Statistica Sinica **17**(2007), 635-666

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

Marianna Pensky, Brani Vidakovic and Daniela De Canditiis

University of Central Florida, Georgia Institute of Technology and I.A.C. CNR Italy

Abstract: The problem of estimating the log-spectrum of a stationary time series by Bayesian 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 sufficiently noise-free. It is also shown that the resulting Bayes estimators are asymptotically optimal (in the mean-squared error sense).

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

Key words and phrases: Log-spectral density, spectral density, wavelets.

1. Introduction

Any statistical inference in time series can be conducted in time or frequency domains. The methods are complementary and provide different insights. Spectral analysis and, in particular, estimation of the spectral density is indispensable 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, Pesquet, Bercher and Louveau (1993), Moulin (1994), Gao (1993a,b, 1997) and Walden, Percival and McCoy (1998), from the classical viewpoint. In this paper we propose a novel wavelet-shrinkage method, based on an 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 real, weakly stationary time series with zero mean and autocovariance function $\gamma(h) = EX(t+h)X(t)$. An absolutely summable 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 $\gamma(0) = \int_{-\pi}^{\pi} f(\omega) d\omega$.

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

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

where B is the backshift operator, $\mathrm{WN}(0, \sigma^2)$ is 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 in the closed 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 geoscience, biometry, medicine, industrial production, etc., where information about frequency behavior of some phenomena is essential (e.g., Percival and Walden (1993)). Spectral density can 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 (Brockwell and Davis (1991)).

A traditional statistic used as an estimator of the spectral density is the *periodogram* I. Based on a sample $X_1, \ldots, X_n$, it is defined as

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

where ω_j is the Fourier frequency $\omega_j = \frac{2\pi j}{n}$, $j = \lfloor -n/2 \rfloor + 1, \ldots, -1, 0, 1, \ldots, [n/2]$.

Because of the symmetry of I, we will focus only on non-negative Fourier frequencies, $\omega_j = 2\pi j/n$, $j = 0, \ldots, \lfloor n/2 \rfloor$ $(= T)$.

For any set of Fourier frequencies $\omega_1, \omega_2, \ldots, \omega_k$ such that $0 < \omega_1 < \cdots < \omega_k < \pi$, the $I(\omega_i)$'s are asymptotically independent exponential random variables with means $f(\omega_i)$ and variances $(f(\omega_i))^2$, where f is the spectral density. A good discussion about the subtleties of this approximation can be found in Anderson (1997). Consequently the periodogram is not a consistent estimator of f and, citing Wahba (1980), "it will be hopelessly wiggly even when $f(\omega)$ is a smooth function" and $n \to \infty$.

Smoothing the periodogram will not only help in visually extracting significant frequencies, but smoothed periodograms can also be consistent estimators of f. For the standard theory see Brockwell and Davis (1991). There are several approaches to achieving consistency in the spectral density estimators: smoothing the periodogram directly via a weighted local average, smoothing the log-periodogram via traditional regression techniques, Welch's overlapped segment averaging (Welch (1967)), maximizing Whittle's likelihood (or penalized likelihood) of the periodogram (Chow and Grenander (1985) and Pawitan and O'Sullivan (1994)), and the multitaper approach (Riedel and Sidorenko (1995) and Walden, McCoy and Percival (1995)), to list a few. The literature on 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 the log-periodogram. Early references on utilizing splines and Fourier decomposition of log-periodogram are Cogburn and Davis (1974) and Wahba (1980). Fan and Kreutzberger (1998) investigate local polynomial smoothers of log-periodogram, demonstrate the consistency of the local linear estimators, and find the optimal bandwidth. Lee and Wong (2003) propose disconnected regression splines and a genetic algorithm in non-parametric smoothing of the log-periodogram.

The idea of using wavelets in smoothing log-periodograms was announced in Donoho (1993) and fully developed by Gao (1993a,b, 1997). Moulin (1994) applies saddle point estimation to tail probabilities of distributions of wavelet coefficients to exhibit thresholds for a log-periodogram. Walden, McCoy and Percival (1995) and Walden, Percival and McCoy (1998) propose a multitaper spectrum estimator to address complicated wavelet-based thresholding schemes since the distribution of the log periodogram is markedly non-Gaussian. The logarithm of the multitaper estimator is close to Gaussian provided a moderate number of tapers is used. In contrast to the log periodogram, log multitaper estimates are not approximately pairwise uncorrelated at the Fourier frequencies, however the form of the correlation can be accurately and simply approximated.

Bayesian approaches to spectral time series analysis include Choudhuri, Ghosal and Roy (2004), Gangopadhyay, Mallick and Denison (1998), Müller and Vidakovic (1999) 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 (see e.g., Abramovich, Sapatinas and Silverman (1998), Chipman, Kolaczyk and McCullach (1997), Clyde, Parmigiani and Vidakovic (1998) or Vidakovic (1999)). This is the core of Bayesian wavelet shrinkage.

It is certainly desirable for selected models to well-describe our empirical observations, and to perform well in terms of mean square error for the majority of signals and images. At the same time, in spite of high dimensionality of wavelet descriptions, the calculational 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 the 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 the (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 that

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

where $\overset{i.i.d.}{\approx}$ means "approximately i.i.d.", for the "inner" non-zero Fourier frequencies ω_ℓ. For $\omega = 0$ and extreme Fourier frequencies when the sample size n is even, the right-hand side of (2.1) is $f(\omega)\chi_1^2$. We ignore this difference since its effect is negligible for large n. We also assume that i.i.d. 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 ($\gamma \approx 0.577126$). The exact distribution of the log-periodogram can be found in Whittwer (1986). In the interest of practicality, we note that use of ω_0 can be problematic if the time series is centered (mean subtracted). This commonly performed operation has $\ln I(\omega_0)$ at $-\infty$.

The following lemma describes the distribution of the error term in (2.2). Assume that, WLOG, sample size n is even and that $T = n/2$. Because of subsequent wavelet analysis, we assume that T is power of 2, say $T = 2^J$ for some integer J.

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

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

where $\gamma^* = e^{-\gamma} \approx 0.546146$. Also, $E\varepsilon_l = 0$ and $\operatorname{Var} \varepsilon_l = \sigma^2 = \pi^2/6$. The skewness of ϵ is $\gamma_1 = -2\zeta(3)/(\pi^2/6)^{3/2} \approx -1.14$, where ζ is Riemann's zeta function.

Proof. Easy, as $\varepsilon_l - \gamma \stackrel{d}{=} \ln(\chi_2^2/2)$.

In the wavelet domain (2.2) becomes

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

where

$$\begin{aligned} \underline{d}^* &= W\underline{z}; \quad \underline{z} = (z_1, z_2, \ldots, z_T); \\ \underline{\theta}^* &= W\underline{y}; \quad \underline{y} = (\ln f(\omega_1), \ln f(\omega_2), \ldots, \ln f(\omega_T)); \\ \underline{\delta}^* &= W\underline{\varepsilon}; \quad \underline{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_T); \end{aligned}$$

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

$$\underline{\delta}^* = (\delta_0^*, \underline{\delta}_0^*, \underline{\delta}_1^*, \ldots, \underline{\delta}_{J-1}^*), \tag{2.5}$$

where $\underline{\delta}_j^* = (\delta_{j,0}^*, \delta_{j,1}^*, \ldots, \delta_{j,2^j-1}^*)$ is the jth sub-vector associated with the multiresolution analysis. Here, coefficients δ_0^* of level $j = 0$ correspond to the smooth part represented by the scaling function while $j = J-1$ is the finest resolution level. Similarly, $\underline{d}^* = (d_0^*, \underline{d}_0^*, \underline{d}_1^*, \ldots, \underline{d}_{J-1}^*)$ and $\underline{\theta}^* = (\theta_0^*, \underline{\theta}_0^*, \underline{\theta}_1^*, \ldots, \underline{\theta}_{J-1}^*)$.

The exact distribution for the vector $\underline{\delta}^*$ can be found, since the transformation matrix W can be written in an explicit form. By central limit theorem arguments (see Moulin (1994) and Gao (1993a)) it follows that the density function of a component $\delta^*_{j,k}$ can be well approximated by a mixture

$$\zeta_j(x) = (1-\lambda_j)\eta(x) + \lambda_j \mu(x), \tag{2.6}$$

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

$$\eta(x) = (\sigma\sqrt{2\pi})^{-1} \exp\left(\frac{-x^2}{2\sigma^2}\right), \tag{2.7}$$

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

Representation (2.6) is motivated by the fact that at the coarse resolution levels $\delta^*_{j,k}$'s are approximately normally distributed, while at the fine resolution levels they are affected by the p.d.f. $\mu(x)$ defined by Lemma 2.1. Since the normal and μ have similar shapes (see Figure 2.1), approximating the density of $\delta^*_{j,k}$ by the mixture in (2.6) is a simplifying compromise that works well in practice.

Figure 2.1 shows three densities and the histogram. The log-chisquare μ and the normal η densities are depicted in dotted and dashed lines and their mixture $\zeta(x)$ (solid line) is obtained from (2.6) 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.

Table 2.1 provides weights λ_j for the highest resolution levels. The table is obtained by matching the skewness of the likelihood mixture (2.6) and the empirical distributions of $\underline{\delta}^*_j$. The wavelet used was Coiflet 3, but the weights are quite robust with respect to selection of the wavelet basis. For other standard bases such as Symmlets and Daubechies', we noted minor deviations in λ_j.

Table 2.1. The weights λ_j in the likelihood approximation (2.6).

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	$\le J-8$	≈ 0

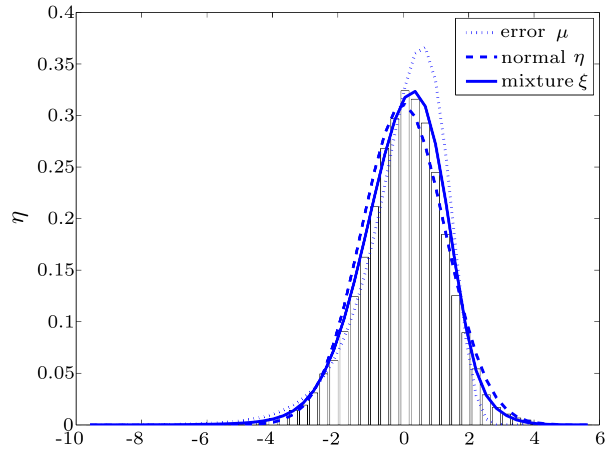


Figure 2.1. The log-chisquare μ and normal η densities (dotted and dashed lines) and their mixture ζ (solid line) obtained with weight $\lambda = 0.355$. The bar plot is the empirical distribution of wavelet coefficients at the finest level of detail.

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}:

$$\theta^*_{jk} \sim \pi_j \delta(0) + (1 - \pi_j)\tau_j \xi(\tau_j x), \tag{2.8}$$

where $0 \leq \pi_j \leq 1$, $\delta(0)$ 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. Note that majority of priors used previously for Bayesian wavelet inference follow the model (2.8) (see e.g., Abramovich, Sapatinas and Silverman (1998), Clyde, Parmigiani and Vidakovic (1998) or Vidakovic (1998)). The factor π_j is the prior probability that a coefficient θ^*_{jk} at level j is zero. In what follows, however, we impose all conditions on the prior odds ratio:

$$\beta_j = \frac{\pi_j}{1 - \pi_j}. \tag{2.9}$$

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

$$g(x) = \frac{1}{\sqrt{\pi}}\theta_0\varphi\Big(\frac{x}{\pi}\Big) + \frac{1}{\sqrt{\pi}}\sum_{j=0}^{\infty}\sum_{k=0}^{2^j-1}\theta_{jk}\psi_{jk}\Big(\frac{x}{\pi}\Big), \quad 0 \le x < \pi, \tag{2.10}$$

with $\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k)$, $\theta_0 = (\sqrt{\pi})^{-1}\int_0^\pi \varphi(x/\pi)g(x)dx$ and $\theta_{jk} = (\sqrt{\pi})^{-1}\int_0^\pi \psi_{jk}(x/\pi)g(x)dx$. Here $\varphi(x)$ is the scaling function and $\psi(x)$ is the corresponding wavelet function periodized on the interval $[0,1]$. 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: θ^* approximates θ only when the sampling interval is 1. The relation $\approx$ in $\theta^*_{jk} \approx \sqrt{T}\theta_{jk}$ can 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.

Let

$$d_{jk} = \frac{d^*_{jk}}{\sqrt{T}}, \quad \nu_j = \sqrt{T}\tau_j. \tag{2.11}$$

Taking into account the relation between θ^*_{jk} and θ_{jk} and (2.6)−(2.11), we find 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 (2.6). Choosing the posterior mean as an estimator,

$$\hat{\theta}_{jk} = \frac{(1-\lambda_j)I_1(d_{jk}) + \lambda_j I_1^*(d_{jk})}{(1-\lambda_j)I_0(d_{jk}) + \lambda_j I_0^*(d_{jk}) + \beta_j\sqrt{T}\zeta_j(\sqrt{T}d_{jk})}, \quad 0 \le j \le J-1, \tag{2.12}$$

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

$$I_i(d) = \int_{-\infty}^{\infty} x^i\sqrt{T}\eta(\sqrt{T}(d-x))\nu_j\xi(\nu_j x)dx, \quad i = 0,1, \tag{2.13}$$

$$I_i^*(d) = \int_{-\infty}^{\infty} x^i\sqrt{T}\mu(\sqrt{T}(d-x))\nu_j\xi(\nu_j x)dx, \quad i = 0,1, \tag{2.14}$$

with $\eta(x)$ and $\mu(x)$ given by (2.3) and (2.7), respectively. The shrinkage rule in (2.12) is shown in Figure 2.2, for some selections 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

$$\hat{g}(x) = \frac{1}{\sqrt{\pi}}\hat{\theta}_0\varphi\Big(\frac{x}{\pi}\Big) + \frac{1}{\sqrt{\pi}}\sum_{j=0}^{J-1}\sum_{k=0}^{2^j-1}\hat{\theta}_{jk}\psi_{jk}\Big(\frac{x}{\pi}\Big), \tag{2.15}$$

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

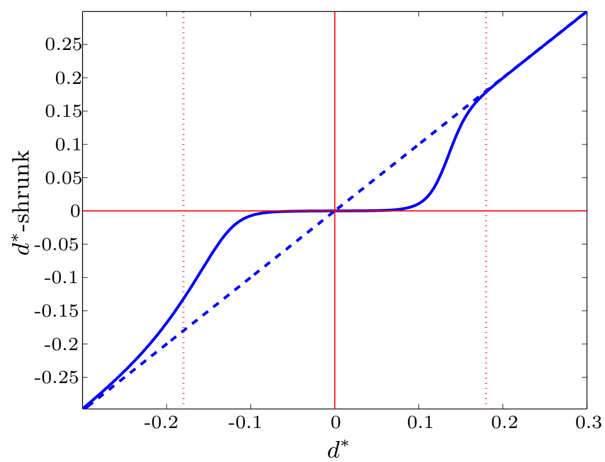


Figure 2.2. Shrinkage rule in (2.12), BAMS-LP rule, for a selection of hyperparameters. The vertical dotted lines are plotted to emphasize the asymmetry of the rule.

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 $I_i(d)$, $i = 0, 1$, can be expressed in terms of the normal cdf (see Johnstone and Silverman (2005, Sec. 2.3)). The integrals $I_i^*(d)$, $i = 0, 1$, can be expressed in terms of incomplete gamma functions $\gamma(\alpha, x)$ and $\Gamma(\alpha, x)$ defined by formula (8.350) of Gradshtein and Ryzhik (1980):

$$
\begin{aligned}
I_0^*(d) &= \nu_j \exp\Big(-\nu_j d + \gamma \frac{\nu_j}{\sqrt{T}}\Big) \gamma\Big(1 + \frac{\nu_j}{\sqrt{T}}, \exp\{d\sqrt{T} - \gamma\}\Big) \\
&\quad + \nu_j \exp\Big(\nu_j d - \gamma \frac{\nu_j}{\sqrt{T}}\Big)\Gamma\Big(1 - \frac{\nu_j}{\sqrt{T}}, \exp\{d\sqrt{T} + \gamma\}\Big), \\
I_1^*(d) &= d I_0^*(d) - \gamma^* \nu_j/\sqrt{T} \exp(-\nu_j d) \frac{d}{ds}\Big[e^{(s+1)\gamma}\gamma(s+1, e^{\sqrt{T}d - \gamma})\Big]_{s=\frac{\nu_j}{\sqrt{T}}} \\
&\quad - \exp(\nu_j d) \frac{d}{ds}\Big[e^{-(s-1)\gamma}\Gamma(-s+1, e^{\sqrt{T}d - \gamma})\Big]_{s=\frac{\nu_j}{\sqrt{T}}}.
\end{aligned}
$$

When $\xi(\cdot)$ is a normal pdf, the values of $I_i(d)$, $i = 0, 1$, are well known (see e.g., Abramovich, Sapatinas and Silverman (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 $\hat{\theta_{jk}}$ efficiently without resorting to numerical integration. Our simulations in Section 4 have been done using the double exponential density.

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 function space. Standard restrictions require the target function to belong to one of a range of smoothness spaces for which the wavelets are unconditional bases.

In order to assess the error of the BAMS-LP estimator $\hat{g}$, we assume that g belongs to a ball $B^r_{p,q}(A)$ in the Besov space $B^r_{p,q}$. The Besov ball $B^r_{p,q}(A)$ can be characterized in terms of wavelet coefficients:

$$g \in B^r_{p,q}(A) \Longleftrightarrow \theta_0^2 + \left\{ \sum_{j=0}^{\infty} 2^{j(r+\frac{1}{2}-\frac{1}{p})q} \left(\sum_{k=0}^{2^j-1} |\theta_{jk}|^p \right)^{\frac{q}{p}} \right\}^{\frac{1}{q}} \leq A. \tag{3.1}$$

Moreover, it is known (see e.g., Johnstone (2002, Lemma 9.3)) that whenever $g \in B^r_{p,q}(A)$ with $r > 1/p$ we have

$$\sum_{k=0}^{2^j-1} \theta_{jk}^2 \leq \begin{cases} B_1 2^{-2jr}, & \text{if } \ p \geq 2, \\ B_1 2^{-2j(r+\frac{1}{2}-\frac{1}{p})}, & \text{if } \ 1 \geq p < 2, \end{cases} \tag{3.2}$$

for some $B_1 > 0$. The cases $p \geq 2$ and $1 \leq p < 2$ apply to spatially homogeneous and nonhomogeneous functions, respectively.

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

$$R(T, \tilde{g}, \mathcal{F}) = \sup_{g \in \mathcal{F}} E\|\tilde{g} - g\|^2_{L^2[0,\pi]}. \tag{3.3}$$

In what follows we establish convergence rates for $R(T, \hat{g}, B^r_{p,q}(A))$ as $T \to \infty$, where $\hat{g}$ is given by (2.15). Moreover, we show that $R(T, \hat{g}, B^r_{p,q}(A))$ could deviate from the optimal rate $O(T^{-2r/(2r+1)})$ by just a logarithmic factor.

Although, to the best of our knowledge, no lower bounds for $R(T, \tilde{g}, B^r_{p,q}(A))$ are available in the case of estimation of the log-spectral density, the rate of $O(T^{-2r/(2r+1)})$ 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 $B^r_{p,q}(A)$ in the Besov space $B^r_{p,q}$, then

$$\inf_{\tilde{g}} R(n, \tilde{g}, B^s_{p,q}(A)) = O\left(n^{-\frac{2r}{2r+1}}\right) \quad (n \to \infty) \tag{3.4}$$

provided $r > \max(0, 1/p - 1/2)$ and $p, q \geq 1$. Since, by (2.6), for the majority of resolution levels ($j \leq J_0$), the errors $\delta^*_{j,k}$ are close to normal, we can expect to achieve a convergence rate close to (3.4) as $T \to \infty$ for some choices of ξ in (2.8).

Note that convergence rate (3.4) cannot be attained by a linear estimator (see e.g., Mallat (1999)). Linear estimator is known to achieve optimal convergence rates only in spacially homogeneous Besov spaces ($p \geq 2$).

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 ξ in the prior (2.8) is three times differentiable, at least in a piecewise sense, has a finite fourth moment, and satisfies the conditions

$$\Big|\frac{\xi^{(k)}(x)}{\xi(x)}\Big| \leq C_{\xi,1}(1+|x|^{\lambda})^k, \quad k=1,2,3, \lambda \geq 0, \tag{3.5}$$

$$C_{\xi,2}\exp(-\frac{x^2}{2\sigma^2}) \leq \xi(x) \leq \xi_0, \tag{3.6}$$

for some positive $C_{\xi,1}$ and $C_{\xi,2}$. Let also the integrals $I_i(d)$ defined in (2.13) be such that

$$\Big|\frac{I_1(d)}{I_0(d)} - d\Big| = O\Big(\frac{|d|\nu_j^2}{T}\Big) \quad \text{if } \frac{\nu_j}{\sqrt{T}} \to 0, \nu_j|d| \to \infty, \tag{3.7}$$

$$I_0(d) \sim \nu_j \xi(\nu_j\theta), \qquad \text{if } \frac{\nu_j}{\sqrt{T}} \to 0, \nu_j|d| \to \infty. \tag{3.8}$$

$$\Big|\frac{I_1(d)}{I_0(d)}\Big| = O\Big(\frac{|d|T}{\nu_j^2}\Big) \qquad \text{if } \frac{\nu_j}{\sqrt{T}} \to \infty. \tag{3.9}$$

We let $p^* = \min(p,2)$,

$$j_0 = (2r+1)^{-1}\log_2 T, \quad jos = 2r[(2r+1)(2r+1-\frac{2}{p^*})]^{-1}\log_2 T, \tag{3.10}$$

$$J_0 = \Big(\frac{1}{2} + r[(2r+1)(2r+1-\frac{2}{p^*})]^{-1}\Big)\log_2 T, \tag{3.11}$$

$$r^* = 1/2\Big[(\frac{1}{p^*}-\frac{1}{2}) + \sqrt{(\frac{1}{p^*}-\frac{1}{2})^2 + 2(\frac{1}{p^*}-\frac{1}{2})}\Big],$$

and assume that

$$\nu_j = 2^{\mu j} \text{ with } \mu = \begin{cases} \mu_1 = r + \frac{1}{2} + \frac{1}{4} - (2p^*)^{-1}, & j \leq j_0, \\ \mu_2 = (r+\frac{1}{2}-\frac{1}{p}) + (\frac{r}{2}+\frac{1}{2}-\frac{1}{p})r^{-1}, & j_0 < j \leq j_0^*, \\ \mu_3 = r + \frac{1}{2}, & j > j_0^*, \end{cases} \tag{3.12}$$

$$\beta_j^2 = O\Big(2^{(4r+1+3(\frac{1}{2}-\frac{1}{p^*}))j}T^{-\frac{4r+1}{2r+1}}\Big), \text{ if } j \leq j_0, \tag{3.13}$$

$$\beta_j^{-2} = O\Big(T^{-\frac{\frac{1}{p}-\frac{1}{2}+\varepsilon}{(r+\frac{1}{2})(r+\frac{1}{2}-\frac{1}{p})}}\Big), \text{ if } j_0+1 \le j \le j_0^* \tag{3.14}$$

for some $\varepsilon > 0$.

Theorem 1. *Let $r > r^*$, (3.5)−(3.8), and (3.12)-(3.14) be valid. Then*

$$R(T, \hat{g}, B_{p,q}^r(A)) = O\Big(T^{-\frac{2r}{2r+1}}[\ln T]^{\alpha}\Big),\ T \to \infty, \tag{3.15}$$

where $\alpha = 1/(2r+1)$ if $p \ge 2$ and $\alpha = 2r/(2r+1)$ if $1 \le p < 2$. If, moreover, (3.9) holds and $p \ge 2$, then $\alpha = 0$.

Corollary 1. *Let $r > r^*$ and (3.12)−(3.14) be valid. If ξ is a normal pdf, then $R(T, \hat{g}, B_{p,q}^r(A))$ is of the form (3.15) with $\alpha = 0$ if $p \ge 2$, and $\alpha = 2r/(2r+1)$ if $1 \le p < 2$.*

Corollary 2. *Let $r > r^*$ and (3.12)−(3.14) be valid. If ξ is a double-exponential or a t density, then $R(T, \hat{g}, B_{p,q}^r(A))$ is of the form (3.15) with $\alpha = 1/(2r+1)$ if $p \ge 2$, and $\alpha = 2r/(2r+1)$ if $1 \le p < 2$.*

Remark 1. Assumptions about ν_j can be translated to restrictions on τ_j using (2.11), namely $\tau_j^2 = C_1 2^{2\mu j} T^{-1}$.

Remark 2. Existence of the fourth moment is a purely technical condition, used for derivation of asymptotic expansions of the integrals $I_i(d)$, $i = 0, 1$. This condition can be dropped and replaced by (6.1)−(6.4) and (6.6)−(6.7).

Remark 3. Condition (3.11) is quite realistic, and agrees with the Central Limit Theorem. Note that if $r > r^*$, we have an infinite number,

$$J - J_0 = \log_2 T(4r^2 + 2r(1 - \frac{2}{p^*}) + 1 - \frac{2}{p^*})/[(2r+1)(2r+1-\frac{2}{p^*})],$$

of resolution levels until the Central Limit Theorem takes over. In practice, the normality assumption can be verified via level-by-level testing.

Remark 4. Corollaries 1 and 2 show that the estimator achieves the optimal convergence rate exactly if the prior distribution is normal and $p \ge 2$, and up to a logarithmic factor otherwise. This does not contradict Johnstone and Silverman (2004, 2005) who studied adaptive empirical Bayes estimators that were mainly based on the posterior median. Johnstone and Silverman (2004, 2005) showed that when the errors are normally distributed and priors are heavy-tailed, the convergence rates (3.15) hold for an adaptive estimator (i.e., for r unknown). Their difficulties with the normal prior are due to adaptivity issues.

4. Simulations and Comparisons

Implementation of the proposed Bayesian wavelet shrinkage can be described algorithmically. Here is description.

1. Calculate the log-periodogram of time series at the 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 the Bayes shrinkage rule (2.12) on all detail coefficients.
3. Transform the data back and take the subsequence corresponding to the unextended log periodogram from 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 the 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 a 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: MA(4), AR(12), AR(24), MA(15000), that can be found in Wahba (1980), Gao (1993a,b, 1997), Moulin (1994), Walden, Percival and McCoy (1998), among others. For MA(4) and AR(12)processes row log-periodograms and the theoretical spectral densities (superimposed in white) are given in Figure 4.1(a, b).

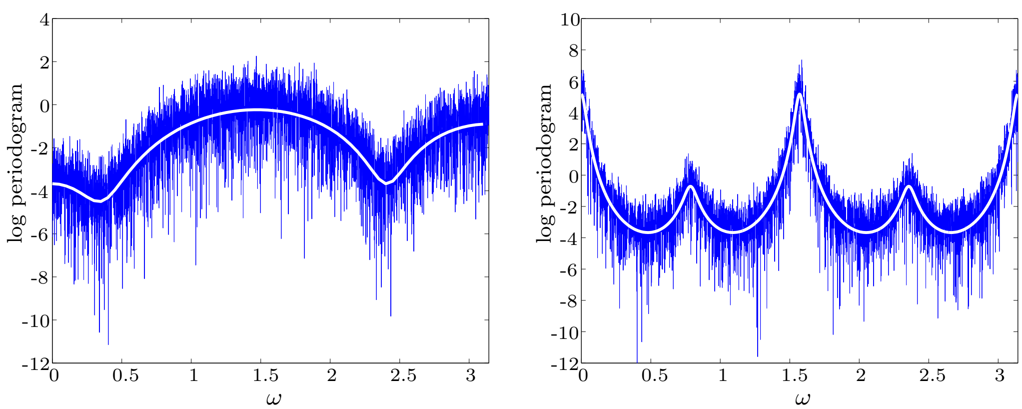


Figure 4.1. Log-periodogram and theoretical log spectral density of (a) the MA(4) process $X_t = Z_t - 0.3Z_{t-1} - 0.6Z_{t-2} - 0.3Z_{t-3} + 0.6Z_{t-4}$, and (b) the AR(12) process $X_t = 0.9X_{t-4} + 0.7X_{t-8} - 0.63X_{t-12} + Z_t$.

4.1. Sunspot data analysis

An 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 the 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 4.2(a), data from Tong (1996), p.471. Because of wavelet data processing, we selected 256 observations from 1,733 to 1,998. The square root transformation was applied to symmetrize the Wolf's counts. The panel (b) in Figure 4.2 shows the BAMS-LP estimator calculated using Daubechies 16 tap filter.

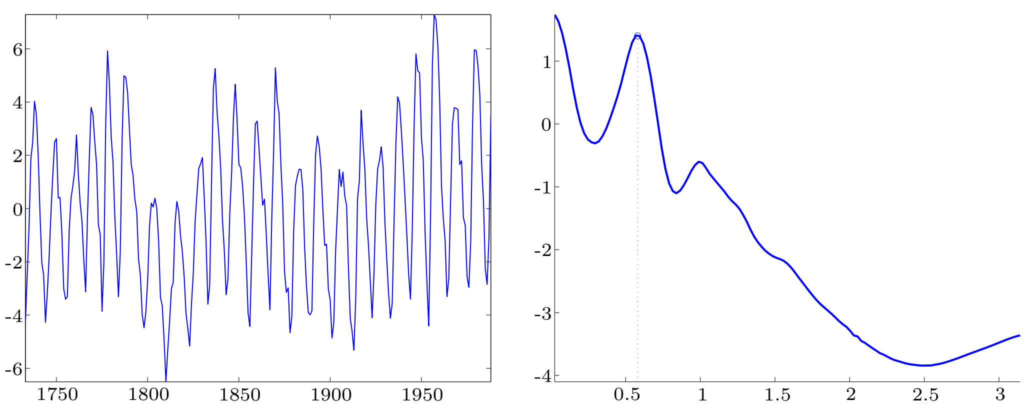


Figure 4.2. (a) Square roots of Wolf's yearly sunspot numbers from 1,732-1,988 (256 observations); (b) BAMS-LP estimator of the log-spectra. The frequency $\omega^* \approx 0.58$ corresponds to Schwabe's period of 10.8 (years).

The estimator reveals a peak at frequency $\omega^* \approx 0.58$, corresponding to Schwabe's cycle that ranges from 9 to 11.5 years, with an average of $2\pi/0.58 \approx 10.8$ years. The Schwabe cycle is the period between two subsequent maxima or minima of 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 2.2), we propose a modification to 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. Steps 1 and 3 in GAO and the BAMS-LP algorithm coincide. 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 ($j = J-1, J-2, \ldots$), then the threshold
   $$\lambda_{j,T} = \alpha_j \ln \frac{T}{2} \tag{4.1}$$
   is selected. The typical values of α_j, robust for commonly used wavelet bases such as Coiflets, Daubechies', and Symmlets, are given in Gao (1993b).
   (b) If the resolution level is coarse, that is, if $j \ll J-1$, then use
   $$\lambda_T = \sqrt{2\ln\frac{T}{2}\cdot\frac{\pi^2}{6}} \approx \sqrt{3.29\cdot\ln\frac{T}{2}}. \tag{4.2}$$

The threshold justification is based on the distribution of the error, as in (2.3) Since $E\epsilon_\ell = 0$ and $Var\ \epsilon_\ell = \pi^2/6$, the threshold (4.2) is simply the universal threshold. The noise at fine levels of detail has non-Gaussian character and the threshold in (4.1) is a result of an analysis of such noise. Details can be found in Gao (1993b).

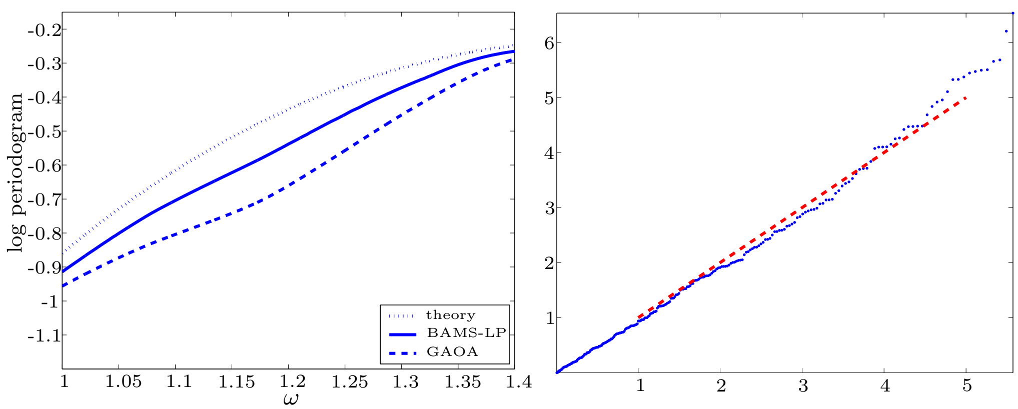
Motivated by the fact that the hard-thresholding policy is superior to the soft in wavelet-smoothing of log-spectral density and by the apparent asymmetry of the BAMS-LP rule (2.12), a modification of the original Gao's algorithm is proposed.

As the asymmetry of the error distribution propagates through the several fine levels of wavelet decomposition, the Bayes rule is asymmetric and shrinks the negative values of the error more, as can be concluded from Figure 2.2.

To improve GAO, 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 $\lambda : -\lambda_1 \geq \lambda$. Simulations show that an appropriate asymmetry is $\lambda_1 = -(1+\rho)\lambda$, with ρ between 0 and 0.1, and λ as in GAO. The shrinkage policy is hard-thresholding. In fact, an extensive simulational study shows that GAO with a hard thresholding policy consistently outperforms (in terms of overall MSE) the original GAO algorithm that uses a soft thresholding policy for a variety of test spectral densities and sample sizes. We also implemented and tested translation invariant versions of the algorithms BAMS-LP and GAOA, and found performance comparable to the originally proposed algorithms. The translation invariant versions produced smoother estimators, however.

4.3. Comparisons

As an illustration of the developed algorithm, we applied the BAMS-LP to 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) with its reconstruction by GAOA (dashed line) and BAMS-LP (solid line); panel (b) 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(18tap filter).



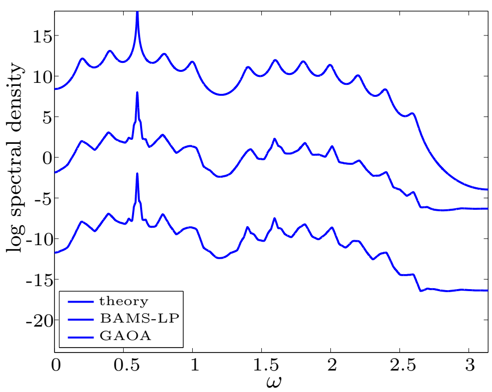
(a) (b)

Figure 4.3. (a) A detail of a single run of smoothing. (b) 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).

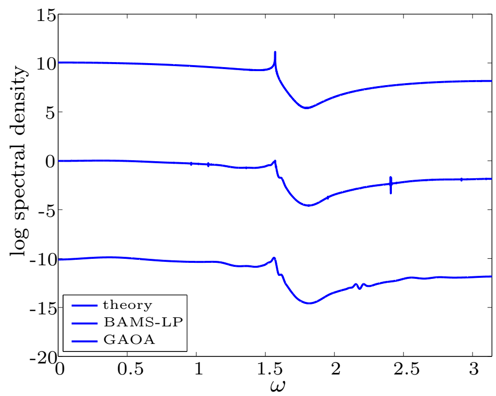
We also explored behavior of BAMS-LP and GAOA on several template spectral densities. For example, the AR(12), AR(24), and MA(15000) result in challenging log-spectral densities, with several, hard to fit, sharp peaks (Wahba (1980), Gao (1993a), Moulin (1994) and Walden, Percival and McCoy (1998)). The AR (12) process is $X_t = 0.9X_{t-4} + 0.8X_{t-8} - 0.63X_{t-12} + Z_t$; coefficients for

AR(24) process are given in Gao (1993b); the template MA(15000) process has θ coefficients $\theta_0 = 1$, $\theta_1 = \pi/4$, $\theta_{k+1} = k^{-1}\sin(\pi k/2)$, $k = 1, 14999$.

Figure 4.4 in Panel (a) shows the BAMS-LP and GAOA estimators of the template AR(24) spectral density. The theoretical density and the GAOA estimator are displaced vertically to make this figure clearer. Panel (b) compares BAMS-LP and GAOA estimators on MA(15000).



(a)



(b)

Figure 4.4. Log-periodograms of (a) Gao's AR(24) spectral density and (b) MA(15000) spectral density. Theoretical log spectral densities are shifted for +10, BAMS-LP estimator is at the place, and GAOA estimator is shifted for -10.

The sample size in both cases was $T = 2^{14}$ with the Coiflet 2 (12 tap) wavelet filter. At this point we note that both estimators, BAMS-LP and GAOA, are quite biased at high frequencies, say $[2.7, \pi]$, due to leakage. This is a common problem for all spectral density estimators based on periodogram (Brillinger (1981)). Such high frequency biases can be taken care of as in Walden, Percival and McCoy (1998), and our methodology could, in principle, be extended to other inconsistent estimators of spectral density, beyond the periodogram.

For the default selection of parameters, various wavelet bases, and various sample sizes, the AMSE of BAMS-LP and GAOA is comparable, as can be seen in Figure 4.5. There we depict boxplots of MSE based on 100 simulations for BAMB-LP (denoted as B) and GAOA (denoted as G). The sample sizes range from $T = 2^{11}$ to $T = 2^{15}$ and the Coiflet 2 (12 tap) wavelet filter was used.

Choudhuri, Ghosal and Roy (2004) provide a table of performance of four competing rules discussed in their paper. All shrinkage methods in their comparison are concerned with the smoothing of the periodogram, and although

BAMS-LP was not designed to estimate the periodogram directly, its exponential performed comparably to the investigated methods.

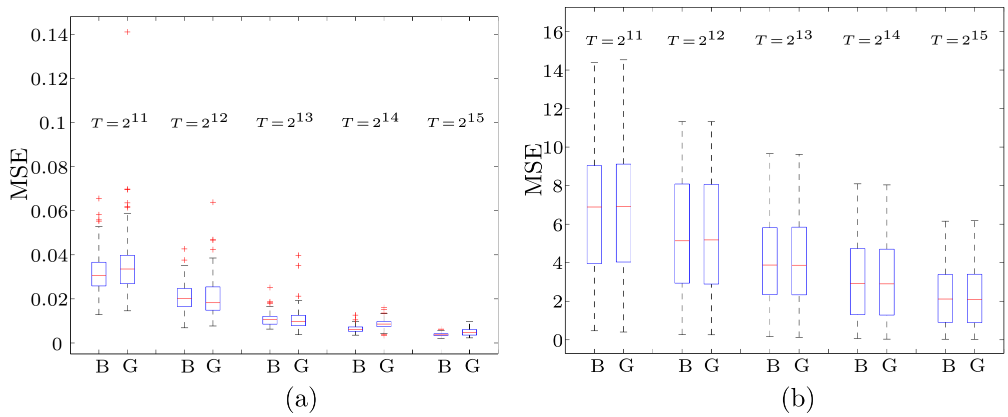


Figure 4.5. For AR(4) and AR(24), boxplots of MSE based on 100 simulations for BAMB-LP (B in graph) and GAOA (G in graph) are plotted. The sample sizes range from $T = 2^{11}$ to $T = 2^{15}$.

4.4. Selection of hyperparameters

Selection of hyperparameters is critical for the success 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 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 rather than at the coarse levels, and this feature contributes to the smooth appearance of the estimator.

The likelihood-mixing coefficients, λ_j, have been previously discussed and may be found in Table 2.1.

The hyperparameter ν_j is proportional to the scale factor τ_j in the spread part of the prior (2.8), $\tau_j\xi(\tau_j\theta)$. We suggest an automatic choice of $\nu_j = (1-\lambda_j)(j+2)$. When going from fine to coarse levels, ν_j will decrease almost as j, making the prior more spread out at coarse levels, thus allowing for prior modeling of large coefficients.

The proposal for the hyperparameters is a common sense reflection on how such parameters should influence the model, but it does not blindly follow the large sample choices; in the ordinary analysis of log spectra, the number of levels in a wavelet decomposition 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 modification.

Matlab routines and scripts used in this paper for shrinkage and figures can be found at `http://www.isye.gatech.edu/∼brani/wavelets.html`. The programs can be freely used and modified. We emphasize that the goal of this paper is not to provide comprehensive simulations and comparison of BAMS-LP and the modified GAO estimator with other available estimators, due to theoretical nature of the paper and space limitations. We hope to present comprehensive simulations elsewhere.

Acknowledgement

Peter Müller gave useful suggestions at an early stage of this project. Ed George also commented on the results of this research when they were communicated at the ISBA2003 meeting in Santa Cruz. The comments and suggestions by the Editor and two anonymous referees further improved the presentation of the paper. We thank them all.

Appendix: Proofs

The 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 $\nu_j/\sqrt{T} \to 0$,*

$$I_0(d) = \nu_j \xi(\nu_j d)\Big[1 + O(T^{-1}\nu_j^2 |\nu_j d|^{2\lambda})\Big], \tag{6.1}$$

$$\frac{I_1(d)}{I_0(d)} = d - \frac{\nu_j \sigma^2}{T}\frac{\xi'(\nu_j d)}{\xi(\nu_j d)}\Big[1 + O\Big(\frac{\nu_j^2}{T}|\nu_j d|^{2\lambda}\Big)\Big] = d - O\Big(\frac{\nu_j}{T}|\nu_j d|^{\lambda}\Big). \tag{6.2}$$

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

$$I_0(d) \sim \frac{\sqrt{T}}{\sigma\sqrt{2\pi}} \exp\Big(-\frac{Td^2}{2\sigma^2}\Big)\Big[1 + O\Big(\frac{T^2 d^2}{\nu_j^2}\Big)\Big], \tag{6.3}$$

$$\frac{I_1(d)}{I_0(d)} \sim -\frac{Td}{\nu_j^2}\int_{-\infty}^{\infty} x^2\xi(x)dx\Big[1+O\Big(\frac{T^2d^2}{\nu_j^2}\Big)\Big] = O\Big(\frac{T|d|}{\nu_j^2}\Big). \tag{6.4}$$

Proof of Lemma 1. We give the proofs of (6.1) and (6.2); the proofs of (6.3) and (6.4) are conducted in a similar manner. Change variables $y = \sqrt{T}(d-x)$ in (2.13) and use the Taylor series expansion

$$\begin{aligned} I_i(d) &= \frac{\nu_j}{\sqrt{T}}\int_{-\infty}^{\infty}\Big(d-\frac{y}{\sqrt{T}}\Big)^i \frac{\exp(\frac{-y^2}{2\sigma^2})}{\sigma\sqrt{2\pi}}\xi\Big(d\nu_j - y\frac{\nu_j}{\sqrt{T}}\Big)dy \\ &= \frac{\nu_j}{\sqrt{T}}\int_{-\infty}^{\infty}\Big(d-\frac{y}{\sqrt{T}}\Big)^i \frac{\exp(\frac{-y^2}{2\sigma^2})}{\sigma\sqrt{2\pi}}\Big[\xi(d\nu_j) - y\frac{\nu_j}{\sqrt{T}}\xi'(d\nu_j). \\ &\quad + y^2\frac{\nu_j^2}{2T}\xi''(d\nu_j) - y^3\frac{\nu_j^3}{6T\sqrt{T}}\xi'''(d\nu_j)+\cdots\Big]dy. \end{aligned} \tag{6.5}$$

Integrating in (6.5) with $i=0$ and $i=1$, we obtain (6.1) and (6.2).

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 $T/\nu_j^2\to 0$,*

$$I_0^*(d) \sim \sqrt{T}\mu(\sqrt{T}d)\Bigg\{1+O\Big(\frac{T}{\nu_j^2}\Big[1+e^{2\sqrt{T}d}\Big]\Big)\Bigg\}, \tag{6.6}$$

$$I_1^*(d) \sim -\nu_j^{-2}\sqrt{T}\Big[1-\gamma^* e^{\sqrt{T}d}\Big]\int_{-\infty}^{\infty} x^2\xi(x)dx\Big[1+O\Big(\frac{T(1+e^{2\sqrt{T}d})}{\nu_j^2}\Big)\Big]. \tag{6.7}$$

Proof of Lemma 2. It 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 at* (2.6). *Then for any positive a and b as $T\to\infty$,*

$$E(d_{jk}-\theta_{jk})^{2i} = O(T^{-i}), \quad i=1,2, \tag{6.8}$$

$$P(\sqrt{T}|d_{jk}-\theta_{jk}| > a\sqrt{\ln T}) = o\Big(T^{\frac{-a^2}{(2\sigma^2)}}\Big), \quad j<J_0, \tag{6.9}$$

$$P(|d_{jk}-\theta_{jk}| > a\ln T) = \lambda_j O(T^{-a}) + (1-\lambda_j)o(T^{-a}), \tag{6.10}$$

$$P(\sqrt{T}(d_{jk}-\theta_{jk}) > a\ln T) = o(T^{-b}). \tag{6.11}$$

Proof of Lemma 3. The validity of Lemma 3 follow directly from the fact that (compare with (2.6))

$$\sqrt{T}(d_{jk}-\theta_{jk}) \sim (1-\lambda_j)(\sqrt{2\pi}\sigma)^{-1}\exp\{-\frac{x^2}{(2\sigma^2)}\} + \lambda_j\mu(x). \tag{6.12}$$

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

$$\Big|\frac{I_1(d)}{I_0(d)}\Big| = O(|d|) \text{ if } \sqrt{T}|d| \to \infty, \tag{6.13}$$

$$\Big|\frac{I_1^*(d)}{I_0^*(d)}\Big| = O(|d|) \text{ if } \frac{\nu_j}{\sqrt{T}} \to \infty, \quad \sqrt{T}|d| \to \infty. \tag{6.14}$$

Proof of Lemma 4. Using the fact that $I_0(d)$ is an even and $I_1(d)$ an odd function of d, we give the proof of (6.13) 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, $\xi(\nu_j(y - d/2)) \leq \xi(\nu_j y)$ for $y < 0$, we find

$$\begin{aligned}
|I_{12}(d)| &\leq \int_{-\infty}^{0} \Big|y - \frac{d}{2}\Big| \frac{\sqrt{T}}{\sqrt{2\pi}\sigma} e^{-\frac{T(\frac{3d}{2}-y)^2}{2\sigma^2}} \nu_j \xi(\nu_j y) dy \\
&= O\Big(e^{-\frac{5d^2T}{8\sigma^2}} d\Big) \int_{-\infty}^{0} \frac{\sqrt{T}}{\sqrt{2\pi}\sigma} e^{-\frac{T(d-y)^2}{2\sigma^2}} \nu_j \xi(\nu_j y) dy.
\end{aligned} \tag{6.15}$$

Here we took into account that $T(3d/2 - y)^2 - T(d - y)^2 = 5Td^2/4 - Tyd$, and, for negative y, $|y - d/2| \exp\{Tyd/2\sigma^2\} = O(|d|) + O(|Td|^{-1}) = O(|d|)$ as $\sqrt{T}|d| \to \infty$. Formula (6.15) implies $|I_{12}(d)/I_0(d)| = O(|d|)$.

To prove (6.14), partition the integral $I_1^*(d)$ as $I_{11}^*(d)$, $I_{12}^*(d)$ and $I_{13}^*(d)$ where I_{1i}^*, $i = 1, 2, 3$, are the integrals 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

$$\begin{aligned}
\Big|\frac{I_{12}^*(d)}{I_0^*(d)}\Big| &\leq \frac{\int_0^\infty (z + d)\mu(\sqrt{T}(z+2d))\xi(\nu_j(z + d))dz}{\int_0^\infty \mu(\sqrt{T}(z + d))\xi(\nu_j z)dz} \\
&\leq \frac{\int_0^\infty (z + d)\mu(\sqrt{T}(z + d))e^{-dT(z+d)}\xi(\nu_j z)dz}{\int_0^\infty \mu(\sqrt{T}(z+d))\xi(\nu_j z)dz} \\
&= O\Big((dT)^{-1} + d\Big) = O(d).
\end{aligned}$$

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

$$I_{13}^*(d) = \int_{-\infty}^{0} z\mu(\sqrt{T}z)\xi(\nu_j(d - z))dz - d\int_{-\infty}^{0} \mu(\sqrt{T}z)\xi(\nu_j(d - z))dz$$

$$= I^*_{131}(d) - dI^*_{132}(d), \tag{6.16}$$

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

$$\begin{aligned} 0 \le \int_{-\infty}^0 (-z) e^{\sqrt{T}z} \xi(\nu_j(d-z))dz &= T^{-\frac{1}{2}} \int_{-\infty}^0 (-z)\xi(\nu_j(d-z))d(e^{\sqrt{T}z}) \\ &= T^{-\frac{1}{2}} \int_{-\infty}^0 e^{\sqrt{T}z}\xi(\nu_j(d-z))dz - \nu_j T^{-\frac{1}{2}} \int_{-\infty}^0 z e^{\sqrt{T}z}\xi'(\nu_j(d-z))dz. \end{aligned} \tag{6.17}$$

Taking into account that both integrals in (6.17) are positive, we obtain

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

which implies that $|I^*_{131}(d)/I^*_0(d)| = O(1/\sqrt{T}) = O(|d|)$.

Lemma 5. *If $f \in B^r_{p,q}(A)$, then*

$$\sum_j \sum_{k=0}^{2^j-1} \theta^2_{jk}\ I(\sqrt{T}|\theta_{jk}| \le b\sqrt{\ln T}) = O\Big(T^{-\frac{2r}{2r+1}} [\ln n]^{\frac{2r}{2r+1}}\Big),$$

$$\sum_j \sum_{k=0}^{2^j-1} T^{-1} I(\sqrt{T}|\theta_{jk}| > b\sqrt{\ln T}) = O\Big(T^{-\frac{2r}{2r+1}} [\ln n]^{-\frac{p}{2}}\Big).$$

Proof of Lemma 5. This can be found in Donoho et al. (1996).

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

$$R(T, \hat{g}, B^r_{p,q}(A)) = E(\hat{\theta}_0 - \theta_0)^2 + \sum_{j=0}^{J-1}\sum_{k=0}^{2^j-1} E(\hat{\theta}_{jk} - \theta_{jk})^2 + \sum_{j=J}^{\infty}\sum_{k=0}^{2^j-1} \theta^2_{jk}. \tag{6.18}$$

Observe that the first term in (6.18) is $O(T^{-1})$ while the last term is bounded by $2^{-2rJ}A = O(T^{-2r})$ due to (3.2), i.e., the main contribution to $R(T, \hat{g}, B^r_{p,q}(A))$ is made by the second term. Divide all resolution levels into the low ($0 \le j \le j_0$), the intermediate ($j_0 + 1 \le j \le j^*_0$), the high ($j^*_0 + 1 \le j \le J_0$), and the highest ($J_0 + 1 \le j \le J - 1$), and partition the second sum in (6.18) accordingly into R_1, R_2, R_3 and R_4. Note that in case $p \ge 2$, $j_0 = j^*_0$ and the intermediate resolution levels vanish.

Low resolution levels: $0 \le j \le j_0$, $\lambda_j = 0$. Let

$$A_{jT}(d) = \frac{\beta_j \sqrt{T}\eta(\sqrt{T}d)}{I_0(d)}, \tag{6.19}$$

and note that $R_1 \leq 2(R_{11} + R_{12})$ where

$$R_{11} = \sum_{j=0}^{j_0} \sum_{k=0}^{2^j-1} E\Big(\frac{I_1(d_{jk})}{I_0(d_{jk})} - \theta_{jk}\Big)^2, \tag{6.20}$$

$$R_{12} = \sum_{j=0}^{j_0} \sum_{k=0}^{2^j-1} E\Big(\frac{\frac{I_1(d_{jk})}{I_0(d_{jk})}}{1 + A_{jT}(d_{jk})} A_{jT}(d_{jk})\Big)^2. \tag{6.21}$$

To establish an asymptotic upper bound for R_{11}, observe that by the combination of Lemma 1 and (3.7), as $j \leq j_0$,

$$\begin{aligned} E(\frac{I_1(d_{jk})}{I_0(d_{jk})} - \theta_{jk})^2 &\leq 2[E(\frac{I_1(d_{jk})}{I_0(d_{jk})} - d_{jk})^2 + E(d_{jk} - \theta_{jk})^2] \\ &= O\Big(E(\frac{\nu_j^4 d_{jk}^2}{T^2}) + \frac{\nu_j^2}{T^2} + \frac{\sigma^2}{T}\Big) \\ &= O\Big(\frac{\nu_j^4 \theta_{jk}^2}{T^2} + \frac{\nu_j^2}{T^2} + \frac{\sigma^2}{T}\Big), \end{aligned} \tag{6.22}$$

so that by (3.2) and the choice of ν_j,

$$R_{11} = O\left(\sum_{j=0}^{j_0} \Big[2^{-2j(r+\frac{1}{2}-\frac{1}{p^*})} \frac{\nu_j^4}{T^2} + 2^j \frac{\nu_j^2}{T^2} + \frac{2^j}{T}\Big]\right) = O\Big(T^{-\frac{2r}{2r+1}}\Big). \tag{6.23}$$

To derive an asymptotic expression for R_{12} note that by (3.8) and Lemma 4,

$$\Big|\frac{\frac{I_1(d_{jk})}{I_0(d_{jk})}}{1 + A_{jT}(d_{jk})} - \frac{I_1(d_{jk})}{I_0(d_{jk})}\Big| = O\Big(\frac{\beta_j |d_{jk}| \sqrt{T} \eta(\sqrt{T} d_{jk})}{I_0(d_{jk})}\Big) = O\Big(\frac{\beta_j |d_{jk}| \sqrt{T}}{\nu_j}\Big),$$

where $\eta(x)$ is the normal pdf (2.7), so that (3.13) implies that

$$\begin{aligned} R_{12} &= O\Big(\sum_{j=0}^{j_0} \nu_j^{-2} \beta_j^2 T \sum_{k=0}^{2^j-1} (\theta_{jk}^2 + T^{-1})\Big) \\ &= O\Big(\sum_{j=0}^{j_0} 2^{-2j\mu_1} \beta_j^2 \Big[T 2^{-j(r+\frac{1}{2}-\frac{1}{p^*})} + 2^j\Big]\Big) = O\Big(T^{-\frac{2r}{2r+1}}\Big). \end{aligned} \tag{6.24}$$

High resolution levels: $j_0^*+1 \leq j \leq J_0$, $\lambda_j = 0$. Since $|\hat{\theta}_{jk}| \leq |I_1(d_{jk})/I_0(d_{jk})|$, we have

$$R_3 = O\left(\sum_{j=j_0^*+1}^{J_0} \sum_{k=0}^{2^j-1} \Big[E\Big(\theta_{jk}^2 + \frac{I_1(d_{jk})}{I_0(d_{jk})}\Big)^2\Big]\right) = R_{31} + R_{32}. \tag{6.25}$$

Note that by the choice of j_0^*

$$R_{31} = \sum_{j=j_0^*+1}^{J_0} \sum_{k=0}^{2^j-1} \theta_{jk}^2 = O\Big(T^{-\frac{2r}{2r+1}}\Big). \tag{6.26}$$

If (3.9) is valid, then combining Lemma 1 and (3.9), we derive $|I_1(d)/I_0(d)| = O(\sqrt{T}/\nu_j^2) + O(|d|T/\nu_j^2)$. Hence,

$$\begin{aligned} R_{32} &= O\Bigg(\sum_{j=j_0^*+1}^{J_0} \sum_{k=0}^{2^j-1} E(\frac{I_1(d_{jk})}{I_0(d_{jk})})^2 \Bigg) \\ &= O\Bigg(\sum_{j=j_0^*+1}^{J_0} \sum_{k=0}^{2^j-1} \Big[\frac{T^2 E(d_{jk}-\theta_{jk})^2}{\nu_j^4} + \frac{\theta_{jk}^2 T^2}{\nu_j^4} + \frac{T}{\nu_j^4}\Big] \Bigg) rkr = O\Big(T^{-\frac{2r}{2r+1}}\Big) \end{aligned} \tag{6.27}$$

and $R_3 = O(T^{-2r/(2r+1)})$ in this case. If (3.9) does not hold, then divide R_{32} into two portions:

$$R_{321} = \sum_{j=j_0^*+1}^{J_0} \sum_{k=0}^{2^j-1} E(\frac{I_1(d_{jk})}{I_0(d_{jk})})^2 I(\sqrt{T}|\theta_{jk}| \to \infty),$$

$$R_{322} = \sum_{j=j_0^*+1}^{J_0} \sum_{k=0}^{2^j-1} E(\frac{I_1(d_{jk})}{I_0(d_{jk})})^2 I(\sqrt{T}|\theta_{jk}| \le M),$$

where $\sqrt{T}|\theta_{jk}| \le M$ means that the $\sqrt{T}|\theta_{jk}|$ are bounded by some constant M. First consider R_{321}. Note that if the $\sqrt{T}|d_{jk}|$ 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. Then we obtain

$$\begin{aligned} R_{321} &= O\Bigg(\sum_{j=j_0^*+1}^{J_0} \sum_{k=0}^{2^j-1} \Big[E(d_{jk}-\theta_{jk})^2 I(\sqrt{T}|\theta_{jk}| \to \infty) + \frac{T}{\nu_j^4} + \theta_{jk}^2\Big] \Bigg) \\ &= O\Big(T^{-\frac{2r}{2r+1}}\Big), \end{aligned} \tag{6.28}$$

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

Now, represent $R_{322} = R_{3221} + R_{3222}$ with

$$R_{3221} = \sum_{j=j_0^*+1}^{J_0} \sum_{k=0}^{2^j-1} E\Big[\Big(\frac{I_1(d_{jk})}{I_0(d_{jk})}\Big)^2 I(|d_{jk}-\theta_{jk}|\sqrt{T} > a\sqrt{\ln T})\Big] I(\sqrt{T}|\theta_{jk}| \le M),$$

$$R_{3222} = \sum_{j=j_0^*+1}^{J_0} \sum_{k=0}^{2^j-1} E\Big[\Big(\frac{I_1(d_{jk})}{I_0(d_{jk})}\Big)^2 I(|d_{jk}-\theta_{jk}|\sqrt{T} \le a\sqrt{\ln T})\Big] I(\sqrt{T}|\theta_{jk}| \le M),$$

where $a^2 \geq 4\sigma^2$. Then, using (6.8) and (6.9) similarly to (6.28), we arrive at

$$\begin{aligned} R_{3221} &= O\left(\sum_{j=j_0^*+1}^{J_0}\sum_{k=0}^{2^j-1}\left[\sqrt{E(d_{jk}-\theta_{jk})^4}\sqrt{P(|d_{jk}-\theta_{jk}|\sqrt{T}>a\sqrt{\ln T})}+\theta_{jk}^2+\frac{T}{\nu_j^4}\right]\right) \\ &= O\left(\sum_{j=j_0^*+1}^{J_0}\sum_{k=0}^{2^j-1}\left[T^{-2}+\theta_{jk}^2+\frac{T}{\nu_j^4}\right]\right) = O\left(T^{-\frac{2r}{2r+1}}\right), \end{aligned} \tag{6.29}$$

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 $M_1 > 0$ and large T we have

$$\begin{aligned} &I(\sqrt{T}|\theta_{jk}|\leq M)I(|d_{jk}-\theta_{jk}|\sqrt{T}\leq a\sqrt{\ln T}) \\ &\quad\leq I(\sqrt{T}|\theta_{jk}|\leq M)I(\sqrt{T}|d_{jk}|\leq 2a\sqrt{\ln T}) \\ &\quad\leq I(\sqrt{T}|\theta_{jk}|\leq M)\left[I(\sqrt{T}|d_{jk}|\leq 2a\sqrt{\ln T})I\left(\frac{\sqrt{T}}{\nu_j}\sqrt{\ln T}\to 0\right)\right. \\ &\qquad\qquad\left.+I\left(\frac{\nu_j}{\sqrt{T}\sqrt{\ln T}}\leq M_1\right)\right] \\ &\quad\leq I(\sqrt{T}|\theta_{jk}|\leq M)\left[I\left(\frac{\sqrt{T}}{\nu_j}(\sqrt{T}|d_{jk}|)\to 0\right)+I\left(\frac{2^{j(2r+1)}}{T(\ln T)}\leq M_1\right)\right]. \end{aligned} \tag{6.30}$$

Note that Lemma 1 and $\sqrt{T}\nu_j^{-1}(\sqrt{T}|d_{jk}|)\to 0$ imply that

$$E\left(\frac{I_1(d_{jk})}{I_0(d_{jk})}\right)^2 = O\left(T\nu_j^{-4}(\sqrt{T}|d_{jk}|)^2+\theta_{jk}^2\right) = O\left(T\nu_j^{-4}+\theta_{jk}^2\right),$$

since $E(Td_{jk}^2) \leq 2T[E(d_{jk}-\theta_{jk})^2+\theta_{jk}^2] = O(1)$. Therefore, the portion of R_{3222} corresponding to the fis t term in (6.30) is $(T^{-2r/(2r+1)})$. By (6.13), $E(I_1(d_{jk})/I_0(d_{jk}))^2 = O(E[d_{jk}-\theta_{jk}]^2+\theta_{jk}^2)$, so the second term in R_{3222} is

$$O\left(\sum_{j=j_0^*+1}^{J_0}\sum_{k=0}^{2^j-1}[T^{-1}+\theta_{jk}^2]I\left[2^j = O(T^{\frac{1}{2r+1}}(\ln T)^{\frac{1}{2r+1}})\right]\right) = O\left(T^{-\frac{2r}{2r+1}}(\ln T)^{\frac{1}{2r+1}}\right).$$

Consequently,

$$R_{3222} = O\left(T^{-\frac{2r}{2r+1}}(\ln T)^{\frac{1}{2r+1}}\right), \tag{6.31}$$

and formulas (6.28)–(6.31) imply that $R_3 = O\left(T^{-2r/(2r+1)}(\ln T)^{1/(2r+1)}\right)$ whenever (3.9) is invalid.

Highest resolution levels: $J_0+1 \le j \le J-1$, $\lambda_j > 0$. To derive an asymptotic upper bound for R_4 note that

$$|\hat{\theta}_{jk} - \theta_{jk}| \le |\frac{I_1(d_{jk})}{I_0(d_{jk})}| + |\frac{I_1^*(d_{jk})}{I_0^*(d_{jk})}| + \theta_{jk}^2. \tag{6.32}$$

Hence $R_4 = O(\sum_{i=1}^{6} R_{4i})$, with

$$R_{41} = \sum_{j=J_0+1}^{J-1} \sum_{k=0}^{2^j-1} \theta_{jk}^2,$$

$$R_{42} = \sum_{j=J_0+1}^{J-1} \sum_{k=0}^{2^j-1} E\Big[[\frac{I_1(d_{jk})}{I_0(d_{jk})}]^2 + [I_1^*(d_{jk})/I_0^*(d_{jk})]^2\Big] I(|\theta_{jk}|\sqrt{T} \to \infty),$$

$$R_{43} = \sum_{j=J_0+1}^{J-1} \sum_{k=0}^{2^j-1} E\Big[[\frac{I_1(d_{jk})}{I_0(d_{jk})}]^2 I(|d_{jk} - \theta_{jk}|\sqrt{T} > 2\ln T)\Big],$$

$$R_{44} = \sum_{j=J_0+1}^{J-1} \sum_{k=0}^{2^j-1} E\Big[[\frac{I_1(d_{jk})}{I_0(d_{jk})}]^2 I(|d_{jk} - \theta_{jk}|\sqrt{T} \le 2\ln T)\Big] I(\sqrt{T}|\theta_{jk}| \le M),$$

$$R_{45} = \sum_{j=J_0+1}^{J-1} \sum_{k=0}^{2^j-1} E\Big[[\frac{I_1^*(d_{jk})}{I_0^*(d_{jk})}]^2 I((d_{jk} - \theta_{jk})\sqrt{T} > 0.25r\ln T)\Big],$$

$$R_{46} = \sum_{j=J_0+1}^{J-1} \sum_{k=0}^{2^j-1} E\Big[[\frac{I_1^*(d_{jk})}{I_0^*(d_{jk})}]^2 I((d_{jk} - \theta_{jk})\sqrt{T} \le 0.25r\ln T)\Big] I(\sqrt{T}|\theta_{jk}| \le M).$$

Note that similarly to (6.27), $R_{41} = O(T^{-2r/(2r+1)})$. Using Lemma 4, we derive that

$$R_{42} = O\Bigg(\sum_{j=J_0+1}^{J-1} \sum_{k=0}^{2^j-1} E d_{jk}^2 I(|\theta_{jk}|\sqrt{T} \to \infty)\Bigg) = O\Big(T^{-\frac{2r}{2r+1}}\Big) \tag{6.33}$$

by a calculation similar to that at (6.28). Repeating (6.29) with $a\sqrt{\ln T}$ replaced with $2\ln T$, and using (6.10) instead of (6.9), we obtain $R_{43} = O(2^J T^{-1} T^{-2r/(2r+1)})$ $= O(T^{-2r/(2r+1)})$. To find an upper bound for R_{44}, note that $I(|d_{jk} - \theta_{jk}|\sqrt{T} \le 2\ln T) \le I(\sqrt{T}|d_{jk}| \le 2\ln T + \sqrt{T}|\theta_{jk}|)$, and

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

hence $T\nu_j^{-2}(\sqrt{T}|d_{jk}|)^2 \to 0$. Consequently, by Lemma 1,

$$R_{44} = O\Bigg(\sum_{j=J_0+1}^{J-1} \sum_{k=0}^{2^j-1} \Big[\nu_j^{-4} T \ln^2 T + \nu_j^{-4} T\Big]\Bigg) = O\Big(\nu_j^{-2}\Big) = o(T^{-1}). \tag{6.34}$$

To examine R_{45}, use (6.14), repeat (6.29) with $a\sqrt{\ln T}$ replaced by $0.25r\ln T$, and apply (6.11) instead of (6.9). Hence, $R_{45} = o(T^{-2r/(2r+1)})$. Finally, in the case of R_{46}, note that when $\sqrt{T}(d_{jk} - \theta_{jk}) < (4r+1)^{-1}\ln T$, we have $\sqrt{T}\nu_j^{-1}\exp(\sqrt{T}d_{jk}) = o(1)$. Thus, Lemma 2 is applicable and $|I_1^*(d_{jk})/I_0^*(d_{jk})| = O(\sqrt{T}\nu_j^{-2}[\exp(\sqrt{T}d_{jk})+1]) = O(\nu_j^{-1})$, so that $R_{46} = o(T^{-2r/(2r+1)})$. Combining all the R_{4i}-terms together, we arrive at $R_4 = O(T^{-2r/(2r+1)})$.

Intermediate resolution levels: $j_0 \le j \le j_0^*$, $\lambda_j = 0$, $1 \le p < 2$. Partition R_2 into

$$R_{21} = \sum_{j=j_0+1}^{j_0^*}\sum_{k=0}^{2^j-1} E(\hat\theta_{jk} - \theta_{jk})^2 I\Big(|\theta_{jk}| > \frac{\sqrt{\ln T}}{\sqrt{T}}\Big),$$

$$R_{22} = \sum_{j=j_0+1}^{j_0^*}\sum_{k=0}^{2^j-1} E(\hat\theta_{jk} - \theta_{jk})^2 I\Big(|\theta_{jk}| \le \frac{\sqrt{\ln T}}{\sqrt{T}}\Big). \tag{6.35}$$

Then further partition R_{21}, similarly to (6.21), into

$$R_{211} = \sum_{j=0}^{j_0}\sum_{k=0}^{2^j-1} E\Big(\frac{I_1(d_{jk})}{I_0(d_{jk})} - \theta_{jk}\Big)^2 I\Big(|\theta_{jk}| > \frac{\sqrt{\ln T}}{\sqrt{T}}\Big), \tag{6.36}$$

$$R_{212} = \sum_{j=0}^{j_0}\sum_{k=0}^{2^j-1} E\Big(\frac{A_{jT}(d_{jk})I_1(d_{jk})/I_0(d_{jk})}{1 + A_{jT}(d_{jk})}\Big)^2 I\Big(|\theta_{jk}| > \frac{\sqrt{\ln T}}{\sqrt{T}}\Big). \tag{6.37}$$

Repeat (6.22) and note that μ_2 is chosen so that

$$\sum_{j=j_0+1}^{j_0^*}\sum_{k=0}^{2^j-1}\Big[\nu_j^2 T^{-2} + \theta_{jk}^2\nu_j^4 T^{-2}\Big] = O\Big(T^{-\frac{2r}{2r+1}}\Big). \tag{6.38}$$

Then, by Lemma 5,

$$\begin{aligned} R_{211} &= O\bigg(\sum_{j=j_0+1}^{j_0^*}\sum_{k=0}^{2^j-1}\Big[\frac{\nu_j^2}{T^2} + \frac{\theta_{jk}^2\nu_j^4}{T^2} + \frac{\sigma^2}{T} I\Big(|\theta_{jk}| > \frac{\sqrt{\ln T}}{\sqrt{T}}\Big)\Big]\bigg) \\ &= O\Big(T^{-\frac{2r}{2r+1}}\Big). \end{aligned} \tag{6.39}$$

For examination of R_{212}, note that $A_{jT}(d_{jk})/(1+A_{jT}(d_{jk})) \sim \min(1, A_{jT}(d_{jk}))$ and that, due to the condition (3.6),

$$d_{jk}^2 < \gamma_j \equiv 2\sigma^2 T^{-1}\ln(\sqrt{T}\beta_j/(\nu_j\xi_0)) \Longrightarrow A_{jT}(d_{jk}) > 1. \tag{6.40}$$

We partition R_{212} into R_{2121} and R_{2122} according to whether the inequality on the left hand side of (6.40 is valid. Then, since $\ln(\sqrt{T}\beta_j/(\nu_j\xi_0)) \sim \ln T$, we have

$$\begin{aligned}
R_{2121} &= O\Big(\sum_{j=j_0+1}^{j_0^*}\sum_{k=0}^{2^j-1} E\Big[d_{jk}^2 I(d_{jk}^2 < \gamma_j)\Big] I(|\theta_{jk}| > \sqrt{\ln\frac{T}{T}})\Big) \\
&= O\Big(\sum_{j=j_0+1}^{j_0^*}\sum_{k=0}^{2^j-1} \frac{\ln T}{T} I\Big(|\theta_{jk}| > \frac{\sqrt{\ln T}}{\sqrt{T}}\Big)\Big) = O\Big(T^{-\frac{2r}{2r+1}}[\ln T]^{1-\frac{p}{2}}\Big). \qquad (6.41)
\end{aligned}$$

The second term R_{2122} is of the form

$$R_{2122} = O\Big(\sum_{j=j_0+1}^{j_0^*}\sum_{k=0}^{2^j-1} E\Big[\frac{d_{jk}^2\beta_j^2 T e^{-Td_{jk}^2/\sigma^2}}{\nu_j^2\xi^2(\nu_j d_{jk})} I(d_{jk}^2 > \gamma_j)\Big] I\Big(|\theta_{jk}| > \frac{\sqrt{\ln T}}{\sqrt{T}}\Big)\Big). \qquad (6.42)$$

Note that by condition (3.6),

$$\frac{\exp\Big(-\frac{Td_{jk}^2}{\sigma^2}\Big)}{\xi^2(\nu_j d_{jk})} = O\Big(\exp\Big\{-(T-\nu_j^2)\frac{d_{jk}^2}{\sigma^2}\Big\}\Big). \qquad (6.43)$$

Moreover, one sees that

$$\max_x \Big[x^2 e^{-\frac{(T-\nu_j^2)x^2}{\sigma^2}} I\Big(x^2 > \gamma_j\Big)\Big] = \gamma_j\Big(\frac{\sqrt{T}\beta_j}{\nu_j\xi_0}\Big)^{-\frac{2(T-\nu_j^2)}{T}} = O\Big(\frac{\nu_j^2\ln T}{\beta_j^2 T^2}\Big),$$

since $\ln(\sqrt{T}\beta_j/\nu_j) \sim \ln T$ by (3.14), and

$$\Big(\nu_j^{-1}\sqrt{T}\beta_j\Big)^{\nu_j^2/T} = \exp\Big(CT^{-1}\ln T\nu_j^2\Big) \sim 1$$

since $\nu_j^2 \ln T/T = o(1)$ as $T \to \infty$. Hence,

$$\begin{aligned}
R_{2122} &= O\Big(\sum_{j=j_0+1}^{j_0^*}\sum_{k=0}^{2^j-1} \frac{\beta_j^2 T}{\nu_j^2}\frac{\nu_j^2\ln T}{\beta_j^2 T^2} I\Big(|\theta_{jk}| > \frac{\sqrt{\ln T}}{\sqrt{T}}\Big)\Big) \\
&= O\Big(\sum_{j=j_0+1}^{j_0^*}\sum_{k=0}^{2^j-1} \frac{\ln T}{T} I\Big(|\theta_{jk}| > \frac{\sqrt{\ln T}}{\sqrt{T}}\Big)\Big) \\
&= O\Big(T^{-\frac{2r}{2r+1}}[\ln T]^{1-\frac{p}{2}}\Big) \qquad (6.44)
\end{aligned}$$

by Lemma 5. Now note that $2r/(2r+1) > 1 - p/2$ for $r > r^*$, hence $R_{21} = O([\ln T/T]^{2r/(2r+1)})$.

To complete the proof we consider the term R_{22} given by (6.35). Since

$$|\hat{\theta}_{jk}-\theta_{jk}| \le \frac{|I_1(d_{jk})/I_0(d_{jk})-d_{jk}|}{1+A_{jT}(d_{jk})}+\frac{|d_{jk}-\theta_{jk}|}{1+A_{jT}(d_{jk})}+\frac{|\theta_{jk}|}{1+A_{jT}(d_{jk})}+|\theta_{jk}|$$
$$= O\Big(\frac{\nu_j}{T}+\frac{\nu_j^2|d_{jk}|T^{-1}}{1+A_{jT}(d_{jk})}+\frac{|d_{jk}-\theta_{jk}|}{1+A_{jT}(d_{jk})}+|\theta_{jk}|\Big)$$
$$= O\Big(|\theta_{jk}|+\frac{\nu_j}{T}+\frac{|d_{jk}-\theta_{jk}|}{1+A_{jT}(d_{jk})}\Big),$$

we partition R_{22} into R_{221}, R_{222} and R_{223}. Here, by Lemma 5 and (6.38),

$$R_{221} = O\left(\sum_{j=j_0+1}^{j_0^*}\sum_{k=0}^{2^j-1}\theta_{jk}^2 I(\sqrt{T}|\theta_{jk}|\le\sqrt{\ln T})\right) = O\Big([\ln T/T]^{\frac{2r}{2r+1}}\Big), \tag{6.45}$$

$$R_{222} = O\left(\sum_{j=j_0+1}^{j_0^*}\sum_{k=0}^{2^j-1}\nu_j^2/T^2\right) = \Big(T^{-2r/(2r+1)}\Big). \tag{6.46}$$

The third term, R_{223} we partition into R_{2231} and R_{2232} where

$$R_{2231} = O\left(\sum_{j=j_0+1}^{j_0^*}\sum_{k=0}^{2^j-1}E\Big[\frac{(d_{jk}-\theta_{jk})^2}{(1+A_{jT}(d_{jk}))^2}\Big]I(1\le\sqrt{T}|\theta_{jk}|\le\sqrt{\ln T})\right),$$
$$R_{2232} = O\left(\sum_{j=j_0+1}^{j_0^*}\sum_{k=0}^{2^j-1}E\Big[\frac{(d_{jk}-\theta_{jk})^2}{(1+A_{jT}(d_{jk}))^2}\Big]I(\sqrt{T}|\theta_{jk}|\le 1)\right). \tag{6.47}$$

By Lemma 5,

$$R_{2231} = O\left(\sum_{j=j_0+1}^{j_0^*}\sum_{k=0}^{2^j-1}T^{-1}I(T\theta_{jk}^2>1)I(T\theta_{jk}^2\le\ln T)\right)$$
$$= O\left(\Big[\frac{\ln T}{T}\Big]^{\frac{2r}{2r+1}}\right). \tag{6.48}$$

For an upper bound for R_{2232}, partition it into R_{22321} and R_{22322} depending on the value of $I(\sqrt{T}|d_{jk}-\theta_{jk}|>a\sigma\sqrt{2}\sqrt{\ln T})$. Note that

$$E\Big[(d_{jk}-\theta_{jk})^2 I(\sqrt{T}|d_{jk}-\theta_{jk}|>a\sigma\sqrt{2}\sqrt{\ln T})\Big]$$
$$= \frac{2\sigma^2}{T}\int_{a^2\ln T}^{\infty}e^{-z}\sqrt{z}dz \sim T^{-1-a^2}\sqrt{\ln T}$$

by 8.357 of Gradshtein and Ryzhik (1980). Therefore,

$$R_{22321} = O\left(\sum_{j=j_0+1}^{j_0^*}\sum_{k=0}^{2^j-1}T^{-1-a^2}\sqrt{\ln T}\right) = O\Big(T^{-2r/(2r+1)}\Big) \tag{6.49}$$

provided $a^2 > [(2r+1)(r+1/2-1/p)]^{-1}(1/p-1/2)$. The last term,

$$R_{22322} = O\Bigg(\sum_{j=j_0+1}^{j_0^*} \sum_{k=0}^{2^j-1} E\Big[\frac{(d_{jk}-\theta_{jk})^2}{(1+A_{jT}(d_{jk}))^2} I(\sqrt{T}|d_{jk}-\theta_{jk}| \le a\sigma\sqrt{2}\sqrt{\ln T}) \Big] I(\sqrt{T}|\theta_{jk}| \le 1) \Bigg)$$

$$= O\Bigg(\sum_{j=j_0+1}^{j_0^*} \sum_{k=0}^{2^j-1} E\Big[\frac{(d_{jk}-\theta_{jk})^2 \nu_j^2 e^{T d_{jk}^2/\sigma^2}}{\beta_j^2 T} I\Big(|d_{jk}-\theta_{jk}| \le \frac{a\sigma\sqrt{2\ln T}}{\sqrt{T}}\Big) \Big] I(\sqrt{T}|\theta_{jk}| \le 1) \Bigg). \tag{6.50}$$

Noting that in the last expression $\sqrt{T}|d_{jk}| \le a\sigma\sqrt{2\ln T}+1$, we find that the expectation in (6.50) is bounded by

$$e^{\frac{(a\sigma\sqrt{2\ln T}+1)^2}{\sigma^2}} \frac{\sigma^2}{T} \int_0^{a^2 \ln T} \sqrt{z} e^{-z} dz = O\Big(T^{2a^2-1} e^{2a\sqrt{2\ln T/\sigma}}\Big).$$

Hence,

$$R_{22322} = O\Bigg(\sum_{j=j_0+1}^{j_0^*} \frac{2^j \nu_j^2 T^{2a^2-1} e^{2a\sqrt{2\ln T/\sigma}}}{T\beta_j^2} \Bigg) = O\Big(T^{-2r/(2r+1)}\Big) \tag{6.51}$$

provided $\exp(a\sqrt{2\ln T}+\sigma^{-1})^2 \beta_j^{-2} = O(1)$, which is valid whenever $a^2 = (1/p - 1/2+\varepsilon)/[(r+1/2)(r+1/2-1/p)]$, by (3.14). Now to complete the proof combine (6.39), (6.45)–(6.49) and (6.51).

Proof of Corollary 1. It is easy to verify by direct calculations that in the case of the normal distribution, conditions (3.5)–(3.8) and (3.9) are valid.

Proof of Corollary 2. Validity of Corollary 2 follows from the fact that $\lambda = 0$ and conditions (3.7) and (3.8) hold, due to Lemma 1.

References

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

Anderson, T. W. (1997). Discussion on: Diggle, P. J. and Al Wasel, I. Spectral analysis of replicated biomedical time series. *Appl. Statist.* **46**, 65-66.

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

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

Choudhuri, N., Ghosal, S. and Roy, A. (2004). Bayesian estimation of the spectral density of a time series. *J. Amer. Statist. Assoc.* **99**, 1050-1059.

Chipman, H. A., Kolaczyk, E. D. and McCulloch, R. E. (1997). Adaptive Bayesian wavelet shrinkage. *J. Amer. Statist. Assoc.* **92**, 1413-1421.

Chow, Y. S. and Grenander, U. (1985). A sieve method for the spectral density. *Ann. Statist.* **13**, 998-1010.

Clyde, M., Parmigiani, G. and Vidakovic, B. (1998). Multiple shrinkage and subset selection in wavelets. *Biometrika* **85**, 391-401.

Cogburn, R. and Davis, H. T. (1974). Periodic splines and spectral estimation. *Ann. Statist.* **2**, 1108-1126.

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), 109-128. Frontières Ed.

Donoho, D. L., Johnstone, I. M., Kerkyacharian, G. and Picard, D. (1996). Density estimation by wavelet thresholding. *Ann. Statist.* **24**, 508-539.

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

Fan, J. and Kreutzberger, E. (1998). Automatic local smoothing for spectral density estimation. *Scand. J. Statist.* **25**, 359-369.

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.

Gao, H.-Y. (1993a). *Wavelet Estimation of Spectral Densities in Time Series analysis.* Ph.D. dissertation, University of California, Berkeley.

Gao, H.-Y. (1993b). Spectral density estimation via wavelet shrinkage. Tech. Rept., Statistics, Stanford.

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

Gradshtein, I. S. and Ryzhik, I. M. (1980). *Tables of Integrals, Series, and Products.* Academic Press, New York.

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

Huerta, G. and West, M. (1999). Bayesian inference on periodicities and component spectral structure in time series. *J. Time Ser. Anal.* **20**, 401-416.

Johnstone, I. M. (2002). Function estimation in Gaussian noise: sequence models. Manuscript. (available at http://www-stat.stanford.edu/~imj)

Johnstone, I. M. and Silverman, B. W. (2004). Needles and straw in haystacks: empirical Bayes estimates of possibly sparse sequences. *Ann. Statist.* **32**, 1594-1649.

Johnstone, I. M. and Silverman, B. W. (2005). Empirical Bayes selection of wavelet thresholds. To appear in *Ann. Statist.*

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

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.

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.) 285-290. Gif-sur-Yvette: Editions Frontières.

Mallat, S. G. (1999). *A Wavelet Tour of Signal Processing.* 2nd edition. Academic Press, San Diego.

Moulin, P. (1994). Wavelet Thresholding Techniques for Power Spectrum Estimation, *IEEE Trans. Signal Process.* **42**, 3126-3136.

Müller, P. and Vidakovic, B. (1999). MCMC methods in wavelet shrinkage: Non-equally spaced regression, density and spectral density estimation. In *Bayesian Inference in Wavelet Based Models*, (Eds. P. Müller and B. Vidakovic), 187-202. Lecture Notes in Statistics, **141** Springer-Verlag, New York.

Pawitan, Y. and O'Sullivan, F. (1994). Nonparametric spectral density estimation using Whittle's likelihood, *J. Amer. Statist. Assoc.* **98**, 600-610.

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

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

Riedel, K. S. and Sidorenko, A. (1995). Minimum biased multitaper spectral estimation. *IEEE Trans. Signal Process.* **43**, 188-195.

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

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

Vidakovic, B. (1998). Nonlinear wavelet shrinkage with Bayes rules and Bayes factors. *J. Amer. Statist. Assoc.*, 93, 173-179.

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

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

Wahba, G. (1980). Automatic smoothing of the log periodogram. *J. Amer. Statist. Assoc.* **75**, 122-132.

Welch, P. D. (1967). The use of Fast Fourier Transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Trans. Audio and Electroacoustics* **15**, 70-73.

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

Walden, A. T., Percival, D. B. and McCoy, E. J. (1998). Spectrum estimation by wavelet thresholding of multitaper estimators. *IEEE Trans. Signal Process.* **46**, 3153-3180.

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

Department of Mathematics, University of Central Florida, Orlando, FL 32816-1364, U.S.A.

E-mail: mpensky@pegasus.cc.ucf.edu

School of Biomedical Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0535, U.S.A.

E-mail: brani@bme.gatech.edu

I.A.C. "M. Picone", V.le del Policlinico 137, 00161 Rome, Italy.

E-mail: d.decanditiis@iac.cnr.it

(Received September 2003; accepted April 2006)