Nonlinear wavelet-based estimation to spectral density for stationary non-Gaussian linear processes

Linyuan Li* and Biao Zhang
Department of Mathematics and Statistics, University of New Hampshire, Durham, NH, 03824, USA

October 4, 2019

Abstract

Nonlinear wavelet-based estimators for spectral densities of non-Gaussian linear processes are considered. The convergence rates of mean integrated squared error (MISE) for those estimators over a large range of Besov function classes are derived, and it is shown that those rates are identical to minimax lower bounds in standard nonparametric regression model within a logarithmic term. Thus, those rates could be considered as nearly optimal. Therefore, the resulting wavelet-based estimators outperform traditional linear methods if the degree of smoothness of spectral densities varies considerably over the interval of interest, such as sharp spike, cusp, bump, etc, since linear estimators are not able to attain these rates. Unlike in classical nonparametric regression with Gaussian noise errors where thresholds are determined by normal distribution, we determine the thresholds based on a Bartlett type approximation of a quadratic form with dependent variables by its corresponding quadratic form with independent identically distributed (i.i.d.) random variables and Hanson-Wright inequality for quadratic forms in sub-gaussian random variables. The theory is illustrated with some numerical examples, and our simulation studies show that our proposed estimators are comparable to the current ones.

Short title: Spectral density estimation
2010 Mathematics Subject Classification: Primary: 62M10; Secondary: 62G05, 62M15, 62G20
Keywords: Nonlinear wavelet shrinkage; linear processes; periodogram; spectral density; Besov spaces.

1 Introduction

Spectral density estimation is a long standing problem in both theoretical and applied time series analysis. It is very useful for studying the periodicity of stationary time series. There exists a large literature on this topic. For a systematic and comprehensive discussion, see books, e.g., Hannan (1970), Brillinger (1981), Priestley (1981), Brockwell and Davis (1991) and Shumway

*Corresponding author: Linyuan Li, Phone: (603) 862-4592; Fax: (603) 862-4096; E-mail: linyuan.li@unh.edu

and Stoffer (2011). The estimation of spectral densities can be obtained parametrically based on certain time series models, such as autoregression moving average (ARMA) or other parametric models. Typically maximum likelihood method provides efficient estimates for parameters if model assumption is correct. Nevertheless, some spectral densities could not efficiently be approximated by parametric models. In that case, one uses periodogram of data to estimate spectral. It is well-known that periodogram is an unbiased estimator for spectral density, but it is not consistent. Thus, in order to obtain a consistent estimate, one has to smooth the periodogram using methods, such as, kernel, spline or wavelet.

Kernel and spline approaches are linear methods, they are effective when underlying spectral density is highly regular or smooth. However, when spectral density has significant spatial inhomogeneity, such as seasonal or business cycle periodicities, which are often encountered in economic and financial time series, linear smoothing methods are incapable of achieving optimal mean square convergence rates. In that case, wavelet method appears to improve the estimation. Wavelet estimate can automatically adapt to erratic behavior and local degree of smoothness of underlying spectral density because of the local property of wavelet basis. These wavelet estimators typically achieve optimal convergence rates over exceptionally large function spaces.

Many wavelet-based spectral estimates have been proposed since introduction of wavelet method to nonparametric regression by Donoho and Johnstone (1994, 1995, 1998). Here we only cite a very few closely related work. Moulin (1994), Gao (1997) and Pensky et al. (2007) proposed wavelet estimates based on log transform of periodogram in order to stabilize the variance and apply thresholding method alike in the standard Gaussian regression model proposed by Donoho and Johnstone (1994, 1995). As Fryzlewicz et al. (2008) pointed out, the downside for using the log transform is that it flattens out the data, and the resulting estimates often obscure interesting features, such as peaks or troughs. Thus, one may prefer to work on periodogram directly in order to preserve those features. In this paper, we consider wavelet estimate for spectral density directly using periodogram as what Neumann (1996) and Fryzlewicz et al. (2008) did. The main difference between ours and theirs is that the time series we considered here are non-Gaussian linear processes with certain conditions imposed on those coefficients and innovations opposed to their general non-Gaussian time series satisfying certain cumulants assumption. Our assumption on time series seems to be more informative, easier to verify and typically satisfied for most of ARMA time series. While their thresholds are determined from uniform estimates of the cumulants of empirical wavelet coefficients, we determine thresholds based on a Bartlett type approximation of our empirical wavelet coefficients by quadratic forms with independent identically distributed (i.i.d.) random variables and Hanson-Wright inequality for quadratic forms in sub-gaussian random variables. Our empirical wavelet coefficients are proposed in terms of Fourier transform of wavelet, which are same as those in Lee and Hong (2001). One advantage of this approach is that one could obtain explicit expressions for those empirical coefficients for arbitrary sample size.

The paper is organized as follows. In next section, we introduce linear time series, the elements of wavelets, spectral density function space and non-linear wavelet-based estimators. The main result on optimal mean-square rates of convergence for spectral estimators is stated in

Section 3. In Section 4, we provide some simulation studies and demonstrate that our proposed estimates are comparable to current estimates, while all technical proofs are provided in Section 5.

2 Preliminaries

This section contains some facts and assumptions about linear processes, wavelets, spectral density function spaces $B^s_{p,q}$ and wavelet-based estimators that will be used in the sequel.

2.1 Linear processes

In this paper, we assume that time series X_t, $t = 1, 2, \cdots, n$ is a realization of a linear process

$$X_t = \sum_{j=0}^{\infty} \varphi_j Z_{t-j}, \tag{2.1}$$

where $\{Z_t\}$ is a sequence of white noise, $E(Z_t) = 0, E(Z_t^2) = 1$, and coefficients φ_j satisfy assumption $\sum_{j=0}^{\infty} \varphi_j^2 < \infty$. The class of above linear processes $\{X_t\}$ is very large. From Giraitis et al. (2012, p.38), every stationary process $\{X_t\}$ with zero mean, whose spectral density f satisfies $\int_{[-\pi,\pi]} \log f(u)du > -\infty$, can be represented as a linear process as above.

In order to control the bound on a Bartlett type approximation of our time series below, we need to impose some conditions on those coefficients φ_j in (2.1). In specific, we require following assumption in order to simplify some technical arguments below.

A1: There exists a number $a > 1$ such that $\sum_{j=n}^{\infty} |\varphi_j| = O(n^{-a})$.

Above assumption **A1** is assumed in Bhansali et al. (2007b) in order to control the bounds for Bartlett type approximation of a quadratic form with dependent variables by its corresponding quadratic form with i.i.d. random variables. If time series X_t follows a commonly used causal and invertible ARMA process, then above coefficients satisfy $|\varphi_j| \le C\rho^j$ for some positive $\rho < 1$. Thus above assumption **A1** is satisfied. If one assumes a stronger condition $|\varphi_j| = O(j^{-(1+a)})$ for some number $a > 1$, then assumption **A1** is also satisfied. The assumption **A1** is slightly stronger than $\sum_j |j|^{1/2}|\varphi_j| < \infty$, which is usually assumed for short memory linear processes.

In order to derive an exponential inequality for our empirical wavelet coefficients, we need to assume that random variables $\{Z_t\}$ are i.i.d. (stronger than above white noise assumption) and form sub-Gaussian. This can be expressed by following assumption.

A2: There exists some $b > 0$ such that, for every $t \in \mathbb{R}$, one has $E\exp(tZ_1) \le \exp(b^2t^2/2)$, or equivalently there is some $K > 0$ such that $E\exp(Z_1^2/K^2) \le 2$.

Above assumption **A2** states that the Laplace transform of a sub-Gaussian random variable is dominated by the Laplace transform of a Gaussian random variable with mean zero and variance b^2. Sub-Gaussian assumption is weaker than Gaussian assumption, and it is usually needed in order to obtain an exponential inequality. Commonly used distributions, such as Gamma, Weibull, Bernoulli, and all centered and bounded random variables, are sub-Gaussian. For discussion on concentration results for quadratic forms in sub-Gaussian random variables, see, e.g., Wright (1973) and Rudelson and Vershynin (2013).

2.2 Wavelets

Let $\phi(x)$ and $\psi(x)$ be father and mother wavelets. We call a wavelet ψ *r-regular* if ψ has r vanishing moments (i.e., $\int x^k \psi(x)dx = 0, k = 0, 1, \ldots, r$) and r continuous derivatives. For more on the existence of these wavelets with high order vanishing moments and continuous derivative, see Daubechies (1992). Let

$$\phi_{j_0 k}(x) = 2^{j_0/2}\phi(2^{j_0}x - k), \quad \psi_{jk}(x) = 2^{j/2}\psi(2^j x - k), \quad x \in \mathbb{R},\ j_0,\ j,\ k \in \mathbb{Z}.$$

Then the collection $\{\phi_{j_0 k},\, \psi_{jk},\, j \geq j_0,\, k \in \mathbb{Z}\}$ is an orthonormal basis (ONB) of $L^2(\mathcal{R})$. Furthermore, let V_{j_0} and W_j be linear subspaces of $L^2(\mathcal{R})$ with the ONB $\phi_{j_0 k},\ k \in \mathbb{Z}$ and $\psi_{jk},\ k \in \mathbb{Z}$, respectively, we have the following decomposition

$$L^2(\mathcal{R}) = V_{j_0} \oplus W_{j_0} \oplus W_{j_0+1} \oplus W_{j_0+2} \oplus \cdots .$$

Therefore, for all $f \in L^2(\mathcal{R})$,

$$f(x) = \sum_{k\in\mathbb{Z}} \alpha_{j_0 k}\phi_{j_0 k}(x) + \sum_{j \geq j_0}\sum_{k\in\mathbb{Z}} \beta_{jk}\psi_{jk}(x),$$

where

$$\alpha_{j_0 k} = \int f(x)\phi_{j_0 k}(x)\, dx, \quad \beta_{jk} = \int f(x)\psi_{jk}(x)\, dx.$$

The orthogonality properties of ϕ and ψ imply:

$$\int \phi_{j_0 k_1}\phi_{j_0 k_2} = \lambda_{k_1 k_2}, \quad \int \psi_{j_1 k_1}\psi_{j_2 k_2} = \lambda_{j_1 j_2}\lambda_{k_1 k_2}, \quad \int \phi_{j_0 k_1}\psi_{j k_2} = 0,\ \forall j_0 \leq j, \tag{2.2}$$

where λ_{jk} denotes the Kronecker delta, i.e., $\lambda_{jk} = 1$, if $j = k$; and $\lambda_{jk} = 0$, otherwise. Since spectral density function $f(\cdot)$ is a 2π-periodic function over $\mathbb{R}$, one needs to use 2π-periodic wavelet basis to avoid any boundary correction of wavelet basis. Given any orthonormal wavelet basis $\{\phi_{jk}(\cdot),\, \psi_{jk}(\cdot)\}$ of $L_2(\mathbb{R})$, an orthonormal 2π-periodic wavelet basis $\{\Phi_{jk}(\cdot),\ \Psi_{jk}(\cdot)\}$ for $L_2(\Pi)$-space of 2π-periodic functions, where $\Pi = [-\pi,\ \pi]$, could be constructed by periodizing $\{\phi_{jk}(\cdot),\ \psi_{jk}(\cdot)\}$ via following expressions

$$\begin{aligned} \Phi_{jk}(w) &= \frac{1}{(2\pi)^{1/2}} \sum_{m=-\infty}^{\infty} \phi_{jk}\Big(\frac{w}{2\pi} + m\Big), \\ \Psi_{jk}(w) &= \frac{1}{(2\pi)^{1/2}} \sum_{m=-\infty}^{\infty} \psi_{jk}\Big(\frac{w}{2\pi} + m\Big), \end{aligned} \tag{2.3}$$

where $-\infty < w < \infty$. Both $\Phi_{jk}(\cdot)$ and $\Psi_{jk}(\cdot)$ are real-valued and periodic functions with period 2π. An example is the Haar wavelet, which is defined as: $\phi(x) = 1, x \in [0, 1]$ and

$$\psi(x) = \begin{cases} -1, & x \in [0, 1/2], \\ 1, & x \in (1/2, 1). \end{cases} \tag{2.4}$$

The function ϕ is father wavelet (or scaling function), while ψ is called mother wavelet. Let $\hat{\psi}(w)$ be Fourier transform of $\psi(\cdot)$ defined as

$$\hat{\psi}(w) = (2\pi)^{-1/2} \int_{-\infty}^{\infty} \psi(x) e^{-\mathbf{i}wx} dx, \ \mathbf{i} = \sqrt{-1},$$

then $\hat{\psi}(0) = 0$. For $w \neq 0$, through simple calculation, one obtains $\hat{\psi}(w) = -\mathbf{i}e^{-\mathbf{i}w/2}(2\pi)^{-1/2}$ $\sin^2(w/4)/(w/4)$. This shows that $|\hat{\psi}(w)|$ decays to zero as $|w| \to \infty$ at the rate of $|w|^{-1}$. Because of a technical reason in the proof below for the limit distribution for empirical wavelet coefficients, we require that wavelet $\psi(w)$ decays in frequency domain faster than $|w|^{-1}$. In specific, we need a following assumption on our wavelet.

A3: $|\hat{\psi}(w)| \leq \gamma_0(1 + |w|)^{-\gamma}$ for some contants $\gamma_0 > 0$ and $\gamma > 1$.

Most commonly used wavelets satisfy this condition, e.g., Daubechies, Meyer and spline wavelets with positive order. For further details, see, e.g., Hernández and Weiss (1996), Vidakovic (1999), Daubechies (1992), among others.

From Mallat (1998, p.282), one has that $\{\Phi_{Jk}, 0 \leq k \leq 2^J - 1;\ \Psi_{jk},\ j \geq J,\ 0 \leq k \leq 2^j - 1\}$ for any $J \geq 0$ is an orthogonal basis of $L_2(\Pi)$-space. Thus, a spectral density has following wavelet expansion

$$f(w) = \alpha_{00}\Phi_{00}(w) + \sum_{j=0}^{\infty} \sum_{k=0}^{2^j-1} \beta_{jk}\Psi_{jk}(w), \ w \in [-\pi, \pi],$$

where $\alpha_{00} = \int_{-\pi}^{\pi} f(w)\Phi_{00}(w)dw$ and $\beta_{jk} = \int_{-\pi}^{\pi} f(w)\Psi_{jk}(w)dw$ for all $j \geq 0$ and $k = 0, 1, \cdots, 2^j - 1$. Since $\sum_{m=-\infty}^{\infty} \phi_{00}(w + m) = 1$ for all w, we find a useful relation $\Phi_{00}(w) = (2\pi)^{-1/2}$ for all $w \in [-\pi, \pi]$. Therefore the spectral density can be written as

$$f(w) = \alpha_{00}\frac{1}{\sqrt{2\pi}} + \sum_{j=0}^{\infty} \sum_{k=0}^{2^j-1} \beta_{jk}\, \Psi_{jk}(w), \ w \in [-\pi, \pi]. \tag{2.5}$$

Above relation (2.5) is a general wavelet expansion of a spectral density.

2.3 Spectral density function spaces

As those in wavelet literature, we investigate asymptotic convergence rates of wavelet-based estimators over a large range of Besov function classes $B^s_{p,q},\ s > 0,\ 1 \leq p,\ q \leq \infty$. The parameter s is an index of regularity or smoothness and parameters p and q are used to specify the type of norm. Besov spaces contain many traditional function spaces such as the well-known Sobolev and Hölder spaces of smooth functions H^m and C^s (which are $B^m_{2,2}$ and $B^s_{\infty,\infty}$, respectively). Besov spaces also include significant spatial inhomogeneity function classes, such as the Bump Algebra and Bounded Variations Classes. For a more detailed study, we refer to Triebel (1992).

For a given *r-regular* mother wavelet ψ with $r > s$, define a sequence norm of wavelet coefficients of a function $f \in B^s_{p,q}$ by

$$|f|_{B^s_{p,q}} = \left(\sum_k |\alpha_{j_0 k}|^p\right)^{1/p} + \left\{\sum_{j=j_0}^{\infty}\left[2^{j\sigma}\left(\sum_k |\beta_{jk}|^p\right)^{1/p}\right]^q\right\}^{1/q}, \tag{2.6}$$

where $\sigma = s + 1/2 - 1/p$. Meyer (1992) showed that Besov function norm $\|f\|_{B^s_{p,q}}$ is equivalent to the sequence norm $|f|_{B^s_{p,q}}$ of the wavelet coefficients of f. Therefore we will use the sequence norm to calculate the Besov norm $\|f\|_{B^s_{p,q}}$ in the sequel. For any constant $M > 0$, define the standard Besov function space $B^s_{p,q}(M)$ by

$$B^s_{p,q}(M) = \left\{f \in B^s_{p,q} : \|f\|_{B^s_{p,q}} \le M,\, 1 \le p,\, q \le \infty,\, s > 1/p,\, \operatorname{supp} f \subseteq [-\pi,\, \pi]\right\}.$$

From function space embedding theorem, one has $B^s_{p,q} \subseteq B^{s-1/p}_{\infty,\infty}$. Thus, the space of functions which we consider in this paper is a subset of a space of bounded continuous functions. Because of a technical reason, we assume following assumption on the spectral density.

A4: There exists some arbitrary small but fixed number τ_0 such that $s \ge \tau_0$ for all $f \in B^s_{p,q}(M)$.

This assumption is mainly used in Lemma 5.7 below. Similar assumptions to above **A4** are imposed in literature, e.g., Fryzlewicz et al. (2008) assumes that $s \ge 1$ and Brown et al. (2008, p.2065) assumes that $s > 1/6$. Since we assume $s > 1/p$ in $B^s_{p,q}(M)$, one has $s > 1/p_0 (=: \tau_0)$ for all $1 \le p \le p_0$ for any fixed p_0. Thus, we consider this assumption is very weak, since τ_0 could be arbitrary small.

Analogous to Fryzlewicz et al. (2008) and Von sachs and Macgibbon (2000), we assume spectral density is bounded below. Since it is continuous and bounded, we assume following condition.

A5: Spectral density functions we considered here satisfy $\tau_1 \le f(w) \le \tau_2$, $w \in [-\pi,\, \pi]$, for some constants $0 < \tau_1, \tau_2 < \infty$.

2.4 Wavelet-based estimators

In this paper, we use wavelets to estimate spectral density. Formally, let $X = \{X_t,\, t \in \mathbb{Z}\}$ be a covariance stationary real-valued time series with spectral density $f(w)$, $w \in [-\pi, \pi]$. Assuming $\sum_{h=-\infty}^{\infty} |r(h)| < \infty$, where lag-$h$ autocovariance is defined by $r(h) = \mathrm{Cov}(X_t, X_{t-|h|})$, $h \in \mathbb{Z}$, the spectral density can be written as

$$f(w) = \frac{1}{2\pi}\sum_{h=-\infty}^{\infty} r(h) e^{-\mathbf{i}hw}, \quad w \in [-\pi, \pi]. \tag{2.7}$$

Assume one has a realization of $\{X_t, t = 1, 2, \cdots, n\}$. From $\alpha_{00} = \int_{-\pi}^{\pi} f(w)\Phi_{00}(w)dw = (2\pi)^{-1/2}\int_{-\pi}^{\pi} f(w)dw = (2\pi)^{-1/2} r(0)$, one proposes

$$\hat{\alpha}_{00} = \frac{1}{\sqrt{2\pi}}\hat{r}(0) = \frac{1}{\sqrt{2\pi}\, n}\sum_{t=1}^{n}(X_t - \bar{X})^2, \tag{2.8}$$

where $\bar{X} = n^{-1}\sum_{t=1}^{n} X_t$.

From (2.7), we have

$$\beta_{jk} = \int_{-\pi}^{\pi} f(w)\Psi_{jk}(w)dw = \frac{1}{(2\pi)^{1/2}} \sum_{h=-\infty}^{\infty} r(h)\hat{\Psi}_{jk}(h), \tag{2.9}$$

where $\hat{\Psi}_{jk}(\cdot)$ is the Fourier transform of $\Psi_{jk}(\cdot)$, defined as

$$\hat{\Psi}_{jk}(h) = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} \Psi_{jk}(w)e^{-\mathbf{i}wh}dw. \tag{2.10}$$

Thus, one proposes estimators for β_{jk}'s as

$$\hat{\beta}_{jk} = \frac{1}{\sqrt{2\pi}} \sum_{h=-(n-1)}^{n-1} \hat{r}(h)\hat{\Psi}_{jk}(h), \tag{2.11}$$

for all $j = 0, 1, \ldots, J_n$ and $k = 0, 1, \ldots, 2^j - 1$, where J_n satisfies $2^{J_n+1} \le n < 2^{J_n+2}$ with $\hat{r}(h) = n^{-1}\sum_{t=|h|+1}^{n}(X_t - \bar{X})(X_{t-|h|} - \bar{X})$ and $\bar{X} = n^{-1}\sum_{t=1}^{n} X_t$.

Our first-step wavelet estimator $\bar{f}$ of f is defined as

$$\bar{f}(w) = \hat{\alpha}_{00}\frac{1}{\sqrt{2\pi}} + \sum_{j=0}^{j_0-1}\sum_{k=0}^{2^j-1} \hat{\beta}_{jk}\Psi_{jk}(w) + \sum_{j=j_0}^{j_1}\sum_{k=0}^{2^j-1} \hat{\beta}_{jk} I\big(|\hat{\beta}_{jk}| > \delta_{jk}\big)\Psi_{jk}(w), \tag{2.12}$$

where above smoothing parameters j_0 and j_1 satisfy $2^{j_0} \simeq \ln n$ and $2^{j_1} \simeq n^{1-\tau}$ for some small positive number τ. The notation $2^{j(n)} \simeq h(n)$ means that $j(n)$ is chosen to satisfy the inequalities $2^{j(n)} \le h(n) < 2^{j(n)+1}$. The above τ plays the same roles of α in Neumann (1996, p.607) and δ in Fryzlewicz et al. (2008, p.873). In this paper, we require that τ satisfies $0 < \tau \le 2\tau_0/(2\tau_0+1)$ (the positive number τ_0 is provided in assumption **A4**. If one limits $\tau_0 \le 1/4$, then one has $\tau \le 1/3$ which will be used in Lemma 5.7 below). The empirical wavelet coefficient $\hat{\alpha}_{00}$ is defined in (2.8) and $\hat{\beta}_{jk}$'s are defined in (2.11). As in Neumann (1996), we define $\Lambda = \Lambda(n) = \{(j,k)|j_0 \le j \le j_1, k = 0, 1, \cdots, 2^j - 1\}$, the set of indices (j,k) whose corresponding empirical wavelet coefficients $\hat{\beta}_{jk}$ are under thresholding. Then, the number of those coefficients is $\#\Lambda = 2^{j_1+1} - 2^{j_0}$. Thresholding constants δ_{jk} ($j_0 \le j \le j_1$) are level and location dependent satisfying $\delta_{jk}^2 = K_1\ln(\#\Lambda)\cdot(\sigma_{jk}^2 \vee \delta_0 n^{-1})$, where $a \vee b = \max\{a, b\}$ and $\sigma_{jk}^2 = \mathrm{Var}(\hat{\beta}_{jk})$. The calculation of the variance of $\hat{\beta}_{jk}$ is provided in Lemma 5.5 below, see (5.26). The constant K_1 is sufficiently large and constant δ_0 is sufficiently small for a technical reason. How to determine the values of those constants in practice will be discussed in implementation section below.

In view of (5.26), the variance $\sigma_{jk}^2 = \mathrm{Var}(\hat{\beta}_{jk})$ could be expressed as

$$\sigma_{jk}^2 = \frac{1}{n}\left\{\pi\int_{-\pi}^{\pi}\big[\Psi_{jk}(w) + \Psi_{jk}(-w)\big]^2 f^2(w)dw + \frac{(\mu_4 - 3)\beta_{jk}^2}{(2\pi)^2}\right\}\big(1 + o(1)\big), \tag{2.13}$$

where $\mu_4 = E(Z_1^4)$. The rational for the above proposed thresholds δ_{jk}^2 $(= K_1\ln(\#\Lambda)\cdot(\sigma_{jk}^2 \vee \delta_0 n^{-1}))$ is similar to that of Neumann (1996), which requires that $n\delta_{jk}^2$ are bounded from below uniformly for $(j,k) \in \Lambda$ because of some technical reason (see Lemma 5.5 below). First we note that, from (2.13), we have upper bound $n\sigma_{jk}^2 \le C_0 < \infty$ for all j and k from assumption **A5**. Second, from (5.27) below, we have

$$\begin{aligned}\sigma_{jk}^2 &\ge \frac{C}{n}\int_{-\pi}^{\pi}\big[\Psi_{jk}(w)+\Psi_{jk}(-w)\big]^2 f^2(w)dw \\ &= \frac{2C}{n}\left\{\int_{-\pi}^{\pi}\Psi_{jk}^2(w)f^2(w)dw + \int_{-\pi}^{\pi}\Psi_{jk}(w)\Psi_{jk}(-w)f^2(w)dw\right\},\end{aligned}$$

the last equality follows from the symmetry of $f(w)$ about zero. If one uses compactly supported wavelet $\psi(\cdot)$, then one has $\int_{-\pi}^{\pi}\Psi_{jk}(w)\Psi_{jk}(-w)f^2(w)dw = 0$ except for a finite number of k's (independent of j) at each scale j. Now, from the above assumption **A5** such that $f(w) \ge \tau_1 > 0$, we have $n\sigma_{jk}^2 \ge 2C\int_{-\pi}^{\pi}\Psi_{jk}^2(w)f^2(w)dw \ge 2C\tau_1^2 =: c_0$ for some positive number c_0. Thus, there exist at most finite number of k's such that $n\sigma_{jk}^2 \le c_0$. If one chooses the constant δ_0 in thresholding constants small enough such as $\delta_0 < c_0$ (it is always possible, since c_0 is a fixed positive number), then one obtains that thresholding constants $\delta_{jk}^2 = K_1 \ln(\#\Lambda)\,\delta_0 n^{-1}$ for at most finite number of k's at each level j and $\delta_{jk}^2 = K_1\ln(\#\Lambda)\,\sigma_{jk}^2$ for all other k's. In specific, we define Γ_j^c as the set which includes above finite k's at level j and Γ_j as the complement set of Γ_j^c (i.e., $\Gamma_j = \{k;\, 0 \le k \le 2^j - 1 \,\text{and}\, k \notin \Gamma_j^c\}$). Thus, we have, for all $j_0 \le j \le j_1$,

$$n\sigma_{jk}^2 \ge \delta_0 > 0,\ k \in \Gamma_j \quad \text{and} \quad n\sigma_{jk}^2 < \delta_0,\ k \in \Gamma_j^c. \tag{2.14}$$

Denote $\#\{\Gamma_j^c\}$ as the number of k's in the set Γ_j^c, then $\#\{\Gamma_j^c\} < C$ for some constant C and for all j's. This property will be used in the proof to the main result below.

The above estimator $\bar{f}(w)$ defined as in (2.12) is of purely theoretical interest, because those thresholding constants $\delta_{jk}^2 = K_1\ln(\#\Lambda)\cdot(\sigma_{jk}^2 \vee \delta_0 n^{-1})$ are unknown. In order to make the estimate implementable, one must estimate those σ_{jk} first. First, we note that the second term $(\mu_4 - 3)\beta_{jk}^2/(2\pi)^2$ in RHS of (2.13) in σ_{jk}^2 is dominated by the first term. From Hall and Patil (1995, p.917), if $f(\cdot)$ has s-times continuous derivatives, then β_{jk}^2 is at the order of $O(2^{-j(2s+1)})$. Thus, the second term is at most at the order $O(\beta_{jk}^2) = O(2^{-j})$. In view of our estimator $\bar{f}(\cdot)$ in (2.12), we consider those j's such that $j \ge j_0$ and $2^{j_0} \simeq \ln n$. Thus, we have that the second term $(\mu_4 - 3)\beta_{jk}^2/(2\pi)^2$ is at most $O((\ln n)^{-1})$, which goes to zero when sample size goes to infinity.

Based on above discussion, we propose $\hat{\delta}_{jk}^2 = K_1\ln(\#\Lambda)\cdot(\hat{\sigma}_{jk}^2 \vee \delta_0 n^{-1})$, where

$$\hat{\sigma}_{jk}^2 = \frac{\pi}{n}\int_{-\pi}^{\pi}\big[\Psi_{jk}(w)+\Psi_{jk}(-w)\big]^2\tilde{f}^2(w)dw \tag{2.15}$$

and $\tilde{f}(\cdot)$ is an estimator for $f(\cdot)$ satisfying certain properties. For specific, in this paper, we consider a kernel estimator $\tilde{f}(\cdot)$. For details, see (3.1) below.

Now we are ready to propose corresponding estimator for $f(\cdot)$, which is defined as

$$\hat{f}(w) = \hat{\alpha}_{00}\frac{1}{\sqrt{2\pi}} + \sum_{j=0}^{j_0-1}\sum_{k=0}^{2^j-1}\hat{\beta}_{jk}\Psi_{jk}(w) + \sum_{j=j_0}^{j_1}\sum_{k=0}^{2^j-1}\hat{\beta}_{jk} I\big(|\hat{\beta}_{jk}| > \hat{\delta}_{jk}\big)\Psi_{jk}(w), \tag{2.16}$$

where $\hat{\delta}_{jk}^2 = K_1 \ln(\#\Lambda) \cdot (\hat{\sigma}_{jk}^2 \vee \delta_0 n^{-1})$ and $\hat{\sigma}_{jk}^2$ is defined in (2.15). As long as one selects appropriate levels j_0 and j_1, and constants K_1 and δ_0 in $\hat{\delta}_{jk}^2$, one could implement above estimate. For further discussion on choices of those constants, see the implementation section below.

Throughout this paper, we use $C, C_1, C_2, \cdots$ to denote positive and finite constants whose value may change from line to line. Specific constants are denoted by $a, b, \tau, \tau_0, \tau_1, \gamma, \gamma_0, \gamma_1, K, K_1$ and so on.

3 Main results and discussions

In order to make the proof to the main result on estimator $\hat{f}(\cdot)$ easier to read, we first state a theorem on the theoretical or impractical estimator $\bar{f}(\cdot)$ defined as in (2.12).

Theorem 3.1 *Let time series $\{X_t\}$ satisfy the assumptions* **A1** *and* **A2**, *their spectral density functions satisfy the assumptions* **A4** *and* **A5**, *and our wavelet ψ be r-regular and satisfy the assumption* **A3**. *Then, there exists a constant C which does not depend on n such that*

$$\sup_{f \in B_{p,q}^s(M)} E \int_{-\pi}^{\pi} (\bar{f} - f)^2 \le C \Big(\frac{\ln n}{n}\Big)^{2s/(2s+1)},$$

where $\bar{f}$ is the wavelet estimator given in (2.12).

Let $\tilde{f}(\cdot)$ be any kernel estimator for $f(\cdot)$ defined as

$$\tilde{f}(w) = \frac{1}{h_n} \int_{-\pi}^{\pi} K\Big(\frac{w - \lambda}{h_n}\Big) I_{n,X}(\lambda)\, d\lambda, \tag{3.1}$$

where $I_{n,X}(\lambda)$ is the periodogram of the sequence $\{X_t\}$ (for the definition, see (5.8) below), $K(\cdot)$ is a kernel function and h_n is a bandwidth or smoothing parameter. In order to obtain desired convergence rates, we need following assumption:
A6: (i). Kernel function $K(\cdot)$ has compact support. (ii). $\int K(x)\, dx = 1$. (iii). $K(\cdot)$ is continuously differentiable. (iv). Bandwidth $h_n = Cn^{-d}$ for some $0 < d < (1-\tau)/2$, where τ is provided in $\bar{f}$ in (2.12).

Above assumption **A6** is very standard in kernel estimation literature. Once one uses above estimate $\tilde{f}(\cdot)$ in (2.15), one obtains implementable thresholds $\hat{\delta}_{jk}$ and implementable estimator $\hat{f}(\cdot)$ in (2.16). For this practical estimator, we have the following result:

Theorem 3.2 *Let time series $\{X_t\}$ satisfy the assumptions* **A1** *and* **A2**, *their spectral density functions satisfy the assumptions* **A4** *and* **A5**, *our wavelet ψ be r-regular and satisfy the assumption* **A3**, *and the kernel estimator $\tilde{f}$ defined in (3.1) satisfy the assumption* **A6**. *Then, there exists a constant C which does not depend on n such that*

$$\sup_{f \in B_{p,q}^s(M)} E \int_{-\pi}^{\pi} (\hat{f} - f)^2 \le C \Big(\frac{\ln n}{n}\Big)^{2s/(2s+1)},$$

where $\hat{f}$ is the wavelet estimator given in (2.16).

Remark 3.1 There exists a large literature on spectral density estimation based on kernel and spline methods, which typically assume that the spectral density is s-times continuously differentiable over its whole domain in order to derive their MISE expressions. Nevertheless, our spectral density could be spatially inhomogeneous, which is more suitable for certain time series whose spectral may present various forms of irregularities such as sharp peaks and high frequency alternations. In light of rate-of-convergence result, our wavelet-based estimator for spectral attains convergence rate at order $(\ln n/n)^{2s/(2s+1)}$, which is very close to optimal rate $n^{-2s/(2s+1)}$ in standard nonparametric regression. For linear estimation procedures, such as kernel and spline estimators, their convergence rates typically are at order $n^{-\alpha_0}$, where $\alpha_0 = [2s + 2(1/(p \vee 2) - 1/p)]/[2s + 1 + 2(1/(p \vee 2) - 1/p)]$. Thus, when $p < 2$, the above rate $n^{-\alpha_0}$ is slower than that of wavelet-based method. The above theorem shows that our wavelet-based estimators defined in (2.16), based on simple thresholding of empirical wavelet coefficients, attain nearly optimal convergence rates over a large range of Besov function classes and behave themselves as if they know in advance in which class the functions lie.

4 Implementation and Simulation Studies

Our nonlinear wavelet-based estimator $\hat{f}$ is defined in (2.16) with $2^{j_0} \simeq \ln n$ and $2^{j_1} \simeq n^{1-\tau}$, where τ satisfies $0 < \tau \le 2\tau_0/(2\tau_0 + 1)$ for any small number $\tau_0 > 0$. This theoretical result on τ is not helpful for practical applications. In general, the larger value of τ is used for larger value of smooth parameter s (i.e., the smoother the spectral density function is). Fryzlewicz et al. (2008) suggested that $\tau = 0.01$ works very well from their empirical study. In this paper, as suggested in Neumann (1996), we simply let the empirical wavelet coefficients from the highest (or together with the second highest) resolution levels be zeroed, since we consider the same time series model here. The thresholds are defined as $\hat{\delta}_{jk}^2 = K_1 \ln(\#\Lambda) \times (\hat{\sigma}_{jk}^2 \vee \delta_0 n^{-1})$, where $\hat{\sigma}_{jk}^2$ is defined in (2.15) and δ_0 is an arbitrary small positive number. As we mentioned before, the way that we proposed the above thresholds is mainly of technical reason (such that $n\hat{\delta}_{jk}^2 \ge K_1 \ln(\#\Lambda)\delta_0$ for all j, k, so they are bounded from below uniformly). In practice, since $n\hat{\sigma}_{jk}^2 > 0$ for all practical real data, we always can obtain $n\hat{\sigma}_{jk}^2 > \delta_0$ (or equivalently $\hat{\sigma}_{jk}^2 > \delta_0 n^{-1}$) for a sufficiently small δ_0 and large n. Hence, in practice, we could simply use thresholds $\hat{\delta}_{jk}^2 = K_1 \ln(\#\Lambda)\, \hat{\sigma}_{jk}^2$, ignoring the term $\delta_0 n^{-1}$ (again, δ_0 is introduced mainly for theoretical rigorous argument). The requirement that constant K_1 is very large also is mainly for theoretical justification. For details, see Lemma 5.5 below. Li and Lu (2018) showed that

$$n^{1/2}\big(\hat{\beta}_{jk} - \beta_{jk}\big) \Longrightarrow_d N\big(0, \sigma_{jk}^2\big), \quad n \to \infty, \tag{4.1}$$

where σ_{jk}^2 is defined in (2.13). From the "universal" threshold theory first developed by Donoho and Johnstone (1994) in the Gaussian regression case, the thresholds typically are set equal to $\delta_{jk} = \{2\ln(\#\text{coefficients})\,\text{var}(\hat{\beta}_{jk})\}^{1/2}$. Therefore, in view of (4.1) and our estimator $\hat{f}$, it is reasonable for us to propose new thresholds as $\hat{\delta}_{jk,1}^2 = 2\ln(\#\Lambda)\,\hat{\sigma}_{jk}^2$, i.e., replacing constant K_1 with a constant 2 and $\hat{\sigma}_{jk}^2$ is defined in (2.15). Neumann (1996) showed that threshold $\hat{\delta}_{jk,1}$ seems a little bit large and it over-thresholds empirical wavelet coefficients. He suggested another smaller threshold as $\hat{\delta}^2_{jk,2} = 2\ln(\#\Lambda/2^{j_0})\,\hat{\sigma}^2_{jk}$, which provides better numerical results. If one traces the proof for Theorems 3.1 and 3.2, then one could derive that those two theoretical results still hold for estimators with $\ln(\#\Lambda)$ replaced by $\ln(\#\Lambda/2^{j_0})$. In following simulation study, we tried both above thresholds $\hat{\delta}_{jk,1}$ and $\hat{\delta}_{jk,2}$.

Our theoretical results show that performances of wavelet-based estimators do not depend on the choice of wavelets in view of asymptotical convergence rates. Nevertheless, the choice of wavelets may influence the performances of the estimators in their finite numerical calculation. Because of that, we conduct simulation studies for estimators with several wavelets including Daubechies wavelets. Since Haar, Franklin and Meyer wavelets have explicit expressions, we have further expressions for those empirical wavelet coefficients below.

The Franklin wavelet with the first order is given in frequency domain by

$$\begin{aligned}\hat{\phi}(w) &= (2\pi)^{-1/2}\,\frac{\sin^2(w/2)}{(w/2)^2}\Big(1-\frac{2}{3}\sin^2(w/2)\Big)^{-1/2},\\ \hat{\psi}(w) &= e^{\mathbf{i}w/2}(2\pi)^{-1/2}\,\frac{\sin^4(w/4)}{(w/4)^2}\bigg(\frac{1-\frac{2}{3}\cos^2(w/4)}{\big(1-\frac{2}{3}\sin^2(w/2)\big)\big(1-\frac{2}{3}\sin^2(w/4)\big)}\bigg)^{1/2}.\end{aligned} \tag{4.2}$$

For details, see Hernández and Weiss (1996, p.148). Thus, we have $|\hat{\psi}(w)| \le C(1+|w|)^{-2}$ as $|w| \to \infty$, which satisfies our assumption **A3** on wavelets.

From the definition of $\hat{\beta}_{jk}$ in (2.11), using $\hat{\Psi}_{jk}(h) = \sqrt{2\pi}\,\hat{\psi}_{jk}(2\pi h)$, $\hat{\psi}_{jk}(2\pi h) = e^{-\mathbf{i}2\pi hk/2^j}2^{-j/2}\hat{\psi}(2\pi h/2^j)$, the Fourier transform $\hat{\psi}$ in (4.2) and through straightforward but tedious algebra, we have

$$\begin{aligned}\hat{\beta}_{jk} &= \sum_{h=1}^{n-1}\hat{r}(h)\big\{\hat{\psi}_{jk}(2\pi h)+\hat{\psi}_{jk}(-2\pi h)\big\}\\ &= \frac{2^{\frac{3j}{2}+5}}{\sqrt{2\pi}}\sum_{h=1}^{n-1}\hat{r}(h)\cos\Big(\frac{2\pi h(k-1/2)}{2^j}\Big)\frac{\sin^4\big(\frac{2\pi h}{2^{j+2}}\big)}{(2\pi h)^2}\left\{\frac{1-\frac{2}{3}\cos^2\big(\frac{2\pi h}{2^{j+2}}\big)}{\Big(1-\frac{2}{3}\sin^2\big(\frac{2\pi h}{2^{j+1}}\big)\Big)\Big(1-\frac{2}{3}\sin^2\big(\frac{2\pi h}{2^{j+2}}\big)\Big)}\right\}^{1/2}.\end{aligned} \tag{4.3}$$

Meyer wavelet is given in frequency domain by

$$\hat{\psi}(w) = \begin{cases} e^{\mathbf{i}w/2}(2\pi)^{-1/2}\sin\big\{\frac{\pi}{2}\nu(\frac{3}{2\pi}|w|-1)\big\}, & |w| \in [\frac{2\pi}{3},\frac{4\pi}{3}],\\ e^{\mathbf{i}w/2}(2\pi)^{-1/2}\cos\big\{\frac{\pi}{2}\nu(\frac{3}{4\pi}|w|-1)\big\}, & |w| \in (\frac{4\pi}{3},\frac{8\pi}{3}],\\ 0, & \text{otherwise,}\end{cases} \tag{4.4}$$

where we use following function $\nu(x)$,

$$\nu(x) = \begin{cases} 0, & x \le 0,\\ x^2(3-2x), & 0 \le x \le 1,\\ 1, & x \ge 1.\end{cases} \tag{4.5}$$

Since $\hat{\psi}$ has compact support, it satisfies our assumption **A3**. From above explicit expression of $\hat{\psi}(w)$, one could derive corresponding expression of $\hat{\beta}_{jk}$ as in (4.3).

Although Haar wavelet, given in (2.4), does not satisfy our theoretical assumption **A3**, we provide its finite sample performance also because of its simplicity.

Our test process is the same as that in Neumann (1996) and Fryzlewicz et al. (2008), which is defined by

$$X_t = Y_t + \frac{1}{2} Z_t,$$

where

$$Y_t + \frac{1}{5} Y_{t-1} + \frac{9}{10} Y_{t-2} = \varepsilon_t + \varepsilon_{t-2},$$

and $\{\varepsilon_t\}$, $\{Z_t\}$ are independent Gaussian white noise processes with mean zero and variance one. All our simulations are based on sample size $n = 1024$. From $2^9 = 1024$ and $2^{j_0} \simeq \ln n$ and $2^{j_1} \simeq n^{1-\tau}$ for some $\tau > 0$, one could select $j_1 \leq 8$, i.e., we keep empirical wavelet coefficients at low levels $j = 0, 1, \cdots, j_0 - 1$ intact, threshold coefficients at higher levels $j = j_0, j_0 + 1, \cdots, j_1$, and let empirical wavelet coefficients from higher levels $j > j_1$ be zero. For this particular example, empirical study shows that j_0 could be selected as 3 or 4 and j_1 could be selected as 5 or 6 which provide similar smaller risk errors compared to other parameter values.

We compare empirical performance of our wavelet estimators with different combinations of levels (j_0, j_1) and above two thresholds ($\hat{\delta}_{jk,1}$ and $\hat{\delta}_{jk,2}$) to the kernel estimators with optimally chosen global and locally varying bandwidths. For the latter two methods, the routines **glkerns** and **lokerns** are called respectively from the R package **lokern**. This simulation study also provides an indirect comparison with the methods of Neumann (1996) and Fryzlewicz et al. (2008). For numerical comparisons, we consider the L_2 risk of those estimators on the basis of $N = 500$ simulation runs,

$$L_2 = \frac{1}{N} \sum_{l=1}^{N} \left[\frac{2\pi}{512} \sum_{i=0}^{512} \big(\hat{f}_l(w_i) - f(w_i) \big)^2 \right],$$

where $w_i = 2\pi i/n$ is the Fourier frequencies, $\hat{f}_l(w_i)$ is the estimate of $f(w_i)$ in l-th replication and $N = 500$ is the total number of replications. The simulation results for different levels and thresholds are summarized in Table 1.

Table 1: L_2 risks for Wavelet estimators and Kernel estimators

	Daub8	Haar	Franklin	Meyer	glkerns	lokerns
$j_0 = 3,\ j_1 = 5;\ \hat{\delta}_{jk,1}$	0.0289	0.0299	0.0314	0.0335	0.0331	0.0292
$j_0 = 3,\ j_1 = 5;\ \hat{\delta}_{jk,2}$	0.0286	0.0295	0.0306	0.0340	0.0331	0.0292
$j_0 = 4,\ j_1 = 5;\ \hat{\delta}_{jk,1}$	0.0284	0.0294	0.0316	0.0340	0.0331	0.0292
$j_0 = 4,\ j_1 = 5;\ \hat{\delta}_{jk,2}$	0.0283	0.0292	0.0306	0.0329	0.0331	0.0292

Table 1 shows the empirical L_2 risks of our wavelet estimators and their competitors (results for other combinations are slightly worse and we do not report them here). The L_2 risks of estimator with Meyer wavelet are comparable to Kernel global estimator, both of them have larger risks than other estimators. Estimators with Daublets with index 8 and Haar wavelet have smaller risks, and appear very comparable to local kernel estimator. In addition, estimators with smaller threshold $\hat{\delta}_{jk,2}$ seem to provide slightly smaller L_2 risk. Those findings are consistent to what Neumann (1996) and Fryzlewicz et al. (2008) found.

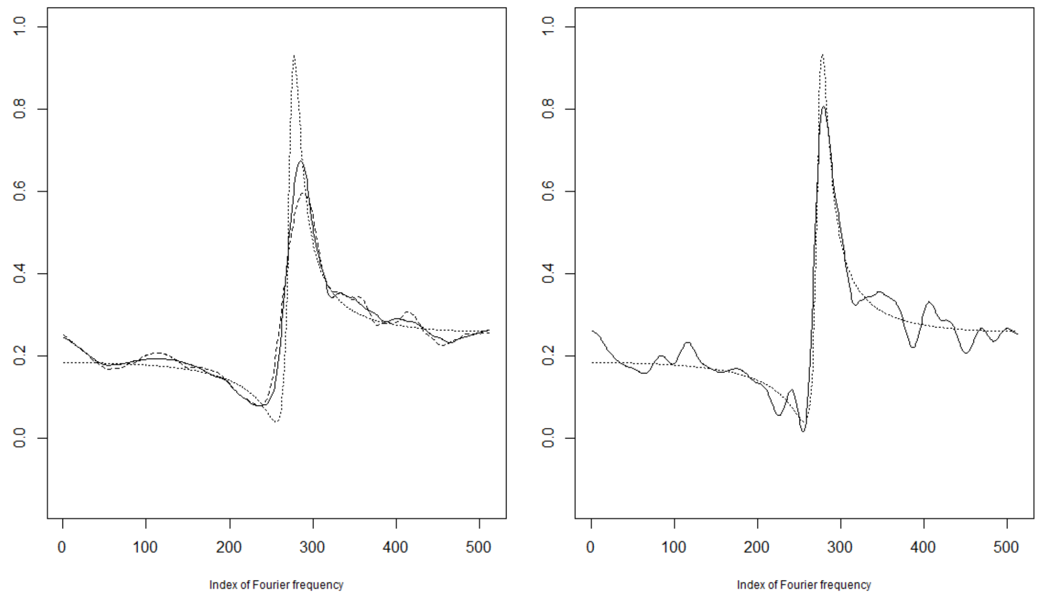

Figure 1: Both: true spectral density (dotted). Left: reconstructions using glkerns (dashed); reconstructions using lokerns (solid). Right: reconstructions using proposed wavelet estimator (solid).

Figure 1 shows the true spectral density and the sample reconstructions using the two kernel estimators and one wavelet estimate with Daubechies wavelets Daub8. The wavelet estimate is constructed with $j_0 = 3$, $j_1 = 5$ and threshold $\hat{\delta}_{jk,2}$. It appears that our new wavelet estimator captures the peak better than its competitors. On the other hand, two kernel estimators achieve a better job at reconstructing the smooth parts. These features are well expected effect.

Overall, the finite sample performance of our new wavelet estimator does depend on the use of the wavelet. Given the above results, it is recommended to use Daubechies wavelets estimator with smaller threshold $\hat{\delta}_{jk,2}$ in practice.

5 Proofs

Proofs to Theorems 3.1 and 3.2 are very similar, and the proof to Theorem 3.2 is more involved because of random thresholds $\hat{\delta}_{jk}$ used in $\hat{f}(\cdot)$. We first provide a proof to Theorem 3.1 mainly for theoretical purpose. After that, we explain where we need to do extra work to prove Theorem 3.2. To proceed, we break the proof to Theorem 3.1 into several parts. In view of (2.5) and

(2.16), and observing the orthogonality of $\Phi_{jk}(\cdot)$ and $\Psi_{jk}(\cdot)$ as those in (2.2), we have

$$E\int_{-\pi}^{\pi}\big(\bar{f}(w)-f(w)\big)^2dw =: I_0+I_1+I_2+I_3+I_4+I_5, \tag{5.1}$$

where

$$I_0 := E\big(\hat{\alpha}_{00}-\alpha_{00}\big)^2, \qquad I_3 := \sum_{j=j_0}^{j_s}\sum_{k\in\Gamma_j} E\Big[\hat{\beta}_{jk}I\big(|\hat{\beta}_{jk}|>\delta_{jk}\big)-\beta_{jk}\Big]^2,$$

$$I_1 := \sum_{j=j_1+1}^{\infty}\sum_{k=0}^{2^j-1}\beta_{jk}^2, \qquad I_4 := \sum_{j=j_s+1}^{j_1}\sum_{k\in\Gamma_j} E\Big[\hat{\beta}_{jk}I\big(|\hat{\beta}_{jk}|>\delta_{jk}\big)-\beta_{jk}\Big]^2,$$

$$I_2 := \sum_{j=0}^{j_0-1}\sum_{k=0}^{2^j-1} E\big(\hat{\beta}_{jk}-\beta_{jk}\big)^2, \qquad I_5 := \sum_{j=j_0}^{j_1}\sum_{k\in\Gamma_j^c} E\Big[\hat{\beta}_{jk}I\big(|\hat{\beta}_{jk}|>\delta_{jk}\big)-\beta_{jk}\Big]^2,$$

where j_s satisfies $j_s = j_s(n)$ such that $2^{j_s}\simeq (n/\ln n)^{1/(2s+1)}$ and Γ_j is defined as in (2.14). In order to prove Theorem 3.1, it suffices to show that $I_i \le C(\ln n/n)^{2s/(2s+1)}$, $i=0,1,\cdots,5$. Those bounds are obtained in following Lemmas 5.6 to 5.11. In order to do that, we need some preparation.

There exists an extensive literature providing MISE calculation for nonparametric mean regression functions and density functions. However, for spectral density case, empirical wavelet coefficients $\hat{\beta}_{jk}$ is a quadratic form of random vector with dependent components, instead of a sum of random variables in regression or density estimation cases. In order to derive an exponential inequality for the difference between the empirical wavelet coefficients $\hat{\beta}_{jk}$ and true wavelet coefficients β_{jk}, we would borrow a result from Bhansali (2007a) to approximate $\hat{\beta}_{jk}$'s by its corresponding quadratic form with i.i.d. random variables Z_i's. From there, we could apply Hanson-Wright inequality to obtain desired rates. For this purpose, we provide some preparatory results.

The following lemma is very useful for estimating the sequence norm of a function $f\in B^s_{p,q}$ for different values of p, which is used in Cai (1999).

Lemma 5.1 *Let $u\in\mathbb{R}^n$, $\|u\|_p = \big(\sum_i |u_i|^p\big)^{1/p}$, and $0<p_1\le p_2\le\infty$. Then the following inequalities hold:*

$$\|u\|_{p_2}\le\|u\|_{p_1}\le n^{\frac{1}{p_1}-\frac{1}{p_2}}\|u\|_{p_2}.$$

The following lemma provides an exponential inequality for sub-gaussian random vectors. In specific, for an $m\times n$ matrix $A=(a_{ij})$, the operator norm of A is $\|A\| = \max_{\|x\|_2\le 1}\|Ax\|_2$, and the Hilbert-Schmidt (or Frobenius) norm of A is $\|A\|_{HS} = (\sum_{i,j}|a_{ij}|^2)^{1/2}$. The following inequality is the Theorem 1.1 in Rudelson and Vershynin (2013, p.2).

Lemma 5.2 *Let $Z=(Z_1,Z_2,\cdots,Z_n)^T\in\mathbb{R}^n$ be a random vector with independent components Z_i which satisfy assumption* **A2**. *Let A be an $n\times n$ matrix. Then, for every $t\ge 0$, there exists a positive constant C which does not depend on n,*

$$P\Big\{\big|Z^TAZ-E(Z^TAZ)\big|>t\Big\}\le 2\exp\Big[-C\min\Big(\frac{t^2}{K^4\|A\|_{HS}^2},\ \frac{t}{K^2\|A\|}\Big)\Big].$$

In order to obtain an exponential inequality for empirical wavelet coefficients $\hat{\beta}_{jk}$, we first need some notation. In view of those $\hat{\beta}_{jk}$ defined in (2.11), substituting the sample covariance $\hat{r}(h)$ into $\hat{\beta}_{jk}$, using $\sum_{t=|h|+1}^{n}(X_t - \bar{X})(X_{t-|h|} - \bar{X}) = \sum_{t=|h|+1}^{n} X_t X_{t-|h|} - \bar{X}\sum_{t=|h|+1}^{n} X_t - \bar{X}\sum_{t=|h|+1}^{n} X_{t-|h|} + (n-|h|)\bar{X}^2$, we have

$$\begin{aligned}\hat{\beta}_{jk} =& \frac{1}{\sqrt{2\pi}n}\sum_{|h|<n}\sum_{t=|h|+1}^{n} X_t X_{t-|h|}\,\hat{\Psi}_{jk}(h) - \frac{1}{\sqrt{2\pi}n}\bar{X}\sum_{|h|<n}\sum_{t=|h|+1}^{n} X_t\,\hat{\Psi}_{jk}(h)\\ &- \frac{1}{\sqrt{2\pi}n}\bar{X}\sum_{|h|<n}\sum_{t=|h|+1}^{n} X_{t-|h|}\,\hat{\Psi}_{jk}(h) + \frac{1}{\sqrt{2\pi}n}\bar{X}^2\sum_{|h|<n}(n-|h|)\,\hat{\Psi}_{jk}(h).\end{aligned} \tag{5.2}$$

Using $\hat{\Psi}_{jk}(0) = 0$, we write the first term in $\hat{\beta}_{jk}$ as a quadratic term in X_i's.

$$\begin{aligned}\sqrt{2\pi}\sum_{|h|<n}\sum_{t=|h|+1}^{n} X_t X_{t-|h|}\,\hat{\Psi}_{jk}(h) &= \sqrt{2\pi}\sum_{t=1}^{n}\sum_{h=-(t-1)}^{t-1} X_t X_{t-|h|}\,\hat{\Psi}_{jk}(h)\\ &= \sqrt{2\pi}\sum_{t=1}^{n}\sum_{h=1}^{t-1} X_t X_{t-h}[\hat{\Psi}_{jk}(-h) + \hat{\Psi}_{jk}(h)]\\ &= \sqrt{2\pi}\sum_{t=1}^{n}\sum_{s=1}^{t-1} X_t X_s[\hat{\Psi}_{jk}(-(t-s)) + \hat{\Psi}_{jk}(t-s)]\\ &= \sqrt{2\pi}\sum_{t=1}^{n}\sum_{s=1}^{n} X_t X_s\big\{[\hat{\Psi}_{jk}(-(t-s)) + \hat{\Psi}_{jk}(t-s)]/2\big\}\\ &=: \sum_{t=1}^{n}\sum_{s=1}^{n} d_n(t-s) X_t X_s,\end{aligned}$$

where

$$\begin{aligned}&d_n(t) = \sqrt{2\pi}\,[\hat{\Psi}_{jk}(-t) + \hat{\Psi}_{jk}(t)]/2 =: \int_{-\pi}^{\pi}\eta_n(w)e^{\mathbf{i}wt}dw, \text{ where}\\ &\eta_n(w) = [\Psi_{jk}(-w) + \Psi_{jk}(w)]/2.\end{aligned} \tag{5.3}$$

Denote

$$Q_{n,X} := \sum_{t=1}^{n}\sum_{s=1}^{n} d_n(t-s)X_t X_s, \quad Q_{n,Z} := \sum_{t=1}^{n}\sum_{s=1}^{n} e_n(t-s) Z_t Z_s, \tag{5.4}$$

where

$$e_n(t) = 2\pi\int_{-\pi}^{\pi}\eta_n(w)f(w)e^{\mathbf{i}wt}dw. \tag{5.5}$$

In view of (5.2), let

$$\begin{aligned}R_{n,X} :=& \bar{X}\sum_{|h|<n}\sum_{t=|h|+1}^{n} X_t\,\hat{\Psi}_{jk}(h) + \bar{X}\sum_{|h|<n}\sum_{t=|h|+1}^{n} X_{t-|h|}\,\hat{\Psi}_{jk}(h) - \bar{X}^2\sum_{|h|<n}(n-|h|)\,\hat{\Psi}_{jk}(h)\\ &=: R_{1n} + R_{2n} - R_{3n}.\end{aligned} \tag{5.6}$$

Then, combine (5.2), (5.4), (5.6), we have

$$
\begin{aligned}
\hat{\beta}_{jk} &= \frac{1}{2\pi n} Q_{n,X} - \frac{1}{\sqrt{2\pi n}} R_{n,X} \\
&= \frac{1}{2\pi n}\Big(Q_{n,X} - Q_{n,Z}\Big) - \frac{1}{\sqrt{2\pi n}} R_{n,X} + \frac{1}{2\pi n} Q_{n,Z} \\
&=: \hat{\beta}_{jk,1} + \hat{\beta}_{jk,2} + \hat{\beta}_{jk,3}.
\end{aligned} \tag{5.7}
$$

In following, we want to show that $\hat{\beta}_{jk,1}$ and $\hat{\beta}_{jk,2}$ in (5.7) are negligible and $\hat{\beta}_{jk,3}$ is a leading term. In specific, we have the following result.

Lemma 5.3 *Under the assumptions of Theorem 3.1, we have, for any integer $m \geq 1$,*

$$
E\Big(\big|\hat{\beta}_{jk,1}\big|^{2m}\Big) \leq C_m 2^{j_1 m} n^{-2m},
$$

where the constant C_m depends on m but not on n.

Proof: In view of definition of $\hat{\beta}_{jk,1}$, we only need to show that $E\big(|Q_{n,X} - Q_{n,Z}|^{2m}\big) \leq C_m 2^{j_1 m}$, which is a bound on Bartlett approximation of a quadratic form $Q_{n,X}$ with dependent variables X_i's by the corresponding quadratic form $Q_{n,Z}$ with i.i.d. random variables Z_i's. In view of (5.4), write $Q_{n,X}$ and $Q_{n,Z}$ in terms of periodogram of sequence $\{X_t\}$ and of noise $\{Z_t\}$,

$$
\begin{aligned}
Q_{n,X} :&= \sum_{k,t=1}^{n} d_n(k-t) X_k X_t = (2\pi n) \int_{-\pi}^{\pi} \eta_n(\lambda) I_{n,X}(\lambda) d\lambda, \\
Q_{n,Z} :&= \sum_{k,t=1}^{n} e_n(k-t) Z_k Z_t = (2\pi n) \int_{-\pi}^{\pi} \eta_n(\lambda) 2\pi f(\lambda) I_{n,Z}(\lambda) d\lambda,
\end{aligned} \tag{5.8}
$$

where

$$
\begin{aligned}
I_{n,X}(\lambda) &= \frac{1}{2\pi n}\Big|\sum_{j=1}^{n} X_j e^{\mathbf{i}\lambda j}\Big|^2 = |w_X(\lambda)|^2, \quad w_X(\lambda) = \frac{1}{\sqrt{2\pi n}} \sum_{j=1}^{n} X_j e^{\mathbf{i}\lambda j}, \\
I_{n,z}(\lambda) &= \frac{1}{2\pi n}\Big|\sum_{j=1}^{n} Z_j e^{\mathbf{i}\lambda j}\Big|^2 = |w_Z(\lambda)|^2, \quad w_Z(\lambda) = \frac{1}{\sqrt{2\pi n}} \sum_{j=1}^{n} Z_j e^{\mathbf{i}\lambda j}.
\end{aligned}
$$

Following similar argument and using the same notation $L_n(\lambda)$ as that in Bhansali, et al (2007b, in (5.8)), we have

$$
Q_{n,X} - Q_{n,Z} = n \int_{-\pi}^{\pi} \eta_n(\lambda) L_n(\lambda) d\lambda, \text{ where } L_n(\lambda) = 2\pi\Big[I_n(\lambda) - |\Psi(\lambda)|^2 I_{n,Z}(\lambda)\Big]. \tag{5.9}
$$

Above $L_n(\lambda)$ is a Bartlett approximation, which also could be written as

$$
L_n(\lambda) = n^{-1}\Big\{|\Delta_n(\lambda)|^2 + 2\mathcal{R}\Big(\Delta_n(\lambda)\bar{\Psi}(\lambda)\big(\sum_{l=1}^{n} Z_l e^{\mathbf{i}\lambda l}\big)\Big)\Big\}, \tag{5.10}
$$

where $\Delta_n(\lambda) = \sum_{k=-\infty}^{n} e^{-\mathbf{i}\lambda k} d_k(\lambda) Z_k - c_n(\lambda) \sum_{k=1}^{n} e^{-\mathbf{i}\lambda k} Z_k$, $c_n(\lambda) = \sum_{j=n+1}^{\infty} \varphi_j e^{-\mathbf{i}\lambda j}$, and $d_k(\lambda) = \sum_{j=1-k}^{n-k} \varphi_j e^{-\mathbf{i}\lambda j}$ for $k \leq 0$ and $d_k(\lambda) = -\sum_{j=n-k+1}^{n} \varphi_j e^{-\mathbf{i}\lambda j}$ for $1 \leq k \leq n$. Term

$\bar{\Psi}(\lambda)$ is a conjugate of $\Psi(\lambda)$, where $\Psi(\lambda) = \sum_{j=0}^{\infty} \varphi_j e^{-\mathbf{i}\lambda j}$. For further details, see Bhansali, et al (2007b, p.82). A consequence from our assumption **A1** is that, for all λ,

$$|c_n(\lambda)| \leq Cn^{-a}; \ |d_k(\lambda)| \leq C|k|_+^{-a}, \text{ if } k \leq 0; \ |d_k(\lambda)| \leq C|n-k+1|_+^{-a}, \text{ if } 1 \leq k \leq n. \quad (5.11)$$

Above results (see Lemmas 6.1 to 6.4 in Bhansali, et al (2007b) for special case $d = 0$, $\beta = 0$) will be used in following proof. From the expression of $L_n(\lambda)$ in (5.10), we could write above $Q_{n,X} - Q_{n,Z}$ in (5.9) as

$$\begin{aligned} Q_{n,X} - Q_{n,Z} &= \int_{-\pi}^{\pi} \eta_n(\lambda)|\Delta_n(\lambda)|^2 d\lambda + \int_{-\pi}^{\pi} \eta_n(\lambda)\Delta_n(\lambda)\bar{\Psi}(\lambda)\sum_{l=1}^{n} Z_l e^{\mathbf{i}\lambda l} d\lambda \\ &\quad + \int_{-\pi}^{\pi} \eta_n(\lambda)\bar{\Delta}_n(\lambda)\Psi(\lambda)\sum_{l=1}^{n} Z_l e^{-\mathbf{i}\lambda l} d\lambda \\ &=: Q_{1n} + Q_{2n} + Q_{3n}. \end{aligned} \tag{5.12}$$

Hence, in order to prove the Lemma, it suffices to show $E(|Q_{ln}|^{2m}) \leq C_m 2^{j_1 m}$ for $l = 1, 2, 3$.

Let's consider Q_{1n} first. We write that term further as

$$\begin{aligned} Q_{1n} &= \sum_{k=-\infty}^{n}\sum_{t=-\infty}^{n}\int_{-\pi}^{\pi} e^{\mathbf{i}(t-k)\lambda} d_k(\lambda)\bar{d}_t(\lambda)\eta_n(\lambda)d\lambda\, Z_k Z_t \\ &\quad - \sum_{k=-\infty}^{n}\sum_{t=1}^{n}\int_{-\pi}^{\pi} e^{\mathbf{i}(t-k)\lambda}\bar{c}_n(\lambda)d_k(\lambda)\eta_n(\lambda)d\lambda\, Z_k Z_t \\ &\quad - \sum_{k=1}^{n}\sum_{t=-\infty}^{n}\int_{-\pi}^{\pi} e^{\mathbf{i}(t-k)\lambda}c_n(\lambda)\bar{d}_t(\lambda)\eta_n(\lambda)d\lambda\, Z_k Z_t \\ &\quad + \sum_{k=1}^{n}\sum_{t=1}^{n}\int_{-\pi}^{\pi} e^{\mathbf{i}(t-k)\lambda}|c_n(\lambda)|^2\eta_n(\lambda)d\lambda\, Z_k Z_t \\ &=: \sum_{k=-\infty}^{n}\sum_{t=-\infty}^{n}\xi_{1n}(k,t)\, Z_k Z_t - \sum_{k=-\infty}^{n}\sum_{t=1}^{n}\xi_{2n}(k,t)\, Z_k Z_t \\ &\quad - \sum_{k=1}^{n}\sum_{t=-\infty}^{n}\xi_{3n}(k,t)\, Z_k Z_t + \sum_{k=1}^{n}\sum_{t=1}^{n}\xi_{4n}(k,t)\, Z_k Z_t \\ &=: Q_{11n} - Q_{12n} - Q_{13n} + Q_{14n}. \end{aligned} \tag{5.13}$$

Again, in order to prove $E(|Q_{1n}|^{2m}) \leq C_m 2^{j_1 m}$, it suffices to show $E(|Q_{1ln}|^{2m}) \leq C_m 2^{j_1 m}$, $l = 1, 2, \cdots, 4$. Since proofs to four terms are similar, we only provide a proof to the first term $E(|Q_{11n}|^{2m}) \leq C_m 2^{j_1 m}$.

Recall $\xi_{1n}(k,t) = \int_{-\pi}^{\pi} e^{\mathbf{i}(t-k)\lambda} d_k(\lambda)\bar{d}_t(\lambda)\eta_n(\lambda)d\lambda$, use above results in (5.11) and $|\eta_n(\lambda)| \leq C2^{j_1/2}$, we have

$$\sum_{k=-\infty}^{n}\sum_{t=-\infty}^{n}\xi_{1n}^2(k,t) \leq C2^{j_1}. \tag{5.14}$$

Write Q_{11n} further as

$$
\begin{aligned}
Q_{11n} &= \sum_{k=-\infty}^{n} \sum_{t=-\infty}^{n} \xi_{1n}(k,t)\big(Z_k Z_t - E(Z_k Z_t)\big) + \sum_{k=-\infty}^{n} \sum_{t=-\infty}^{n} \xi_{1n}(k,t) E(Z_k Z_t) \\
&=: Q_{11n,1} + Q_{11n,2}.
\end{aligned} \tag{5.15}
$$

Apply Rosenthal's inequality to term $Q_{11n,1}$ and combine (5.14) and assumption **A2**, we have, for all $m \geq 1$, there is a constant C_m such that

$$
E(|Q_{11n,1}|^{2m}) \leq C_m \Big(\sum_{k=-\infty}^{n} \sum_{t=-\infty}^{n} \xi_{1n}^2(k,t) \Big)^m \leq C_m 2^{j_1 m}. \tag{5.16}
$$

As to $Q_{11n,2}$, we have $Q_{11n,2} = C \sum_{k=-\infty}^{n} \xi_{1n}(k,k)$. In view of $\xi_{1n}(k,t)$, (5.11), and $|\eta_n(\lambda)| \leq C 2^{j_1/2}$, we obtain $|Q_{11n,2}| \leq C 2^{j_1/2}$. Thus, we have $|Q_{11n,2}|^{2m} \leq C_m 2^{j_1 m}$. Combine with (5.15) and (5.16), we prove that $E(|Q_{11n}|^{2m}) \leq C_m 2^{j_1 m}$. Thus, in view of (5.13), we prove the first term Q_{1n} in (5.12).

Since proofs to terms Q_{2n} and Q_{3n} are same, we only provide a proof to Q_{2n} below. In view of definition of $\Delta_n(\lambda)$, we write that term as

$$
\begin{aligned}
Q_{2n} &= \sum_{k=-\infty}^{n} \sum_{t=1}^{n} \beta_{1n}(k,t)\, Z_k Z_t - \sum_{k=1}^{n} \sum_{t=1}^{n} \beta_{2n}(k,t)\, Z_k Z_t \\
&=: Q_{21n} - Q_{22n},
\end{aligned} \tag{5.17}
$$

where

$$
\begin{aligned}
\beta_{1n}(k,t) &= \int_{-\pi}^{\pi} e^{\mathbf{i}(t-k)\lambda} d_k(\lambda) \Psi(\lambda) \eta_n(\lambda) d\lambda, \\
\beta_{2n}(k,t) &= \int_{-\pi}^{\pi} e^{\mathbf{i}(t-k)\lambda} c_n(\lambda) \Psi(\lambda) \eta_n(\lambda) d\lambda.
\end{aligned} \tag{5.18}
$$

As to Q_{21n}, we first have

$$
\sum_{k=-\infty}^{n} \sum_{t=1}^{n} \beta_{1n}^2(k,t) \leq \sum_{k=-\infty}^{n} \sum_{t=-\infty}^{\infty} \beta_{1n}^2(k,t) \leq C \sum_{k=-\infty}^{n} \int_{-\pi}^{\pi} |d_k(\lambda)\Psi(\lambda)\eta_n(\lambda)|^2 d\lambda \leq C 2^{j_1}, \tag{5.19}
$$

where the 2nd inequality follows from Parseval's equality, and 3rd inequality follows from (5.11), $|\Psi(\lambda)| \leq C$, and $|\eta_n(\lambda)| \leq C 2^{j_1/2}$. Now, follow the similar argument to term Q_{11n}, apply Rosenthal's inequality to term $Q_{21n,1}$ and use (5.19), we are able to show that $E(|Q_{21n}|^{2m}) \leq C_m 2^{j_1 m}$. As to Q_{22n}, we first have

$$
\sum_{k=1}^{n} \sum_{t=1}^{n} \beta_{2n}^2(k,t) \leq \sum_{k=1}^{n} \sum_{t=-\infty}^{\infty} \beta_{2n}^2(k,t) \leq C \sum_{k=1}^{n} \int_{-\pi}^{\pi} |c_n(\lambda)\Psi(\lambda)\eta_n(\lambda)|^2 d\lambda \leq C 2^{j_1} n^{-2a+1}, \tag{5.20}
$$

where the last inequality follows from (5.11) and $|\eta_n(\lambda)| \leq C 2^{j_1/2}$. From **A1** with $a > 1$, we see that term Q_{22n} is much smaller than Q_{21n}. Thus we prove term Q_{2n}. Combine with term Q_{1n}, we complete the proof to Lemma.

Lemma 5.4 *Under the assumptions of Theorem 3.1, we have, for any integer $m \geq 1$,*

$$E\Big(\big|\hat{\beta}_{jk,2}\big|^{2m}\Big) \leq C_m 2^{j_1 m} n^{-2m},$$

where the constant C_m depends on m but not on n.

Proof: In view of $\hat{\beta}_{jk,2}$ in (5.7), we only need to show that $E(|R_{n,X}|^{2m}) \leq C_m 2^{j_1 m}$. It is equivalent to show $E(|R_{ln}|^{2m}) \leq C_m 2^{j_1 m}$ for $l = 1, 2, 3$. Since proofs to three terms are similar, we only provide a proof to first term R_{1n} given in (5.6). After exchange the order of two summations, we write it as

$$R_{1n} = \bar{X} \sum_{t=1}^{n} b_t X_t, \quad \text{where} \quad b_t = \sum_{|h|<t} \hat{\Psi}_{jk}(h). \tag{5.21}$$

Now apply Cauchy-Schwarz inequality, for any integer $m \geq 1$, we have

$$E\big(R_{1n}^{2m}\big) \leq \Big[E(\bar{X}^{4m})\Big]^{1/2} \Big[E\Big(\sum_{t=1}^{n} b_t X_t\Big)^{4m}\Big]^{1/2}. \tag{5.22}$$

From a result such as $E|S_n|^p \leq C_p(ES_n^2)^{p/2}$ for $p \geq 2$, where $S_n = \sum_{t=1}^n X_t$, $X_t = \sum_{j=0}^\infty \varphi_j Z_{t-j}$ and C_p depends on p and $E|Z_0|^p$ only (see Proposition 4.4.3. in Gitaitis, 2012, p.79), we have $E(\bar{X}^{4m}) \leq C_m n^{-4m}(ES_n^2)^{2m} \leq C_m n^{-2m}$. The last inequality follows from $E(S_n^2) \leq Cn$. Note that C_m may have different values at different places, but they do not depend on φ_j's and n. As to the second term, from $X_t = \sum_{j=0}^\infty \varphi_j Z_{t-j} = \sum_{k=-\infty}^{t} \varphi_{t-k} Z_k$, we write

$$S := \sum_{t=1}^{n} b_t X_t = \sum_{t=1}^{n} \sum_{k=-\infty}^{t} b_t \varphi_{t-k} Z_k = \sum_{k=-\infty}^{n} \sum_{t=k\vee 1}^{n} b_t \varphi_{t-k} Z_k =: \sum_{k=-\infty}^{n} d_k Z_k. \tag{5.23}$$

Now apply a result such as $E|S|^p \leq C_p(ES^2)^{p/2}$ for $p \geq 2$ where $S = \sum_{k\in\mathbb{Z}} d_k Z_k$ and C_p depends on p and $E|Z_0|^p$ but not on d_k's (see Corollary 2.5.1. in Gitaitis, 2012, p.30), we have, for $p = 4m$,

$$E\Big(\sum_{t=1}^{n} b_t X_t\Big)^{4m} \leq C_m \Big[E\Big(\sum_{k=-\infty}^{n} d_k Z_k\Big)^2\Big]^{2m} = C_m \Big[E\Big(\sum_{t=1}^{n} b_t X_t\Big)^2\Big]^{2m}.$$

Furthermore, we have

$$E\Big(\sum_{t=1}^{n} b_t X_t\Big)^2 = \sum_{j.k=1}^{n} b_j r(j-k) b_k, \quad \text{where} \quad r(h) = \mathrm{Cov}(X_1, X_{1+h}).$$

As to the above b_j's, we want to show that $|b_t| \leq C 2^{j_1/2}$ for all $t \geq 1$ below. Recall $b_t = \sum_{|h|<t} \hat{\Psi}_{jk}(h)$, From our assumption **A3**, our wavelet satisfies $|\hat{\psi}(w)| \leq \gamma_0(1+|w|)^{-\gamma}, \gamma > 1$. Combining with $\hat{\Psi}_{jk}(h) = \sqrt{2\pi}\,\hat{\psi}_{jk}(2\pi h)$ and $\hat{\psi}_{jk}(2\pi h) = e^{-\mathrm{i}2\pi hk/2^j} 2^{-j/2} \hat{\psi}(2\pi h/2^j)$, we have $|\hat{\psi}_{jk}(2\pi h)| \leq C\, 2^{-j/2} 2^{\gamma j}(2^j + 2\pi h)^{-\gamma}$. Thus we have $|\hat{\Psi}_{jk}(h)| \leq C\, 2^{-j/2} 2^{\gamma j}(2^j + 2\pi h)^{-\gamma}$. From this, we have $\sum_{|h|<\infty} |\hat{\Psi}_{jk}(h)| \leq (\sum_{|h|\leq 2^j} + \sum_{|h|>2^j}) C\, 2^{-j/2} 2^{\gamma j}(2^j + 2\pi h)^{-\gamma} \leq C 2^{j/2}$ for all

$j \le j_1$. Therefore, we have that $|b_t| \le \sum_{|h|<\infty} |\hat{\Psi}_{jk}(h)| \le C2^{j_1/2}$ for all $t \ge 1$. Combine with $\sum_{j.k=1}^{n} |r(j-k)| \le Cn$ from $\sum_h |r(h)| < \infty$, we have $E(\sum_{t=1}^n b_t X_t)^2 \le C2^{j_1} n$. Thus, we have $E(\sum_{t=1}^n b_t X_t)^{4m} \le C_m 2^{2j_1 m} n^{2m}$. Therefore, from (5.22), we obtain $E(R_{1n}^{2m}) \le C_m n^{-m} 2^{j_1 m} n^m = C_m 2^{j_1 m}$, which proves the Lemma.

Combining above Lemmas 5.3 and 5.4, we are able to obtain an exponential inequality for the difference between empirical wavelet coefficients $\hat{\beta}_{jk}$ and true wavelet coefficients β_{jk}.

Lemma 5.5 *Let empirical wavelet coefficients $\hat{\beta}_{jk}$ be defined in (2.11). Under the assumptions of Theorem 3.1, we have*

$$P\Big\{\Big|\hat{\beta}_{jk} - \beta_{jk}\Big| > \delta_{jk}\Big\} \le Cn^{-2}, \quad \forall j \in [j_0,\, j_1] \;\; \textit{and} \;\; k = 0, 1, \cdots, 2^j - 1.$$

Proof: We consider two cases $k \in \Gamma_j$ and $k \in \Gamma_j^c$ separately. When $k \in \Gamma_j$, we have $\delta_{jk}^2 = K_1 \ln(\#\Lambda)\, \sigma_{jk}^2$ and $n\sigma_{jk}^2 \ge \delta_0 > 0$ from (2.14). In view of (5.7), we have

$$P\Big\{\Big|\hat{\beta}_{jk} - \beta_{jk}\Big| > \delta_{jk}\Big\} \le P\Big\{|\hat{\beta}_{jk,1}| > \gamma_1 \delta_{jk}\Big\} + P\Big\{|\hat{\beta}_{jk,2}| > \gamma_2 \delta_{jk}\Big\} + P\Big\{|\hat{\beta}_{jk,3} - \beta_{jk}| > \gamma_3 \delta_{jk}\Big\},$$

where $\gamma_1, \gamma_2, \gamma_3$ are any fixed positive numbers such as $\gamma_1 + \gamma_2 + \gamma_3 = 1$. From Markov inequality, we have

$$P\Big\{|\hat{\beta}_{jk,1}| > \gamma_1 \delta_{jk}\Big\} \le \frac{E\big(|\hat{\beta}_{jk,1}|^{2m}\big)}{\gamma_1^{2m} \delta_{jk}^{2m}} \le \frac{C_m 2^{j_1 m} n^{-2m}}{\gamma_1^{2m} K_1^m (\ln(\#\Lambda))^m \sigma_{jk}^{2m}} = \frac{C_m 2^{j_1 m} n^{-m}}{\gamma_1^{2m} K_1^m (\ln(\#\Lambda))^m (n\sigma_{jk}^2)^m} \le Cn^{-2}.$$

The above second inequality follows from Lemma 5.3 and our definition of threshold δ_{jk}, and the last inequality follows from $\#\Lambda = 2^{j_1+1} - 2^{j_0}$, our selection of j_1 such as $2^{j_1} \simeq n^{1-\tau}$, $n\sigma_{jk}^2 \ge \delta_0 > 0$ for $k \in \Gamma_j$ and m is large enough such as $m \ge 2/\tau$.

Similarly, we have

$$P\Big\{|\hat{\beta}_{jk,2}| > \gamma_2 \delta_{jk}\Big\} \le \frac{E\big(|\hat{\beta}_{jk,2}|^{2m}\big)}{\gamma_2^{2m} \delta_{jk}^{2m}} \le \frac{C_m 2^{j_1 m} n^{-2m}}{\gamma_2^{2m} K_1^m (\ln(\#\Lambda))^m \sigma_{jk}^{2m}} \le Cn^{-2},$$

where the second inequality follows from above Lemma 5.4.

In view of (5.4) and (5.5), the last term could be written as

$$\begin{aligned} P\Big\{|\hat{\beta}_{jk,3} - \beta_{jk}| > \gamma_3 \delta_{jk}\Big\} &= P\Big\{|Q_{n,Z} - 2\pi n \beta_{jk}| > 2\pi \gamma_3 n \delta_{jk}\Big\} \\ &= P\Big\{\Big|\sum_{k,t=1}^{n} e_n(k-t) Z_k Z_t - E\sum_{k,t=1}^{n} e_n(k-t) Z_k Z_t\Big| > 2\pi\gamma_3 n \delta_{jk}\Big\} \\ &= P\Big\{\big|Z^T A Z - E(Z^T A Z)\big| > 2\pi\gamma_3 n\delta_{jk}\Big\}, \end{aligned}$$

where above matrix $A = (a_{n;tk})_{n\times n}$ is defined as $a_{n;tk} := e_n(t-k) = \int_{-\pi}^{\pi} 2\pi \eta_n(w) f(w) e^{\mathbf{i}(t-k)w} dw$. We write $a_{n;tk}$ further as $a_{n;tk} = \int_{-\pi}^{\pi} g_n(w) e^{\mathbf{i}(t-k)w} dw$, where $g_n(w) := 2\pi\eta_n(w) f(w)$. From the definition of $\eta_n(w)$, we have $|\eta_n(w)| \le C2^{j/2}$. Together with $f(w)$ is bounded, we obtain that $|g_n(w)| \le C2^{j/2}$ for some finite constant C. Apply Theorem 2.2 with $\alpha = 0$ in Bhansali et al. (2007a, p.728), we have the operator norm of A is $\|A\| \le C2^{j_1/2}$ for all $j_0 \le j \le j_1$.

Next, we calculate its Hilbert-schmidt norm $\|A\|_{HS}$ of A. In view of (5.4), (5.5) and (5.3), apply a change of variable ($h = t-s$) and exchange the order of summation for h and t, we have

$$\|A\|_{HS}^2 = \sum_{t=1}^{n}\sum_{s=1}^{n} |e_n(t-s)|^2 = \sum_{h=-(n-1)}^{(n-1)} (n-|h|)\,|e_n(h)|^2.$$

Thus, one has

$$n^{-1}\|A\|_{HS}^2 \to \sum_{h=-\infty}^{\infty} e_n(h)\,\overline{e_n(h)}, \quad n\to\infty, \tag{5.24}$$

where $\overline{e_n(h)}$ stands for the conjugate of $e_n(h)$. Above $e_n(h)$ could be written as $e_n(h) = (2\pi)^{-1/2}\int_{-\pi}^{\pi}(2\pi)^{3/2}\eta_n(w)f(w)e^{-\mathbf{i}hw}dw$. It means that $e_n(h)$'s are Fourier coefficients of the real function $(2\pi)^{3/2}\eta_n(w)f(w)$. Therefore, from Parseval's Theorem, one has

$$\begin{aligned}\sum_{h=-\infty}^{\infty} e_n(h)\,\overline{e_n(h)} &= \int_{-\pi}^{\pi}(2\pi)^3\eta_n^2(w)f^2(w)dw \\ &= \frac{(2\pi)^3}{4}\int_{-\pi}^{\pi}\Big[\Psi_{jk}(-w)+\Psi_{jk}(w)\Big]^2 f^2(w)dw.\end{aligned} \tag{5.25}$$

Now, let's calculate the variance $\sigma_{jk}^2 = \mathrm{Var}(\hat\beta_{jk})$. In view of (5.7) and above discussion, we have $\mathrm{Var}(\hat\beta_{jk}) = \mathrm{Var}(\hat\beta_{jk,3})(1+o(1)) = (2\pi n)^{-2}\mathrm{Var}(Q_{n,Z})(1+o(1)) = (2\pi n)^{-2}\mathrm{Var}(Z^TAZ)(1+o(1))$. Through a direct calculation, we have $\mathrm{Var}(Z^TAZ) = 2\sum_{s,t=1:t\neq s}^{n} e_n^2(t-s) + \mathrm{Var}(Z_1^2)$ $\sum_{t=1}^{n} e_n^2(t-t) = 2\|A\|_{HS}^2 + (\mu_4-3)ne_n^2(0) = n\{2n^{-1}\|A\|_{HS}^2 + (\mu_4-3)\beta_{jk}^2\}$, where $\mu_4 = E(Z_1^4)$. Thus, we have

$$\sigma_{jk}^2 = \mathrm{Var}(\hat\beta_{jk}) = \frac{1}{n}\Big\{\pi\int_{-\pi}^{\pi}\big[\Psi_{jk}(w)+\Psi_{jk}(-w)\big]^2 f^2(w)dw + \frac{(\mu_4-3)\beta_{jk}^2}{(2\pi)^2}\Big\}\big(1+o(1)\big). \tag{5.26}$$

Above variance expression is used in our spectral density estimator $\bar f(\cdot)$ in (2.12). One also has $\min\{2,\mathrm{Var}(Z_1^2)\}\|A\|_{HS}^2 \le \mathrm{Var}(Z^TAZ) \le \max\{2,\mathrm{Var}(Z_1^2)\}\|A\|_{HS}^2$, i.e., $\mathrm{Var}(Z^TAZ)$ and $\|A\|_{HS}^2$ have the same order. In particular, one has the low bound

$$\sigma_{jk}^2 \ge \frac{\min\{2,\mathrm{Var}(Z_1^2)\}\cdot\pi}{2n}\int_{-\pi}^{\pi}\big[\Psi_{jk}(w)+\Psi_{jk}(-w)\big]^2 f^2(w)dw\cdot\big(1+o(1)\big). \tag{5.27}$$

From (5.24), (5.25), (2.14) and $n\sigma_{jk}^2 \le C_0$ for all (j,k), one obtains that $C_1 n \le \|A\|_{HS}^2 \le C_2 n$ for some finite and positive constants C_1 and C_2. Now, In view of above Lemma 5.2 with $t = 2\pi\gamma_3 n\delta_{jk} \le C(n\ln n)^{1/2}$, together with $C_1 n \le \|A\|_{HS}^2 \le C_2 n$, $\|A\| \le C2^{j_1/2}$ with $2^{j_1} \simeq n^{1-\tau}$, it is not hard to obtain that the second term (i.e., $t/(K^2\|A\|)$) dominates the first term (i.e., $t^2/(K^4\|A\|_{HS}^2)$) in the right side of the inequality in Lemma 5.2. Now, apply above Lemma 5.2 with the constant K_1 in the threshold δ_{jk} large enough, one obtains $P\{|\hat\beta_{jk,3}-\beta_{jk}| > \gamma_3\delta_{jk}\} \le Cn^{-2}$. Combine with the other two terms, we prove inequality claimed in Lemma for the first case $k\in\Gamma_j$.

For the second case, $k \in \Gamma_j^c$, we have $\delta_{jk}^2 = K_1 \ln(\#\Lambda)\, \delta_0 n^{-1}$ and $n\sigma_{jk}^2 < \delta_0$. We follow the same steps.

$$P\Big\{|\hat{\beta}_{jk,1}| > \gamma_1 \delta_{jk}\Big\} \le \frac{E\big(|\hat{\beta}_{jk,1}|^{2m}\big)}{\gamma_1^{2m}\delta_{jk}^{2m}} = \frac{C_m 2^{j_1 m} n^{-2m}}{\gamma_1^{2m} K_1^m (\ln(\#\Lambda))^m \delta_0^m (n)^m} \le C n^{-2},$$

as long as m is large enough such as $m \ge 2/\tau$. Similarly, we can obtain the second inequality for $\hat{\beta}_{jk,2}$. For the last term, similarly we write

$$P\{|\hat{\beta}_{jk,3} - \beta_{jk}| > \gamma_3 \delta_{jk}\} = P\Big\{\big|Z^T A Z - E(Z^T A Z)\big| > 2\pi\gamma_3 n \delta_{jk}\Big\}$$

with $2\pi\gamma_3 n\delta_{jk} = 2\pi\gamma_3 (K_1 \delta_0 \ln(\#\Lambda)\, n)^{1/2} =: t$. In view of above Lemma 5.2, we have the first term is at the order $t^2/(K^4\|A\|_{HS}^2) = (2\pi)^2 \gamma_3^2 K^{-4} K_1 \ln(\#\Lambda)\, \delta_0 (n\sigma_{jk}^2)^{-1} \ge C K_1 \ln n$ for some constant C from $n\sigma_{jk}^2 < \delta_0$. The second term is at the order $t/(K^2\|A\|) = 2\pi\gamma_3 K_1^{1/2} K^{-2} \delta_0^{1/2}$ $(\ln(\#\Lambda))^{1/2} n^{1/2} \|A\|^{-1}$. Since $\|A\| \le C 2^{j_1/2}$ with $2^{j_1} \simeq n^{1-\tau}$, one has $n^{1/2}\|A\|^{-1} \ge C^{-1} n^{\tau/2}$. Thus, the second term $t/(K^2\|A\|) \ge C n^{\tau/2}$ for some constant C. Now apply Lemma 5.2 either with the first term $t^2/(K^4\|A\|_{HS}^2)$ with sufficient large K_1 or the second term $t/(K^2\|A\|)$, we could obtain the desired rate. Thus, we complete the proof to the lemma.

Now, we are in a position to bound above five terms I_0, $I_1, \cdots, I_5$ in the desired rate.

Lemma 5.6 *Under the assumptions of Theorem 3.1, we have*

$$I_0 := E\big(\hat{\alpha}_{00} - \alpha_{00}\big)^2 = o\bigg(\Big(\frac{\ln n}{n}\Big)^{2s/(2s+1)}\bigg).$$

Proof: Recall $\hat{\alpha}_{00} = (2\pi)^{-1/2}\hat{r}(0) = (2\pi)^{-1/2} n^{-1} \sum_{t=1}^n (X_t - \bar{X})^2$ and $\alpha_{00} = (2\pi)^{-1/2} r(0)$, we have $\sqrt{2\pi}\,(\hat{\alpha}_{00} - \alpha_{00}) = \hat{r}(0) - r(0)$. Now applying the Proposition 7.3.4 from Brockwell and Davis (1991, p.229), we obtain $I_0 \le C n^{-1}$, which proves the Lemma.

Lemma 5.7 *Under the assumptions of Theorem 3.1, we have*

$$I_1 := \sum_{j=j_1+1}^{\infty} \sum_{k=0}^{2^j-1} \beta_{jk}^2 = o\bigg(\Big(\frac{\ln n}{n}\Big)^{2s/(2s+1)}\bigg).$$

Proof: Let's consider $1 \le p < 2$ first. From Lemma 5.1 and (2.6), one has $\|\beta_{j\cdot}\|_2 \le \|\beta_{j\cdot}\|_p \le M 2^{-j\sigma}$, where $\sigma = s + 1/2 - 1/p$. Thus $\sum_k \beta_{jk}^2 \le M^2 2^{-2j\sigma}$. Thus $I_1 \le M^2 \sum_{j=j_1}^{\infty} 2^{-2j\sigma} \le C 2^{-2j_1\sigma}$. From our choice of j_1 satisfying $2^{j_1} \simeq n^{1-\tau}$, one has $I_1 \le C n^{-2\sigma(1-\tau)}$. Recall τ satisfies $0 < \tau \le 2\tau_0/(2\tau_0+1)$ where τ_0 is a arbitrary small number provided in assumption **A4**. Let $\tau_0 \le 1/4$, then one has $0 < \tau \le 1/3$. Then, one has $I_1 \le C n^{-4\sigma/3}$. When $1 \le p < 2$ and $s > 1/p$, through simple algebra, one could show $4\sigma/3 \ge 2s/(2s+1)$. Thus, one obtains $I_1 \le C n^{-2s/(2s+1)}$. Therefore we obtain that $I_1 = o((\ln n/n)^{2s/(2s+1)})$.

For $p \ge 2$, again from Lemma 5.1 and (2.6), one has $\|\beta_{j\cdot}\|_2 \le (C 2^j)^{\frac{1}{2}-\frac{1}{p}} \|\beta_{j\cdot}\|_p \le C 2^{-js}$. Thus $I_1 \le C \sum_{j_1}^{\infty} 2^{-2js} \le C 2^{-2j_1 s}$. Again, based on our choice of j_1 such as $2^{j_1} \simeq n^{1-\tau}$, $0 < \tau \le 2\tau_0/(2\tau_0+1)$ and $s \ge \tau_0$, one could show that $I_1 \le C n^{-2s/(2s+1)}$ and thus $I_1 = o((\ln n/n)^{2s/(2s+1)})$. Therefore, one proves the Lemma.

Lemma 5.8 *Under the assumptions of Theorem 3.1, we have*

$$I_2 := \sum_{j=0}^{j_0-1} \sum_{k=0}^{2^j-1} E\big(\hat{\beta}_{jk} - \beta_{jk}\big)^2 = o\Big(\Big(\frac{\ln n}{n}\Big)^{2s/(2s+1)}\Big).$$

Proof: Write $\hat{\beta}_{jk} - \beta_{jk} = \hat{\beta}_{jk,1} + \hat{\beta}_{jk,2} + \hat{\beta}_{jk,3} - \beta_{jk}$. Apply Lemma 5.3 with $m = 2$, we obtain $E(\hat{\beta}^2_{jk,1}) \le Cn^{-1}$. Similarly, from Lemma 5.4 with $m = 2$, we obtain $E(\hat{\beta}^2_{jk,2}) \le Cn^{-1}$. From the discussion of Lemma 5.5, we have $E(\hat{\beta}_{jk,3} - \beta_{jk})^2 \le Cn^{-1}$. Now, from simple inequality, we have $E(\hat{\beta}_{jk} - \beta_{jk})^2 \le Cn^{-1}$. Thus, we have $I_2 \le C2^{j_0}n^{-1}$. Hence we have $I_2 = o((\ln n/n)^{2s/(2s+1)})$ from our choice of j_0 in (2.16).

Lemma 5.9 *Under the assumptions of Theorem 3.1, we have*

$$I_3 := \sum_{j=j_0}^{j_s} \sum_{k\in\Gamma_j} E\Big[\hat{\beta}_{jk} I\big(|\hat{\beta}_{jk}| > \delta_{jk}\big) - \beta_{jk}\Big]^2 \le C\Big(\frac{\ln n}{n}\Big)^{2s/(2s+1)},$$

where $j_s = j_s(n)$ such that $2^{j_s} \simeq (n/\ln n)^{1/(2s+1)}$.

Proof: From a simple inequality, we have

$$\begin{aligned} I_3 &\le 2\sum_{j=j_0}^{j_s} \sum_{k\in\Gamma_j} E\Big[\beta^2_{jk} I\big(|\hat{\beta}_{jk}| \le \delta_{jk}\big)\Big] + 2\sum_{j=j_0}^{j_s} \sum_{k\in\Gamma_j} E\Big[\big(\hat{\beta}_{jk} - \beta_{jk}\big)^2 I\big(|\hat{\beta}_{jk}| > \delta_{jk}\big)\Big] \\ &=: 2(I_{31} + I_{32}). \end{aligned} \tag{5.28}$$

Also,

$$\begin{aligned} I_{31} &\le \sum_{j=j_0}^{j_s} \sum_{k\in\Gamma_j} \beta^2_{jk} I\big(|\beta_{jk}| \le 2\delta_{jk}\big) + \sum_{j=j_0}^{j_s} \sum_{k\in\Gamma_j} \beta^2_{jk} P\big(|\hat{\beta}_{jk} - \beta_{jk}| > \delta_{jk}\big) \\ &=: I_{311} + I_{312}. \end{aligned} \tag{5.29}$$

From $\delta_{jk} \le C(\ln n/n)^{1/2}$, we have

$$I_{311} \le C\frac{\ln n}{n} \sum_{j=j_0}^{j_s} \sum_{k\in\Gamma_j} 1 \le C\Big(\frac{\ln n}{n}\Big)^{2s/(2s+1)}.$$

As to the term I_{312}, from Lemma 5.5, we have $I_{312} \le Cn^{-1} \sum_{j=j_0}^{j_s} \sum_{k=0}^{2^j-1} \beta^2_{jk}$. Now apply the arguments in Lemma 5.7 for $1 \le p < 2$ and $p \ge 2$, it is easy to have

$$I_{312} \le Cn^{-1} = o\Big(\Big(\frac{\ln n}{n}\Big)^{2s/(2s+1)}\Big). \tag{5.30}$$

Now let's consider the second term I_{32}. We have $I_{32} \le 2\sum_{j=j_0}^{j_s} \sum_{k=0}^{2^j-1} E\big(\hat{\beta}_{jk} - \beta_{jk}\big)^2$. Now from the proof to Lemmas 5.8, we have $E\{\big(\hat{\beta}_{jk} - \beta_{jk}\big)^2\} \le Cn^{-1}$. Thus, we have $I_{32} \le C(\ln n/n)^{2s/(2s+1)}$. Combined with the proof to I_{31}, this completes the proof of the lemma.

Lemma 5.10 *Under the assumptions of Theorem 3.1, we have*

$$I_4 := \sum_{j=j_s+1}^{j_1} \sum_{k\in\Gamma_j} E\Big[\hat{\beta}_{jk} I\big(|\hat{\beta}_{jk}| > \delta_{jk}\big) - \beta_{jk}\Big]^2 \le C\Big(\frac{\ln n}{n}\Big)^{2s/(2s+1)}.$$

where $j_s = j_s(n)$ such that $2^{j_s} \simeq (n/\ln n)^{1/(2s+1)}$.

Proof: As in Lemma 5.9, we have

$$\begin{aligned} I_4 &\le 2\sum_{j=j_s+1}^{j_1} \sum_{k\in\Gamma_j} E\Big[\beta_{jk}^2 I\big(|\hat{\beta}_{jk}| \le \delta_{jk}\big)\Big] + 2\sum_{j=j_s+1}^{j_1} \sum_{k\in\Gamma_j} E\Big[\big(\hat{\beta}_{jk} - \beta_{jk}\big)^2 I\big(|\hat{\beta}_{jk}| > \delta_{jk}\big)\Big] \\ &=: 2(I_{41} + I_{42}). \end{aligned} \tag{5.31}$$

Also,

$$\begin{aligned} I_{41} &\le \sum_{j=j_s+1}^{j_1} \sum_{k\in\Gamma_j} \beta_{jk}^2 P\big(|\hat{\beta}_{jk}| \le \delta_{jk}, |\beta_{jk}| \le 2\delta_{jk}\big) + \sum_{j=j_s+1}^{j_1} \sum_{k\in\Gamma_j} \beta_{jk}^2 P\big(|\hat{\beta}_{jk}| \le \delta_{jk}, |\beta_{jk}| > 2\delta_{jk}\big) \\ &=: I_{411} + I_{412}. \end{aligned} \tag{5.32}$$

Let's consider the term I_{411} first. We need to consider $p \ge 2$ and $1 \le p < 2$ separately. When $p \ge 2$, using argument in Lemma 5.7, we have

$$I_{411} \le \sum_{j=j_s+1}^{j_1} \sum_{k=0}^{2^j-1} \beta_{jk}^2 \le C \sum_{j=j_s+1}^{j_1} 2^{-2sj} = C2^{-2sj_s} \le C\Big(\frac{\ln n}{n}\Big)^{2s/(2s+1)}. \tag{5.33}$$

If $1 \le p < 2$, we have

$$I_{411} \le \sum_{j=j_s+1}^{j_1} \sum_{k=0}^{2^j-1} \beta_{jk}^2 \big(4\delta_{jk}^2\,\beta_{jk}^{-2}\big)^{1-p/2} \le C(\ln n/n)^{1-p/2} \sum_{j=j_s+1}^{j_1} \sum_{k=0}^{2^j-1} |\beta_{jk}|^p, \tag{5.34}$$

from $\delta_{jk}^2 \le Cn^{-1}\ln n$. Now, applying the argument in Lemma 5.7, we have $\sum_{k=0}^{2^j-1} |\beta_{jk}|^p \le CM2^{-j\sigma p}$. Through some algebra, we can show that

$$I_{411} \le C\Big(\frac{\ln n}{n}\Big)^{2s/(2s+1)}.$$

Thus, we have bounded the term I_{411} for all $1 \le p < \infty$.

As to the term I_{412}, we have $I_{412} \le C\sum_{j=j_s+1}^{j_1} \sum_{k\in\Gamma_j} \beta_{jk}^2 P(|\hat{\beta}_{jk} - b_{jk}| > \delta_{jk})$. Now applying Lemma 5.5 and the same argument to I_{312}, we obtain $I_{412} \le Cn^{-1}$. Together with (5.32) and (5.33), we prove $I_{41} \le C(\ln n/n)^{2s/(2s+1)}$.

As to the term I_{42}, let $a_1 > 0, a_2 > 0, a_1 + a_2 = 1$, we have

$$\begin{aligned} I_{42} &\le \sum_{j=j_s+1}^{j_1} \sum_{k\in\Gamma_j} E\Big[\big(\hat{\beta}_{jk} - \beta_{jk}\big)^2 I\big(|\beta_{jk}| > a_1\delta_{jk}\big)\Big] + \sum_{j=j_s+1}^{j_1} \sum_{k\in\Gamma_j} E\Big[\big(\hat{\beta}_{jk} - \beta_{jk}\big)^2 I\big(|\hat{\beta}_{jk} - \beta_{jk}| > a_2\delta_{jk}\big)\Big] \\ &=: I_{421} + I_{422}. \end{aligned} \tag{5.35}$$

As to the term I_{421}, like the term I_{411}, we need to consider $p \geq 2$ and $1 \leq p < 2$ respectively. When $p \geq 2$, we have

$$I_{421} \leq \sum_{j=j_s+1}^{j_1} \sum_{k\in\Gamma_j} \beta_{jk}^2 a_1^{-2} \delta_{jk}^{-2} E\big(\hat{\beta}_{jk} - \beta_{jk}\big)^2 \leq C \sum_{j=j_s+1}^{j_1} \sum_{k=0}^{2^j-1} \beta_{jk}^2 \leq C(\ln n/n)^{2s/(2s+1)}.$$

When $1 \leq p < 2$, we have

$$I_{421} \leq C \sum_{j=j_s+1}^{j_1} \sum_{k\in\Gamma_j} |\beta_{jk}|^p a_1^{-p} \delta_{jk}^{-p} E\big(\hat{\beta}_{jk} - \beta_{jk}\big)^2. \tag{5.36}$$

Again, like the term I_{411}, it is not difficult to bound the above term when $1 \leq p < 2$. Thus we prove $I_{421} \leq C(\ln n/n)^{2s/(2s+1)}$ for all $p \geq 1$.

To complete the proof of the Lemma, we need to bound the last term I_{422}. Now we apply Cauchy-Schwarz inequality, we have its bound

$$\sum_{j=j_s+1}^{j_1} \sum_{k\in\Gamma_j} \Big[E\big(\hat{\beta}_{jk} - \beta_{jk}\big)^4\Big]^{1/2} \Big[P\big(|\hat{\beta}_{jk} - \beta_{jk}| > a_2\delta_{jk}\big)\Big]^{1/2}. \tag{5.37}$$

From Lemma 5.5, let $a_2 > 0$ be large and close to 1 enough, we derive $\big[P(|\hat{\beta}_{jk} - \beta_{jk}| > a_2\delta_{jk})\big]^{1/2} = O(n^{-1})$. For the expectation term, recall $\hat{\beta}_{jk} - \beta_{jk} = \hat{\beta}_{jk,1} + \hat{\beta}_{jk,2} + \hat{\beta}_{jk,3} - \beta_{jk}$. Apply Lemma 5.3 with $m = 2$, we obtain $E(\hat{\beta}_{jk,1}^4) \leq C2^{2j_1}n^{-4}$. Similarly, from Lemma 5.4 with $m = 2$, we obtain $E(\hat{\beta}_{jk,2}^4) \leq C2^{2j_1}n^{-4}$. Apply Rosenthal's inequality, we could have $E(\hat{\beta}_{jk,3} - \beta_{jk})^4 \leq Cn^{-2}$. Since $2^{j_1} < n$, from simple inequality, we have $E\big(\hat{\beta}_{jk} - \beta_{jk}\big)^4 \leq Cn^{-2}$. Putting this together we see that (5.37) is at most $C\sum_{j=j_s+1}^{j_1}\sum_{k=0}^{2^j-1} n^{-1}\, n^{-1} = C\sum_{j=j_s+1}^{j_1} 2^j\, n^{-2} \leq Cn^{-1}$. Therefore we obtain that $I_{422} \leq C(\ln n/n)^{2s/(2s+1)}$.

Lemma 5.11 *Under the assumptions of Theorem 3.1, we have*

$$I_5 := \sum_{j=j_0}^{j_1} \sum_{k\in\Gamma_j^c} E\Big[\hat{\beta}_{jk} I\big(|\hat{\beta}_{jk}| > \delta_{jk}\big) - \beta_{jk}\Big]^2 \leq C\Big(\frac{\ln n}{n}\Big)^{2s/(2s+1)}.$$

Proof: The proof is very similar to I_3 and I_4 with difference that $k \in \Gamma_j^c$. We write

$$\begin{aligned} I_5 &\leq 2\sum_{j=j_0}^{j_1} \sum_{k\in\Gamma_j^c} E\Big[\beta_{jk}^2 I\big(|\hat{\beta}_{jk}| \leq \delta_{jk}\big)\Big] + 2\sum_{j=j_0}^{j_1} \sum_{k\in\Gamma_j^c} E\Big[\big(\hat{\beta}_{jk} - \beta_{jk}\big)^2 I\big(|\hat{\beta}_{jk}| > \delta_{jk}\big)\Big] \\ &=: 2(I_{51} + I_{52}). \end{aligned} \tag{5.38}$$

From the proof to Lemmas 5.8, we have $E\{(\hat{\beta}_{jk} - \beta_{jk})^2\} \leq Cn^{-1}$. Since $\#\{\Gamma_j^c\} \leq C$ for all j's from (2.14), we have $I_{52} \leq Cj_1 n^{-1} \leq C\ln n/n$, which is within the desired rate. As to the first term I_{51}, we write

$$I_{51} \leq \sum_{j=j_0}^{j_1} \sum_{k\in\Gamma_j^c} \beta_{jk}^2 I\big(|\beta_{jk}| \leq 2\delta_{jk}\big) + \sum_{j=j_0}^{j_1} \sum_{k\in\Gamma_j^c} \beta_{jk}^2 P\big(|\hat{\beta}_{jk} - \beta_{jk}| > \delta_{jk}\big). \tag{5.39}$$

Now, the above first term is bounded by $Cj_1\delta_{jk}^2 = Cj_1K_1 \ln(\#\Lambda)\,\delta_0 n^{-1} \le C(\ln n)^2 n^{-1}$. The second term is bounded with Cn^{-2} from Lemma 5.5 and $\sum\sum_{jk}\beta_{jk}^2 < \infty$. Thus, we complete the proof to the Lemma.

Now we are ready to outline the proof to Theorem 3.1.

Proof to Theorem 3.1. In view of (5.1), we have

$$E\int_{-\pi}^{\pi}\big(\bar f(w) - f(w)\big)^2 dw = I_0 + I_1 + I_2 + I_3 + I_4 + I_5.$$

For details on I_0 to I_5, see expressions below (5.1). Thus, in order to prove Theorem 3.1, it suffices to show that $I_i \le C(\ln n/n)^{2s/(2s+1)}$, $i = 0, 1, \cdots, 5$, which were proved from the above Lemmas 5.6 to 5.10.

Theorem 3.1 is mainly of theoretical interest, because estimator $\bar f(\cdot)$ depends on unknown thresholds. It is introduced as a device to facilitate the proof to a data-driven estimator $\hat f(\cdot)$. In order to prove Theorem 3.2, we still need some preparation.

Lemma 5.12 *Let $\tilde f(\cdot)$ be a kernel estimator defined as in (3.1). Then, under Assumption* **A6***, we have, for any arbitrary small number $\epsilon > 0$,*

$$P\Big\{\big|\tilde f(w) - f(w)\big| > \epsilon\Big\} = O(n^{-\eta}), \tag{5.40}$$

for all $w \in (-\pi, \pi)$ and any positive number $\eta > 0$.

Proof: The proof is similar to that of Lemma 5.3. Write

$$\tilde f(w) = \frac{1}{h_n}\int_{-\pi}^{\pi} K\Big(\frac{w-\lambda}{h_n}\Big) I_{n,X}(\lambda)\,d\lambda =: (2\pi n)^{-1} Q_{n,X}^*, \tag{5.41}$$

where

$$Q_{n,X}^* = (2\pi n)\int_{-\pi}^{\pi}\eta_n^*(\lambda) I_{n,X}(\lambda) d\lambda, \quad \eta_n^*(\lambda) = \frac{1}{h_n}K\Big(\frac{w-\lambda}{h_n}\Big). \tag{5.42}$$

Further, we can write it in terms of a quadratic form just as that of $Q_{n,X}$ in (5.8), i.e., $Q_{n,X}^* = \sum_{k,t=1}^n d_n^*(k-t) X_k X_t$, where $d_n^*(t) = h_n^{-1}\int K\big(\frac{w-\lambda}{h_n}\big) e^{-\mathbf{i}t\lambda}\,d\lambda$. In view of (5.8), we also define $Q_{n,Z}^* := \sum_{k,t=1}^n e_n^*(k-t) Z_k Z_t = (2\pi n)\int_{-\pi}^{\pi}\eta_n^*(\lambda) 2\pi f(\lambda) I_{n,Z}(\lambda) d\lambda$, where $e_n^*(t) = h_n^{-1}\int K\big(\frac{w-\lambda}{h_n}\big) 2\pi f(\lambda) e^{-\mathbf{i}t\lambda}\,d\lambda$. Let

$$f^*(w) = (2\pi n)^{-1} E\big(Q_{n,Z}^*\big) = \int_{-\pi}^{\pi}\frac{1}{h_n}K\Big(\frac{w-\lambda}{h_n}\Big) f(\lambda)\,d\lambda. \tag{5.43}$$

Then, in view of (5.41) and (5.43), we have

$$\begin{aligned}\tilde f(w) - f(w) &= (2\pi n)^{-1}\big[Q_{n,X}^* - Q_{n,Z}^*\big] + (2\pi n)^{-1}\big[Q_{n,Z}^* - E(Q_{n,Z}^*)\big] + \big[f^*(w) - f(w)\big] \\ &=: P_{1n}(w) + P_{2n}(w) + P_{3n}(w).\end{aligned}$$

We apply Lemma 5.3 to term P_{1n} with $|\eta_n^*(\lambda)| \le Ch_n^{-1} = Cn^d$ from (5.42). Considering that $\eta_n^*(\lambda)$ plays the role of $\eta_n(\lambda)$ there, we have

$$E\big(P_{1n}^{2m}(w)\big) \le \frac{C_m h_n^{-2m}}{(2\pi n)^{2m}} = C(nh_n)^{-2m} = Cn^{-2m(1-d)} \le Cn^{-\eta},$$

for any $\eta > 0$ as long as $m > \eta/(1+\tau)$, where $0 < d < (1-\tau)/2$ from assumption **A6**. Then, apply Markov inequality, we obtain $P\{|P_{1n}(w)| > \gamma_1 \epsilon\} \le C n^{-\eta}$ for any positive $\gamma_1 > 0$.

For second term P_{2n}, for any positive number $\gamma_2 > 0$, we have $P\{|P_{2n}(w)| > \gamma_2 \epsilon_n\} = P\{|Q^*_{n,Z} - E(Q^*_{n,Z})| > 2\pi n \gamma_2 \epsilon\} \le P\{|Q^*_{n,Z} - E(Q^*_{n,Z})| > 2\pi n \gamma_2 \delta_{jk}\} \le C n^{-\eta}$. Above second inequality follows from that $2\pi n \gamma_2 \epsilon$ is at the order of n, which is much larger than $2\pi n \gamma_2 \delta_{jk}$ considered in Lemma 5.5. The last inequality follow from Lemma 5.5.

As to the last term P_{3n}, in view of (5.43), we have $|P_{3n}(w)| \le \int_{-\pi}^{\pi} |K(u)||f(w - uh_n) - f(w)| du$. From that kernel $K(\cdot)$ is continuously differentiable and compactly support, $h_n = C n^{-d}$, and f is uniformly continuous, we have $|P_{3n}(w)| \to 0$ uniformly for w as $n \to \infty$. Thus, this term is negligible compared to any fixed $\epsilon > 0$. Combine three terms together, we have

$$\begin{aligned} P\big\{\big|\tilde{f}(w) - f(w)\big| > \epsilon\big\} &\le P\big\{\big|P_{1n}(w) + P_{2n}(w)\big| > \epsilon/2\big\} \\ &\le P\big\{|P_{1n}(w)| > \gamma_1\epsilon/2\big\} + P\big\{|P_{2n}(w)| > \gamma_2\epsilon/2\big\} \le C n^{-\eta}, \end{aligned}$$

where γ_1, γ_2 are any fixed positive numbers such as $\gamma_1 + \gamma_2 = 1$.

Lemma 5.13 *Let $\hat{\sigma}^2_{jk}$ be the estimator defined as in (2.15). Then, under assumption* **A6***, we have, for any small positive $\epsilon > 0$,*

$$P\Big\{\big|\hat{\sigma}_{jk} - \sigma_{jk}\big| > \epsilon \sigma_{jk}\Big\} = O(n^{-2}), \quad \forall j \in [j_0,\, j_1] \;\; \textit{and} \;\; k \in \Gamma_j.$$

Proof: In view of (2.14), when $k \in \Gamma_j$, one has $n\sigma^2_{jk} \ge \delta_0$. From $\hat{\sigma}_{jk} \ge 0$, we have $P\{|\hat{\sigma}_{jk} - \sigma_{jk}| > \epsilon\sigma_{jk}\} \le P\{|n\hat{\sigma}^2_{jk} - n\sigma^2_{jk}| > \epsilon n \sigma^2_{jk}\}$. In view of (2.13), the second term within brace of RHS of (2.13) is at order $O((\ln n)^{-1})$, which is negligible compared to $n\sigma^2_{jk}$ (since $n\sigma^2_{jk} \ge \delta_0 > 0$). Thus we obtain that above probability $P\{|n\hat{\sigma}^2_{jk} - n\sigma^2_{jk}| > \epsilon n\sigma^2_{jk}\}$ is bounded above by

$$\begin{aligned} &P\Big\{\int_{-\pi}^{\pi} \pi\big[\Psi_{jk}(w) + \Psi_{jk}(-w)\big]^2 \big|\tilde{f}^2(w) - f^2(w)\big| dw + \frac{|\mu_4 - 3|\beta^2_{jk}}{(2\pi)^2} > \epsilon n \sigma^2_{jk}\Big\} \\ &\qquad\qquad \le P\Big\{\int_{-\pi}^{\pi} \pi\big[\Psi_{jk}(w) + \Psi_{jk}(-w)\big]^2 \big|\tilde{f}^2(w) - f^2(w)\big| dw > 2^{-1}\epsilon n \sigma^2_{jk}\Big\}. \end{aligned}$$

In view of (5.27), we have $n\sigma^2_{jk} \ge c_0 \pi \int_{-\pi}^{\pi} \big[\Psi_{jk}(w) + \Psi_{jk}(-w)\big]^2 f^2(w) dw \cdot \big(1 + o(1)\big) \ge c_0 \tau_1^2$ $\int_{-\pi}^{\pi} \pi\big[\Psi_{jk}(w) + \Psi_{jk}(-w)\big]^2 dw$ for some positive $c_0 > 0$. Thus, above probability is further bounded above by

$$P\Big\{\int_{-\pi}^{\pi} \pi\big[\Psi_{jk}(w) + \Psi_{jk}(-w)\big]^2 \big|\tilde{f}^2(w) - f^2(w)\big| dw > 2^{-1}\epsilon c_0 \tau_1^2 \int_{-\pi}^{\pi} \pi\big[\Psi_{jk}(w) + \Psi_{jk}(-w)\big]^2 dw\Big\}. \tag{5.44}$$

Use Riemann sum to approximate above integrals, above probability is bounded above by

$$P\Big\{\sum_l \pi\big[\Psi_{jk}(w_l) + \Psi_{jk}(-w_l)\big]^2 \big|\tilde{f}^2(w_l) - f^2(w_l)\big| > 4^{-1}\epsilon c_0 \tau_1^2 \sum_l \pi\big[\Psi_{jk}(w_l) + \Psi_{jk}(-w_l)\big]^2\Big\},$$

where $w_l = 2\pi l/n$ are Fourier frequencies. Now, above probability is bounded above by

$$P\Big\{\bigcup_l \Big\{|\tilde{f}^2(w_l) - f^2(w_l)| > 4^{-1}\epsilon c_0 \tau_1^2\Big\}\Big\} \le \sum_l P\Big\{|\tilde{f}^2(w_l) - f^2(w_l)| > 4^{-1}\epsilon c_0 \tau_1^2\Big\} \le C n^{-2},$$

where the last inequality follows from Lemma 5.12 with $4^{-1}\epsilon c_0 \tau_1^2$ is a fixed positive number.

Lemma 5.14 *Let $\hat\sigma^2_{jk}$ be an estimator defined as in (2.15). Then, under assumption* **A6**, *we have, for any small positive $\epsilon > 0$,*

$$P\Big\{\big|\hat\sigma_{jk} - \sigma_{jk}\big| > \epsilon\delta_0^{1/2} n^{-1/2}\Big\} \le C n^{-2}, \quad \forall j \in [j_0,\, j_1] \ \ \textit{and} \ \ k \in \Gamma_j^c.$$

Proof: The proof is similar to that to Lemma 5.13. The difference is that, when $k \in \Gamma_j^c$, we have $n\sigma^2_{jk} \le \delta_0$. For the brevity of exposition, we introduce some further notations. In view of (2.13), ignoring a negligible term, we have

$$n\sigma^2_{jk} = \pi \int_{-\pi}^{\pi} \big[\Psi_{jk}(w) + \Psi_{jk}(-w)\big]^2 f^2(w) dw + \frac{(\mu_4 - 3)\beta^2_{jk}}{(2\pi)^2} =: A_{jk} + R_{jk}. \tag{5.45}$$

Now above probability $P\{|\hat\sigma_{jk} - \sigma_{jk}| > \epsilon\delta_0^{1/2} n^{-1/2}\}$ is bounded above by $P\{|n^{1/2}\hat\sigma_{jk} - A_{jk}^{1/2}| + |A_{jk}^{1/2} - (A_{jk} + R_{jk})^{1/2}| > \epsilon\delta_0^{1/2}\}$. Next, we want to show that $|A_{jk}^{1/2} - (A_{jk} + R_{jk})^{1/2}| \to 0$ as $n \to \infty$. From previous discussion, we have $|R_{jk}| = O((\ln n)^{-1})$. Thus, if $A_{jk} \le C|R_{jk}|$, then it is easy to see $|A_{jk}^{1/2} - (A_{jk} + R_{jk})^{1/2}| \le C_1(\ln n)^{-1/2} \to 0$. If $A_{jk} \ge C|R_{jk}|$, we have $|A_{jk}^{1/2} - (A_{jk} + R_{jk})^{1/2}| \le |R_{jk}|(A_{jk}^{1/2} + (A_{jk} + R_{jk})^{1/2})^{-1} \le (|R_{jk}|/A_{jk})^{1/2}|R_{jk}|^{1/2} \le C^{-1}|R_{jk}|^{1/2} \le C_2(\ln n)^{-1/2} \to 0$. Thus, above probability $P\{|\hat\sigma_{jk} - \sigma_{jk}| > \epsilon\delta_0^{1/2} n^{-1/2}\}$ is bounded by $P\{|n^{1/2}\hat\sigma_{jk} - A_{jk}^{1/2}| > 2^{-1}\epsilon\delta_0^{1/2}\}$. It is further bounded by $P\{|n\hat\sigma^2_{jk} - A_{jk}| > 2^{-1}\epsilon\delta_0^{1/2} A_{jk}^{1/2}\}$. From (5.27), we have $n\sigma^2_{jk} \ge c_1 A_{jk}$ for some positive $c_1 > 0$. Together with $n\sigma^2_{jk} \le \delta_0$, we have $\delta_0 \ge c_1 A_{jk}$ and thus $\delta_0^{1/2} A_{jk}^{1/2} \ge c_1^{1/2} A_{jk}$. Thus, we have

$$\begin{aligned}
&P\Big\{\big|\hat\sigma_{jk} - \sigma_{jk}\big| > \epsilon\delta_0^{1/2} n^{-1/2}\Big\} \le P\Big\{\big|n\hat\sigma^2_{jk} - A_{jk}\big| > 2^{-1}\epsilon c_1^{1/2} A_{jk}\Big\} \\
&\quad \le P\Big\{\int_{-\pi}^{\pi} \pi\big[\Psi_{jk}(w) + \Psi_{jk}(-w)\big]^2 \big|\tilde f^2(w) - f^2(w)\big| dw > 2^{-1}\epsilon c_1^{1/2}\tau_1^2 \int_{-\pi}^{\pi} \pi\big[\Psi_{jk}(w) + \Psi_{jk}(-w)\big]^2 dw\Big\}.
\end{aligned}$$

Now, the rest of proof simply follows from (5.44) in previous Lemma 5.13.

Now, we are ready to provide a proof to Theorem 3.2.

Proof to Theorem 3.2. The proof is very similar to that for Theorem 3.1. The main difference between them is that we use data-driven thresholds $\hat\delta_{jk}$ in Theorem 3.2 instead of theoretical thresholds δ_{jk} in Theorem 3.1. Therefore, we only provide the difference here. Similar to the MISE for $\bar f$, we have following expression for $\hat f(\cdot)$:

$$E\int_{-\pi}^{\pi} \big(\hat f(w) - f(w)\big)^2 dw =: J_0 + J_1 + J_2 + J_3 + J_4 + J_5,$$

where J_0, J_1 and J_2 are same as those I_0, I_1 and I_2, because they do not involve with any random thresholds $\hat\delta_{jk}$. The other three terms are defined as

$$J_3 := \sum_{j=j_0}^{j_s} \sum_{k\in\Gamma_j} E\Big[\hat\beta_{jk} I\big(|\hat\beta_{jk}| > \hat\delta_{jk}\big) - \beta_{jk}\Big]^2,$$

$$J_4 := \sum_{j=j_s+1}^{j_1} \sum_{k\in\Gamma_j} E\Big[\hat\beta_{jk} I\big(|\hat\beta_{jk}| > \hat\delta_{jk}\big) - \beta_{jk}\Big]^2,$$

$$J_5 := \sum_{j=j_0}^{j_1} \sum_{k\in\Gamma_j^c} E\Big[\hat\beta_{jk} I\big(|\hat\beta_{jk}| > \hat\delta_{jk}\big) - \beta_{jk}\Big]^2.$$

As before, we need to show that $J_i \le C(\ln n/n)^{2s/(2s+1)}$ for $i = 3, 4$ and 5. For brevity of exposition, denote a set $I(V) = \{|\hat\sigma_{jk} - \sigma_{jk}| \le \epsilon\sigma_{jk}\}$, then we have $P(V) = 1 + O(n^{-2})$ or $P(V^c) = O(n^{-2})$ from Lemma 5.13 for $k \in \Gamma_j$. Note on set $I(V)$, we have $(1-\epsilon)^2\delta_{jk}^2 \le \hat\delta_{jk}^2 \le (1+\epsilon)^2\delta_{jk}^2$. So random thresholds are extremely close to theoretical thresholds because of arbitrary small ϵ. Now we write

$$\begin{aligned} J_3 &\le 2\sum_{j=j_0}^{j_s} \sum_{k\in\Gamma_j} E\Big[\beta_{jk}^2 I\big(|\hat\beta_{jk}| \le \hat\delta_{jk}\big)\Big] + 2\sum_{j=j_0}^{j_s} \sum_{k\in\Gamma_j} E\Big[\big(\hat\beta_{jk} - \beta_{jk}\big)^2 I\big(|\hat\beta_{jk}| > \hat\delta_{jk}\big)\Big] \\ &=: 2(J_{31} + J_{32}). \end{aligned} \tag{5.46}$$

Also,

$$J_{31} \le \sum_{j=j_0}^{j_s} \sum_{k\in\Gamma_j} E\Big[\beta_{jk}^2 I\big(|\hat\beta_{jk}| \le \hat\delta_{jk}\big) I(V)\Big] + \sum_{j=j_0}^{j_s} \sum_{k\in\Gamma_j} E\Big[\beta_{jk}^2 I\big(|\hat\beta_{jk}| \le \hat\delta_{jk}\big) I(V^c)\Big].$$

Now above first term basically could be treated as term I_{31}. The rate for second term in J_{31} is obtained by $P(V^c) = O(n^{-2})$ and $\sum\sum_{jk}\beta_{jk}^2 < \infty$. As to J_{32}, similarly we have

$$J_{32} \le \sum_{j=j_0}^{j_s} \sum_{k\in\Gamma_j} E\Big[\big(\hat\beta_{jk}-\beta_{jk}\big)^2 I\big(|\hat\beta_{jk}| > \hat\delta_{jk}\big) I(V)\Big] + \sum_{j=j_0}^{j_s} \sum_{k\in\Gamma_j} E\Big[\big(\hat\beta_{jk}-\beta_{jk}\big)^2 I\big(|\hat\beta_{jk}| > \hat\delta_{jk}\big) I(V^c)\Big].$$

Now above first term basically could be treated as term I_{32}. The rate for the second term in J_{32} is proved by applying Cauchy-Schwarz inequality, $P(V^c) = O(n^{-2})$ and the same argument to term I_{422}. The proof to term J_4 is almost the same as that to term J_3, which is skipped here. Now we apply Lemma 5.14 to prove term J_5. Denote a set $I(U) = \{|\hat\sigma_{jk} - \sigma_{jk}| \le \epsilon\delta_0^{1/2} n^{-1/2}\}$, then we have $P(U) = 1 + O(n^{-2})$ or $P(U^c) = O(n^{-2})$ from Lemma 5.14. We write

$$\begin{aligned} J_5 &\le 2\sum_{j=j_0}^{j_1} \sum_{k\in\Gamma_j^c} E\Big[\beta_{jk}^2 I\big(|\hat\beta_{jk}| \le \hat\delta_{jk}\big)\Big] + 2\sum_{j=j_0}^{j_1} \sum_{k\in\Gamma_j^c} E\Big[\big(\hat\beta_{jk} - \beta_{jk}\big)^2 I\big(|\hat\beta_{jk}| > \hat\delta_{jk}\big)\Big] \\ &=: 2(J_{51} + J_{52}). \end{aligned} \tag{5.47}$$

Apply the same argument to I_{52}, we have $J_{52} \le Cj_1 n^{-1} \le C\ln n/n$. As to the first term J_{51}, we write

$$J_{51} \le \sum_{j=j_0}^{j_1} \sum_{k\in\Gamma_j^c} \beta_{jk}^2 E\Big[I\big(|\hat\beta_{jk}| \le \hat\delta_{jk}\big) I(U)\Big] + \sum_{j=j_0}^{j_1} \sum_{k\in\Gamma_j^c} \beta_{jk}^2 E\Big[I\big(|\hat\beta_{jk}| \le \hat\delta_{jk}\big) I(U^c)\Big]. \tag{5.48}$$

Now, above first term can be treated as term I_{51} using $n\sigma_{jk}^2 \le \delta_0$, which is bounded by $Cj_1\delta_{jk}^2 = Cj_1K_1\ln(\#\Lambda)\,\delta_0 n^{-1} \le C(\ln n)^2 n^{-1}$. The second term is bounded with Cn^{-2} from Lemma 5.14 and $\sum\sum_{jk}\beta_{jk}^2 < \infty$. Thus, we complete the proof to Theorem 3.2.

REFERENCES

Bhansali, R. J., Giraitis, L. and Kokoszka, P. S. (2007a). Convergence of quadratic forms with nonvanishing diagonal. *Statist. Probab. Lett.* **77**, 726–734.

Bhansali, R. J., Giraitis, L. and Kokoszka, P. S. (2007b). Approximations and limit theory for quadratic forms of linear processes. *Stochastic Process. Appl.* **117**, 71–95.

Brillinger, D. R. (1981). *Time series. Data analysis and theory.* Second edition. Holden-Day Series in Time Series Analysis. Holden-Day, Inc., Oakland, Calif..

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods.* Second edition. Springer Series in Statistics. Springer-Verlag, New York.

Brown, L. D., Cai, T. T. and Zhou, H. H. (2008). Robust nonparametric estimation via wavelet median regression. *Ann. Statist.* **36**, 2055–2084.

Cai, T. T. (1999). Adaptive wavelet estimation: a block thresholding and oracle inequality approach. *Ann. Statist.* **27**, 898–924.

Daubechies, I. (1992). *Ten Lectures on Wavelets.* SIAM, Philadelphia. lengths. *J. Time Ser. Anal.* **36**, 189–208.

Donoho, D. L. and Johnstone, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **81**, 425–455.

Donoho, D. L. and Johnstone, I. M. (1995). Adapting to unknown smoothness via wavelet shrinking. *J. Amer. Statist. Assoc.* **90**, 1200–1224.

Donoho, D. L. and Johnstone, I. M. (1998). Minimax estimation via wavelet shrinkage. *Ann. Statist.* **26**, 879–921.

Fryzlewicz, P., Nason, G. P. and von Sachs, R. (2008). A wavelet-Fisz approach to spectrum estimation. *J. Time Ser. Anal.* **29**, 868–880.

Gao, H. (1997). Choice of thresholds for wavelet shrinkage estimate of the spectrum. *J. Time Ser. Anal.* **18**, 231–251.

Giraitis, L., Koul, H. L. and Surgailis, D. (2012). *Large sample inference for long memory processes.* Imperial College Press, London.

Hall, P. and Patil, P. (1995). Formulae for mean integrated squared error of nonlinear wavelet-based density estimators. *Ann. Statist.* **23**, 905–928.

Hannan, E. J. (1970). *Multiple time series.* John Wiley and Sons, Inc., New York-London-Sydney.

Hernández, E. and Weiss, G. (1996). *A first course on wavelets.* CRC Press, Boca Raton, FL.

Lee, J. and Hong, Y. (2001). Testing for serial correlation of unknown form using wavelet methods. *Econometric Theory* **17**, 386–423.

Li, L. and Lu, K. (2018). Tests for the equality of two processes' spectral densities with unequal lengths using wavelet methods. *Journal of Time Series Analysis*, 39, 4–27.

Mallat, S. (1998). *A wavelet tour of signal processing.* Academic Press, Inc., San Diego, CA.

Meyer, Y. (1992). *Wavelets and Operators*, Cambridge University Press, Cambridge.

Moulin, P. (1994). Wavelet thresholding techniques for power spectrum estimation. *IEEE Transactions on Signal Process* **42**, 3126–3136.

Neumann, M. H. (1996). Spectral density estimation via nonlinear wavelet methods for stationary non-Gaussian time series. *J. Time Ser. Anal.* **17**, 601–633.

Pensky, M., Vidakovic, B. and De Canditiis, D. (2007). Bayesian decision theoretic scale-adaptive estimation of a log-spectral density. *Statist. Sinica* **17**, 635–666.

Priestley, M. B. (1981). *Spectral analysis and time series.* Academic Press, London.

Rudelson, M. and Vershynin, R. (2013). Hanson-Wright inequality and sub-Gaussian concentration. *Electron. Commun. Probab.*, **18**, no. 82, 1–9.

Shumway, R. H. and Stoffer, D. S. (2011). *Time series analysis and its applications. With R examples.* Third edition. Springer Texts in Statistics. Springer, New York.

Triebel, H. (1992). *Theory of Function Spaces II.* Birkhäuser, Basel.

Vidakovic, B. (1999). *Statistical Modeling by Wavelets.* John Wiley & Sons, Inc., New York.

von Sachs, R. and Macgibbon, B. (2000). Non-parametric curve estimation by wavelet thresholding with locally stationary errors. *Scandinavian J. Statist.* **27**, 475–499.

Wright, F. T. (1973). A bound on tail probabilities for quadratic forms in independent random variables whose distributions are not necessarily symmetric. *Ann. Probability*, **1**, 1068–1070.