mathrm{d}\theta \approx \frac{1}{\pi} \int_0^\pi f(u,\theta)\mathrm{d}\theta = g_0(u). \qquad (29)$$

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

Example 5.2 (Testing stationarity). Under the null hypothesis that the time series is stationary, it is equivalent to testing

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

for some function $h(\cdot)$. Consider the following optimization problem:

$$h_0(\theta) := \arg\min_{\tilde{h}} \int_0^1 |f(u,\theta) - \tilde{h}(\theta)|^2 \mathrm{d}u. \qquad (31)$$

That is, we would like to find a function of θ which is closest to $f(u,\theta)$ in L_2 distance. Direct calculations show that $h_0(\theta) = \int_0^1 f(u,\theta)\mathrm{d}u$. Therefore, under the null hypothesis, we can estimate the function h in Equation (30) by

$$\hat{h}(\theta) := \int_0^1 \hat{f}_n(u,\theta)\mathrm{d}u \approx \int_0^1 f(u,\theta)\mathrm{d}u = h_0(\theta). \qquad (32)$$

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

Example 5.3 (Testing time-frequency separability or correlation stationarity). We call a nonstationary time series time-frequency separable if $f(u,\theta) = g(u)h(\theta)$ for some functions $g(\cdot)$ and $h(\cdot)$. If a nonstationary time series is time-frequency separable, the frequency curves across different times are parallel to each other. Similarly, the time curves across different frequencies are parallel to each other as well. Therefore, the property of time-frequency separability enables one to model the temporal and spectral behaviors of the time-frequency function separately. Furthermore, it can be verified that testing time-frequency separability is equivalent to testing correlation stationarity for locally stationary time series, that is, $\mathrm{corr}(X_i, X_{i+k}) = l(k)$, for some function $l(\cdot)$. Without loss of generality, we can formulate the null hypothesis as

$$H_0: \quad f(u,\theta) = C_0 g(u)h(\theta), \qquad (33)$$

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

$$\hat{C}_0 := \int_0^\pi \int_0^1 \hat{f}_n(u,\theta)\mathrm{d}u\mathrm{d}\theta \approx \int_0^\pi \int_0^1 f(u,\theta)\mathrm{d}u\mathrm{d}\theta = C_0, \qquad (34)$$

$$\hat{g}(u) := \frac{1}{\hat{C}_0} \int_0^\pi \hat{f}_n(u,\theta)\mathrm{d}\theta \approx \frac{1}{C_0} \int_0^\pi f(u,\theta)\mathrm{d}\theta = g(u), \qquad (35)$$

$$\hat{h}(\theta) := \frac{1}{\hat{C}_0} \int_0^1 \hat{f}_n(u,\theta)\mathrm{d}u \approx \frac{1}{C_0} \int_0^1 f(u,\theta)\mathrm{d}u = h(\theta), \qquad (36)$$

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

Example 5.4 (Validating time-varying ARMA models). Consider the null hypothesis that the time series follows the following time-varying ARMA model

$$H_0: \sum_{i=0}^p a_i(t/N)X_{t-i} = \sum_{j=0}^q b_j(t/N)\epsilon_{t-j}, \qquad (37)$$

where $a_0(u) = 1$, $a_i(\cdot), b_i(\cdot) \in \mathcal{C}^1[0,1]$, and ϵ_i are uncorrelated random variables with mean 0 and variance 1. Under the null hypothesis, $\{X_i\}$ is a locally stationary time series with spectral density

$$f(u,\theta) = \frac{1}{2\pi} \frac{\left|\sum_{j=0}^q b_j(u) \exp(\sqrt{-1}2\pi\theta j)\right|^2}{\left|\sum_{i=0}^p a_i(u) \exp(\sqrt{-1}2\pi\theta i)\right|^2}. \qquad (38)$$

The spectral density can be fitted using the generalized Whittle's method (Dahlhaus 1997), 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 $\mathcal{O}_\mathbb{P}(\sqrt{\log n}/\sqrt{n})$ which is faster than the rate $\mathcal{O}_\mathbb{P}(\sqrt{\log n}/\sqrt{n/B_n})$ by the STFT-based methods without prior information. Therefore, to test the fitted nonparametric time-varying ARMA model, we can plot the nonparametric spectral density using the estimated time-varying parameters $a_i(\cdot)$ and $b_i(\cdot)$. Under the null hypothesis, the nonparametric spectral density should fall within our SCR with the prescribed probability asymptotically.

The benefits of spectral domain approach to various hypothesis testing problems depend on the specific application. For example, for tests of stationarity, the test based on evolutionary spectral density is technically easier than the corresponding tests in the time domain. The main reason is that the time domain test needs to consider time-invariance of $r(u,k)$ for a diverging number of k and hence is a high-dimensional problem. On the other hand, the spectral domain test of stationarity only needs to check that $f(u,\theta)$ does not depend on u. Similar arguments

apply to the test of white noise. For another example, we proposed a frequency domain method for the problem of model validation of nonstationary linear models. However, technically it is difficult to approach this problem from the time domain. Furthermore, for many time series signals in engineering applications, the most important information is embedded in the frequency domain. Therefore, in engineering and signal processing applications, frequency domain methods are typically more favorable and are widely used. Therefore, frequency-domain-based tests are preferable in many such applications.

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 5.1. We discuss methods for tuning parameter selection in Section 6.2.

6.1. The Bootstrap Procedure

Although Theorem 5.1 shows that SCR can be constructed using the Gumbel distribution, the convergence rate in Theorem 5.1 is too slow to be useful in moderate samples. 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 iid $\mathcal{N}(0, 1)$ random variables. Defining

$$\hat{r}^\epsilon(u, k) := \frac{1}{n} \sum_{i=1}^{N} \tau\left(\frac{i - \lfloor uN \rfloor}{n}\right) \tau\left(\frac{i + k - \lfloor uN \rfloor}{n}\right) \epsilon_i \epsilon_{i+k} \tag{39}$$

and

$$\hat{f}_n^\epsilon(u, \theta) := \frac{1}{2\pi} \sum_{k=-B_n}^{B_n} \hat{r}^\epsilon(u, k) a(k/B_n) \exp(\sqrt{-1} k\theta), \tag{40}$$

it can be easily verified that the following analogy of Theorem 5.1 holds.

$$\mathbb{P}\left[\max_{(u,\theta)\in\mathcal{G}_N} \frac{n}{B_n} \frac{|\hat{f}_n^\epsilon(u,\theta) - \mathbb{E}(\hat{f}_n^\epsilon(u,\theta))|^2}{[f^\epsilon(u,\theta)]^2 \int_{-1}^{1} a^2(t)\mathrm{d}t} \right.$$
$$\left. - 2\log B_n - 2\log C_n + \log(\pi \log B_n + \pi \log C_n) \le x \right]$$
$$\to e^{-e^{-x/2}}. \tag{41}$$

Therefore, we propose to construct the 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 $\bar{f}_n^\epsilon(u, \theta)$ be the sample mean of $\{\hat{f}_{n,m}^\epsilon(u, \theta), m = 1, \ldots, N_{\mathrm{MC}}\}$ from the N_{MC} Monte Carlo experiments. Then we compute the empirical distribution of

$$\max_{(u,\theta)\in\mathcal{G}_N} \frac{|\hat{f}_{n,m}^\epsilon(u,\theta) - \bar{f}_n^\epsilon(u,\theta)|^2}{[\bar{f}_n^\epsilon(u,\theta)]^2}, \quad m = 1, \ldots, N_{\mathrm{MC}} \tag{42}$$

to approximate the distribution of

$$\max_{(u,\theta)\in\mathcal{G}_N} \frac{|f(u,\theta) - \hat{f}_n(u,\theta)|^2}{[\hat{f}_n(u,\theta)]^2}, \tag{43}$$

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

$$\mathbb{P}\left(\max_{(u,\theta)\in\mathcal{G}_N} \frac{|f(u,\theta) - \hat{f}_n(u,\theta)|^2}{[\hat{f}_n(u,\theta)]^2} \le \gamma_{1-\alpha}^2 \right) = 1 - \alpha. \tag{44}$$

Therefore, the constructed SCR is

$$\max\{0, (1 - \gamma_{1-\alpha})\hat{f}(u,\theta)\} \le f(u,\theta) \le (1 + \gamma_{1-\alpha})\hat{f}(u,\theta),$$
$$\forall (u, \theta) \in \mathcal{G}_N. \tag{45}$$

Note that in small sample cases, the lower bound 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

$$\frac{|f(u,\theta) - \hat{f}_n(u,\theta)|^2}{[\hat{f}_n(u,\theta)]^2} \approx [\log(f(u,\theta)/\hat{f}_n(u,\theta))]^2. \tag{46}$$

Then the SCR can be constructed as

$$\exp(-\gamma_{1-\alpha})\hat{f}_n(u,\theta) \le f(u,\theta) \le \exp(+\gamma_{1-\alpha})\hat{f}_n(u,\theta),$$
$$\forall (u, \theta) \in \mathcal{G}_N. \tag{47}$$

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 Equation (17).
4. Compute the SCR defined in Section 6.1 using the spectral density estimates and the critical value obtained by the bootstrap.

Note that the validity of the proposed bootstrap procedure is asymptotically justified by Theorem 5.1. On the other hand, theoretical justification for the superiority of the bootstrap procedure for moderate samples is extremely difficult, as it requires deriving higher order asymptotics of the maximum deviation of the time-varying spectral densities. We will investigate this problem in some future work.

6.2. Tuning Parameter Selection

Choosing B_n and n in practice is a nontrivial problem. In our Monte Carlo experiments and real data analysis, we find that the minimum volatility (MV) method (Politis, Romano, and Wolf 1999; Zhou 2013) 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 $N \leq 1000$, $n_l = 2.5N^\eta$ and $n_r = 4N^\eta$ if $1000 < N \leq 2000$ and $n_l = 3N^\eta$ and $n_r = 5N^\eta$ if $N > 2000$, where $\eta = 0.48$. Although the rule for setting n_l and n_r is ad-hoc, it works well in our simulations and data analysis. In practice, one can also either choose n_l and n_r based on prior knowledge of the data, or select them by visually evaluating the fitted evolutionary spectral densities. A reasonable value of n should not produce too rough or too smooth estimates of the spectral density. We remark that n_l and n_u are only upper and lower bounds of the candidate bandwidths. Simulations show that the simulated coverage probabilities are typically not sensitive to the choices of n_l and n_r. 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. We refer to Politis, Romano, and Wolf (1999) and Zhou (2013) for more detailed discussions of the MV method.

Note that cross-validation is another popular method for choosing bandwidths (Dahlhaus and Richter 2019). However, it is a difficult task to implement cross-validation in the context of time-varying spectral density estimation. Finally, it is well-known that choosing theoretically optimal bandwidths is an extremely difficult problem. We hope to investigate this problem in some future work.

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 perform real data analysis in Section 7.4. Throughout this section, the kernel $\tau(\cdot)$ is chosen to be a rescaled Epanechnikov kernel such that $\int \tau^2(x)dx = 1$, and the kernel $a(\cdot)$ is a rescaled tri-cube kernel such that $a(0) = 1$. The two kernel functions are defined as follows.

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

$$a(x) := \begin{cases} (1 - |x|^3)^3, & \text{if } |x| < 1, \\ 0, & \text{otherwise.} \end{cases} \tag{48}$$

In all the simulations, we ran Monte Carlo experiments for $N_{\text{MC}} = 10000$. The results for SCR and hypothesis testing are obtained by averaging over 1000 independent datasets.

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 Examples 7.1–7.5.

Example 7.1 (Time-varying AR model). We have

$$X_i = a(i/N)X_{i-1} + \epsilon_i, \tag{49}$$

where $\{\epsilon_i\}$ are iid $\mathcal{N}(0, 1)$. 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 $[0, 1]$. The simulated uncoverage probabilities of the SCR are shown in Table 1.

Example 7.2 (Time-varying ARCH model). Consider the following time-varying ARCH(1) model:

$$X_i = \epsilon_i \sqrt{a_0(i/N) + a_1(i/N)X_{i-1}^2}, \tag{50}$$

where $\{\epsilon_i\}$ are iid standard normally distributed random variables, $a_0(u) > 0, a_1(u) > 0$ and $a_0(u) + a_1(u) < 1$. Note that $\{X_i\}$ is a white noise sequence. In this example, we choose $a_0(u) = 0.6$ and $a_1(u) = 0.3\sin(\pi u)$. The simulated uncoverage probabilities of the SCR 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

$$X_i = \begin{cases} b_1(i/n)X_{i-1} + \epsilon_i, & \text{if } S_i = 0, \\ b_2(i/n)X_{i-1} + \epsilon_i, & \text{if } S_i = 1, \end{cases} \tag{51}$$

where $\{\epsilon_i\}$ are iid standard normally distributed random variables, $|b_1| < 1$, and $|b_2| < 1$. In this example, we choose $P = \begin{bmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{bmatrix}$, $b_1(u) = 0.4\cos(2\pi u)$, and $b_2(u) = 0.1\sin(2\pi u)$. The simulated uncoverage probabilities of the SCR are shown in Table 3.

Table 1. Simulated uncoverage probabilities 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

Table 2. Simulated uncoverage probabilities 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	0.02	0.05
72	32	0.04	0.08	0.04	0.08
72	28	0.05	0.10	0.05	0.11
54	36	0.03	0.06	0.03	0.07
54	32	0.05	0.09	0.04	0.09
54	28	0.05	0.10	0.04	0.10
36	32	0.06	0.12	0.06	0.12
36	28	0.08	0.15	0.08	0.14

Table 3. Simulated uncoverage probabilities 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	0.04	0.09	0.04	0.09
72	32	0.05	0.11	0.05	0.12
72	28	0.06	0.12	0.06	0.13
54	36	0.06	0.10	0.05	0.10
54	32	0.06	0.11	0.06	0.11
54	28	0.06	0.12	0.07	0.13
36	32	0.07	0.14	0.08	0.14
36	28	0.07	0.14	0.08	0.15

Table 4. Simulated uncoverage probabilities 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 uncoverage probabilities 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	0.07	0.14
36	28	0.09	0.16	0.09	0.17

Example 7.4 (Time-varying threshold AR model). Suppose $\{\epsilon_i\}$ are iid standard normally distributed random variables and consider the following threshold AR model

$$X_i = a(i/N) \max(0, X_{i-1}) + b(i/N) \max(0, -X_{i-1}) + \epsilon_i, \quad (52)$$

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 uncoverage probabilities of the SCR are shown in Table 4.

Example 7.5 (Time-varying bilinear process). Let $\{\epsilon_i\}$ be iid standard normally distributed random variables and consider

the following model

$$X_i = b(i/N)X_{i-1} + \epsilon_i + c(i/N)X_{i-1}\epsilon_{i-1}, \quad (53)$$

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 uncoverage probabilities of the SCR are shown in Table 5.

According to the results in Tables 1–5, one can see that the proposed bootstrap works well when B_n and n are chosen in a relatively wide range. 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 described in Section 6.2 to select the tuning parameters for Examples 7.1–7.5. For all examples, $N = 400$ and $N = 800$ are considered. The bootstrap accuracy is shown in Table 6. Furthermore, according to Equation (45), we also included the average width of the SCR over $(u, \theta) \in \mathcal{G}_N$ in Table 6. From Table 6, we can see that the coverage probabilities of the SCR with bandwidths selected by the MV method are accurate. Furthermore, the average width of the SCR decreases as N increases.

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 nonparametric ARMA model validation, which is given in Example 7.8.

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

$$X_i = \sigma_i \epsilon_i, \quad \sigma_i^2 = a_0(i/N) + a_1(i/N)X_{i-1}^2, \quad (54)$$

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 $N = 400$ and the other with $N = 800$, where n and B_n are selected by the MV method. We have shown the simulated Type I error rates of the SCR in Table 7. Next, we study the power of the hypothesis testing for stationarity using the proposed SCR by increasing δ. We study both 0.05 and 0.1 level tests. The simulated powers of the SCR for $N = 800$ and $N = 600$ are shown in Figure 1. One can see that, for both $N = 800$ and $N = 600$, the simulated Type I error rates of the SCR (when $\delta = 0$) are accurate. Furthermore, the simulated power of the SCR increases with N.

Table 6. Simulated uncoverage probabilities with tuning parameters selected by the MV method (numbers in the parentheses represent the average width of the SCR).

	N = 400				N = 800			
	n	B_n	$\alpha = 0.05$	$\alpha = 0.1$	n	B_n	$\alpha = 0.05$	$\alpha = 0.1$
Example 7.1 (Table 1)	54	30	0.06 (0.51)	0.10 (0.45)	72	30	0.05 (0.47)	0.09 (0.43)
Example 7.2 (Table 2)	52	32	0.05 (0.37)	0.09 (0.33)	69	30	0.04 (0.35)	0.09 (0.30)
Example 7.3 (Table 3)	54	32	0.06 (0.52)	0.11 (0.47)	72	32	0.05 (0.49)	0.12 (0.43)
Example 7.4 (Table 4)	50	32	0.06 (0.53)	0.12 (0.45)	70	32	0.06 (0.48)	0.12 (0.43)
Example 7.5 (Table 5)	52	32	0.04 (0.52)	0.09 (0.46)	69	31	0.06 (0.44)	0.11 (0.40)

Table 7. Simulated 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
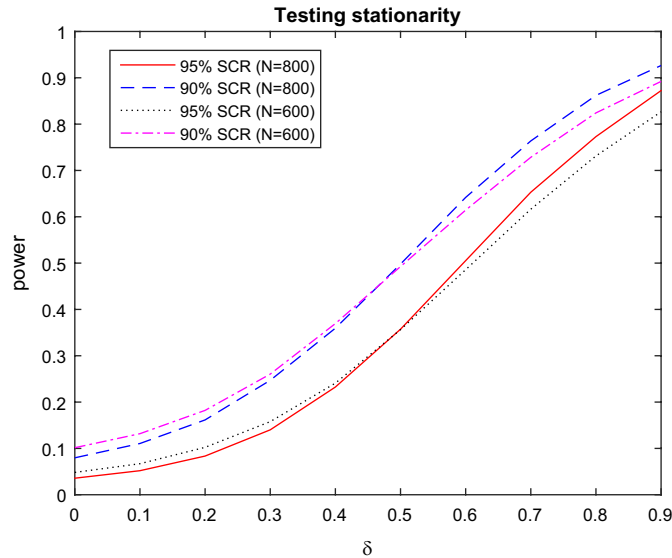
Figure 1. Simulated powers for testing stationarity for Example 7.6.

Figure 2. Simulated powers for testing TV white noise for Example 7.7.

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

$$X_i = a_0(i/N)\epsilon_i + a_1(i/N)\epsilon_{i-1}, \qquad (55)$$

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 a 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 $N = 800$ and the other with $N = 1200$. The tuning parameters n and B_n are selected by the MV method. We then test time-varying white noise using our proposed SCR by increasing δ for $N = 800$ and $N = 600$. The simulated powers of the SCR are shown in Figure 2. According to Figure 2, one can see that the simulated coverage probabilities for $N = 600$ are slightly below the nominal level. This is because the structure of the time series in this example is complicated. A sample size with $N = 600$ is not large enough for the local stationarity of the time series to be fully captured statistically. On the other hand, for sample size $N = 800$, the simulated Type I error rates are accurate and the powers are significantly higher than the case of $N = 600$.

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

$$\sum_{j=0}^{p} a_j(i/N)X_{i-j} = \sigma(i/N)\epsilon_i, \qquad (56)$$

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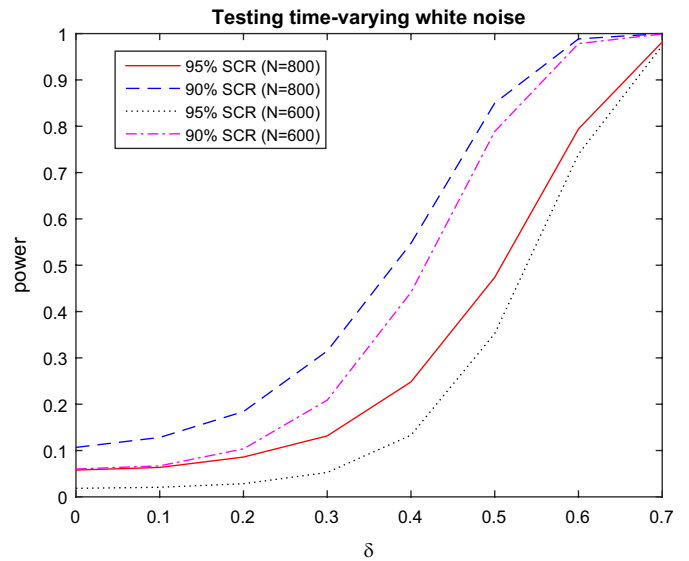
time series with spectral density

$$f(u, \theta) = \frac{\sigma^2(u)}{2\pi} \left| \sum_{j=0}^{p} a_j(u) \exp(\sqrt{-12}\pi\theta j) \right|^{-2}. \qquad (57)$$

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

For each generated time series, we fit a time-varying AR model with $p = 1$ by minimizing the local Whittle likelihood (Dahlhaus 1997). We can then test if the spectral density of the fitted nonparametric time-varying AR model falls into the proposed SCR. The simulated coverage probabilities of the SCR are shown in Table 7, where n and B_n are selected by the MV method. We can see that, under the null hypothesis, the nonparametric time-varying AR model is validated since the simulated Type I error rates match quite well with the nominal levels of the proposed SCR test.

7.4. Real Data Analysis

In this subsection, we present some real data analysis. We study an earthquake and explosion dataset from seismology in Example 7.9 and then daily SP500 return from finance in Example 7.10. Observe that all time series are relatively long with $N > 2000$. For tuning parameter selection, we use the MV method to search (n, B_n) within the region $B_n < n/\log(n)$. Hypothesis tests are performed, including testing stationarity, time-varying white noise, and time-frequency separability on all the datasets.

Example 7.9 (Earthquakes and explosions (Shumway and Stoffer 2017)). In this example, we study an earthquake signal and an explosion signal from a seismic recording station (Shumway and Stoffer 2017). The recording instruments in Scandinavia are observing earthquakes and mining explosions with one of each shown in Figures 3 and 4, respectively. The two time series (see Figures 3 and 4) each has length $N = 2048$ representing
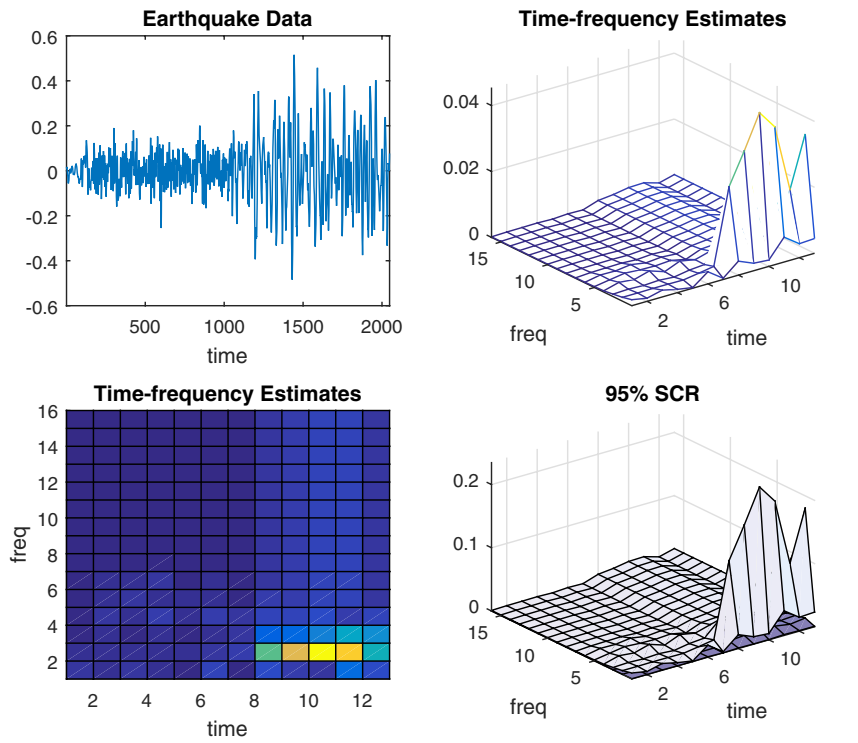
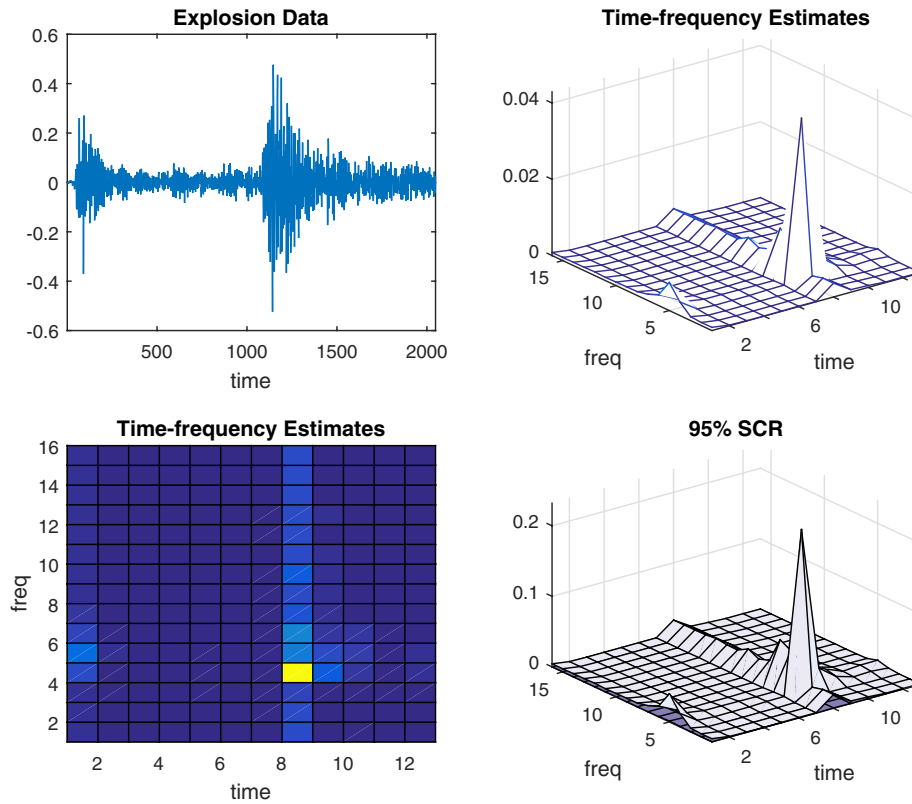**Figure 3.** Analysis of earthquake data.

Figure 4. Analysis of explosion data.

two phases or arrivals along the surface, denote by phase P: $\{X_i : i = 1, \ldots, 1024\}$ and phase S: $\{X_i : i = 1025, \ldots, 2048\}$. The general problem of interest is in distinguishing or discriminating between waveforms generated by earthquakes and those generated by explosions. The original data came from the technical report by Blandford (1993). According to Blandford (1993), the original earthquake and explosion signals have been filtered with a 3-pole highpass Butterworth filter with the corner frequency at 1 Hz to improve the signal-to-noise ratio. Then the amplitudes of the waveforms have been rescaled so the
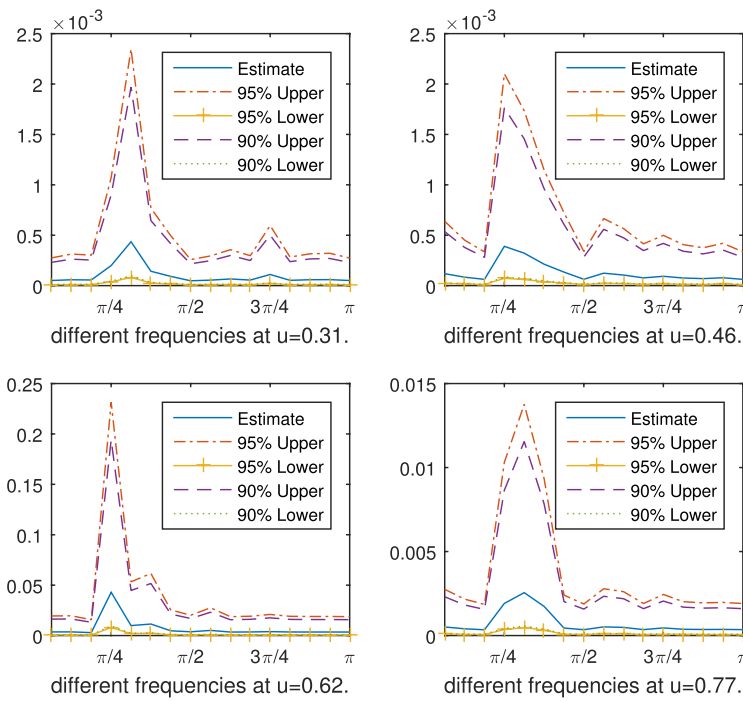
Figure 5. Explosion data: selected times.
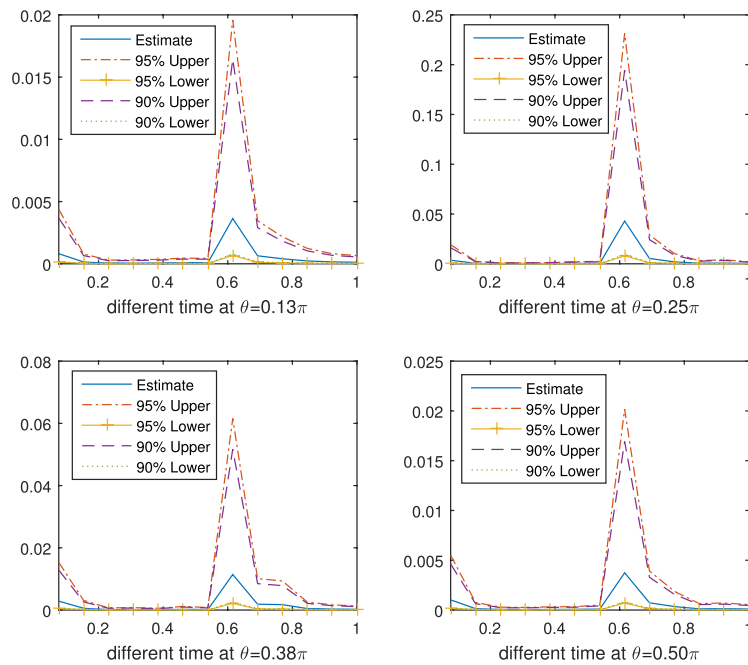
Figure 6. Explosion data: selected frequencies.

maximum amplitude for each signal is equal. According to Blandford (1993, Figure 2a and b), the unit for time is 0.02 second and the values of the earthquake and explosion data are rescaled to be no more than 1.

From the time domain (see Figures 3 and 4), one can observe that rough amplitude ratios of the first phase P to the second phase S are different for the two datasets, 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 Figure 3) shows power at the low frequencies only, and the power remains strong for a long time.

In contrast, the explosion (see Figure 4) 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 times and frequencies that the spectral density of explosion has the similar shape at different times, as well as at different frequencies (see Figures 5 and 6). However, the spectral density of earthquakes does not seem to have this property (see Figures 7 and 8). This may suggest that the explosion data are correlation stationary or time-frequency separable. We further perform hypothesis tests on both datasets to confirm our observation
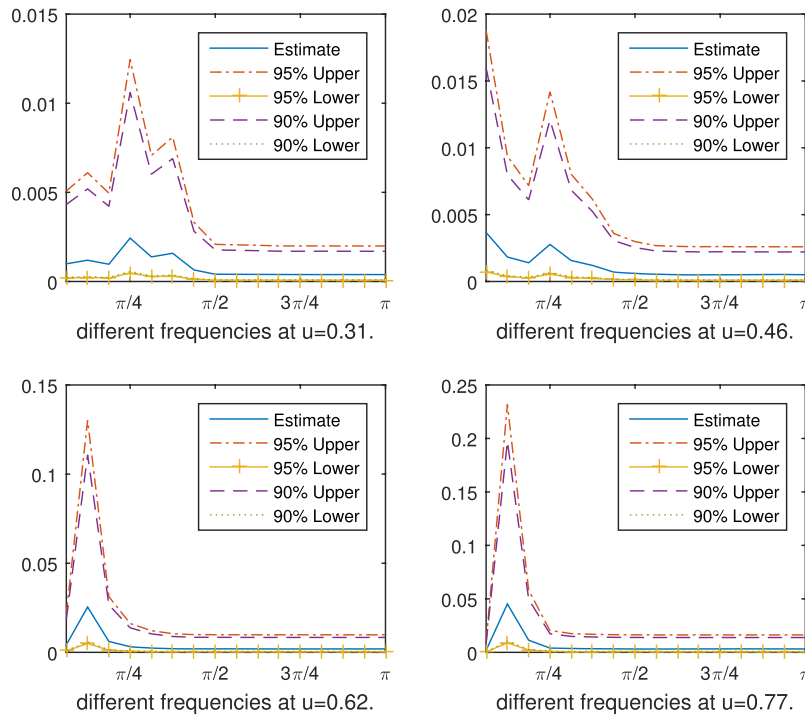
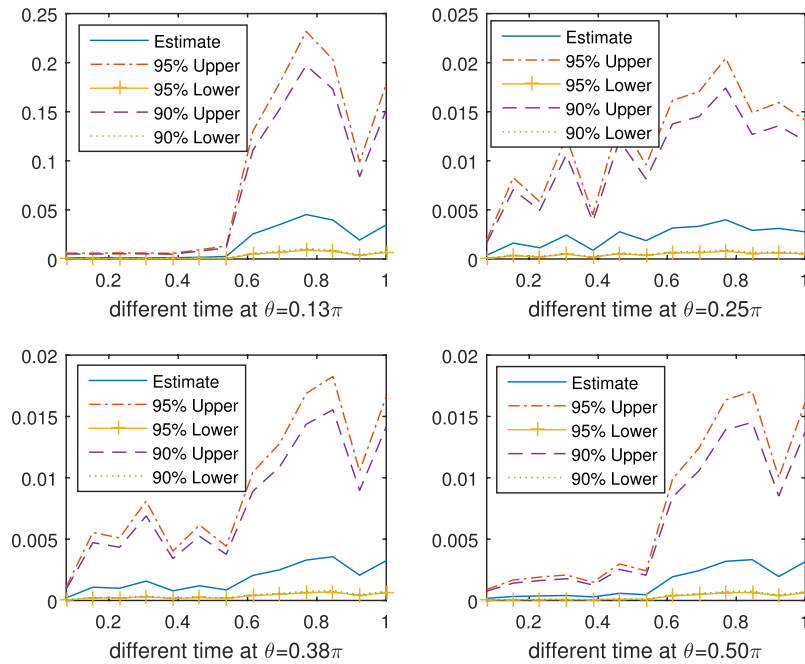**Figure 7.** Earthquake data: selected times.

Figure 8. Earthquake data: selected frequencies.

(see Table 8). The *p*-values for testing stationarity and time-varying white noise for both earthquake and explosion are quite small, which implies that 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.

There are two main benefits from knowing that explosion time series are time-frequency separable but earthquake time series are not. First, this reveals an important structural property of the time-frequency behavior for explosion signals. Since time-frequency separability implies the time curves for different frequencies are parallel and the frequency curves for different times are parallel as well, this directly suggests a parsimonious model for explosion time series using two one-dimensional models. Second, for the classification of earthquake and explosion signals, time-frequency separability provides a nonlinear

Table 8. Real data: p-values for testing (a) stationarity, (b) time-varying white noise, (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: $(***) < 0.001 \leq (**) < 0.01 \leq (*) < 0.05 \leq (+) < 0.1$.

feature of the explosion that could potentially serve the purpose. Since most commonly used features for classification are linear features, time-frequency separability is potentially important for feature extraction to improve the accuracy in classification tasks. However, since we only have analyzed one pair of earthquake and explosion signals, further studies with a large database of earthquake and explosion signals are needed to confirm this property for explosions which we leave to a future work.

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 Figure 9. Observing that the SCR in Figure 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 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, that is, 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 way is to fit the absolute returns by an appropriate nonstationary linear model, then apply the fitted model to forecast the future volatility. To date, to our knowledge, there exists no methodology for validating nonstationary linear models. In the following, we demonstrate that the proposed SCR is a useful tool for validating nonstationary linear models for absolute SP500 daily returns.

We first remove the local mean of the original SP500 time series by kernel smoothing. The spectral density estimates and the SCRs are shown in Figures 10–12. We observe from the plots that the spectral density of the absolute SP500 returns behaves quite differently from the original SP500 time series. For example, unlike the case for the original SP500 time series, the SCR for the absolute SP500 in Figure 11 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-value for testing time-varying white noise is 0.037, which is much smaller than that of the original SP500 time series. 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.
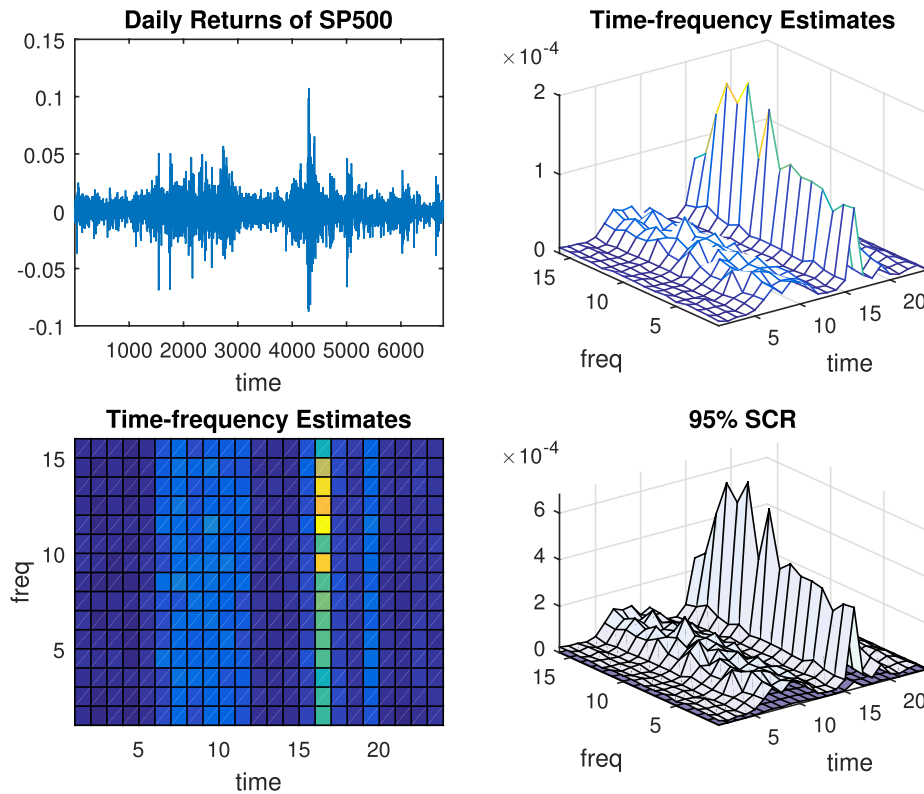

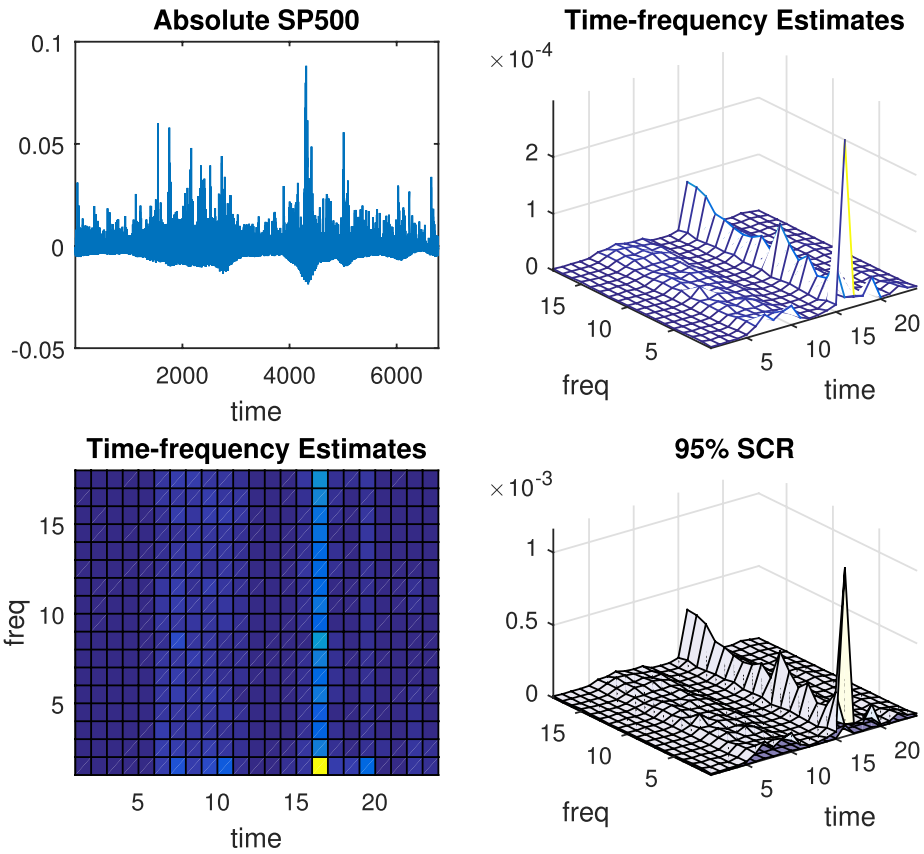**Figure 9.** Analysis of daily returns of SP500.

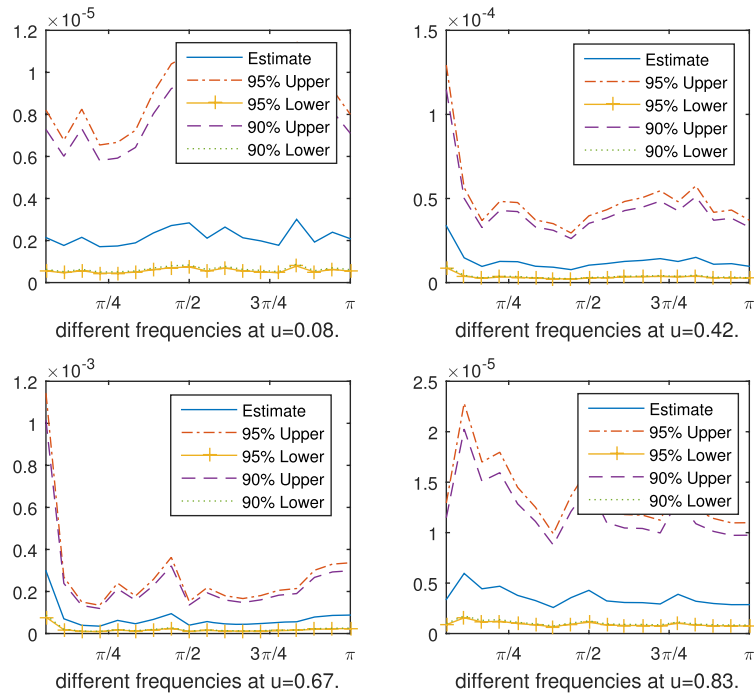**Figure 10.** Analysis of absolute SP500 returns.

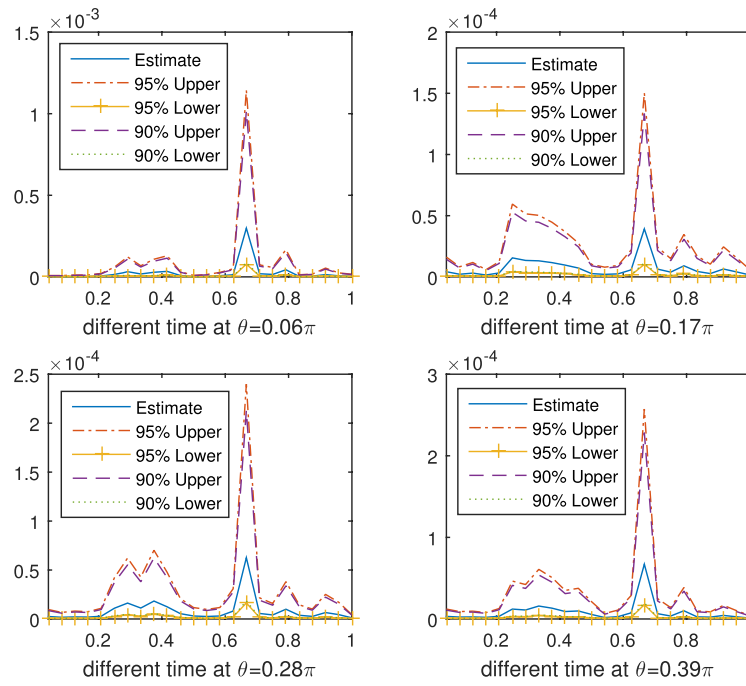**Figure 11.** Absolute SP500 return: selected times.

Figure 12. Absolute SP500 return: selected frequencies.

Table 9. p-values for validating time-varying ARMA 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: $(***) < 0.001 \leq (**) < 0.01 \leq (*) < 0.05 \leq (+) < 0.1$.

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

$$\sum_{i=0}^{p} a_i(t/N)X_{t-i} = \sum_{j=0}^{q} b_j(t/N)\epsilon_{t-j} \qquad (58)$$

to the absolute returns by minimizing the local Whittle likelihood (Dahlhaus 1997). We then validate if the fitted spectral densities from the time-varying AR or ARMA models fall 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 the tv-ARMA(1, 1) model equals 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 the tv-ARMA(2, 1) model is 0.79. This interesting observation suggests that the tv-ARMA(2, 1) model may be appropriate to fit the absolute returns. We further plot the 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 Figure 13. From Figure 13, one can see that the fitted spectral 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 the tv-ARMA(1, 1) model are not close to the STFT-based spectral density estimates either. Therefore, based on the proposed SCR, we conclude that the tv-ARMA(2, 1) model is an appropriate candidate for the analyzed data and can be used for short-term future volatility forecasting.

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 $\Omega_p = \{c \in \mathbb{R}^p : |c| = 1\}$, $Z_{u,J} = (Z_{u,j_1}^{(n)}, \ldots, Z_{u,j_p}^{(n)})^T$ for $J = (j_1, \ldots, j_p)$ satisfies $1 \leq j_1, \ldots, j_p \leq 2m$ (recall that $m = \lfloor (n-1)/2 \rfloor$). We prove for any fixed $p \in \mathbb{N}$, as $n \to \infty$, we have that

$$\sup_u \sup_J \sup_{c \in \Omega_p} \sup_x |P(c^T Z_{u,J} \leq x) - \Phi(x)| = o(1). \qquad (59)$$

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.

8.1.1. Proof of Equation (59)

We denote $2\pi j/n$ by θ_j in this proof. Without 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 $\mu_{u,k} \leq \sum_{\ell=1}^{p} \frac{|c_\ell|}{\sqrt{\pi f_*}} \leq$
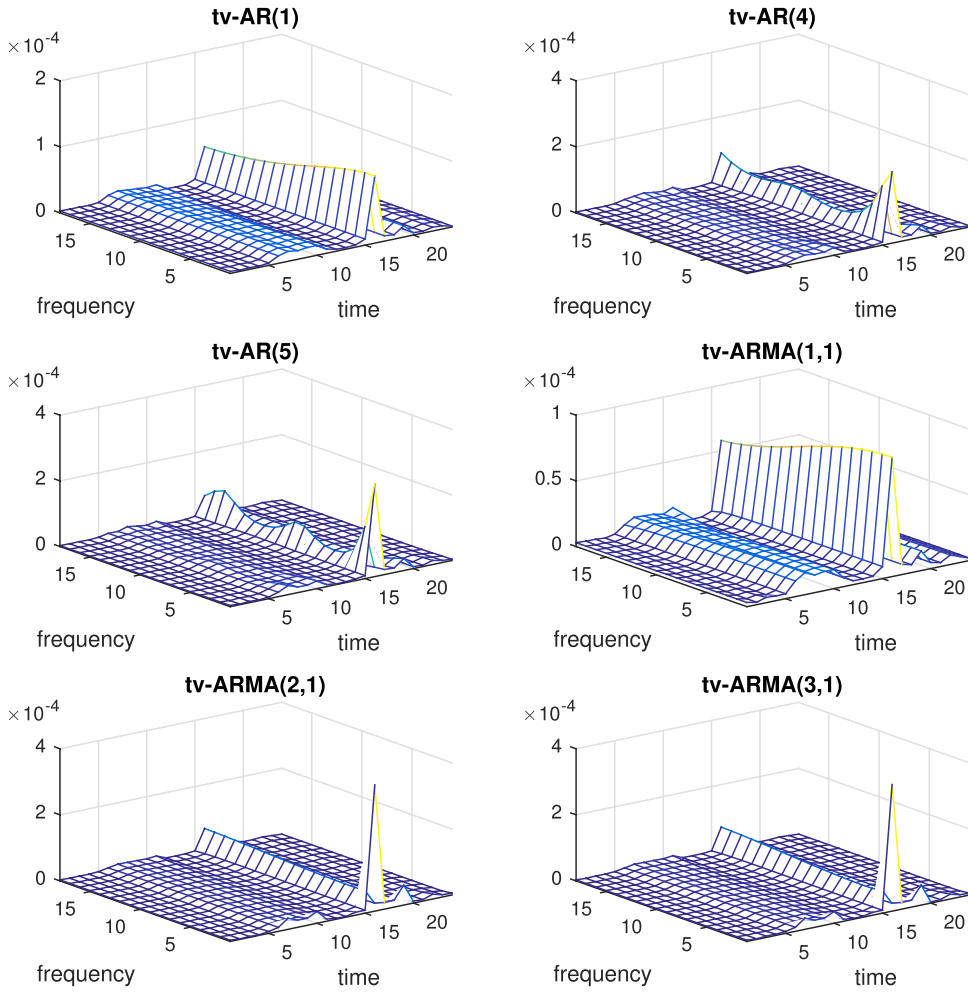
Figure 13. Fitting absolute SP500 daily returns to time-varying ARMA models.

$\frac{p}{\sqrt{\pi f_*}} =: \mu_*, \quad \forall c \in \Omega_p, \quad \forall J.$ Furthermore, defining

$$T_{u,n} := \sum_{k=1}^{N} \mu_{u,k} \tau \left(\frac{k - \lfloor uN \rfloor}{n} \right) X_k,$$

$$\tilde{T}_{u,n} := \sum_{k=1}^{N} \mu_{u,k} \tau \left(\frac{k - \lfloor uN \rfloor}{n} \right) \tilde{X}_k^{[\ell]}, \tag{60}$$

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. Under the assumptions of Theorem 3.1, we have

$$\lim_{n \to \infty} \sup_J \sup_c \sup_u \left| \frac{\| T_{u,n} \|^2}{n} - 1 \right|^2 = 0. \tag{61}$$

Proof. See Appendix A.1 of the supplementary materials. ☐

Lemma 8.2. Under the assumptions of Theorem 3.1, we have

$$\lim_{\ell \to \infty} \sup_J \sup_c \sup_u \frac{\| T_{u,n} - \tilde{T}_{u,n} \|}{\sqrt{n}} = 0. \tag{62}$$

Proof. See Appendix A.2 of the supplementary materials. ☐

Lemma 8.3. Under the assumptions of Theorem 3.1, we have

$$\sup_x \left| \mathbb{P} \left(\frac{T_{u,n}}{\sqrt{n}} \leq x \right) - \Phi \left(\frac{x}{\| T_{u,n} \| / \sqrt{n}} \right) \right|$$
$$= \mathcal{O} \left(\mathbb{P} \left(\left| \frac{T_{u,n} - \tilde{T}_{u,n}}{\sqrt{n}} \right| \geq \eta \right) + \delta_n + \eta^2 \right), \tag{63}$$

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

Proof. See Appendix A.3 of the supplementary materials. ☐

Using the above results, we can then prove Equation (59) as follows. First, by Chebyshev inequality and $\eta = \left(\frac{\| T_{u,n} - \tilde{T}_{u,n} \|}{\sqrt{n}} \right)^{1/2}$, we have

$$\mathbb{P} \left(\left| \frac{T_{u,n} - \tilde{T}_{u,n}}{\sqrt{n}} \right| \geq \eta \right) \leq \frac{\mathbb{E}(T_{u,n} - \tilde{T}_{u,n})^2 / n}{\eta^2} = \eta^2. \tag{64}$$

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

$$\sup_x \left| \mathbb{P} \left(\frac{T_{u,n}}{\sqrt{n}} \leq x \right) - \Phi \left(\frac{x}{\| T_{u,n} \| / n} \right) \right|$$
$$\to \sup_x \left| \mathbb{P} \left(\frac{T_{u,n}}{\sqrt{n}} \leq x \right) - \Phi(x) \right|. \tag{65}$$

By Lemma 8.3, we have that

$$\sup_x \left| \mathbb{P}\left(\frac{T_{u,n}}{\sqrt{n}} \leq x \right) - \Phi\left(\frac{x}{\|T_{u,n}\|/n} \right) \right| = \mathcal{O}(2\eta^2 + \delta_n). \quad (66)$$

Note that $\delta_n \to 0$ as $n \to \infty$. Also, by Lemma 8.2, 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 that $\sup_x \left| \mathbb{P}\left(\frac{T_{u,n}}{\sqrt{n}} \leq x \right) - \Phi(x) \right| \to 0$, 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 Equation (60) as $T_{u_i,n,c^{(i)}}$ and $\tilde{T}_{u_i,n,c^{(i)}}$. Then by Lemma 8.2, it suffices to show that $\{\tilde{T}_{u_i,n,c^{(i)}}, i = 1, \ldots, q\}$ 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 $Y_{u,i} := Y_{u,i}(\theta) = \frac{1}{2\pi} \sum_{k=-B_n}^{B_n} X_{u,i} X_{u,i+k} a(k/B_n) \cos(k\theta)$, $g_n(u, \theta) := \sum_{i=1}^n Y_{u,i}(\theta)$, and $h_n(u, \theta) := \frac{1}{\sqrt{nB_n}} g_n(u, \theta) - \sqrt{n/B_n} \hat{f}_n(u, \theta)$, we have that

$$\sqrt{n/B_n}\{\hat{f}_n(u, \theta) - \mathbb{E}(\hat{f}_n(u, \theta))\}$$
$$= \frac{g_n(u, \theta) - \mathbb{E}(g_n(u, \theta))}{\sqrt{nB_n}} - h_n(u, \theta) + \mathbb{E}(h_n(u, \theta)). \quad (67)$$

Next, denote $\tilde{X}_k^{[\ell]}$ as the ℓ-dependent conditional expectation of X_k, $\tilde{X}_{u,i}^{[\ell]}$ as the ℓ-dependent conditional expectation of $X_{u,i}$, and $\tilde{Y}_{u,i}$ as the correspondence of sum using $\tilde{X}_{u,i}^{[\ell]}$ 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 $0 < \rho < 1$, then

$$\sup_\theta \sup_u \|h_n(u, \theta)\| = (nB_n)^{-1/2} \mathcal{O}(B_n), \quad (68)$$

$$\sup_\theta \sup_u \sup_i \|Y_{u,i} - \tilde{Y}_{u,i}\| = \mathcal{O}(B_n \rho^{\ell/4}), \quad (69)$$

$$\sup_\theta \sup_u \|g_n(u, \theta) - \tilde{g}_n(u, \theta)\| = o(1). \quad (70)$$

Proof. See Appendix A.4 of the supplementary materials. □

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

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

$$V_{u,r}(\theta) := \sum_{i=(r-1)(p_n+q_n)+p_n+1}^{r(p_n+q_n)} \tilde{Y}_{u,i}(\theta), \quad 1, \ldots, k_n, \quad (71)$$

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 $U_{u,r}(\theta), r = 1, \ldots, k_n$ are independent (not identically distributed) block sums with block length p_n, and $V_{u,r}(\theta), r = 1, \ldots, k_n - 1$ are independent block sums with block length q_n. Define $U'_{u,r}(\theta) := U_{u,r}(\theta)\mathbf{1}(|U_{u,r}(\theta)| \leq d_n)$ 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 that

$$\sup_u \mathbb{E}(\max_\theta |V_{u,k_n}(\theta)|) = \mathcal{O}(\sqrt{p_n \ell_n} B_n), \quad (72)$$

$$\sup_u \mathbb{E}(\max_\theta |h_n(u, \theta)|) = o(1), \quad (73)$$

$$\sup_u \max_r \max_\theta \text{var}(U_{u,r}(\theta)) = \mathcal{O}(p_n B_n). \quad (74)$$

Furthermore, we have that

$$\text{var}(U'_{u,r}(\theta)) = \text{var}(U_{u,r}(\theta))[1 + o(1)], \quad (75)$$

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

Proof. See Appendix A.5 of the supplementary materials. □

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

$$\sup_u \sup_i \sup_\theta \|U_{u,i}(\theta)\|_{2+\delta/2} = \mathcal{O}(\ell_n \sqrt{p_n B_n}). \quad (76)$$

Proof. See Appendix A.6 of the supplementary materials. □

Using the previous results Equations (68), (70), and (73), we have that

$$\sup_u \max_\theta \sqrt{n/B_n} |\hat{f}_n(u, \theta) - \mathbb{E}(\hat{f}_n(u, \theta))|$$
$$\leq \frac{\sup_u \max_\theta |\tilde{g}_n(u, \theta) - \mathbb{E}(\tilde{g}_n(u, \theta))| + o(1)}{\sqrt{nB_n}}$$
$$+ \mathcal{O}_{\mathbb{P}}(\sqrt{B_n/n}) + o_{\mathbb{P}}(1)$$
$$\leq \frac{\sup_u \max_\theta |\sum_{r=1}^{k_n} U_{u,r}(\theta) - \mathbb{E}(\sum_{r=1}^{k_n} U_{u,r}(\theta))|}{\sqrt{nB_n}}$$
$$+ \frac{\sup_u \max_\theta |\sum_{r=1}^{k_n-1} V_{u,r}(\theta) - \mathbb{E}(\sum_{r=1}^{k_n-1} V_{u,r}(\theta))|}{\sqrt{nB_n}}$$
$$+ \frac{\sup_u \max_\theta |V_{u,k_n}(\theta) - \mathbb{E}(V_{u,k_n}(\theta))|}{\sqrt{nB_n}}$$
$$+ \mathcal{O}_{\mathbb{P}}(\sqrt{B_n/n}) + o_{\mathbb{P}}(1).$$
$$(77)$$

First, we can show that the third term of the right-hand side of Equation (77) is $o_\mathbb{P}(\sqrt{\log n})$. This is because by Equation (72), it suffices to show $\frac{\sqrt{p_n \ell_n B_n}}{\sqrt{nB_n}} = o(\sqrt{\log n})$ and this can be easily verified using $p_n = n^{1-4\eta/\delta}(\log n)^{-8/\delta-4}$, $B_n = \mathcal{O}(n^\eta)$ and $\delta \leq 4$.

Next, we show that the right-hand side of the first two terms of Equation (77) have an order of $\mathcal{O}_\mathbb{P}(\sqrt{\log n})$. 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 Woodroofe and Ness (1967, Corollary 2.1).

Lemma 8.7. Let $p(\lambda) = \sum_{v=-k}^{k} \alpha_v \exp(iv\lambda)$ be a trigonometric polynomial. Let $\lambda_i = \pi(i/rk), |i| \leq rk$. Then $\max_{|\lambda| \leq \pi} |p(\lambda)| \leq \max_{|i| \leq rk} |p(\lambda_i)/(1 - 3\pi r^{-1})|$.

Proof. See Woodroofe and Ness (1967, Corollary 2.1). □

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

$$\max_\theta |H_{u,n}(\theta)| \leq \frac{1}{1 - 3\pi/\log(B_n)} \max_{j \leq t_n} |H_{u,n}(\theta_j)|. \tag{78}$$

By Equations (74) and (75), there exists a constant C_1 such that

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

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

$$\mathbb{P}(\max_{0 \leq j \leq t_n} |H'_{u,n}(\theta_j)| \geq 4\alpha_n) \leq \sum_{j=0}^{t_n} \mathbb{P}(|H'_{u,n}(\theta_j)| \geq 4\alpha_n). \tag{79}$$

Then we apply Bernstein's inequality (see Lemma A.3 in the supplementary materials) to $\mathbb{P}(|H'_{u,n}(\theta_j)| \geq 4\alpha_n)$. This leads to, uniformly over u and θ_j,

$$\mathbb{P}(|H'_{u,n}(\theta_j)| \geq 4\alpha_n) \leq \exp\left(\frac{-16\alpha_n^2}{2k_n C_1 p_n B_n + \frac{8}{3}d_n\alpha_n}\right)$$
$$\leq C \exp\left(-\frac{nB_n \log n}{nB_n}\right). \tag{80}$$

Therefore, uniformly over u, we have that $\mathbb{P}(\max_{0 \leq j \leq t_n} |H'_{u,n}(\theta_j)| \geq 4\alpha_n) = \mathcal{O}(t_n)\mathcal{O}(1/n) = o(1)$. 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

$$\mathbb{P}(\max_{0 \leq j \leq t_n} |H^*_{u,n}(\theta_j)| \geq 4\alpha_n) \leq \sum_{j=0}^{t_n} \mathbb{P}(|H^*_{u,n}(\theta_j)| \geq 4\alpha_n)$$
$$\leq \sum_{j=0}^{t_n} \frac{\sum_{i=1}^{k_n} \text{var}(U^*_{u,i}(\theta_j))}{16\alpha_n^2}. \tag{81}$$

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

$$\text{var}(U^*_{u,i}\mathbf{1}_{|U^*_{u,i}|>d_n}) = d_n^2\text{var}\left(\frac{U^*_{u,i}}{d_n}\mathbf{1}_{|U^*_{u,i}|>d_n}\right)$$
$$\leq d_n^2\mathbb{E}\left[\left(\frac{U^*_{u,i}}{d_n}\right)^{2+\delta/2}\right], \tag{82}$$

we have that

$$\sum_{j=0}^{t_n} \frac{\sum_{i=1}^{k_n} \text{var}(U^*_{u,i}(\theta_j))}{16\alpha_n^2} = \mathcal{O}\left(\frac{t_n k_n(\sqrt{p_n B_n}\ell_n)^{2+\delta/2}}{\alpha_n^2 d_n^{\delta/2}}\right)$$
$$= \mathcal{O}\left(\frac{(B_n \log B_n)(n/p_n)(\sqrt{p_n B_n} \log n)^{2+\delta/2}}{(nB_n \log n)(nB_n)^{\delta/4}(\log n)^{-\delta/4}}\right) \tag{83}$$
$$= \mathcal{O}\left(\frac{(p_n B_n)^{1+\delta/4}(\log n)^{2+\delta/2}}{p_n(nB_n)^{\delta/4}(\log n)^{-\delta/4}}\right)$$
$$= \mathcal{O}(p_n^{\delta/4}(B_n/n)^{\delta/4}(\log n)^{2+\delta/2+\delta/4}).$$

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,

$$\sum_{j=0}^{t_n} \frac{\sum_{i=1}^{k_n} \text{var}(U^*_{u,i}(\theta_j))}{16\alpha_n^2} = \mathcal{O}\left(\frac{t_n k_n(\sqrt{p_n B_n}\ell_n)^{2+\delta/2}}{\alpha_n^2 d_n^{\delta/2}}\right)$$
$$= \mathcal{O}(n^{-\eta}B_n^{\delta/4}(\log n)^{-\delta/4}). \tag{84}$$

Finally, $B_n = \mathcal{O}(n^\eta), \delta \leq 4$ implies $B_n^{\delta/4} = \mathcal{O}(n^\eta)$, so we have that $\sum_{j=0}^{t_n} \frac{\sum_{i=1}^{k_n} \text{var}(U^*_{u,i}(\theta_j))}{16\alpha_n^2} = o(1)$. 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 $\max_\theta |H_{u,n}(\theta)| = \max_\theta |H'_{u,n}(\theta) + H^*_{u,n}(\theta)| = \mathcal{O}_\mathbb{P}(\alpha_n) = \mathcal{O}_\mathbb{P}(\sqrt{nB_n \log n})$. So Equation (77) has the order of $\mathcal{O}_\mathbb{P}(\sqrt{\log n})$.

8.3. Proof of Theorem 4.2

Throughout the proof, we use $\|\cdot\|$ to denote $\|\cdot\|_2$ for simplicity. We define $Y_{u,i}, g_n, h_n, \tilde{X}_k^{[\ell]}, \tilde{Y}_{u,i}, \tilde{g}_n$ 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

$$U_{u,r}(\theta) := \sum_{i=(r-1)(p_n+q_n)+1}^{(r-1)(p_n+q_n)+p_n} \tilde{Y}_{u,i}(\theta),$$
$$V_{u,r}(\theta) := \sum_{i=(r-1)(p_n+q_n)+p_n+1}^{r(p_n+q_n)} \tilde{Y}_{u,i}(\theta), \quad 1,\ldots,k_n, \tag{85}$$

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 identically 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 that

$$\left\|\sum_{i=-s_n/2}^{s_n/2} \{Y_{u,i}(\theta) - \mathbb{E}(Y_{u,i}(\theta))\}\right\|^2 \sim s_n B_n \sigma_u^2(\theta), \tag{86}$$

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

Proof. See Appendix A.7 of the supplementary materials. ☐

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

$$
\begin{aligned}
\|U_{u,r} - \mathbb{E}(U_{u,r})\| &= \left\| \sum_{j \in \mathcal{L}_r} \{\tilde{Y}_{u,j} - \mathbb{E}(\tilde{Y}_{u,j})\} \right\| \\
&= \left\| \sum_{j \in \mathcal{L}_r} \{Y_{u,j} - \mathbb{E}(Y_{u,j})\} \right\| \\
&\quad + \mathcal{O}\left(\sum_{j \in \mathcal{L}_r} \|Y_{u,j} - \tilde{Y}_{u,j}\| \right) \\
&\sim (p_n B_n \sigma_u^2)^{1/2} + \mathcal{O}(p_n B_n \rho^{\ell_n/4}) \\
&\sim (p_n B_n \sigma_u^2)^{1/2},
\end{aligned}
\tag{87}
$$

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 that $\|V_{u,r} - \mathbb{E}(V_{u,r})\| \sim (q_n B_n \sigma_u^2)^{1/2} + \mathcal{O}(q_n B_n \rho^{\ell_n/4})$. Then, since $q_n = o(p_n)$, we have that

$$
\begin{aligned}
\mathrm{var}\left(\sum_{r=1}^{k_n - 1} V_{u,r} + V_{u,k_n} \right) &= (k_n - 1)\mathcal{O}(q_n B_n \sigma_u^2) \\
&\quad + \mathcal{O}((p_n + q_n)B_n) = o(nB_n),
\end{aligned}
\tag{88}
$$

which implies that $\frac{\sum_r (V_{u,r} - \mathbb{E}(V_{u,r}))}{\sqrt{nB_n}} \Rightarrow 0$. Also, by Equation (68), we have that $\mathrm{var}(h_n(u,\theta)) = \mathcal{O}(B_n/n) = \mathcal{O}((\log n)^{-2-8/\delta})$, which implies that $h_n(u,\theta) - \mathbb{E}(h_n(u,\theta)) \Rightarrow 0$. Therefore, by

$$
\begin{aligned}
\sqrt{n/B_n}\{\hat{f}_n(u,\theta) - \mathbb{E}(\hat{f}_n(u,\theta))\} &= \frac{g_n(u,\theta) - \mathbb{E}(g_n(u,\theta))}{\sqrt{nB_n}} \\
&\quad - h_n(u,\theta) + \mathbb{E}(h_n(u,\theta)),
\end{aligned}
\tag{89}
$$

we only need to show that $\frac{\sum_r (U_{u,r} - \mathbb{E}(U_{u,r}))}{\sqrt{nB_n}} \Rightarrow \mathcal{N}(0, \sigma_u^2)$. We can check the conditions of Lemma A.2 of the supplementary materials (the Berry–Esseen lemma) as follows.

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

$$
\sum_r \frac{\|U_{u,r} - \mathbb{E}(U_{u,r})\|^2}{nB_n} \sim k_n \frac{p_n B_n \sigma_u^2}{nB_n} \sim \sigma_u^2.
\tag{90}
$$

By Lemma 8.6, we know $\|U_{u,r}\|_{2+\delta/2} = \mathcal{O}(\ell_n \sqrt{p_n B_n})$, which implies

$$
\begin{aligned}
\sum_r \frac{\|U_{u,r} - \mathbb{E}(U_{u,r})\|_{2+\delta/2}^{2+\delta/2}}{(nB_n)^{1+\delta/4}} &= \mathcal{O}\left(k_n \frac{(\ell_n \sqrt{p_n B_n})^{2+\delta/2}}{(nB_n)^{1+\delta/4}} \right) \\
&= \mathcal{O}(\ell_n k_n^{-\delta/4}).
\end{aligned}
\tag{91}
$$

Note that $k_n = \lfloor n/(p_n + q_n) \rfloor \sim n\psi^{-2/3}B_n^{-1/3} \sim n^{1/3}(\log n)^{(4/3+16/3\delta)}B_n^{-1/3}$, which implies $k_n^{-1} = \mathcal{O}((\log n)^{-4/3-16/3\delta})$. Then $\ell_n k_n^{-\delta/4} = \mathcal{O}((\log n)(\log n)^{(-\delta/3-4/3)}) = \mathcal{O}((\log n)^{(-\delta/3-1/3)}) \to 0$. Therefore, the result holds by Lemma A.2 of the supplementary materials.

8.4. Proof of Theorem 5.1

Define $D_n = C_n B_n$, $\theta_i = \frac{i\pi}{B_n}, i = 0, \ldots, B_n$, and $\alpha_{n,k} = a(k/B_n)\cos(k\theta)$. We use the previous definitions of $X_{u,k}$ and the ℓ-dependent $\tilde{X}_{u,k}^{[\ell]}$ as 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

$$
\begin{aligned}
&\hat{f}_n(u,\theta) - \mathbb{E}(\hat{f}_n(u,\theta)) \\
&= \frac{1}{2\pi n} \sum_{1 \leq k,k' \leq n} \alpha_{n,k-k'}[X_{u,k}X_{u,k'} - \mathbb{E}(X_{u,k}X_{u,k'})] \\
&= \frac{1}{2\pi n} \left(g_n(u,\theta) + \sum_{k=1}^n (X_{u,k}^2 - \mathbb{E}X_{u,k}^2) \right).
\end{aligned}
\tag{92}
$$

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

$$
\begin{aligned}
\bar{g}_n &= 2 \sum_{1 \leq s < k \leq n} \alpha_{n,k-s}[\bar{X}_{u,k}\bar{X}_{u,s} - \mathbb{E}(\bar{X}_{u,k}\bar{X}_{u,s})] \\
&= 2 \sum_{k=2}^n \bar{X}_{u,k} \sum_{s=1}^{k-1} \alpha_{n,k-s}\bar{X}_{u,s} - 2\mathbb{E} \sum_{k=2}^n \bar{X}_{u,k} \sum_{s=1}^{k-1} \alpha_{n,k-s}\bar{X}_{u,s}.
\end{aligned}
\tag{93}
$$

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

Lemma 8.9. Under the assumptions of Theorem 5.1, we have $\max_{u \in \mathcal{U}} \max_{0 \leq i \leq B_n} \mathbb{E}|g_n(u,\theta_i) - \tilde{g}_n(u,\theta_i)| = o(n^{1+\gamma}\rho^{\lfloor n^\gamma \rfloor})$ and $\max_{u \in \mathcal{U}} \max_{0 \leq i \leq B_n} \frac{|g_n(u,\theta_i) - \tilde{g}_n(u,\theta_i)|}{\sqrt{nB_n}} = o_{\mathbb{P}}(1)$.

Proof. See Appendix A.8 of the supplementary materials. ☐

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

Lemma 8.10. Under the assumptions of Theorem 5.1, we have that

$$
\mathbb{E}\left(\max_{u \in \mathcal{U}} \max_\theta \frac{|\tilde{g}_n(u,\theta) - \bar{g}_n(u,\theta)|}{\sqrt{nB_n}} \right) = o(1).
\tag{94}
$$

Proof. See Appendix A.9 of the supplementary materials. ☐

According to Lemmas 8.9 and 8.10, together with $\max_i |\tilde{g}_n(u,\theta_i) - \bar{g}_n(u,\theta_i)| \leq \max_\theta |\tilde{g}_n(u,\theta) - \bar{g}_n(u,\theta)|$, we have that $\max_{u \in \mathcal{U}}$ $\max_{0 \leq i \leq B_n} \frac{|g_n(u,\theta) - \tilde{g}_n(u,\theta)|^2}{nB_n} = o_{\mathbb{P}}(1)$ and

$$
\begin{aligned}
&\mathbb{P}\left(\max_{u \in \mathcal{U}} \max_{0 \leq i \leq B_n} \frac{|\tilde{g}_n(u,\theta_i) - \bar{g}_n(u,\theta_i)|^2}{nB_n} \geq y \right) \\
&\leq \frac{\mathbb{E}\left(\max_{u \in \mathcal{U}} \max_\theta \frac{|\tilde{g}_n(u,\theta) - \bar{g}_n(u,\theta)|^2}{nB_n} \right)}{y} = o(1).
\end{aligned}
\tag{95}
$$

Since $\max_u \max_i |\mathbb{E}\tilde{g}_n(u,\theta_i) - \mathbb{E}\bar{g}_n(u,\theta_i)| \leq \mathbb{E}(\max_u \max_i |\tilde{g}_n(u,\theta_i) - \bar{g}_n(u,\theta_i)|)$, it suffices to show that, for $D_n = B_n C_n$, we have that

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

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

$$H_j = [(j-1)(p_n + q_n) + 1, jp_n + (j-1)q_n], \quad 1 \leq j \leq k_n,$$
$$I_j = [jp_n + (j-1)q_n + 1, j(p_n + q_n)], \quad 1 \leq j \leq k_n,$$
$$I_{k_n+1} = [k_n(p_n + q_n) + 1, n].$$

$$(96)$$

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 $1 \leq j \leq k_n + 1$, let

$$U_j(u,\theta) := \sum_{k \in H_j} (\bar{Y}_{u,k} - \mathbb{E}\bar{Y}_{u,k}),$$
$$V_j(u,\theta) := \sum_{k \in I_j} (\bar{Y}_{u,k} - \mathbb{E}\bar{Y}_{u,k}). \qquad (97)$$

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

$$\bar{U}_j(u,\theta) := U_j(u,\theta)\mathbf{1}\left(\frac{|U_j(u,\theta)|}{\sqrt{nB_n}} \leq \frac{1}{(\log B_n)^4}\right)$$
$$- \mathbb{E}U_j(u,\theta)\mathbf{1}\left(\frac{|U_j(u,\theta)|}{\sqrt{nB_n}} \leq \frac{1}{(\log B_n)^4}\right). \qquad (98)$$

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. Under the assumptions of Theorem 5.1, we have that

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

$$(99)$$

Proof. See Appendix A.10 of the supplementary materials. □

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. Under the assumptions of Theorem 5.1, we have that

$$\mathbb{P}\left(\max_{u \in \mathcal{U}} \max_{i \notin [(\log B_n)^2, B_n - (\log B_n)^2]} \frac{\left|\sum_{j=1}^{k_n} \bar{U}_j(u,\theta_i)\right|}{\sqrt{nB_n}}\right.$$
$$\left. \geq x\sqrt{\log(B_n C_n)}\right) = o(1). \qquad (100)$$

Proof. See Appendix A.11 of the supplementary materials. □

Finally, we complete the proof of Equation (24) by the following result.

Lemma 8.13. Under the assumptions of Theorem 5.1, we have that

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

$$(101)$$

Proof. See Appendix A.12 of the supplementary materials. □

Supplementary Materials

The supplementary materials contain additional lemmas and detailed proofs of the remarks, lemmas, and theorems of the article.

Acknowledgments

References

Adak, S. (1998), "Time-Dependent Spectral Analysis of Nonstationary Time Series," *Journal of the American Statistical Association*, 93, 1488–1501. [133]

Allen, J. (1977), "Short Term Spectral Analysis, Synthesis, and Modification by Discrete Fourier Transform," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 25, 235–238. [133]

Anderson, T. W. (1971), *The Statistical Analysis of Time Series*, New York: Wiley. [134]

Andrews, D. W. K. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858. [136]

Blandford, R. R. (1993), "Discrimination of Earthquakes and Explosions at Regional Distances Using Complexity," Technical Report. [144,145]

Brillinger, D. R. (1969), "Asymptotic Properties of Spectral Estimates of Second Order," *Biometrika*, 56, 375–390. [134]

Cohen, L. (1995), *Time-Frequency Analysis: Theory and Applications*, Upper Saddle River, NJ: Prentice Hall. [133]

Dahlhaus, R. (1997), "Fitting Time Series Models to Nonstationary Processes," *The Annals of Statistics*, 25, 1–37. [133,135,139,143,149]

Dahlhaus, R., and Richter, S. (2019), "Adaptation for Nonparametric Estimators of Locally Stationary Processes," arXiv no. 1902.10381. [141]

Daubechies, I. (1990), "The Wavelet Transform, Time-Frequency Localization and Signal Analysis," *IEEE Transactions on Information Theory*, 36, 961–1005. [133]

—— (1992), *Ten Lectures on Wavelets*, CBMS-NSF Regional Conference Series in Applied Mathematics (Vol. 61), Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM). [133]

Daubechies, I., Lu, J., and Wu, H.-T. (2011), "Synchrosqueezed Wavelet Transforms: An Empirical Mode Decomposition-Like Tool," *Applied and Computational Harmonic Analysis*, 30, 243–261. [133]

Dette, H., Preuss, P., and Vetter, M. (2011), "A Measure of Stationarity in Locally Stationary Processes With Applications to Testing," *Journal of the American Statistical Association*, 106, 1113–1124. [133]

Dwivedi, Y., and Subba Rao, S. (2011), "A Test for Second-Order Stationarity of a Time Series Based on the Discrete Fourier Transform," *Journal of Time Series Analysis*, 32, 68–91. [133]

Fryzlewicz, P., and Nason, G. P. (2006), "Haar-Fisz Estimation of Evolutionary Wavelet Spectra," *Journal of the Royal Statistical Society*, Series B, 68, 611–634. [133]

Gröchenig, K. (2001), *Foundations of Time-Frequency Analysis*, Boston, MA: Springer. [133]

Hlawatsch, F., and Boudreaux-Bartels, G. F. (1992), "Linear and Quadratic Time-Frequency Signal Representations," *IEEE Signal Processing Magazine*, 9, 21–67. [133]

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N.-C., Tung, C. C., and Liu, H. H. (1998), "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, 903–995. [133]

Jentsch, C., and Subba Rao, S. (2015), "A Test for Second Order Stationarity of a Multivariate Time Series," *Journal of Econometrics*, 185, 124–161. [133]

Liu, W., and Wu, W. B. (2010), "Asymptotics of Spectral Density Estimates," *Econometric Theory*, 26, 1218–1245. [134,136]

Meyer, Y. (1992), *Wavelets and Operators*, Cambridge Studies in Advanced Mathematics (Vol. 37, Translated from the 1990 French original by D. H. Salinger), Cambridge: Cambridge University Press. [133]

Nason, G. (2013), "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, 75, 879–904. [133]

Nason, G. P., von Sachs, R., and Kroisandt, G. (2000), "Wavelet Processes and Adaptive Estimation of the Evolutionary Wavelet Spectrum," *Journal of the Royal Statistical Society*, Series B, 62, 271–292. [133]

Ombao, H. C., Raz, J. A., von Sachs, R., and Malow, B. A. (2001), "Automatic Statistical Analysis of Bivariate Nonstationary Time Series," *Journal of the American Statistical Association*, 96, 543–560. [133]

Paparoditis, E. (2010), "Validating Stationarity Assumptions in Time Series Analysis by Rolling Local Periodograms," *Journal of the American Statistical Association*, 105, 839–851. [133]

Paparoditis, E., and Politis, D. N. (2012), "Nonlinear Spectral Density Estimation: Thresholding the Correlogram," *Journal of Time Series Analysis*, 33, 386–397. [134]

Parzen, E. (1957), "On Consistent Estimates of the Spectrum of a Stationary Time Series," *Annals of Mathematical Statistics*, 28, 329–348. [134]

Politis, D. N., Romano, J. P., and Wolf, M. (1999), *Subsampling*, New York: Springer. [140,141]

Priestley, M. B. (1965), "Evolutionary Spectra and Non-Stationary Processes" (with discussion), *Journal of the Royal Statistical Society*, Series B, 27, 204–237. [133]

Rosenblatt, M. (1984), "Asymptotic Normality, Strong Mixing and Spectral Density Estimates," *The Annals of Probability*, 12, 1167–1180. [134]

Shao, X., and Wu, W. B. (2007), "Asymptotic Spectral Theory for Nonlinear Time Series," *The Annals of Statistics*, 35, 1773–1801. [135,136]

Shumway, R., and Stoffer, D. (2017), *Time Series Analysis and Its Applications, With R Examples* (4th ed.), New York: Springer. [143]

Woodroofe, M. B., and Ness, J. W. V. (1967), "The Maximum Deviation of Sample Spectral Densities," *The Annals of Mathematical Statistics*, 38, 1558–1569. [134,152]

Wu, W. B., and Zaffaroni, P. (2018), "Asymptotic Theory for Spectral Density Estimates of General Multivariate Time Series," *Econometric Theory*, 34, 1–22. [134]

Wu, W. B., and Zhao, Z. (2007), "Inference of Trends in Time Series," *Journal of the Royal Statistical Society*, Series B, 69, 391–410. [134]

Zhou, Z. (2013), "Heteroscedasticity and Autocorrelation Robust Structural Change Detection," *Journal of the American Statistical Association*, 108, 726–740. [140,141]

Zhou, Z., and Wu, W. B. (2009), "Local Linear Quantile Estimation for Nonstationary Time Series," *The Annals of Statistics*, 37, 2696–2729. [134,135]

——— (2010), "Simultaneous Inference of Linear Models With Time Varying Coefficients," *Journal of the Royal Statistical Society*, Series B, 72, 513–531. [134]