BAYESIAN DECISION THEORETIC SCALE-ADAPTIVE ESTIMATION OF A LOG-SPECTRAL DENSITY

MARIANNA PENSKY AND BRANI VIDAKOVIC

Abstract: The problem of estimating the log-spectrum of a stationary Gaussian time series by Bayesianly induced shrinkage of empirical wavelet coefficients is studied. A model in the wavelet domain that accounts for distributional properties of the log-periodogram at levels of fine detail and approximate normality at coarse levels in the wavelet decomposition, is proposed. The smoothing procedure, called BAMS-LP (Bayesian Adaptive Multiscale Shrinker of Log-Periodogram), ensures that the reconstructed log-spectrum is as noise-free as possible. It is also shown that the resulting Bayes estimators are asymptotically optimal (in the frequentist sense).

Comparisons with non-wavelet and wavelet-non-Bayesian methods are discussed.

Key words and phrases: Spectral Density, Log-Spectral Density, Wavelets.

1 Introduction

Any statistical inference in time series can be conducted in time and frequency domains. The methods are complementary and provide different insights. Spectral analysis, and in particular, estimation of spectral density is an indispensable tool for exploring the frequency behavior of a time series.

Wavelet shrinkage methods have successfully been applied to the spectral density estimation in work of Lumeau *et al.* (1993), Moulin (1992, 1994), Gao (1992, 1993a,b) from the classical view-point. In this paper we propose a novel wavelet-shrinkage method, based on intrinsic shrinkage property of Bayes rules. The proposed shrinkage rules resulting from hierarchical Bayes statistical models are both realistic, i.e., describe data accurately, and capable of incorporating the available prior information on smoothness of functions represented by their wavelet coefficients.

Let $\{X_t, t \in Z\}$ be a a real, weakly stationary time series with zero mean and autocovariance function $\gamma(h) = EX(t+h)X(t)$. An absolutely summable complex-valued function $\gamma(\cdot)$ defined on integers is the autocovariance function of X_t if and only if the function

$$f(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) e^{-ih\omega} \tag{1.1}$$

is non-negative for all $\omega \in [-\pi, \pi]$. The function $f(\omega)$ is called the spectral density associated with covariance function $\gamma(h)$. Thus, the spectral density of a stationary process is a symmetric and non-negative function. Given the spectral density, the autocovariance function can uniquely be recovered via inverse Fourier transformation,

$$\gamma(h) = \int_{-\pi}^{\pi} f(\omega) e^{ih\omega} d\omega,\ h = 0, \pm 1, \pm 2, \ldots.$$

In particular, the variance of X_t is equal to $\gamma(0) = \int_{-\pi}^{\pi} f(\omega) d\omega$.

An important class of stationary processes consists of autoregressive-moving average ARMA(p, q) processes defined via

$$\phi(B) X_t = \theta(B) Z_t,\ \ \{Z_t\} \sim \text{WN}(0, \sigma^2), \tag{1.2}$$

where B is the backshift operator, WN$(0, \sigma^2)$ is the white noise with variance σ^2, the polynomials $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$ and $\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$ have no common zeroes, and $\phi(z)$ has no zeroes on the unit circle. The spectral density of X_t in (1.2) is a rational trigonometric function,

$$f_X(\omega) = \frac{\sigma^2}{2\pi} \frac{|\theta(e^{-i\omega})|^2}{|\phi(e^{-i\omega})|^2},\quad -\pi \le \omega \le \pi. \tag{1.3}$$

Estimation of spectral density from the observed data is an important statistical task in a variety of applied fields in which the information about frequency behavior of the phenomena is essential. Spectral density cam be estimated in the time domain by fitting the polynomials $\phi(z)$ and $\theta(z)$ in the representation (1.3), or directly in the frequency domain. It turns out that latter approach is generally superior.

A traditional statistic used as an estimator of the spectral density is the *periodogram.* The periodogram $I(\omega)$, based on a sample $X_0, \ldots, X_{T-1}$ is defined as

$$I(\omega_j) = \frac{1}{2\pi T} \left| \sum_{t=0}^{T-1} X_t e^{-it\omega_j} \right|^2, \tag{1.4}$$

where ω_j is the Fourier frequency $\omega_j = \frac{2\pi j}{T},\ j = [-T/2] + 1, \ldots, -1, 0, 1, \ldots, [T/2]$. By a discrete version of the sampling theorem it holds that $I(\omega)$ is uniquely determined for all $\omega \in [-\pi, \pi]$, given its values at Fourier frequencies. Because of symmetry of $I(\omega)$, we will focus only on non-negative Fourier frequencies, $\omega_j = \frac{2\pi j}{T},\ j = 0, 1, \ldots, [T/2]$.

For any set of Fourier frequencies $\omega_1, \omega_2, \ldots, \omega_n$ such that $0 \le \omega_1 < \cdots < \omega_n < \pi$, $I(\omega_i)$ are asymptotically independent exponential random variables with means $f(\omega_i)$ and variances $(f(\omega_i))^2$, where f is the spectral density. Consequently the periodogram is not a consistent estimator of $f(\omega)$, and citing Wahba (1980), "it will be hopelessly wiggly even when $f(\omega)$ is a smooth function" and $T \to \infty$.

Smoothing the periodogram will not only help in visually extracting significant frequencies, but smoothed periodograms can also be consistent estimators of f. For a standard theory see Brockwell and Davis (1991). There are three approaches in achieving the consistency in the spectral density estimators: (i) smoothing the periodogram directly via a weighted local average, (ii) smoothing the log-periodogram via traditional regression techniques, and (iii) maximizing Whittle's likelihood (or penalized likelihood) of the periodogram (Chow and Grenander, 1985; Pawitan and O'Sullivan,1994). The literature on the smoothing techniques in spectral density estimation is quite rich, see for example Brillinger (1981), Koopmans (1995), Percival and Walden (1993), Priestley (1981), Shumway and Stoffer (2000), and the numerous references therein.

In this paper we focus on the smoothing of log-periodogram. Early reference on utilizing splines and Fourier decomposition of log-periodogram are Cogburn and Davis (1974) and Wahba (1980). Fan and Kreutzberger (1998) investigate likelihood-based spectral density and log-spectral density operators.

The idea of using wavelets in smoothing log-periodograms was announced in Donoho (1993) and fully developed by Hong-Ye Gao in his Berkeley PhD Thesis (Gao, 1993a) and papers (Gao, 1993b; 1993c). Moulin (1994) applies saddle point estimation to tail probabilities of distributions of wavelet coefficients to exhibit thresholds for a log-periodogram.

Bayesian approaches to spectral time series analysis include Choudhuri, Ghosal, and Roy (2003), Gangopadhyay, Mallick and Denison (1998), and Huerta and West (1999), among others.

2 Bayesian Model

It is now standard practice in wavelet shrinkage to specify a location model on the wavelet coefficients, elicit the prior on their locations (the signal part in wavelet coefficients) and other unknown parameters, exhibit the Bayes estimator for the locations and, if the resulting Bayes rules are shrinkage estimators, apply the inverse wavelet transformation to the estimators. This is the core of Bayesian wavelet shrinkage.

It is certainly desirable for selected models to well-describe our empirical observations and perform well in terms of mean square error, for the majority of signals and images. At the same time, usually high dimensions of wavelet descriptions, the calculation complexity of shrinkage rules should remain low. Our experience (Vidakovic and Ruggeri, 2001) is that advanced but complicated models, for which the rules are obtained by, say, extensive MCMC simulations, or genetic algorithms, etc., are seldom accepted by practitioners, despite their reportedly good performance.

We believe that two desirable goals, *simplicity and reality* of a model, can be achieved simultaneously by statistical modeling in the wavelet domain.

As a consequence of decorrelating property of wavelet transformations, simple "independence" models that model each coefficient separately are justified. We adopt a paradigmatic location model in which the empirical wavelet coefficients of (shifted) log-periodogram, d, are modeled via a density (likelihood) $\zeta(d-\theta)$ where θ is the wavelet counterpart of the log-spectrum. The same model can be used with slight scale modifications implied by the prior on θ, for all detail coefficients.

We discuss the model building in stages: the likelihood, the prior, the calculation of the Bayes rule and selection of the hyperparameters. We call the resulting shrinkage algorithm BAMS-LP (short for Bayesian Adaptive Multiscale Shrinker of Log-Periodogram).

2.1 Likelihood

Under mild conditions (Brillinger, 1981; Theorem 5.2.6) it holds

$$I(\omega_\ell) \overset{iid}{\approx} \frac{1}{2} f(\omega_\ell)\chi_2^2, \tag{2.1}$$

where $\overset{iid}{\approx}$ means "approximately iid", for the "inner" non-zero Fourier frequencies ω_ℓ. For $\omega = 0$ and extreme Fourier frequencies when the sample size T is even, the right-hand side of (2.1) becomes $f(\omega)\chi_1^2$. We will ignore this difference since its effect is negligible for T large. We also assume that iid in (2.1) is exact, which is true for only circulant time series (Harvey, 1989). By taking the logarithm in (2.1) we obtain a regression formulation (called Wahba's formulation)

$$z_\ell = \ln f(\omega_\ell) + \varepsilon_\ell, \tag{2.2}$$

where $z_\ell = \ln I(\omega_\ell) + \gamma$ and γ is the Euler-gamma constant ($\gamma = 0.577126$).
The exact distribution of the log-periodogram can be found in Wittwer (1986).

The following lemma describes the distribution of the error term ϵ, in (2.2).

Lemma 2.1. *The random variables* ε_l, $l = 0, \cdots, T-1$, *are approximately independent, identically distributed with the density*

$$\mu(x) = \gamma^* \exp(x - \gamma^* e^x), \tag{2.3}$$

where $\gamma^* = e^{-\gamma} = 0.546146$*. Also,* $E\varepsilon_l = 0$ *and Var* $\varepsilon_l = \sigma^2 = \pi^2/6$*,* $l = 0, \cdots, T-1$*. Skewness of* ϵ *is* $\gamma_1 = -2\zeta(3)/(\frac{\pi^2}{6})^{3/2} \approx -1.14$, where ζ is Reimann's zeta function.

Proof. Easy, as $\varepsilon_l - \gamma \overset{d}{=} \ln\left(\frac{1}{2}\chi_2^2\right)$.

In the wavelet domain (2.2) becomes

$$\underline{d}^* = \underline{\theta}^* + \underline{\delta}^*, \tag{2.4}$$

where

$$\begin{aligned}
\underline{d}^* &= W\underline{z}; \;\; \underline{z} = (z_0, z_1, \cdots, z_{T-1}); \\
\underline{\theta}^* &= W\underline{y}; \;\; \underline{y} = (\ln f(\omega_0), \ln f(\omega_1), \cdots, \ln f(\omega_{T-1})); \\
\underline{\delta}^* &= W\underline{\varepsilon}; \;\; \underline{\varepsilon} = (\varepsilon_0, \varepsilon_1, \cdots, \varepsilon_{T-1});
\end{aligned}$$

Level j	λ_j	Level j	λ_j
$J-1$	0.355	$J-5$	0.060
$J-2$	0.179	$J-6$	0.045
$J-3$	0.127	$J-7$	0.025
$J-4$	0.092	$\leq J-8$	≈ 0

Table 1: The weights λ_j in the likelihood approximation (2.5)

.

and W is an orthogonal matrix of the discrete wavelet transform.

Let J be such that $2^J = T$. Then vector $\underline{\delta}^*$ can be represented as $\underline{\delta} = (\delta_0^*, \underline{\delta}_0^*, \underline{\delta}_1^*, \cdots, \underline{\delta}_{J-1}^*)$ where $\underline{\delta}_j^* = (\delta_{j,0}^*, \delta_{j,1}^*, \cdots, \delta_{j,2^j-1}^*)$ is the j-th sub-vector associated with the multiresolution analysis. Here, $j = 0$ denotes the smooth part corresponding to the scaling function while $j = J-1$ is the finest resolution level. Similarly, $\underline{d}^* = (d_0^*, \underline{d}_0^*, \underline{d}_1^*, \cdots, \underline{d}_{J-1}^*)$ and $\underline{\theta}^* = (\theta_0^*, \underline{\theta}_0^*, \underline{\theta}_1^*, \cdots, \underline{\theta}_{J-1}^*)$.
Exact distribution for vector $\underline{\delta}^*$ can be found since the transformation matrix W can be written in an explicit form. By the central limit theorem (for conditions on wavelet bases for CLT to hold, see Moullin, 1994) it follows that the density function of a component $\delta_{j,k}^*$ can be well approximated by a mixture

$$\zeta_j(x) \quad = \quad (1-\lambda_j)\eta(x) + \lambda_j \mu(x), \tag{2.5}$$

where $\mu(x)$ is defined in (2.3), $\eta(x)$ is the normal pdf

$$\eta(x) = (\sigma\sqrt{2\pi})^{-1} \exp\left(-x^2/2\sigma^2\right). \tag{2.6}$$

and $\sigma^2 = \pi^2/6$. Here, λ_j's are non-zero at fine resolution levels, and equal to zero at coarse resolution levels, namely, $\lambda_j = 0$ if $j \leq J_0$. In theory, we need $J - J_0 \to \infty$, however, in practice, the central limit theorem can be applied for all except a few finest resolution levels.

Figure 2.1 shows three densities and the histogram. The log-chisquare $\mu(x)$ and normal $\eta(x)$ densities are depicted in dotted and dashed lines and their mixture $\zeta(x)$ (solid line) is obtained from (2.5) with weight $\lambda = 0.355$, see Table 2.1. The histogram shows the empirical distribution of wavelet coefficients at the first level of detail. Wavelet is Coiflet 3 (18 tap filter), and the histogram is based on 2^{14} observations ($T = 2^{15}$). Note quite satisfactory approximation of the histogram by the mixture.

The Table 2.1 provides weights λ_j for the highest resolution levels. The table is obtained by matching skewness of the likelihood mixture (2.5) and the empirical distributions of $\underline{\delta}_j^*$. Wavelet used was Coiflet 3, but the weights are quite robust for other standard bases such as Symmlets and Daubechies'.

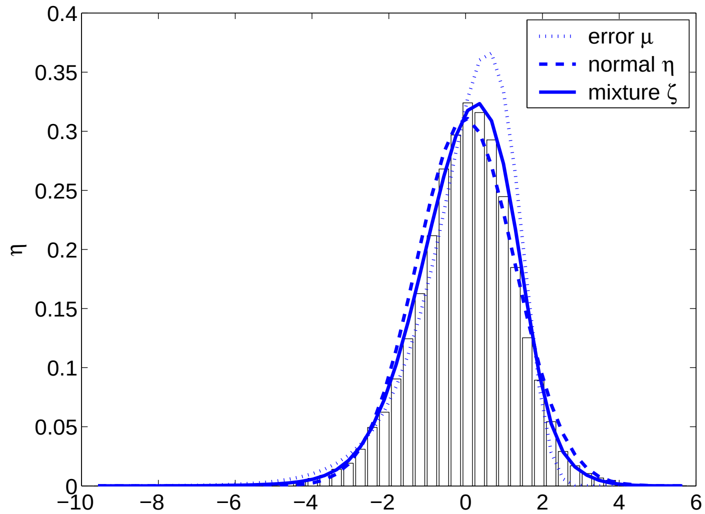


Figure 1: The log-chisquare $\mu(x)$ and normal $\eta(x)$ densities (dotted and dashed lines) and their mixture $\zeta(x)$ (solid line) obtained with weight $\lambda = 0.355$. The bar plot is the empirical distribution of wavelet coefficients at the first level of detail.

2.2 Prior Selection

Since wavelet representations of regular and piecewise-regular functions contain only a few non-negligible coefficients in their expansions, we place the standardly used Berger-Müller prior on the discrete wavelet coefficient θ^*_{jk}:

$$\theta^*_{jk} \sim \pi_j \, \delta(0) + (1-\pi_j) \, \tau_j \xi(\tau_j x), \tag{2.7}$$

where $0 \le \pi_j \le 1$, $\delta(0)$ is a point mass at zero, and the "spread" density $\xi(x)$ is symmetric and unimodal. We also assume that wavelet coefficients θ^*_{jk} are apriori independent. The factor π_j is the prior probability that a coefficient θ^*_{jk} at level j is zero. In what follows, however, we shall impose all conditions on the prior odds ratio:

$$\beta_j = \frac{\pi_j}{1-\pi_j}. \tag{2.8}$$

2.3 Bayes Rule and BAMS-LP Estimator

Our objective is to estimate the location parameter in our model, i.e., the log-spectral density $g(\omega) = \ln f(\omega)$. Denote the wavelet coefficients of g by θ_{jk}, so that g can be reconstructed as

$$g(x) = \sqrt{\pi}\theta_0 \varphi(\pi x) + \sqrt{\pi} \sum_{j=0}^{\infty} \sum_{k=-2^{j-1}}^{2^{j-1}-1} \theta_{jk}\psi_{jk}(\pi x) \tag{2.9}$$

with $\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k)$, $\theta_0 = \sqrt{\pi}\int_{-\infty}^{\infty}\varphi(\pi x)g(x)dx$ and $\theta_{jk} = \sqrt{\pi}\int_{-\infty}^{\infty}\psi_{jk}(\pi x)g(x)dx$. Here $\varphi(x)$ is the scaling function and $\psi(x)$ is the corresponding wavelet function. Recall that θ^*_{jk}

and θ_{jk} are related as $\theta^*_{jk} \approx \sqrt{T}\theta_{jk}$ (see e.g. Vidakovic, 1999)). This rescaling is a consequence of changing the domain of the transformed function, namely, θ^* approximates θ only when the sampling interval is equal to 1. The relation $\approx$ in $\theta^*_{jk} \approx \sqrt{T}\theta_{jk}$ could be replaced by equality only when the wavelet basis is interpolating. The wavelet bases we used in our simulations, symmlets and coiflets, are close to interpolating.

Denote

$$d_{jk} = d^*_{jk}/\sqrt{T}, \;\; \nu_j = \sqrt{T}\tau_j. \tag{2.10}$$

Taking into account the relation between θ^*_{jk} and θ_{jk} and (2.5) – (2.10), we derive that the posterior pdf of θ_{jk} given d_{jk} is of the form

$$P(\theta_{jk}|d_{jk}) = \frac{\sqrt{T}\,\zeta_j(\sqrt{T}(d_{jk}-\theta_{jk}))\;\nu_j\xi(\nu_j\theta_{jk})}{\int_{-\infty}^{\infty}\sqrt{T}\,\zeta_j(\sqrt{T}(d_{jk}-x))\nu_j\xi(\nu_j x)dx + \beta_j\sqrt{T}\,\zeta_j(\sqrt{T}d_{jk})},$$

where $\zeta_j(x)$ is defined in (2.5). Choosing the posterior mean as an estimator we arrive at the following estimator of θ_{jk}:

$$\hat{\theta}_{jk} = \frac{(1-\lambda_j)I_1(d_{jk}) + \lambda_j I^*_1(d_{jk})}{(1-\lambda_j)I_0(d_{jk}) + \lambda_j I^*_0(d_{jk}) + \beta_j\sqrt{T}\,\zeta_j(\sqrt{T}d_{jk})}, \;\; 0 \le j \le J-1, \tag{2.11}$$

and $\hat{\theta}_{jk} = 0$ as $j \ge J$. Here

$$I_i(d) \;\; = \;\; \int_{-\infty}^{\infty} x^i\sqrt{T}\,\eta(\sqrt{T}(d-x))\;\nu_j\xi(\nu_j x)dx, \;\; i = 0, 1, \tag{2.12}$$

$$I^*_i(d) \;\; = \;\; \int_{-\infty}^{\infty} x^i\sqrt{T}\,\mu(\sqrt{T}(d-x))\;\nu_j\xi(\nu_j x)dx, \;\; i = 0, 1, \tag{2.13}$$

with $\eta(x)$ and $\mu(x)$ given by (2.3) and (2.6), respectively. Shrinkage rule in (2.11) is shown in Figure 2.3, for some exemplary selection of hyper-parameters. The vertical dotted lines are plotted to emphasize the asymmetry of the rule

Hence, the BAMS-LP estimator of g is of the form

$$\hat{g}(x) = \sqrt{\pi}\hat{\theta}_0\varphi(\pi x) + \sqrt{\pi}\sum_{j=0}^{J-1}\sum_{k=-2^{j-1}}^{2^{j-1}-1}\hat{\theta}_{jk}\psi_{jk}(\pi x), \tag{2.14}$$

where $\hat{\theta}_0 = d^*_0/\sqrt{T}$.

In spite of having seemingly complex form, the estimators $\hat{\theta_{jk}}$ are easy to compute in a number of cases. For example, if the prior pdf $\xi(\cdot)$ is double exponential, the integrals $I_i(d), i = 0, 1,$ and $I^*_i(d), i = 0, 1,$ can be expressed in terms of normal cdf and incomplete gamma functions, respectively. In the case when $\xi(\cdot)$ is a normal pdf, the values of $I_i(d), i = 0, 1,$ are well known (see e.g. Abramovich *et al.* , 1998). It is somewhat harder to find expressions for functions $I^*_i(d), i = 0, 1,$ however, their Fourier transforms can be written explicitly in terms of gamma functions of complex argument. To summarize, in a number of cases, one can calculate the values $\hat{\theta^2_{jk}}$ efficiently without residing to numerical integration. Our simulations in Section 4 have been done using double exponential density $\xi(\cdot)$.

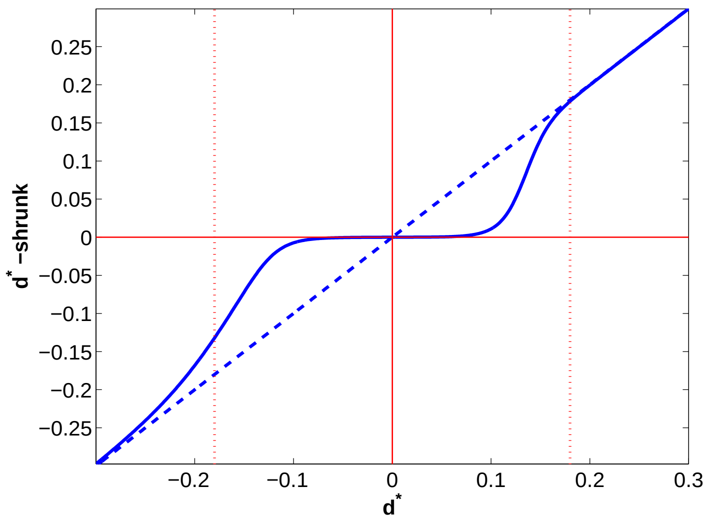


Figure 2: Shrinkage rule in (2.11), BAMS-LP rule, for a selection of hyper-parameters. The vertical dotted lines are plotted to emphasize the asymmetry of the rule.

3 Minimax Convergence Rates for BAMS-LP estimators

It is well known that no function estimation performs well if the function to be estimated belongs to an unrestricted functional space. Standard restrictions require the target function to belong to one of the range of smoothness spaces for which the wavelets are unconditional bases.

In order to assess the error of BAMS-LP estimator $\hat{g}$ we assume that g belongs to a ball $H^r(A)$ in the Sobolev space $H^r, r > 1/2$. Wavelets are unconditional bases for H^r, that is

$$g \in H^r(A) \iff \theta_0^2 + \sum_{j=0}^{\infty} \sum_{k=-2^{j-1}}^{2^{j-1}-1} \theta_{jk}^2 (1 + 2^{2jr}) \leq A, \quad r > 1/2. \tag{3.15}$$

For any possible estimator $\tilde{g}$ of g based on T observations we define the mean integrated square error (MISE) over the set $\mathcal{F}$ as

$$R(T, \mathcal{F}) = \sup_{g \in \mathcal{F}} E \|\tilde{g} - g\|^2_{L^2[-\pi,\pi]}. \tag{3.16}$$

We establish convergence rates for $R(T, H^r(A))$ as $T \to \infty$ and show that $R(T, H^r(A))$ could deviate from the optimal rate $O\left(T^{-\frac{2r}{2r+1}}\right)$ by just a logarithmic factor.

Although, to the best of our knowledge, no lower bounds for $R(T, H^r(A))$ are available in the case of estimation of log-spectral density, the rate of $O\left(T^{-\frac{2r}{2r+1}}\right)$ represents a landmark. Donoho and Johnstone (1998) showed that when the errors $\delta^*_{j,k}$ are independent and normally distributed, $T = n$ and g belongs to a ball $B^s_{p,q}(A)$ in the Besov space $B^s_{p,q}[0, 1]$, then

$$\inf_{\tilde{g}} R(n, B^s_{p,q}(A)) = O\left(n^{-\frac{2s}{2s+1}}\right) \quad (n \to \infty) \tag{3.17}$$

provided $s > \max(0, 1/p - 1/2)$ and $p, q \geq 1$. Since $H^r = B^r_{2,2}$, (3.17) implies that

$$\inf_{\tilde{g}} R(T, H^r(A)) = O\left(T^{-\frac{2r}{2r+1}}\right) \quad (T \to \infty). \tag{3.18}$$

Since, by (2.5), for majority of resolution levels ($j \leq J_0$), the errors $\delta^*_{j,k}$ are close to normal, we can expect to achieve convergence rate close to (3.18) as $T \to \infty$ for some choices of $\xi(\cdot)$ in (2.7).

3.1 Asymptotic Results

Let the multiresolution analysis generating the scaling function $\varphi(x)$ and a corresponding wavelet function $\psi(x)$ be s-regular with $s \geq r$. Assume that the spread density component $\xi(x)$ in the prior (2.7) is three times differentiable at least in a piecewise sense, has a finite fourth moment and satisfies the conditions

$$\begin{aligned}
|\, \xi^{(k)}(x)/\xi(x)| &\leq C_{\xi,1}(1 + |x|^\lambda)^k,\ k = 1, 2, 3,\ \lambda \geq 0, & (3.19)\\
\left|\ \exp\left(-Tx^2/2\sigma^2\right)/\xi(\nu_j x)\right| &\leq C_{\xi,2}\ \text{ if }\ \nu_j/\sqrt{T} \to 0. & (3.20)
\end{aligned}$$

Let also the integrals $I_i(d)$ defined in (2.12) be such that

$$\begin{aligned}
|I_1(d)/I_0(d) - d| &= O\left(|d|\nu_j^2/n\right)\ \text{ if } \nu_j/\sqrt{T} \to 0,\ \nu_j|d| \to \infty, & (3.21)\\
I_0(d) &\sim \nu_j \xi(\nu_j\theta),\ \text{ if } \nu_j/\sqrt{T} \to 0,\ \nu_j|d| \to \infty. & (3.22)\\
|I_1(d)/I_0(d)| &= O\left(|d|T/\nu_j^2\right)\ \text{ if }\ \nu_j/\sqrt{T} \to \infty. & (3.23)
\end{aligned}$$

We denote

$$j_0 = (2r+1)^{-1}\log_2 T \tag{3.24}$$

and assume that parameters ν_j, β_j and J_0 are such that

$$\begin{aligned}
\nu_j^2 &= C_1 2^{(2r+1)j}, & (3.25)\\
2^{J_0} &\geq T^{1/2r} & (3.26)\\
\beta_j^2 &= O\left(2^{(4r+1)j}T^{-(4r+1)/(2r+1)}\right)\ \text{ if }\ j \leq j_0. & (3.27)
\end{aligned}$$

Remark 1. Assumptions about ν_j can be translated into the ones on τ_j using relation (2.10), namely, $\tau_j^2 = C_1^* 2^{(2r+1)j}T^{-1}$.

Remark 2. Condition of the existence of the fourth moment is purely technical and is used for derivation of asymptotic expansions of the integrals $I_i(d)$, $i = 0, 1$. This condition can be dropped and replaced by assumptions (A.1)–(A.4) and (A.6)– (A.7) about integrals $I_j(d)$ and $I_j^*(d)$, $j = 0, 1$.

Remark 3. Condition (3.26) is quite realistic and agree with the central limit theorem. Note that we have an infinite number $J - J_0 = (2r)^{-1}(2r-1)\log_2 T$ resolution levels till the central limit theorem takes place. In practice the normality assumption can be verified via level-by-level testing.

Theorem 1. *Let assumptions (3.19)–(3.22) and (3.25) – (3.27) be valid. Then*

$$R(T, H^r(A)) = O\left(T^{-\frac{2r}{2r+1}}\,[\ln T]^{\frac{1}{2r+1}}\right),\ T \to \infty. \tag{3.28}$$

If, moreover, condition (3.23) holds, then

$$R(T, H^r(A)) = O\left(T^{-\frac{2r}{2r+1}}\right),\ T \to \infty. \tag{3.29}$$

Corollary 1. *Let assumptions (3.25) – (3.27) be valid. If $\xi(x)$ is a normal pdf, then $R(T, H^r(A))$ is of the form (3.28).*

Corollary 2. *Let assumptions (3.25) – (3.27) be valid. If $\xi(x)$ is a double-exponential pdf or a pdf of the t-distribution, then $R(T, H^r(A))$ is of the form (3.29).*

Remark 4. Convergence rates in Theorem 1 differ from optimal rate (3.18) by a logarithmic factor. It is not clear to the authors whether this is due to the nature of the problem or to the deficiency of the proof.

4 Simulations and Comparisons

Implementation of the proposed Bayesian wavelet shrinkage can be described algorithmically. Here is description of BAMS-LP algorithm.

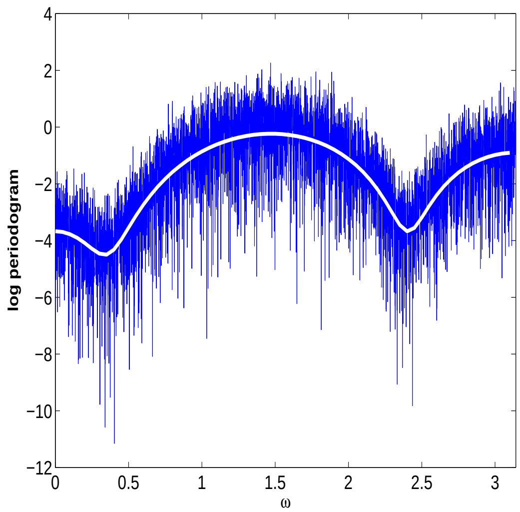
1. Calculate the log-periodogram of time series at for non-negative Fourier frequencies. To avoid boundary effects, the log-periodogram sequence is extended over the boundaries in the mirror-like fashion. The length of the extended sequence should be power of 2.

2. Transform the data in the wavelet domain. Apply Bayes shrinkage rule (2.11) on all detail coefficients.

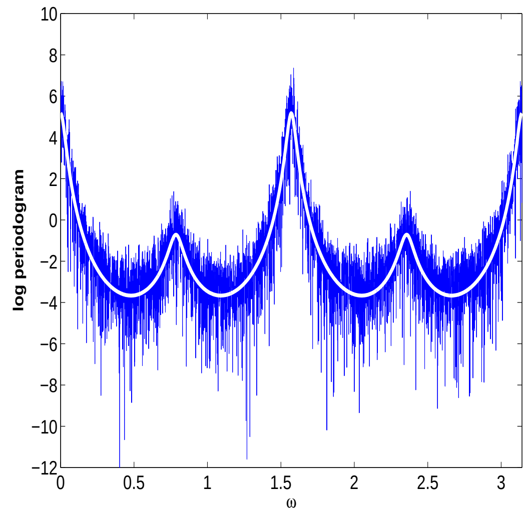
3. Transform back the data and take the subsequence corresponding to the unextended log periodogram from the step **1.** To obtain an estimator of the log-spectral density add the Euler constant γ.

We demonstrate the BAMS-LP on the Sunspot data set. We also briefly review wavelet-based estimator of log-spectral density, GAO, proposed by Gao (1993b), since the developed

Bayesian wavelet shrinkage provides a rationale for its improvements. Finally, we compare the performance of BAMS-LP to the best modification of GAO algorithm and discuss an automatic selection of hyperparameters in the model. The comparison is done on standardly used ARMA template time series: $X_t = Z_t - 0.3Z_{t-1} - 0.6Z_{t-2} - 0.3Z_{t-3} + 0.6Z_{t-4}$, and $X_t = 0.9X_{t-4} + 0.8X_{t-8} - 0.63X_{t-12} + Z_t$. (Wahba, 1980; Gao, 1993b, Moulin, 1994, among others). The row log-periodograms and the theoretical spectral densities (superimposed in white) for these two test examples are given in Figure 4(a, b).



(a)



(b)

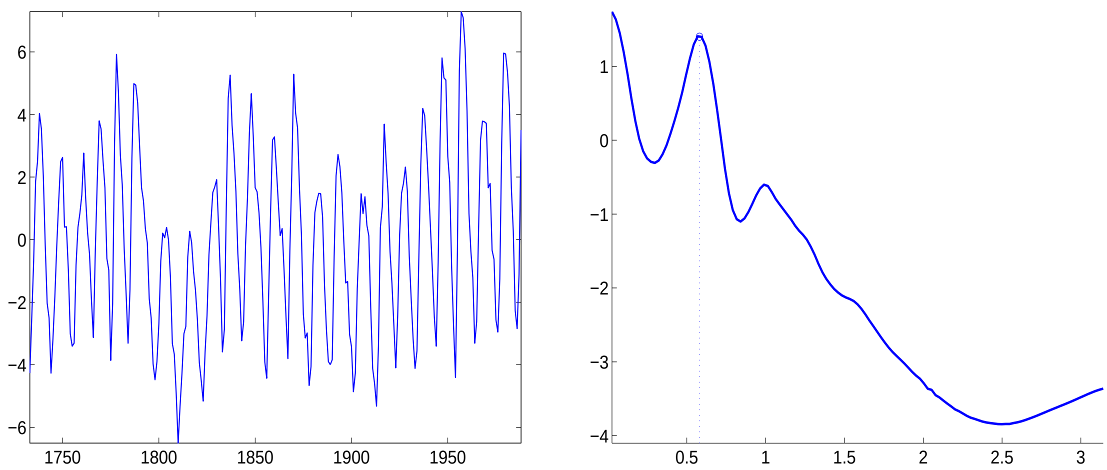
Figure 3: Log-periodogram and theoretical log spectral density of (a) MA(4) process $X_t = Z_t - 0.3Z_{t-1} - 0.6Z_{t-2} - 0.3Z_{t-3} + 0.6Z_{t-4}$, and (b) AR(12) process $X_t = 0.9X_{t-4} + 0.8X_{t-8} - 0.63X_{t-12} + Z_t$.

4.1 Sunspot Data Analysis

A real-life application of spectral and log-spectral estimation involves the processing of Wolf's data set. Although in this situation the statistician does not know the "true" signal, the theory developed by solar scientist helps to evaluate performance of the algorithm.

The Sun's activity peaks every 11 years, creating storms on the surface of our star that disrupt the Earth's magnetic field. These "solar hurricanes" can cause severe problems for electricity transmission systems. An example of influence of such periodic activity to everyday life is 1989 power blackout in the American northeast.

Efforts to monitor the amount and variation of the Sun's activity by counting spots on it have a long and rich history. Relatively complete visual estimates of daily activity date back to 1818, monthly averages can be extrapolated back to 1749, and estimates of annual values can be similarly determined back to 1700. Although Galileo made observations of sunspot numbers in the early 17th century, the modern era of sunspot counting began in the mid-1800s with the research of Bern Observatory director Rudolph Wolf, who introduced what he called the *Universal Sunspot Number* as an estimate of the solar activity. The square root of Wolf's yearly sunspot numbers are given in Figure 4.1(a), data from Tong (1996) p. 471. Because of wavelet data processing we selected a sample of size a power of two, i.e., only 256 observations from 1733 till 1998. The square root transformation was applied to symmetrize and de-trend the Wolf's counts. The panel (b) in Figure 4.1 shows the the BAMS-LP estimator.



(a) (b)

Figure 4: (a) Square roots of Wolf's yearly sunspot numbers from 1732-1988 (256 observations); (b) BAMS-LP estimator of the log-spectra. The frequency $\omega^* \approx 0.58$ corresponds to Schwabe's period of 10.8 (years).

The estimator reveals a peak at frequency $\omega^* \approx 0.58$, corresponding to the Schwabe's cycle ranging from 9 to 11.5 (years), with an average of $\frac{2\pi}{0.58} \approx 10.8$ years. The Schwabe cycle is the period between two subsequent maxima or minima the solar activity, although the solar physicists often think in terms of a 22-year magnetic cycle since the sun's magnetic poles reverse direction every 11 years.

4.2 Gao's Algorithm and Its Modifications

Motivated by the apparent asymmetry of the Bayes shrinkage rules (Figure 2.3), we propose modifications to a Gao's algorithm. For completeness, a brief overview of the original Gao's algorithm (GAO, Gao; 1993b) is provided.

The GAO algorithm for estimating the log-spectral density consists of three steps. The steps 1 and 3 in GAO and BAMS-LP algorithm coincide. The step 2' in which the shrinkage is applied is as follows:

2'. Apply the soft thresholding rule $\delta^s(x,\lambda) = sign(x)\ (|x| - \lambda)_+$, with threshold $\lambda_{j,T}$, depending on the level j and sample size T, as follows:

(a) If the shrinkage is applied to the resolution levels of fine detail ($j = J-1, J-2, \dots$) then the threshold

$$\lambda_{j,T} = \alpha_j \ln \frac{T}{2} \tag{4.30}$$

is selected. The typical values of α_j, robust for commonly used wavelet bases, such as Coiflets, Daubechies', and Symmlets, are given in Table 4.2.

level j	α_j	level j	α_j
$J-1$	1.29	$J-6$	0.54
$J-2$	1.09	$J-7$	0.46
$J-3$	0.92	$J-8$	0.39
$J-4$	0.77	$J-9$	0.32
$J-5$	0.65	$J-10$	0.27

Table 2: Values of multipliers for thresholding scales of fine detail in Gao's Algorithm.

(b) If the resolution level is a coarse, that is, if $j \ll J-1$, then use

$$\lambda_T = \sqrt{2 \ln \frac{T}{2} \cdot \frac{\pi^2}{6}} \approx \sqrt{3.29 \cdot \ln \frac{T}{2}}. \tag{4.31}$$

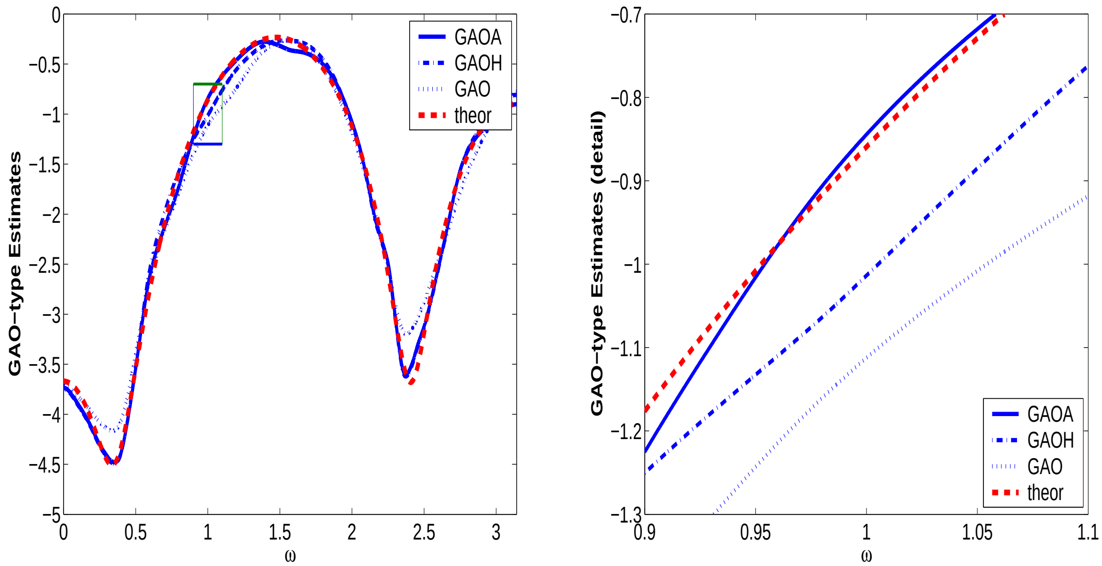
The threshold justification is based the distribution of the error as in (2.3) Since $E\epsilon_\ell = 0$ and $Var\ \epsilon_\ell = \pi^2/6$ the threshold (4.31) is simply the universal threshold. The noise at fine levels of detail has non-Gaussian character and the threshold in (4.30) is a result of an analysis of such noise. Details can be found in Gao (1993b).

Motivated by the fact that hard-thresholding policy is superior to the soft in wavelet-smoothing of log-spectral density and by the apparent asymmetry of the BAMS-LP rule (2.11) two modifications of the original Gao's algorithm are proposed.

We call GAOH the method that retains the threshold values from GAO algorithm but utilizes hard-shrinkage policy, the rule $\delta^h(x,\lambda) = \mathbf{1}(|x| > \lambda)$. An extensive simulational study shows that GAOH consistently outperforms [in terms of overall MSE] the original GAO algorithm for a variety of test spectral densities and sample sizes.

The asymmetry of the error distribution propagates through the several fine levels of wavelet decomposition and the Bayes rule is asymmetric and shrinks more the negative values of the error, as it can be concluded from Figure 2.3

To further improve GAOH, we propose its asymmetric modification GAOA, in which, at the fine level of detail, the negative threshold λ_1 exceeds in absolute value the positive threshold λ, i.e., $-\lambda_1 \geq \lambda$. Simulations show that appropriate asymmetry is $\lambda_1 = -(1+\rho)\lambda$, with ρ between 0 and 0.1. and λ from GAO. The shrinkage policy remains hard-thresholding.



(a) (b)

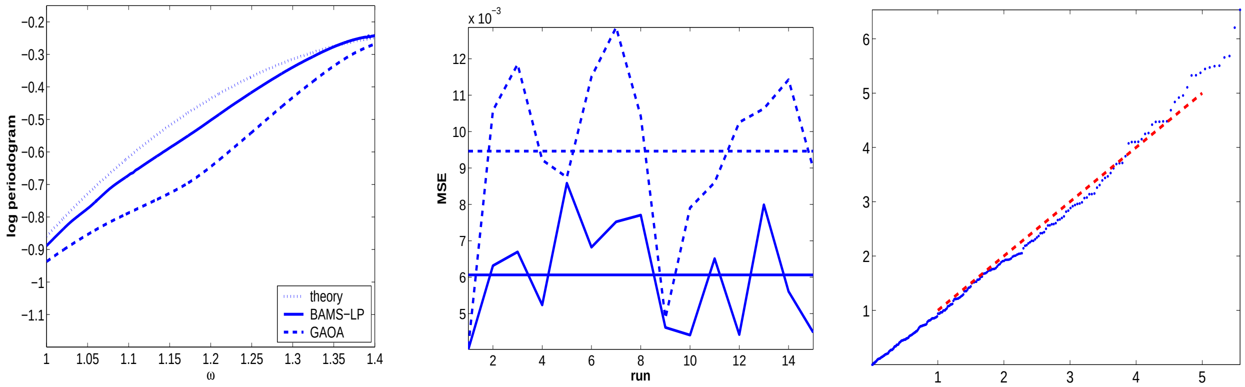
Figure 5: (a) Log-spectral density of MA(4) process $X_t = Z_t - 0.3Z_{t-1} - 0.6Z_{t-2} - 0.3Z_{t-3} + 0.6Z_{t-4}$ estimated by different modifications of GAO algorithm (as in legend) and (b) the area of detail.

Thus, 3 classical methods are compared, GAO, GAOH, and GAOA and Figure 4.2(a) depicts the estimators on Wahba's MA(4) process. Figure 4.2(b) shows the area of detail demonstrating better performance of GAOA algorithm with $\rho = 6\%$.

4.3 Comparisons

As an illustration of the developed algorithm apply the BAMS-LP on the MA(4) template process, $X_t = Z_t - 0.3Z_{t-1} - 0.6Z_{t-2} - 0.3Z_{t-3} + 0.6Z_{t-4}$. Panel (a) in Figure 4.3 gives an area of detail. The theoretical log-spectral density is plotted dotted line) and its reconstruction by GAOA (dashed line) and BAMS-LP (solid line); panel (b) gives the mean square error for the two methods for ten simulational runs; and panel (c) gives the qqplot of the residuals of $exp\{\hat{g}\}$ in the BAMS-LP model against the theoretical quantiles of χ^2_2 distribution, indicating excellent distributional compliance of the residuals with theoretical errors. The sample size was $T = 2^{14}$, and the wavelet used was Coiflet 3.

We also compare the BAMS-LP to GAOA on several template spectral densities. For example, the AR(12) process $X_t = 0.9X_{t-4} + 0.8X_{t-8} - 0.63X_{t-12} + Z_t$ results in a challenging log-spectral density, with several, hard to fit, sharp peaks (Wahba, 1980). For the default selection of parameters, various simulations, and various sample sizes, the AMSE of BAMS-LP is seldom worse than that of GAOA.



(a) (b) (c)

Figure 6: (a) A detail of a single run of smoothing. (b) the AMSE of GAOA (dashed line) and BAMS-LP (solid line). (c) The ordered $2I(\omega)/\ln(\hat{g}(\omega))$ plot against the theoretical χ^2_2 quantiles (the QQ-plot of residuals of the Bayes estimator against χ^2_2 quantiles.)

Choudhuri, Ghosal, and Roy (2003) provide a table of performance of 4 competing rules discussed in their paper. Although all shrinkage methods in their comparison are concerning the smoothing of the periodogram, and BAMS-LP is not designed to estimate the periodigram directly, the exponential of BAMS-LP performs comparable to the investigated methods.

4.4 Selection of Hyperparameters

Selection of hyperparameters is critical for the applicability of BAMS-LP for finite samples. The selection should be automatic, and although a fine tuning can better the performance, such automatic selection should perform well for most of log-spectral densities and for all practicable sample sizes.

The implemented selection of hyperparameters β, λ, and ν, for which all the simulations have been done, is described below.

The hyperparameter β_j is an odds ratio of a coefficient from level j being a priori 0, i.e., $\beta_j = \pi_j/(1-\pi_j)$. Our proposal is $\beta_j = 0.1 + 0.8j/(n-1)$, where j is the level and $n-1$ is the index of the finest level. Thus when going from fine to coarse levels of details, both β_j and π_j decrease. This reflects the fact that more coefficients are a priori zero at fine than at the coarse levels and contributes to the smooth appearance of the estimator.

Likelihood-mixing coefficients, λ_j, had been previously discussed and are provided in Table 2.1.

The hyperparameter ν_j is proportional to the scale factor τ_j in the spread part of the prior (2.7) , $\tau_j\xi(\tau_j\theta)$. We suggested an automatic choice as $\nu_j = (1-\lambda_j)(j+2)$. When going from fine to coarse levels, ν_j will decrease almost as j, making the prior more spread at coarse levels, thus allowing for prior modeling of big coefficients.

The proposal for the hyperparameters is in agreement to common sense of how such parameters should influence the model, but it does not blindly follow the large sample choices; in the common-life analysis of log spectra, the number of levels in wavelet decomposition, n, seldom exceeds 20.

5 Conclusions

In this paper a wavelet based smoothing of log periodogram is proposed. The shrinkage in the wavelet domain is induced by an independence model that assumes mixture likelihood and standard sparseness prior. The Bayes rules produces consistent estimator of the log spectral density and the convergence rates are optimal if the prior is selected in appropriate way.

Motivated by the asymmetry of Bayes rules we propose a modification of Gao's algorithm and compare Bayesian shrinkage with the best version of Gao's algorithm.

Matlab routines and scripts used in this paper for shrinkage and figures could be found at `http://www.isye.gatech.edu/ brani/wavelets.html`. The programs could be freely used and modified in the spirit of Donoho's initiative for reproducible research.

Acknowledgments. Peter Müller gave useful suggestions at early stage of this project. Ed George also commented on the results of this research when they have been communicated at ISBA2003 meeting in Santa Cruz. We thank them both.

A. Appendix

Proofs

Proof of Theorem 1 is based on the following Lemmas.

Lemma 1. *If $|\nu_j d|$ is bounded or $\nu_j |\nu_j d|^\lambda/\sqrt{T} \to 0$, then as $\nu_j/\sqrt{T} \to 0$*

$$I_0(d) = \nu_j \xi(\nu_j d)\left[1 + O\left(T^{-1}\nu_j^2 |\nu_j d|^{2\lambda}\right)\right], \tag{A.1}$$

$$I_1(d)/I_0(d) = d - \frac{\nu_j \sigma^2}{T}\frac{\xi'(\nu_j d)}{\xi(\nu_j d)}\left[1 + O\left(\frac{\nu_j^2}{T}|\nu_j d|^{2\lambda}\right)\right] = d - O\left(\frac{\nu_j}{T}|\nu_j d|^\lambda\right). \tag{A.2}$$

If $\sqrt{T}|d|$ is bounded or $T|d|/\nu_j \to 0$, then as $\sqrt{T}/\nu_j \to 0$

$$I_0(d) \sim \frac{\sqrt{T}}{\sigma\sqrt{2\pi}}\exp\left(-\frac{Td^2}{2\sigma^2}\right)\left[1 + O\left(\frac{T^2 d^2}{\nu_j^2}\right)\right], \tag{A.3}$$

$$I_1(d)/I_0(d) \sim -\frac{Td}{\nu_j^2}\int_{-\infty}^{\infty} x^2 \xi(x)dx\left[1 + O\left(\frac{T^2 d^2}{\nu_j^2}\right)\right] = O\left(\frac{T|d|}{\nu_j^2}\right). \tag{A.4}$$

Proof of Lemma 1. We shall give the proof for (A.1) and (A.2); the proofs of (A.3) and (A.4) are conducted in a similar manner. Change variables $y = \sqrt{T}(d-x)$ in (2.12) and use Taylor series expansion:

$$\begin{aligned} I_i(d) &= \frac{\nu_j}{\sqrt{T}} \int_{-\infty}^{\infty} \left(d - \frac{y}{\sqrt{T}}\right)^i \frac{\exp(-y^2/2\sigma^2)}{\sigma\sqrt{2\pi}}\, \xi\left(d\nu_j - y\frac{\nu_j}{\sqrt{T}}\right) dy \\ &= \frac{\nu_j}{\sqrt{T}} \int_{-\infty}^{\infty} \left(d - \frac{y}{\sqrt{T}}\right)^i \frac{\exp(-y^2/2\sigma^2)}{\sigma\sqrt{2\pi}} \left[\xi(d\nu_j) - y\frac{\nu_j}{\sqrt{T}}\xi'(d\nu_j)\right. \qquad \text{(A.5)} \\ &+ \left. y^2\frac{\nu_j^2}{2T}\xi''(d\nu_j) - y^3\frac{\nu_j^3}{6T\sqrt{T}}\xi'''(d\nu_j) + \cdots\right] dy. \end{aligned}$$

Integrating in (A.5) with $i=0$ and $i=1$, we obtain (A.1) and (A.2).

Lemma 2. *If $\sqrt{T}|d|$ is bounded or $\nu_j^{-1}\sqrt{T}\exp(\sqrt{T}d) \to 0$ when $T\to\infty$, then as $T/\nu_j^2 \to 0$*

$$I_0^*(d) \sim \sqrt{T}\mu(\sqrt{T}d)\left\{1 + O\left(\frac{T}{\nu_j^2}\left[1+e^{2\sqrt{T}d}\right]\right)\right\}, \qquad \text{(A.6)}$$

$$I_1^*(d) \sim -\nu_j^{-2}\sqrt{T}\left[1-\gamma^*\exp(\sqrt{T}d)\right]\int_{-\infty}^{\infty} x^2\xi(x)dx\,\left[1+O\left(\frac{T\left(1+e^{2\sqrt{T}d}\right)}{\nu_j^2}\right)\right] \qquad \text{(A.7)}$$

Proof of Lemma 2 is similar to the proof of Lemma 1. Just note that $\mu'(x) = \mu(x)(1-\gamma^* e^x)$.

Lemma 3. *Let the pdf of d_{jk} given θ_{jk} be of the form $\sqrt{T}\zeta_j(\sqrt{T}(d_{jk}-\theta_{jk}))$ where $\zeta_j(\cdot)$ is defined in (2.5). Then for any positive a and b as $T\to\infty$*

$$\begin{aligned} E(d_{jk}-\theta_{jk})^{2i} &= O\left(T^{-i}\right),\ \ i=1,2, &\text{(A.8)}\\ P(\sqrt{T}|d_{jk}-\theta_{jk}| > a\sqrt{\ln T}) &= o\left(T^{-a^2/(2\sigma^2)}\right),\ \ j<J_0, &\text{(A.9)}\\ P(|d_{jk}-\theta_{jk}| > a\ln T) &= \lambda_j O\left(T^{-a}\right) + (1-\lambda_j)o\left(T^{-a}\right), &\text{(A.10)}\\ P(\sqrt{T}(d_{jk}-\theta_{jk}) > a\ln T) &= o\left(T^{-a}\right), &\text{(A.11)} \end{aligned}$$

Proof of Lemma 3. Validity of Lemma 3 follow directly from the fact that (compare with (2.5))

$$\sqrt{T}(d_{jk}-\theta_{jk}) \sim (1-\lambda_j)(\sqrt{2\pi}\sigma)^{-1}\exp\left\{-x^2/(2\sigma^2)\right\} + \lambda_j\,\mu(x). \qquad \text{(A.12)}$$

Lemma 4. *If $\xi(x)$ is an even unimodal pdf, then*

$$|I_1(d)/I_0(d)| = O\left(|d|\right) \text{ if } T \to \infty, \tag{A.13}$$
$$|I_1^*(d)/I_0^*(d)| = O\left(|d|\right) \text{ if } \nu_j/\sqrt{T} \to \infty, \ \ \sqrt{T}d \to \infty. \tag{A.14}$$

Proof of Lemma 4. Using the fact that $I_0(d)$ is an even and $I_1(d)$ is an odd function of d, we shall conduct the proof of (A.13) for $d > 0$.

Partition $I_1(d)$ into $I_{11}(d)$, $I_{12}(d)$ and $I_{13}(d)$ where $I_{1i}, i = 1, 2, 3,$ are the integrals calculated over the intervals $(-d/2, 3d/2)$, $(-\infty, -d/2)$ and $(3d/2, \infty)$, respectively. It is easy to see that $|I_{11}(d)/I_0(d)| = O(|d|)$. Let us show that $I_{12}(d)/I_0(d) = O(d)$ since the proof for $I_{13}(d)$ can be conducted in a similar manner. Making a change of variable $x = y - d/2$ and taking into account that, since ξ is symmetric unimodal, $\xi(\nu_j(y - d/2)) \le \xi(\nu_j y)$ for $y < 0$, we derive that

$$\begin{aligned} |I_{12}(d)| &\le \int_{-\infty}^{0} \left| y - \frac{d}{2} \right| \frac{\sqrt{T}}{\sqrt{2\pi}\sigma} e^{-\frac{T(3d/2-y)^2}{2\sigma^2}} \nu_j \, \xi(\nu_j y) dy \\ &= O\left(e^{-\frac{5d^2 T}{8\sigma^2}} d \right) \int_{-\infty}^{0} \frac{\sqrt{T}}{\sqrt{2\pi}\sigma} e^{-\frac{T(d-y)^2}{2\sigma^2}} \nu_j \, \xi(\nu_j y) dy. \end{aligned} \tag{A.15}$$

Here we took into account that $T(3d/2 - y)^2 - T(d - y)^2 = 5Td^2/4 - Tyd$, and for negative y

$$|y - d/2| \exp\left\{Tyd/2\sigma^2\right\} = O(|d|) + O(|Td|^{-1}) = O(|d|)$$

as $\sqrt{T}|d| \to \infty$. Formula (A.15) implies $|I_{12}(d)/I_0(d)| = O(|d|)$.

To prove (A.14), partition the integral $I_1^*(d)$ as $I_{11}^*(d)$, $I_{12}^*(d)$ and $I_{13}^*(d)$ where $I_{1i}^*, i = 1, 2, 3,$ the integrals are calculated over the intervals $(-d, d)$, $(-\infty, -d)$ and (d, ∞), respectively. It is easy to see that $|I_{11}^*(d)/I_0^*(d)| = O(|d|)$.

To derive an upper bound for $|I_{12}^*(d)/I_0^*(d)|$, observe that $\mu'(z)/\mu(z) = 1 - \gamma^* e^z < -z$ for $z < -2$. Therefore, changing variables $x = -(z + d)$ and taking into account that $\xi(\cdot)$ is even and unimodal, we obtain

$$\begin{aligned} |I_{12}^*(d)/I_0^*(d)| &\le \int_0^\infty (z+d)\mu(\sqrt{T}(z+2d))\, \xi(\nu_j(z+d)) dz \Big/ \int_0^\infty \mu(\sqrt{T}(z+d))\, \xi(\nu_j z) dz \\ &\le \int_0^\infty (z+d)\mu(\sqrt{T}(z+d)) \exp[-d\sqrt{T}(z+d)\sqrt{T}]\, \xi(\nu_j z) dz \Big/ \int_0^\infty \mu(\sqrt{T}(z+d))\, \xi(\nu_j z) dz \\ &= O\left((dT)^{-1} + d\right) = O\left(d\right). \end{aligned}$$

Finally, in the case of $I_{13}^*(d)$, change variables $z = d - x$:

$$\begin{aligned} I_{13}^*(d) &= \int_{-\infty}^0 z\mu(\sqrt{T}z)\xi(\nu_j(d-z))dz - d\int_{-\infty}^0 \mu(\sqrt{T}z)\xi(\nu_j(d-z))dz \qquad \text{(A.16)} \\ &= I_{131}^*(d) - dI_{132}^*(d), \end{aligned}$$

where $0 \le I_{132}^*(d) \le I_0^*(d)$. To derive an upper bound for $I_{131}^*(d)$, note that $e^{-\gamma^*} \le \mu(x)/(\gamma^* e^x) < 1$ for $x < 0$, so that we can replace $\mu(x)$ by $\gamma^* e^x$ in the expression for the integral. Then, using integration by parts we arrive at

$$\begin{aligned} 0 &\le \int_{-\infty}^0 (-z)e^{\sqrt{T}z}\xi(\nu_j(d-z))dz = (\sqrt{T})^{-1}\int_{-\infty}^0 (-z)\xi(\nu_j(d-z))d(e^{\sqrt{T}z}) \\ &= (\sqrt{T})^{-1}\int_{-\infty}^0 e^{\sqrt{T}z}\xi(\nu_j(d-z))dz - \nu_j(\sqrt{T})^{-1}\int_{-\infty}^0 ze^{\sqrt{T}z}\xi'(\nu_j(d-z))dz. \qquad \text{(A.17)} \end{aligned}$$

Taking into account that both integrals in (A.17) are positive, we obtain

$$\left|\int_{-\infty}^0 ze^{\sqrt{T}z}\xi(\nu_j(d-z))dz\right| \le (\sqrt{T})^{-1}\int_{-\infty}^0 e^{\sqrt{T}z}\xi(\nu_j(d-z))dz,$$

which implies that $|I_{131}^*(d)/I_0^*(d)| = O\left(1/\sqrt{T}\right) = O\left(|d|\right)$.

Proof of Theorem 1. Since the wavelet basis is orthonormal,

$$R(n, H^r(A)) = E(\hat{\theta}_0 - \theta_0)^2 + \sum_{j=0}^{J-1}\sum_{k=-2^{j-1}}^{2^{j-1}-1} E(\hat{\theta}_{jk} - \theta_{jk})^2 + \sum_{j=J}^{\infty}\sum_{k=-2^{j-1}}^{2^{j-1}-1}\theta_{jk}^2. \qquad \text{(A.18)}$$

Observe that the first term in (A.18) is $O(T^{-1})$ while the last term is bounded by $2^{-2rJ}A = O(T^{-2r})$ due to (3.15), i.e. the main contribution to $R(n, H^r(A))$ is made by the second term. Using (2.11), we partition $(\hat{\theta}_{jk} - \theta_{jk})$ as

$$(\hat{\theta}_{jk} - \theta_{jk}) = \Delta_{1jk} - \Delta_{2jk} + \Delta_{3jk},$$

where

$$\begin{aligned} \Delta_{1jk} &= \frac{(1-\lambda_j)[I_1(\theta_{jk}) - \theta_{jk}I_0(\theta_{jk})] + \lambda_j[I_1^*(\theta_{jk}) - \theta_{jk}I_0^*(\theta_{jk})]}{(1-\lambda_j)I_0(d_{jk}) + \lambda_j I_0^*(d_{jk}) + \beta_j\sqrt{T}\,\zeta_j(\sqrt{T}d_{jk})}, \\ \Delta_{2jk} &= \frac{\theta_{jk}\beta_j\sqrt{T}\left[(1-\lambda_j)\eta(\sqrt{T}\theta_{jk}) + \lambda_j\mu(\sqrt{T}\theta_{jk})\right]}{(1-\lambda_j)I_0(d_{jk}) + \lambda_j I_0^*(d_{jk}) + \beta_j\sqrt{T}\,\zeta_j(\sqrt{T}d_{jk})}, \\ \Delta_{3jk} &= \frac{(1-\lambda_j)I_1(d_{jk}) + \lambda_j I_1^*(d_{jk})}{(1-\lambda_j)I_0(d_{jk}) + \lambda_j I_0^*(d_{jk}) + \beta_j\sqrt{T}\,\zeta_j(\sqrt{T}d_{jk})} \\ &- \frac{(1-\lambda_j)I_1(\theta_{jk}) + \lambda_j I_1^*(\theta_{jk})}{(1-\lambda_j)I_0(\theta_{jk}) + \lambda_j I_0^*(\theta_{jk}) + \beta_j\sqrt{T}\,\zeta_j(\sqrt{T}\theta_{jk})}. \end{aligned}$$

Therefore,

$$R = \sum_{j=0}^{J-1} \sum_{k=-2^{j-1}}^{2^{j-1}-1} E(\hat{\theta}_{jk} - \theta_{jk})^2 \le 3(R_1 + R_2 + R_3) \tag{A.19}$$

with

$$R_1 = \sum_{j=0}^{J-1} \sum_{k=-2^{j-1}}^{2^{j-1}-1} \Delta_{1jk}^2, \;\; R_2 = \sum_{j=0}^{J-1} \sum_{k=-2^{j-1}}^{2^{j-1}-1} \Delta_{2jk}^2, \;\; R_3 = \sum_{j=0}^{J-1} \sum_{k=-2^{j-1}}^{2^{j-1}-1} E\Delta_{3jk}^2.$$

Let us examine each of the terms in turn.

Note that $R_1 = R_{11} + R_{12} + R_{13}$ where

$$R_{11} = \sum_{j=0}^{j_0} \sum_{k=-2^{j-1}}^{2^{j-1}-1} \Delta_{1jk}^2, \;\; R_{12} = \sum_{j=j_0+1}^{J_0} \sum_{k=-2^{j-1}}^{2^{j-1}-1} \Delta_{1jk}^2, \;\; R_{13} = \sum_{j=J_0+1}^{J-1} \sum_{k=-2^{j-1}}^{2^{j-1}-1} \Delta_{1jk}^2.$$

To establish an asymptotic upper bound for R_{11} observe that $\lambda_j = 0$ as $j \le j_0$ and

$$|\Delta_{1jk}| \le |I_1(\theta_{jk})/I_0(\theta_{jk}) - \theta_{jk}| = O\left(\nu_j/T\right) + O\left(\nu_j|\theta_{jk}|/T\right) \tag{A.20}$$

by combination of Lemma 1 (with $\nu_j|\theta_{jk}|$ bounded) and assumption (3.21). Since the sum over k contains 2^j terms, (3.15), (3.25) and (A.20) yield

$$R_{11} \;\; = \;\; O\left(\sum_{j=0}^{j_0} \frac{2^j \nu_j^2}{T^2}\right) + O\left(\sum_{j=0}^{j_0} \frac{\nu_j^4}{2^{2rj}T^2} \sum_{k=-2^{j-1}}^{2^{j-1}-1} \theta_{jk}^2 (1 + 2^{2jr})\right) = O\left(T^{-\frac{2r}{2r+1}}\right) \tag{A.21}$$

To obtain the rate of convergence for R_{12} note that $\lambda_j = 0$ as $j_0 < j < J_0$, and by Lemma 4 and the inequality $|\Delta_{1jk}| \le |I_1(\theta_{jk})/I_0(\theta_{jk})| + |\theta_{jk}|$, we derive

$$R_{12} \;\; = \;\; O\left(\sum_{j=j_0+1}^{J_0} \sum_{k=-2^{j-1}}^{2^{j-1}-1} \theta_{jk}^2\right) = O\left(T^{-2r/(2r+1)}\right). \tag{A.22}$$

For the third term R_{13}, note that $|\Delta_{1jk}| \le |I_1(\theta_{jk})/I_0(\theta_{jk})| + |I_1^*(\theta_{jk})/I_0^*(\theta_{jk})| + 2|\theta_{jk}|$. Conditions (3.26) and (3.15) imply that $\sqrt{T}|\theta_{jk}| \le A$ and, since the sum over k contains 2^j terms, by Lemmas 1 and 4,

$$\begin{aligned} R_{13} \;\; &= \;\; O\left(\sum_{j=J_0+1}^{J-1} \sum_{k=-2^{j-1}}^{2^{j-1}-1} \left[\theta_{jk}^2 + \frac{T^2\theta_{jk}^2}{\nu_j^4} + \frac{T}{\nu_j^4}\right]\right) \\ &= \;\; O\left(\sum_{j=J_0+1}^{J-1} \frac{2^j\, T}{\nu_j^4}\right) + O\left(\sum_{j=J_0+1}^{J-1} \sum_{k=-2^{j-1}}^{2^{j-1}-1} \theta_{jk}^2\right) = O\left(T^{-2r/(2r+1)}\right). \end{aligned} \tag{A.23}$$

Combination of (A.21), (A.22) and (A.23) lead to $R_1 = O\left(T^{-2r/(2r+1)}\right)$.

To derive an asymptotic expression for R_2, partition R_2 into R_{21} and R_{22} according to the values of j: $j \le j_0$ and $j_0 < j \le J-1$. Then, since $\lambda_j = 0$ as $j \le j_0$, by assumptions (3.20) and (3.22),

$$|\Delta_{2jk}| \;\; \le \;\; \frac{\beta_j |\theta_{jk}| \sqrt{T}\, \eta(\sqrt{T}\theta_{jk})}{I_0(\theta_{jk})} \sim \frac{\beta_j |\theta_{jk}| \sqrt{T}}{\nu_j} \, \frac{\eta(\sqrt{T}\theta_{jk})}{\xi(\nu_j \theta_{jk})} = O\left(\frac{\beta_j |\theta_{jk}| \sqrt{T}}{\nu_j}\right),$$

where $\eta(x)$ is the normal pdf (2.6), so that formula (3.27) implies that

$$R_{21} \;\; = \;\; O\left(\sum_{j=0}^{j_0} \frac{\beta_j^2 T}{2^{2rj}\nu_j^2} \sum_{k=-2^{j-1}}^{2^{j-1}-1} \theta_{jk}^2 (1 + 2^{2jr})\right) = O\left(T^{-2r/(2r+1)}\right). \qquad \text{(A.24)}$$

For $j > j_0$, we can use the fact that $|\Delta_{2jk}| \le |\theta_{jk}|$, hence

$$R_{22} \;\; = \;\; O\left(\sum_{j=j_0+1}^{J-1} \sum_{k=-2^{j-1}}^{2^{j-1}-1} \theta_{jk}^2\right) = O\left(T^{-2r/(2r+1)}\right). \qquad \text{(A.25)}$$

Therefore, $R_2 = O\left(T^{-2r/(2r+1)}\right)$.

Now, let us examine the variance term R_3. Similarly to R_{11}, R_{12} and R_{13}, let R_{31}, R_{32} and R_{33} represent the portions of R_3 with $j \le j_0$, $j_0 < j \le J_0$ and $J_0 < j \le J-1$, respectively.

Let $j \le j_0$. Note that $\lambda_j = 0$ and $|\Delta_{3jk}| \le |\Delta_{4jk}| + |\Delta_{5jk}|$ where

$$\begin{aligned} |\Delta_{4jk}| \;\; &= \;\; |I_1(d_{jk})/I_0(d_{jk}) - I_1(\theta_{jk})/I_0(\theta_{jk})|, \qquad \text{(A.26)} \\ |\Delta_{5jk}| \;\; &= \;\; \beta_j \left|\frac{I_1(d_{jk})}{I_0(d_{jk})}\right| \frac{\sqrt{T}\, \eta(\sqrt{T}\theta_{jk})}{\nu_j\, \xi(\nu_j \theta_{jk})} + \beta_j \left|\frac{I_1(\theta_{jk})}{I_0(\theta_{jk})}\right| \frac{\sqrt{T}\, \eta(\sqrt{T} d_{jk})}{\nu_j\, \xi(\nu_j d_{jk})}. \end{aligned}$$

where $\xi(\cdot)$ is defined in (2.7) and $\eta(\cdot)$ is the normal pdf (2.6). ¿From Lemma 1 and condition (3.21) it follows that

$$\begin{aligned} |\Delta_{4jk}| \;\; &\le \;\; |I_1(d_{jk})/I_0(d_{jk}) - d_{jk}| + |I_1(\theta_{jk})/I_0(\theta_{jk}) - \theta_{jk}| + |d_{jk} - \theta_{jk}| \\ &= \;\; O\left(\frac{\nu_j}{T}\right) + O\left(\frac{\nu_j^2 |\theta_{jk}|}{T}\right) + O\left(\frac{\nu_j^2 |d_{jk} - \theta_{jk}|}{T}\right) + O(|d_{jk} - \theta_{jk}|), \end{aligned}$$

and, since $E(d_{jk} - \theta_{jk})^2 = O(T^{-1})$ and $\nu_j^2/T \le 1$

$$\begin{aligned} R_{311} \;\; &= \;\; \sum_{j=0}^{j_0} \sum_{k=-2^{j-1}}^{2^{j-1}-1} E\Delta_{4jk}^2 = O\left(\sum_{j=0}^{j_0} \sum_{k=-2^{j-1}}^{2^{j-1}-1} \left[\frac{1}{T} + \frac{\nu_j^2}{T^2} + \frac{\theta_{jk}^2 \nu_j^4}{T^2}\right]\right) \\ &= \;\; O\left(T^{-2r/(2r+1)}\right). \qquad \text{(A.27)} \end{aligned}$$

For Δ_{5jk}, by assumptions (3.20)–(3.22),

$$\begin{aligned}|\Delta_{5jk}| &= O\left(\frac{\beta_j\sqrt{T}}{\nu_j}\left[|d_{jk}|+\frac{\nu_j}{T}\right]\right)+O\left(\frac{\beta_j\sqrt{T}}{\nu_j}\left[|\theta_{jk}|+\frac{\nu_j}{T}\right]\right)\\ &= O\left(\frac{\beta_j\sqrt{T}}{\nu_j}\left[|\theta_{jk}|+|d_{jk}-\theta_{jk}|+\frac{\nu_j}{T}\right]\right)\end{aligned}$$

and, by (3.27),

$$R_{312} = O\left(\sum_{j=0}^{j_0}\sum_{k=-2^{j-1}}^{2^{j-1}-1}\frac{\beta_j^2 T}{\nu_j^2}\left[\frac{1}{T}+\theta_{jk}^2+\frac{\nu_j^2}{T^2}\right]\right)=O\left(T^{-2r/(2r+1)}\right), \qquad \text{(A.28)}$$

so that $R_{31}=O\left(T^{-2r/(2r+1)}\right)$.

To construct an asymptotic upper bound for R_{32} note that for $j_0<j<J_0$ we have $\lambda_j=0$ and

$$|\Delta_{3jk}|\le |I_1(d_{jk})/I_0(d_{jk})|+|I_1(\theta_{jk})/I_0(\theta_{jk})|. \qquad \text{(A.29)}$$

If (3.23) is valid, then, combining Lemma 1 and (3.23), we derive $|I_1(d)/I_0(d)|=O(\sqrt{T}/\nu_j^2)+O(|d|T/\nu_j^2)$. Hence,

$$\begin{aligned}R_{32} &= O\left(\sum_{j=j_0+1}^{J-1}\sum_{k=-2^{j-1}}^{2^{j-1}-1}E\Delta_{3jk}^2\right)\\ &= O\left(\sum_{j=j_0+1}^{J-1}\sum_{k=-2^{j-1}}^{2^{j-1}-1}\left[\frac{T^2\,E(d_{jk}-\theta_{jk})^2}{\nu_j^4}+\frac{\theta_{jk}^2T^2}{\nu_j^4}+\frac{T}{\nu_j^4}\right]\right)=O\left(T^{-\frac{2r}{2r+1}}\right). \qquad \text{(A.30)}\end{aligned}$$

If (3.23) does not hold, then in order to analyze R_{32} divide it into two portions:

$$\begin{aligned}R_{321} &= \sum_{j=j_0+1}^{J_0}\sum_{k=-2^{j-1}}^{2^{j-1}-1}E\Delta_{3jk}^2\;I(\sqrt{T}|\theta_{jk}|\to\infty),\\ R_{322} &= \sum_{j=j_0+1}^{J_0}\sum_{k=-2^{j-1}}^{2^{j-1}-1}E\Delta_{3jk}^2\;I(\sqrt{T}|\theta_{jk}|\;\le M),\end{aligned}$$

where $\sqrt{T}|\theta_{jk}|\;\le M$ means that $\sqrt{T}|\theta_{jk}|$ are bounded by some constant M. First consider R_{321}. Note that if $\sqrt{T}|d_{jk}|$ are bounded, then by Lemma 1, $|I_1(d_{jk})/I_0(d_{jk})|=O(\nu_j^{-2}\sqrt{T})$. Similarly, if $\sqrt{T}|d_{jk}|\to\infty$, then $|I_1(d_{jk})/I_0(d_{jk})|=O(|d_{jk}|)$ by Lemma 4, and by the same argument $|I_1(\theta_{jk})/I_0(\theta_{jk})| = O(|\theta_{jk}|)$. Then, using inequality (A.29), we obtain

$$
\begin{aligned}
R_{321} &= O\left(\sum_{j=j_0+1}^{J_0}\sum_{k=-2^{j-1}}^{2^{j-1}-1}\left[Ed_{jk}^2+\frac{T}{\nu_j^4}+\theta_{jk}^2\right]I(\sqrt{T}|\theta_{jk}|\to\infty)\right)\\
&= O\left(\sum_{j=j_0+1}^{J_0}\sum_{k=-2^{j-1}}^{2^{j-1}-1}\left[E(d_{jk}-\theta_{jk})^2 I(\sqrt{T}|\theta_{jk}|\to\infty)+\frac{T}{\nu_j^4}+\theta_{jk}^2\right]\right) \quad \text{(A.31)}\\
&= O\left(T^{-2r/(2r+1)}\right) \quad \text{(A.32)}
\end{aligned}
$$

since $O(T^{-1}) = O(T^{-1}T\theta_{jk}^2/[T\theta_{jk}^2]) = O(\theta_{jk}^2)$ as $\sqrt{T}|\theta_{jk}|\to\infty$.

Represent $R_{322} = R_{3221} + R_{3222}$ with

$$
\begin{aligned}
R_{3221} &= \sum_{j=j_0+1}^{J_0}\sum_{k=-2^{j-1}}^{2^{j-1}-1} E\left[\Delta_{3jk}^2\, I(|d_{jk}-\theta_{jk}|\sqrt{T} > a\sqrt{\ln T})\right]\; I(\sqrt{T}|\theta_{jk}|\;\le M),\\
R_{3222} &= \sum_{j=j_0+1}^{J_0}\sum_{k=-2^{j-1}}^{2^{j-1}-1} E\left[\Delta_{3jk}^2\, I(|d_{jk}-\theta_{jk}|\sqrt{T} \le a\sqrt{\ln T})\right]\; I(\sqrt{T}|\theta_{jk}|\;\le M),
\end{aligned}
$$

where $a^2 \ge 8\sigma^2 r/(2r+1)$. Then, using (A.8), (A.9) and (A.29) similarly to (A.32), we arrive at

$$
\begin{aligned}
R_{3221} &= O\left(\sum_{j=j_0+1}^{J_0}\sum_{k=-2^{j-1}}^{2^{j-1}-1}\left[\sqrt{E(d_{jk}-\theta_{jk})^4}\sqrt{P(|d_{jk}-\theta_{jk}|\sqrt{T}>a\sqrt{\ln T})}+\theta_{jk}^2+\frac{T}{\nu_j^4}\right]\right)\\
&= O\left(\sum_{j=j_0+1}^{J_0}\sum_{k=-2^{j-1}}^{2^{j-1}-1}\left[T^{-\frac{4r}{2r+1}}+\theta_{jk}^2+\frac{T}{\nu_j^4}\right]\right) = O\left(T^{-\frac{2r}{2r+1}}\right) \quad \text{(A.33)}
\end{aligned}
$$

since $2^{J_0} < T$.

To derive an asymptotic upper bound for R_{3222}, note that since $|\theta_{jk}|\sqrt{T}$ is bounded, for some $M_0 > 0$ and large T we have

$$
\begin{aligned}
& I(\sqrt{T}|\theta_{jk}|\;\le M)\; I(|d_{jk}-\theta_{jk}|\sqrt{T}\le a\sqrt{\ln T}) \le I(\sqrt{T}|\theta_{jk}|\;\le M)\; I(\sqrt{T}|d_{jk}|\le 2a\sqrt{\ln T}) \le\\
& I(\sqrt{T}|\theta_{jk}|\;\le M)\;\left[I(\sqrt{T}|d_{jk}|\le 2a\sqrt{\ln T})I\left(\tfrac{\sqrt{T}}{\nu_j}\sqrt{\ln T}\to 0\right)+I\left(\tfrac{\nu_j}{\sqrt{T}\sqrt{\ln T}}\;\le M_1\right)\right]\\
& \le I(\sqrt{T}|\theta_{jk}|\;\le M)\;\left[I\left(\tfrac{\sqrt{T}}{\nu_j}\left(\sqrt{T}|d_{jk}|\right)\to 0\right)+I\left(\tfrac{2^{j(2r+1)}}{T(\ln T)}\;\le M_1\right)\right].
\end{aligned}
$$

Note that (A.29), Lemma 1 and $\sqrt{T}\nu_j^{-1}\left(\sqrt{T}|d_{jk}|\right)\to 0$ imply that

$$
E\Delta_{3jk}^2 = O\left(T\nu_j^{-4}(\sqrt{T}|d_{jk}|)^2+\theta_{jk}^2\right) = O\left(T\nu_j^{-4}+\theta_{jk}^2\right)
$$

since $E(Td_{jk}^2) \le 2T\left[E(d_{jk}-\theta_{jk})^2+\theta_{jk}^2\right] = O(1)$. Therefore, the portion of R_{3222} corresponding to the fist term in (A.34) is $\left(T^{-2r/(2r+1)}\right)$. By (A.13) and (A.29), $E\Delta_{3jk}^2 = O(E[d_{jk} - \theta_{jk}]^2 + \theta_{jk}^2)$, so the second term in the portion is

$$O\left(\sum_{j=j_0+1}^{J_0}\sum_{k=-2^{j-1}}^{2^{j-1}-1}\left[T^{-1}+\theta_{jk}^2\right] I\left[2^j = O(T^{\frac{1}{2r+1}}(\ln T)^{\frac{1}{2r+1}})\right]\right) = O\left(T^{-\frac{2r}{2r+1}}(\ln T)^{\frac{1}{2r+1}}\right).$$

Consequently,

$$R_{3222} = O\left(T^{-\frac{2r}{2r+1}}(\ln n)^{\frac{1}{2r+1}}\right), \tag{A.34}$$

and formulas (A.32), (A.33) and (A.34) imply that $R_{32} = O\left(T^{-\frac{2r}{2r+1}}(\ln T)^{\frac{1}{2r+1}}\right)$.

To derive an asymptotic upper bound for R_{33} note that for $J_0+1 \le j \le J-1$ we have

$$|\Delta_{3jk}| \;\le\; |\Delta_{6jk}| + |\Delta_{7jk}|, \tag{A.35}$$

with

$$\begin{aligned} |\Delta_{6jk}| &= \left|\frac{I_1(\theta_{jk})}{I_0(\theta_{jk})}\right| + \left|\frac{I_1(d_{jk})}{I_0(d_{jk})}\right| + \left|\frac{I_1^*(\theta_{jk})}{I_0^*(\theta_{jk})}\right|, \\ |\Delta_{7jk}| &= \left|\frac{I_1^*(d_{jk})}{I_0^*(d_{jk})}\right|. \end{aligned}$$

Note that for $j > J_0$, $T\theta_{jk}^2 = O(2^{-2jr}T) = O(2^{-2rJ_0}T)$, i.e. $T\theta_{jk}^2$ are bounded. Partition R_{33} as $R_{33} \le 2(R_{331}+R_{332}+R_{333})$ with

$$\begin{aligned} R_{331} &= \sum_{j=J_0+1}^{J-1}\sum_{k=-2^{j-1}}^{2^{j-1}-1} E\Delta_{6jk}^2\, I(|d_{jk}-\theta_{jk}|\sqrt{T} > 2\ln T), \\ R_{332} &= \sum_{j=J_0+1}^{J-1}\sum_{k=-2^{j-1}}^{2^{j-1}-1} E\Delta_{6jk}^2\, I(|d_{jk}-\theta_{jk}|\sqrt{T} \le 2\ln T), \\ R_{333} &= \sum_{j=J_0+1}^{J-1}\sum_{k=-2^{j-1}}^{2^{j-1}-1} E\Delta_{7jk}^2. \end{aligned}$$

Repeating (A.33) with $a\sqrt{\ln T}$ replaced with $2\ln T$ and using (A.10) instead of (A.9) we obtain $R_{331} = O(2^J T^{-1} T^{-2r/(2r+1)}) = O\left(T^{-2r/(2r+1)}\right)$.

Since $I(|d_{jk}-\theta_{jk}|\sqrt{T} \le 2\ln T) \le I(\sqrt{T}|d_{jk}| \le 2\ln T + \sqrt{T}|\theta_{jk}|)$, and

$$\frac{T(\sqrt{T}d_{jk})^2}{\nu_j^2} = O\left(\frac{T\,\ln^2 T}{\nu_j^2}\right) = O\left(\frac{T\,\ln^2 T}{2^{(2r+1)J_0}}\right) = O\left(\frac{\ln^2 T}{T^{1/2r}}\right) = o(1),$$

we have $T\nu_j^{-2}(\sqrt{T}|d_{jk}|)^2 \to 0$. Consequently, by Lemmas 1 and 2,

$$R_{332} = O\left(\sum_{j=J_0+1}^{J-1}\sum_{k=-2^{j-1}}^{2^{j-1}-1}\left[\frac{T\,\ln^2 T}{\nu_j^4} + \frac{T}{\nu_j^4}\right]\right) = O\left(\frac{\ln^2 T}{T^{1/2r}}T^{-1}\right) = o(T^{-1}).$$

Now let us examine R_{333}. Note that $R_{333} = R_{3331} + R_{3332}$ where

$$R_{3331} = \sum_{j=J_0+1}^{J-1}\sum_{k=-2^{j-1}}^{2^{j-1}-1} E\left[\Delta_{7jk}^2 I\left(\sqrt{T}(d_{jk}-\theta_{jk}) < (4r+1)^{-1}\ln T\right)\right],$$

$$R_{3332} = \sum_{j=J_0+1}^{J-1}\sum_{k=-2^{j-1}}^{2^{j-1}-1} E\left[\Delta_{7jk}^2 I\left(\sqrt{T}(d_{jk}-\theta_{jk}) > (4r+1)^{-1}\ln T\right)\right],$$

Note that when $\sqrt{T}(d_{jk}-\theta_{jk}) < (4r+1)^{-1}\ln T$ we have $\sqrt{T}\nu_j^{-1}\exp(\sqrt{T}d_{jk}) = o(1)$, so that Lemma 2 is applicable and $|I_1^*(d_{jk})/I_0^*(d_{jk})| = O\left(\sqrt{T}\nu_j^{-2}[\exp(\sqrt{T}d_{jk})+1]\right) = O(\nu_j^{-1})$, so that $R_{3331} = o\left(T^{-2r/(2r+1)}\right)$.

To find an asymptotic upper bound for R_{3332}, use (A.14), , repeat (A.33) with $a\sqrt{\ln T}$ replaced with $(4r+1)^{-1}\ln T$ and apply (A.11) instead of (A.9). Hence, $R_{3332} = o\left(T^{-2r/(2r+1)}\right)$. Combining all the R-terms together, we arrive at (3.28) or, when (3.23) is valid, at (3.29).

Proof of Corollary 1. It is easy to verify by direct calculations that in the case of the normal distribution conditions (3.19)–(3.22) and (3.23) are valid. Hence, (3.29) is valid.

Proof of Corollary 2. Validity of Corollary 2 follows from the fact that $\lambda = 0$ and conditions (3.21) and (3.22) hold due to Lemma 1.

References

[1] Abramovich, F., Sapatinas, T., Silverman, B.W. (1998). Wavelet thresholding via a Bayesian approach. *J. Roy. Statist. Soc., Ser. B.*, **60**, 725–749.

[2] Brillinger, D. R. (1981) *Time series: Data Analysis and Theory*, McGraw-Hill Inc., New York.

[3] Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods,* Second Edition, Springer-Verlag, New York.

[4] Choudhuri, N., Ghosal, S., and Roy, A. (2003). Bayesian Estimation of the spectral density of a time series, Preprint at NCSU,// `http://www.stat.ncsu.edu/ sghosal/papers/specden.pdf`.

[5] Donoho, D.L. (1993). Progress in wavelet analysis and WVD: A ten minute tour. In *Progress in wavelet analysis and applications*, Editors Y. Meyer and S. Roques, pp. 109-128. Frontières Ed.

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

[7] Fan, J. and Kreutzberger, E. (1998). Automatic Local Smoothing for Spectral Density Estimation. *Scandinavian Journal of Statistics*, **25**, 359–369.

[8] Gangopadhyay, A. K., Mallick, B. K. and Denison, D. G. T. (1998). Estimation of spectral density of a stationary time series via an asymptotic representation of the periodogram. *J. Statist. Plann. Inference*, **75**, 281–290.

[9] Gao, H.-Ye (1993a). *Wavelet estimation of spectral densities in time series analysis.* Ph.D. dissertation. University of California, Berkeley.

[10] Gao, H.-Ye (1993b). Choice of Thresholds for wavelet estimation of the log-spectrum, Tech. Rept., Statistics, Stanford, 1993.

[11] Gao, H.-Ye (1993c). Spectral Density Estimation via Wavelet Shrinkage, Tech. Rept., Statistics, Stanford, 1993.

[12] Gao, H.-Y. (1997). Choice of thresholds for wavelet shrinkage estimate of the spectrum. *Journal of Time Series Analysis*, **18**, 231–251.

[13] Harvey, A. C. (1989). *Forecasting, Structural Time Series and the Kalman Filter.* Cambridge University Press.

[14] Huerta, G. and West, M. (1999). Bayesian inference on periodicities and component spectral structure in time series. *Journal of Time Series Analysis*, **20**, 401–416.

[15] Johnstone, I.M. and Silverman, B.W. (2002a). Empirical Bayes selection of wavelet thresholds. Unpublished manuscript.

[16] Johnstone, I.M. and Silverman, B.W. (2002b). Finding needles hay in haystacks: Risk bounds for Empirical Bayes estimates of possibly sparse sequences. Manuscript.

[17] Koopmans, L. H. (1995). *The Spectral Analysis of Time Series.* (Second Edition). Academic Press, San Diego.

[18] Lee, T. C. M. and Wong, T. F. (2003). Nonparametric Log-Spectrum Estimation using Disconnected Regression Splines and Genetic Algorithms. *Signal Processing*, **83**, 79–90.

[19] Lumeau, B., Pesquet, J. C., Bercher, J. F., and Louveau, L. (1993). Optimization of bias-variance trade-off in non-parametric spectral analysis by decomposition into wavelet packets. In *Progess in Wavelet Analysis and Applications*, Y. Meyer and S. Roques, Eds. Gif-sur-Yvette: Editions Frontières, 285-290.

[20] Moulin, P. (1994). Wavelet Thresholding Techniques for Power Spectrum Estimation, *IEEE Transactions on Signal Processing,* **42**, 3126–3136.

[21] Pawitan, Y. and O'Sullivan, F. (1994). Nonparametric spectral density estimation using Whittle's likelihood, *Journal of American Statistical Association,* **98**, 600–610.

[22] Percival, D. B. and Walden, A. T. (1993). *Spectral Analysis for Physical Applications.* Cambridge University Press, Cambridge, UK.

[23] Priestley, M. B. (1981). *Spectral Analysis and Time Series*. Academic Press, London.

[24] Riedel, K. S. and Sidorenko, A. (1995). Minimum biased multitaper spectral estimation, *IEEE Trans. Signal Processing,* **43**, 188–195.

[25] Shumway, R. H. and and Stoffer, D. S. (2000). *Time Series Analysis and Its Applications* Springer Texts in Statistics Springer-Verlag, New York, 549 pp.

[26] Tong, H. (1996). *Non-Linear Time Series.* Clarendon Press, Oxford.

[27] Vidakovic, B. (1999). *Statistical Modeling by Wavelets.* Wiley, New York.

[28] Vidakovic, B. and Ruggeri, F. (2001). BAMS Method: Theory and Simulations, Special Issue on Wavelets in Statistics, *Sankhyā*, **63**, 2, 234-249.

[29] Wahba, G. (1980). Automatic smoothing of the log periodogram, *Journal of the American Statistical Association,* **75**, 122–132.

[30] Walden, A. T., McCoy, E. J., and Percival, D. B. (1995). The effective bandwidth of a multitaper spectral estimator. *Biometrika*, **82**, 201–214.

[31] Whittwer, G. (1986). On the distribution of the periodogram for the stationary random sequences (With discussion). *Math. Opera. und Stat.,* **17**, 201–219.