Sankhyā : The Indian Journal of Statistics
2006, Volume 68, Part 4, pp. 513-541

On Pointwise Optimality of Bayes Factor Wavelet Regression Estimators

Natalia Bochkina
Imperial College, London, UK
Theofanis Sapatinas
University of Cyprus, Cyprus

Abstract

We investigate the theoretical performance of Bayes factor estimators at a single point in wavelet regression models with independent and identically distributed errors that are not necessarily normally distributed. We compare these estimators in terms of their frequentist pointwise optimality in Besov spaces for certain combinations of error and prior distributions.

AMS (2000) *subject classification.* Primary 62G08; secondary 62C10, 62C12, 62C20.
Keywords and phrases. Bayesian inference, Besov spaces, nonparametric regression, optimality, pointwise risk, wavelets.

1 Introduction

Over the last decade, the nonparametric regression literature has been dominated by *nonlinear wavelet* methods. These methods are based on the idea of thresholding. This means that if an empirical wavelet coefficient is sufficiently large in magnitude, i.e., if its magnitude exceeds a predetermined threshold, then the corresponding term in the empirical wavelet expansion is retained (or shrunk towards zero); otherwise it is omitted. The resulting term-by-term wavelet thresholding estimators possess optimal or near-optimal convergence rates, and then are typically implemented through fast algorithms, which makes them very appealing in practice (see, e.g., Donoho and Johnstone, 1994, 1995, 1998; Donoho, Johnstone et al., 1995; Vidakovic, 1999; Abramovich et al., 2000; Antoniadis et al., 2001).

Various Bayesian and empirical Bayes approaches for term-by-term wavelet nonlinear shrinkage and wavelet thresholding estimators have been also proposed. (A *shrinkage* rule shrinks empirical wavelet coefficients to zero, whilst a *thresholding* rule shrinks empirical wavelet coefficients and sets all the coefficients below a certain level to zero.) These approaches impose a prior distribution on the wavelet coefficients of the unknown response function, which is designed to capture the sparseness of wavelet expansions common to most applications. The response function is then estimated by applying a suitable Bayes rule to the resulting posterior distribution of the wavelet coefficients. Different choices of loss function lead to different Bayes rules and hence to different, usually *level-dependent*, nonlinear wavelet shrinkage and wavelet thresholding rules (see, e.g., Chipman et al., 1997; Abramovich et al., 1998; Clyde et al., 1998; Vidakovic, 1998; Clyde and George, 2000; Angelini and Sapatinas, 2004; Angelini and Vidakovic, 2004).

However, until recently, their frequentist optimality properties (in the minimax sense) have not been studied. Abramovich et al. (2004) investigated optimality of posterior mean, posterior median and Bayes factor estimators in terms of the *global* L^2-loss function for the combination of normal error and normal prior distributions. Pensky (2006) and Pensky and Sapatinas (2007) studied optimality of posterior mean and Bayes factor estimators respectively with respect to the L^2-loss function for a wide variety of combination of error and prior distributions. Johnstone and Silverman (2005) explored adaptive optimality of *empirical* Bayes posterior mean and posterior median estimators with respect to a wide range of L^r-loss functions ($0 < r \leq 2$) for normal error and some heavy-tailed prior distributions. The adaptive optimality of an *empirical* Bayes procedure for the Bayes factor estimator with respect to the L^2-loss function for normal error and some heavy-tailed prior distributions was considered in Pensky and Sapatinas (2007). Recently, Abramovich et al. (2007) explored the optimality of posterior mean, posterior median and Bayes factor estimators in terms of the *pointwise* l^2-loss function for the combination of normal error and normal prior distributions. They showed that under the considered Bayesian hierarchical model, pointwise optimality is achieved up to a logarithmic factor.

This paper continues the line of investigation of Abramovich et al. (2007). However, our focus will be on investigating optimality of the Bayes factor estimator with respect to the l^2-loss function. The characteristic of this estimator is that it leads to a *hard* thresholding rule, unlike the posterior mean which leads to a nonlinear *shrinkage* rule and the posterior median which leads to a *soft* thresholding rule. Moreover, the Bayes factor estimator is much easier to evaluate in the majority of cases unlike posterior mean or posterior median estimators (see Bochkina and Sapatinas, 2005; Pensky, 2006). As in Pensky and Sapatinas (2007), who studied optimality of Bayes factor estimators with respect to the L^2-loss function, we put very mild restrictions on the errors in the standard nonparametric regression model. Furthermore, we do not assume the distribution of the errors to be known and hence consider a range of error and prior distributions for the wavelet coefficients. Moreover, as we demonstrate below, the use of a more flexible Bayesian hierarchical model improves the pointwise convergence rates and, under certain conditions, achieves pointwise optimality without the extra logarithmic factor that appeared in the results of Abramovich et al. (2007).

The paper is organized as follows. In Section 2, we introduce Bayesian models for the wavelet coefficients, extending the previously considered (in the context of pointwise optimality) normal error and normal prior model in Abramovich et al. (2007) to combinations of error and prior distributions having exponential descents. In Section 3, we provide the formulae for the threshold associated with the Bayes factor estimator for certain combinations of error and prior distributions. In Section 4, we discuss assumptions on the wavelet system, the error and prior distributions as well as their hyperparameters, and provide assertions about pointwise optimality of Bayes factor estimators in Besov spaces for certain combinations of error and prior distributions. Some concluding remarks are made in Section 5. Finally, in the Appendix, we provide some auxiliary statements and the proofs of the theoretical results stated in Section 4.

2 The Bayesian Model

Consider the following nonparametric regression model

$$Y_i = f(t_i) + Z_i, \quad i = 1, \dots, n, \tag{2.1}$$

where $t_i = i/n$, f is the unknown response function that is assumed to belong to the space of square integrable functions on $[0, 1]$, i.e., $f \in L^2[0, 1]$. The Z_i's are assumed to be independent and identically distributed (iid) random variables with $\mathbb{E}(Z_1) = 0$ and $\mathbb{V}(Z_1) = \sigma^2 < \infty$. We also assume that $\mathbb{E}(Z_1^4) < \infty$.

Then, any $f \in L^2[0,1]$ can be represented (in the L^2-sense) by a wavelet series, i.e.,

$$f(t) = \sum_{k \in K_{L-1}} \tilde{\theta}_k \phi_{Lk}(t) + \sum_{j=L}^{\infty} \sum_{k=0}^{2^j-1} \tilde{\theta}_{jk} \psi_{jk}(t),$$

where, for some (fixed) *primary resolution* level $L \geq 0$,

$$\begin{aligned} \phi_{Lk}(t) &= 2^{L/2}\phi(2^L t - k), \\ \psi_{jk}(t) &= 2^{j/2}\psi(2^j t - k), \\ \tilde{\theta}_k &= \int_{-\infty}^{+\infty} \phi_{Lk}(t) f(t) dt, \\ \text{and } \tilde{\theta}_{jk} &= \int_{-\infty}^{+\infty} \psi_{jk}(t) f(t) dt. \end{aligned}$$

Here, ϕ is the *scaling function*, ψ is a corresponding *wavelet function*, and K_{L-1} is the set of indices for which the scaling function ϕ_{Lk} is defined. (Note that, for the standard wavelet transform with boundary corrections, $K_{L-1} = \{0, 1, \ldots, 2^L-1\}$.) For suitable choices of ϕ and ψ, and appropriate boundary treatments, the corresponding set of ϕ_{Lk}'s and ψ_{jk}'s is an orthonormal set in $L^2[0,1]$ (see, e.g., Cohen et al., 1993; Johnstone and Silverman, 2004).

Application of the (boundary corrected) discrete wavelet transform (DWT) to (2.1) yields

$$\begin{aligned} \mathcal{U}_k &= u_k + \epsilon_k, \quad k \in K_{L-1}, \\ \mathcal{W}_{jk} &= w_{jk} + \varepsilon_{jk}, \quad j = L, L+1, \ldots, J-1, \ k = 0, 1, \ldots, 2^j - 1, \end{aligned}$$

where $J = \log_2(n)$ and ϵ_k, ε_{jk} are uncorrelated random variables due to the unitary property of the DWT. Denote $\theta_k = u_k/\sqrt{n}$ and $\theta_{jk} = w_{jk}/\sqrt{n}$ and recall that $\tilde{\theta}_k \approx \theta_k$ and $\tilde{\theta}_{jk} \approx \theta_{jk}$ (see, e.g., Vidakovic, 1999). In the Appendix, we provide a more detailed treatment of this relationship for the boundary coiflets $\{\phi, \psi\}$, a particular case of a wavelet system used to establish the pointwise optimality results given in subsequent sections (see Lemma A.4). In this case, there will be $2^L - 2(S-s-1)$ scaling coefficients at the primary resolution level L, with $K_{L-1} = \{0, 1, \ldots, s-1, S-1, S, \ldots, 2^L - S, 2^L - s, 2^L - s + 1, \ldots, 2^L - 1\}$ (see Johnstone and Silverman, 2004, p. 83).

We use the Bayesian framework to construct estimators $\hat{\theta}_k$ of θ_k (based on $\mathcal{U}_k$) and $\hat{\theta}_{jk}$ of θ_{jk} (based on $\mathcal{W}_{jk}$) in order to estimate the unknown response function f. Since the wavelet representations of a vast majority of functions contain only a few non-negligible wavelet coefficients in their expansions, similar to the priors used previously in the Bayesian wavelet regression literature, we place the following prior on the wavelet coefficient w_{jk}

$$w_{jk} \sim \pi_{j,n}\tau_{j,n}h(\tau_{j,n}\cdot) + (1-\pi_{j,n})\delta(0), \quad j = L, L+1, \ldots, \; k = 0, 1, \ldots, 2^j - 1, \tag{2.2}$$

where $0 \le \pi_{j,n} \le 1$ for $L \le j \le J-1$ and $\pi_{j,n} = 0$ for $j \ge J$, $\tau_{j,n} > 0$, $\delta(0)$ is a point mass at zero, and w_{jk} are independent random variables. For the prior model h, we consider not only the standard normal probability density function (pdf) but also the double-exponential pdf with scale parameter 1. To complete the prior specification of f, we place noninformative priors (e.g., the uniform density on $\mathbb{R} = (-\infty, \infty)$) on the scaling coefficients u_k, $k \in K_{L-1}$.

According to the prior model (2.2), w_{jk} is either zero with probability $1 - \pi_{j,n}$, or with probability $\pi_{j,n}$ is distributed with the pdf h with scale parameter $\tau_{j,n}$. The proportion $\pi_{j,n}$ indicates whether a value is small or large and can be used to 'control' the trade-off between sparse and dense sequences. In what follows, we impose all conditions on the prior odds ratio

$$\beta_{j,n} = (1 - \pi_{j,n})/\pi_{j,n}.$$

Note that we allow dependence of $\pi_{j,n}$ (and hence of $\beta_{j,n}$) not only on the resolution level j but also on n. It is most natural since the proportion of wavelet coefficients we are intending to keep depends not only on the function f itself but also on the amount of data available. When n is large, the estimators of wavelet coefficients become more reliable, and hence, smaller wavelet coefficients can be distinguished from pure noise. Consequently, for larger n, one can keep larger number of wavelet coefficients at a particular resolution level j, which leads to a larger value of $\pi_{j,n}$.

Consider the distribution of the errors ε_{jk}'s. It follows from (2.1) that

$$\varepsilon_{jk} \approx n^{-1/2}2^{j/2} \sum_{i=1}^{n} \psi(2^j i/n - k)Z_i.$$

If the Z_i's are iid random variables with $\mathbb{E}(Z_1^4) < \infty$, it is not difficult to see that the sequence $\{n^{-1/2}2^{j/2}\psi(2^j i/n - k)Z_i\}$ satisfies the Lyapunov condition (see, e.g., Billingsley, 1995, p. 362) provided that $2^j/n \to 0$ as $n \to \infty$. Hence, if the resolution level is reasonably small ($j \le J_0$, where $J - J_0 \to \infty$ as $n \to \infty$), the errors ε_{jk} are asymptotically $N(0, \sigma^2)$ distributed and thus asymptotically independent. For a more detailed treatment of asymptotic normality, the interested reader is referred to, e.g., Neumann and von Sachs (1995).

We assume that the distribution of the errors ε_{jk} is level-dependent

$$\varepsilon_{jk} \sim \varphi_j(\cdot), \quad L \le j \le J-1,$$

with the pdf φ_j having exponential descents, i.e.,

$$\varphi_j(x) = c_j \exp\{-(|x|/\sigma_j)^\beta\}, \quad 0 < \underline{\sigma} \le \sigma_j \le \bar{\sigma} < \infty, \quad c_j > 0, \quad \beta > 0. \tag{2.3}$$

(For the distribution of errors of the scaling coefficients, ϵ_k, we only assume that it has a finite variance σ^2_{L-1}.) As we shall show later, one does not need the knowledge of the true distribution of the errors ε_{jk}, and can achieve pointwise optimality with the choice (2.3) with either β=2 (normal) or β=1 (double-exponential). To keep the exposition simple, we do not consider any heavier-tailed pdf's (e.g., Student-t distributions) for both φ_j and h. In Section 3, we provide some further explanation about the considered choice of error and prior distributions, the choice being combinations of the commonly used distributions of exponential descents, namely normal and double-exponential.

In what follows, we conduct Bayesian inference for each wavelet coefficient separately. Denote

$$d_{jk} = \mathcal{W}_{jk}/\sqrt{n} \quad \text{and} \quad \nu_j = \sqrt{n}\tau_{j,n}. \tag{2.4}$$

Taking into account the relation between w_{jk} and θ_{jk} and (2.2)–(2.4), we derive that the posterior pdf of θ_{jk} given d_{jk} is of the form

$$p(\theta_{jk} \mid d_{jk}) = \frac{\sqrt{n}\,\varphi_j(\sqrt{n}(\theta_{jk} - d_{jk}))\ \nu_j h(\nu_j \theta_{jk}) + \beta_{j,n}\sqrt{n}\,\varphi_j(\sqrt{n} d_{jk})\delta(0)}{\int_{-\infty}^{+\infty} \sqrt{n}\,\varphi_j(\sqrt{n}(x - d_{jk}))\nu_j h(\nu_j x)dx + \beta_{j,n}\sqrt{n}\,\varphi_j(\sqrt{n} d_{jk})}.$$

The Bayes factor estimator of θ_{jk} is derived as follows (see Vidakovic, 1998): after observing d_{jk}, we test the hypothesis

$$H_0 : \theta_{jk} = 0 \quad \text{versus} \quad H_1 : \theta_{jk} \neq 0.$$

If the hypothesis H_0 is rejected, θ_{jk} is estimated by d_{jk}, otherwise $\theta_{jk} = 0$, so that the estimator $\hat{\theta}_{jk}$ is given by

$$\hat{\theta}_{jk} = d_{jk}\, \mathbb{I}\left(\frac{\mathbb{P}(H_1 \mid d_{jk})}{\mathbb{P}(H_0 \mid d_{jk})} > 1\right),$$

where $\mathbb{I}(A)$ denotes the indicator function of set A. Observe that the posterior odds ratio can be rewritten as

$$\frac{\mathbb{P}(H_1 \mid d_{jk})}{\mathbb{P}(H_0 \mid d_{jk})} = \frac{\zeta_{j,n}(d_{jk})}{\beta_{j,n}},$$

where

$$\zeta_{j,n}(d_{jk}) = I_j(d_{jk}) / \left[\sqrt{n}\,\varphi_j(\sqrt{n} d_{jk})\right] \tag{2.5}$$

and

$$I_j(d_{jk}) = \int_{-\infty}^{+\infty} \sqrt{n}\,\varphi_j[\sqrt{n}(x - d_{jk})]\ \nu_j h(\nu_j x) dx. \tag{2.6}$$

Rewriting $\hat{\theta}_{jk}$ in view of (2.5), we obtain

$$\hat{\theta}_{jk} = d_{jk}\ \mathbb{I}(\zeta_{j,n}(d_{jk}) > \beta_{j,n}). \tag{2.7}$$

It is easy to check that $\zeta_{j,n}(d_{jk})$'s are even functions of d_{jk}. If, moreover, the functions $\zeta_{j,n}(d_{jk})$'s are strictly increasing in d_{jk} for $d_{jk} > 0$, then

$$\zeta_{j,n}(d_{jk}) > \beta_{j,n} \quad \text{if and only if} \quad |d_{jk}| > t_{j,n} = \zeta_{j,n}^{-1}(\beta_{j,n}).$$

Hence, (2.7) is a hard thresholding rule with the threshold $t_{j,n}$, i.e.,

$$\hat{\theta}_{jk} = d_{jk}\ \mathbb{I}(|d_{jk}| > t_{j,n}). \tag{2.8}$$

Indeed, in the majority of practical cases, it is true that (2.7) gives rise to a hard thresholding rule as confirmed by the following statement.

PROPOSITION 2.1. *If φ_j is normal or double-exponential pdf and h is a symmetric at zero pdf on $\mathbb{R}$, then $\zeta_{j,n}(d_{jk})$ is strictly increasing in d_{jk} for $d_{jk} > 0$.*

Note that under the considered error model, the noninformative priors for the scaling coefficients u_k's result in their posterior distributions being proper and their estimates being the corresponding empirical scaling coefficients $\mathcal{U}_k$. Thus $\hat{\theta}_k = \mathcal{U}_k/\sqrt{n}$, $k \in K_{L-1}$. Since we assumed that $\pi_{j,n} = 0$ if $j \geq J$, $k = 0, 1, \ldots, 2^j - 1$, it implies that $\hat{\theta}_{jk} = 0$ as $j \geq J$, $k = 0, 1, \ldots, 2^j - 1$, and therefore the estimator $\hat{f}$ of f is of the form

$$\hat{f}(t) = \sum_{k \in K_{L-1}} \hat{\theta}_k \phi_{Lk}(t) + \sum_{j=L}^{J-1} \sum_{k=0}^{2^j-1} \hat{\theta}_{jk} \psi_{jk}(t). \tag{2.9}$$

Coefficients $\hat{\theta}_{jk}$ are found using formula (2.7), where the function $\zeta_{j,n}(d_{jk})$ is defined by (2.5) and (2.6).

In the following section, we give the explicit formulae for the thresholds of the proposed Bayes factor estimators for the combinations of error and prior distributions considered in this paper.

3 Thresholds

In practice, it may be useful to have the formulae for the threshold t_{jn} of the Bayes factor estimator (2.8), particularly if an empirical Bayes approach is employed. These formulae may also be useful to the study of theoretical properties of the Bayes factor estimator. In this section, we give the formulae for the threshold $t_{j,n}$ under the specified choices of pdf's φ_j and h.

1. (Normal φ_j, h) In this case, the threshold $t_{j,n}$ is given by

$$t_{j,n} = \frac{\sigma_j \sqrt{2}}{\sqrt{n}} \sqrt{1 + \frac{\sigma_j^2 \nu_j^2}{n}} \left[\log \left(\frac{\beta_{j,n} \sqrt{n + \nu_j^2 \sigma_j^2}}{\nu_j \sigma_j} \right) \right]^{1/2} \times \mathbb{I} \left\{ \frac{\beta_{j,n} \sqrt{n + \nu_j^2 \sigma_j^2}}{\nu_j \sigma_j} > 1 \right\}$$

(see Abramovich et al., 2004).

2. (Normal φ_j, double-exponential h) In this case, the threshold $t_{j,n}$ is the non-negative solution of the equation

$$\frac{\Phi\left(-\frac{\sigma_j}{\sqrt{n}}(t_{j,n} + \xi_{j,n})\right)}{\varphi\left(-\frac{\sigma_j}{\sqrt{n}}(t_{j,n} + \xi_{j,n})\right)} + \frac{\Phi\left(\frac{\sigma_j}{\sqrt{n}}(t_{j,n} - \xi_{j,n})\right)}{\varphi\left(\frac{\sigma_j}{\sqrt{n}}(t_{j,n} - \xi_{j,n})\right)} = \frac{2\beta_{j,n}\sqrt{n}}{\nu_j \sigma_j},$$

where $\xi_{j,n} = \frac{\nu_j \sigma_j^2}{n}$, and $\Phi(x)$ and $\varphi(x)$ are respectively the distribution function and pdf of the standard normal distribution. If the equation has only a negative solution, then $t_{j,n} = 0$.

3. (Double-exponential φ_j, h) In this case, the threshold $t_{j,n}$ is given by

$$t_{j,n} = \max \left[0, \frac{\log\left(1 + \beta_{j,n}(q_{j,n}^2 - 1)\right) - \log(q_{j,n})}{\nu_j (q_{j,n} - 1)} \right],$$

where $q_{j,n} = \frac{\sqrt{2n}}{\sigma_j \nu_j}$.

The last two formulae follow from the formula for the posterior odds ratio in the considered cases given in Theorem 3.2 and Theorem 3.4 respectively in Bochkina and Sapatinas (2005) (see Section A.4).

4 Pointwise Optimality of Bayes Factor Estimators

To complete the construction of the estimator $\hat{f}$ given in (2.9), we need to specify the error model φ_j, the prior model h, and to choose the values of the parameters ν_j and $\beta_{j,n}$ so that this estimator achieves the optimal pointwise convergence rate over a variety of Besov spaces. This is the objective of this section.

4.1. Assumptions. Now we formulate conditions on the wavelet system $\{\phi, \psi\}$ and the pdf's h and φ_j as well as on the parameters ν_j and $\beta_{j,n}$.

(S1) Let ϕ and ψ be the boundary coiflets introduced in Johnstone and Silverman (2004) possessing s continuous derivatives and $s-1$ vanishing moments ($s \geq 2$), and based on orthonormal coiflets supported in $[-S+1, S]$, $S > s$. Let also $L \geq \log_2(6S-6)$.

The pdf's φ_j and h considered here are defined on $\mathbb{R}$, symmetric at zero, positive-valued and unimodal, and have finite (uniformly bounded over j in the case of φ_j) moments of every (polynomial) order. In addition, we consider only those combinations of pdf's which satisfy the following condition

$$\text{(A1)} \qquad |\varphi_j(x)/h(x)| \leq C_{h,\varphi}.$$

Note that the constant $C_{h,\varphi}$ is assumed to be independent of j, which requires some kind of uniformity for the pdf's φ_j. The consequence of this restriction is that the asymptotic expressions for the thresholds $t_{j,n}$'s will depend on the resolution level j rather than on the particular form of φ_j.

In the subsequent development, we consider the following combinations of error φ_j and prior h

$$\text{normal } \varphi_j \; - \; \text{normal } h, \tag{4.1}$$

$$\text{normal } \varphi_j \; - \; \text{double-exponential } h, \tag{4.2}$$

$$\text{double-exponential } \varphi_j \; - \; \text{double-exponential } h. \tag{4.3}$$

We do not consider the case double-exponential φ_j – normal h, since assumption (A1) does not hold in this case.

Denote

$$j_1 \;=\; \frac{1}{2(r-1/p+1/2)} \log_2(n). \tag{4.4}$$

We assume that the parameter ν_j is of the form

$$\nu_j = C_\nu\, 2^{jm_{(j)}}, \quad \text{where} \quad m_{(j)} = \begin{cases} m_1, & L \le j \le j_1, \\ m_2, & j_1 < j \le J-1, \end{cases} \tag{4.5}$$

and choose $\beta_{j,n}$ such that

$$\beta_{j,n} = (\nu_j/\sqrt{n})^{a_{(j)}}, \quad \text{where} \quad a_{(j)} = \begin{cases} a_1, & L \le j \le j_1, \\ a_2, & j_1 < j \le J-1. \end{cases} \tag{4.6}$$

(Note that we allow both hyperparameters $m_{(j)}$ and $a_{(j)}$ to vary with resolution level j.) We refer to $L \le j \le j_1$ and $j_1 < j \le J-1$ as *low* and *high* resolution levels, respectively.

The Bayesian model considered here does not include the Bayesian model of Abramovich et al. (2007) as a particular case. This is because of the fact that in our model, the parameter $\pi_{j,n}$ depends simultaneously on the sample size n and the resolution level j: if $\pi_{j,n}$ is independent of n (i.e., $a_{(j)} = 0$ for all j), then it is also independent of j, see definition (4.6). However, unlike Abramovich et al. (2007), we allow for different behaviour of the hyperparameters at low and high resolution levels. As we shall see below, under some restrictions on our model, the considered Bayes factor estimators achieve pointwise optimality without the extra logarithmic factor that appeared in the results of Abramovich et al. (2007).

The choices of error and prior models given in (4.1), (4.2) and (4.3) are motivated by the repeated use of these distributions in some practical applications, as well as the asymptotic behaviour of the risk when the pointwise convergence rate is not optimal. For example, as shown in Pensky and Sapatinas (2007), who studied convergence rates of Bayes factor estimators with respect to the L^2-loss function, for distributions having exponential descent the deviation from the optimal behaviour is a factor which grows as a power of the logarithm of the sample size, whereas in the case of the distributions having polynomial descent the deviation is much larger, with a factor that is a power of the sample size. Note also that, when the prior model h has faster descent at $\pm\infty$ than φ_j (i.e., when the assumption (A1) does not hold), sub-optimal convergence rates arise with respect to the L^2-loss function due to the slow convergence of the bias when, e.g., the posterior mean is used as an estimator (see Pensky, 2006). Since we expect to see these types of behaviour for the convergence rates of Bayes factor estimators with respect to the l^2-loss function, in what follows, we restrict ourselves to study the pointwise optimality of Bayes factor estimators *only* for the combination of error and prior models given in (4.1), (4.2) and (4.3).

4.2. Optimal pointwise convergence rate over Besov spaces. Let $r > 0$ and $1 \le p, q \le \infty$. A function f belongs to Besov ball $B^r_{p,q}(A)$ of radius $A > 0$ if and only if its scaling coefficients, $\tilde{\theta}_k$, and wavelet coefficients, $\tilde{\theta}_{jk}$, satisfy the following condition

$$\left(\sum_{k \in K_{L-1}} |\tilde{\theta}_k|^p \right)^{1/p} + \left(\sum_{j=L}^{\infty} 2^{j(r+1/2-1/p)q} \left[\sum_{k=0}^{2^j-1} |\tilde{\theta}_{jk}|^p \right]^{q/p} \right)^{1/q} \le A, \quad (4.7)$$

with respective sum(s) replaced by maximum if $p = \infty$ or $q = \infty$.

For any possible estimator $\tilde{f}$ of f based on n observations from model (2.1), define the maximal pointwise risk, with respect to the l^2-loss function, over a function space $\mathcal{F}$ defined on the unit interval $[0, 1]$, as

$$R_n(t_0, \mathcal{F}, \tilde{f}) = \sup_{f \in \mathcal{F}} \mathbb{E} \left(\tilde{f}(t_0) - f(t_0) \right)^2$$

for any fixed $t_0 \in (0, 1)$. Using the convexity of a Besov ball for $1 \le p, q \le \infty$, Lemma 3 in Donoho and Low (1992), and the optimal pointwise convergence rates obtained by Cai (1993) in the Gaussian white noise model, it easily follows that when the Z_i's in the model (2.1) are iid normal random variables with $\mathbb{E}(Z_1) = 0$ and $\mathbb{V}(Z_1) = \sigma^2 < \infty$, and when f belongs to a ball $B^r_{p,q}(A)$ of radius $A > 0$ in the Besov space $B^r_{p,q}[0, 1]$, then, provided that $r > 1/p$ and $1 \le p, q \le \infty$,

$$\inf_{\tilde{f}} R_n(t_0, B^r_{p,q}(A), \tilde{f}) \asymp n^{-\frac{2(r-1/p)}{2(r-1/p)+1}} \quad \text{as } n \to \infty, \quad (4.8)$$

where the infimum is taken over all estimators $\tilde{f}$ of f. (Here, we write $g_1(n) \asymp g_2(n)$ to denote $0 < \text{liminf}(g_1(n)/g_2(n)) \le \text{limsup}(g_1(n)/g_2(n)) < \infty$ as $n \to \infty$.)

Unlike the global maximal risk with respect to the L^2-loss function (see Donoho and Johnstone, 1998), the pointwise maximal risk with respect to the l^2-loss function depends not only on the smoothness index r, but also on the parameter p. Moreover, it converges at a rate slower than the corresponding global rate.

Since for the majority of resolution levels ($j \leq J_0$ where $J - J_0 \to \infty$ as $n \to \infty$) the errors ε_{jk}'s asymptotically follow the normal distribution, we expect that the pointwise convergence rate over these levels for an arbitrary distribution of errors, satisfying the assumptions stated in Section 2, is not faster than the optimal pointwise convergence rate under the normal errors. As we shall see from the proof of Theorem 4.1, the pointwise convergence rate over the resolution levels $j \leq j_1$ cannot be slower than the rate (4.8) as $n \to \infty$ (since we can take $J_0 > j_1$). Therefore, we can expect that the pointwise convergence rate for the considered choices of the error distributions φ_j to be at least not slower than the pointwise convergence rate (4.8) as $n \to \infty$, which is optimal for normal errors.

4.3. Pointwise optimality of Bayes factor estimators in Besov spaces. The following theorem states under which conditions the considered Bayes factor estimators achieve the optimal pointwise convergence rate under the l^2-loss function.

THEOREM 4.1. *Let the assumptions in Section 4.1 hold, and let* $f \in B^r_{p,q}(A)$ *with* $1 \leq p, q \leq \infty$ *and* $1/p < r < s$. *Assume that the following restrictions hold for* m_1 *and* m_2

$$m_1 < r - 1/p + 1/2, \quad m_2 > r - 1/p + 1/2, \tag{4.9}$$

and that the following restrictions hold for a_1, a_2 *and* φ_j.

(1) $a_1 \geq 1$;

(2) if φ_j *is double-exponential, we have* $a_2 > 0$;

(3) if φ_j *is normal, we have* $a_2 > \dfrac{2(r-1/p)}{(r-1/p+1/2)(2m_2-1)}$.

Then, for any $t_0 \in (0, 1)$,

$$R_n(t_0, B^r_{p,q}(A), \hat{f}) = O\left(n^{-\frac{2(r-1/p)}{2(r-1/p)+1}}\right) \quad \text{as} \quad n \to \infty.$$

REMARK 4.1. The assumption on the hyperparameter a_2 for the double-exponential φ_j is weaker than the corresponding one for the normal φ_j.

Remark 4.2. For values of a_1 or a_2 violating the assumptions of Theorem 4.1, the pointwise rate of convergence is no longer the exactly optimal rate. Following the arguments in the proof of Theorem 4.1, one can show that, for any $t_0 \in (0,1)$,

$$R_n(t_0, B^r_{p,q}(A), \tilde{f}) = O\left(n^{-\frac{2(r-1/p)}{2(r-1/p)+1}} (\log n)^{\Delta}\right) \quad \text{as } n \to \infty,$$

with (i) $\Delta = 1/2$ if $a_1 \geq 1$ and $a_2 = \frac{2(r-1/p)}{(r-1/p+1/2)(2m_2-1)}$,

(ii) $\Delta = 1$ if $a_1 < 1$ and φ_j is normal, and

(iii) $\Delta = 2$ if $a_1 < 1$ and φ_j is double-exponential.

As it is evident from Theorem 4.1, under some restrictions on our model, the considered Bayes factor estimators achieve pointwise optimality without the extra logarithmic factor that appeared in the results of Abramovich et al. (2007). This result is due to the flexibility of our model with respect to allowing the hyperparameters $a_{(j)}$ and $m_{(j)}$ to be different for low and high resolution levels, and the dependence of the prior odds $\beta_{j,n}$ on the sample size n. As one can see from the proof of Theorem 4.1 (see Appendix), the most crucial assumption which allows to achieve pointwise optimality without a logarithmic factor is the separation between low and high resolution levels at the "boundary" level j_1 defined by (4.4).

Remark 4.3. To make the prior model more flexible, we can divide the low resolution levels into low $\{L, \ldots, j_0\}$ and medium $\{j_0+1, \ldots, j_1\}$ levels, with $j_0 = \kappa \log_2(n)$ for arbitrary $\kappa \in (0, 1/[2(r+1/2-1/p)])$, and consider different values of hyperparameters (a_0, m_0) and (a_1, m_1). Then, to satisfy Theorem 4.1, hyperparameters (a_1, m_1) should yield the assumptions of the theorem, and the only additional restriction on the hyperparameters for the low resolution levels is $m_0 < (2\kappa)^{-1}$. Thus, the restrictions on a_1 and m_1 are crucial only for the levels adjacent to the "boundary" level j_1 rather than for all resolution levels coarser than j_1.

5 Conclusions

We investigated the theoretical performance of Bayes factor estimators at a single point in wavelet regression models with independent and identically distributed errors that are not necessarily normally distributed. We compared these estimators in terms of their frequentist pointwise optimality (in the minimax sense) in Besov spaces for some combinations of error and prior distributions. The characteristic of the Bayes factor estimator is that it leads to a *hard* thresholding rule, unlike the recently studied posterior mean and posterior median estimators, which lead to nonlinear *shrinkage* and *soft* thresholding rules respectively. Moreover, the Bayes factor estimator is much easier to evaluate in the majority of cases unlike posterior mean or posterior median estimators.

We extended the normality assumption about the distribution of errors in the standard nonparametric regression model, to include the double-exponential distribution of errors in the wavelet domain. Furthermore, we showed that optimality can be achieved for different error distributions implying that it is not necessary to know the exact error distribution to achieve pointwise optimality. Moreover, as we demonstrated, the use of a more flexible Bayesian hierarchical model, under certain conditions, achieved pointwise optimality without the extra logarithmic factor that appeared in the results of Abramovich et al. (2007).

We conclude this section with some comments on adaptation. Adaptive estimation has become an important part of nonparametric function estimation problems. Adaptation to unknown smoothness is essential because the smoothness parameters of the underlying functions are unknown in virtually all practical situations. In the Gaussian white noise model, Cai (2003) considered adaptation under pointwise risk over Besov spaces; sharp lower bounds on the cost of adaptation were obtained (the minimum cost for adaptation is at least a logarithmic factor) and are shown to be attainable by a (soft-thresholding) wavelet estimator. We have not considered adaptation in our nonparametric regression setup, i.e., to construct a Bayes factor estimator without the knowledge of the parameters of the Besov ball, attaining the adaptive optimal pointwise convergence rate. This is beyond the scope of this article but presents avenues for further research that hopefully will be addressed in the future.

Appendix

Throughout the proof of Theorem 4.1, we use C to denote a generic positive constant, not necessarily the same each time it is used, even within a single equation. (Auxiliary results with proofs are given in the following sections.)

A.1. Proof of Theorem 4.1. Since the wavelet basis is orthonormal, for any fixed $t_0 \in (0,1)$,

$$\begin{aligned} R_n(t_0, B^r_{p,q}(A), \hat{f}) &= \mathbb{E}[\hat{f}(t_0) - f(t_0)]^2 \\ &= \mathbb{E}\left[\sum_{k\in K_{L-1}} (\hat{\theta}_k - \tilde{\theta}_k)\phi_{Lk}(t_0) + \sum_{j=L}^{\infty}\sum_{k=0}^{2^j-1}(\hat{\theta}_{jk} - \tilde{\theta}_{jk})\psi_{jk}(t_0)\right]^2. \end{aligned}$$

Now, we can decompose the risk above into the following terms

$$\begin{aligned} R_n(t_0, B^r_{p,q}(A), \hat{f}) = \mathbb{E}\Bigg[&\sum_{k\in K_{L-1}} (\hat{\theta}_k - \theta_k)\phi_{Lk}(t_0) + \sum_{k\in K_{L-1}} (\theta_k - \tilde{\theta}_k)\phi_{Lk}(t_0) \\ &+ \sum_{j=L}^{J-1}\sum_{k=0}^{2^j-1}(\hat{\theta}_{jk} - \theta_{jk})\psi_{jk}(t_0) + \sum_{j=L}^{J-1}\sum_{k=0}^{2^j-1}(\theta_{jk} - \tilde{\theta}_{jk})\psi_{jk}(t_0) \\ &+ \sum_{j=J}^{\infty}\sum_{k=0}^{2^j-1}\tilde{\theta}_{jk}\psi_{jk}(t_0)\Bigg]^2. \end{aligned}$$

Now, we can apply the elementary inequality

$$\mathbb{E}\left(\sum_{i=1}^{n} X_i\right)^2 \le \left[\sum_{i=1}^{n}(\mathbb{E}|X_i|^2)^{1/2}\right]^2$$

to bound the risk

$$\begin{aligned} &R_n(t_0, B^r_{p,q}(A), \hat{f}) \\ \le\ &\Bigg[\sum_{k\in K_{L-1}(t_0)} 2^{L/2}[\mathbb{E}(\hat{\theta}_k - \theta_k)^2]^{1/2}||\phi||_\infty + \sum_{k\in K_{L-1}(t_0)} 2^{L/2}|\tilde{\theta}_k - \theta_k|\ ||\phi||_\infty \\ &+ \sum_{j=L}^{J-1}\sum_{k\in K_j(t_0)} 2^{j/2}[\mathbb{E}(\hat{\theta}_{jk} - \theta_{jk})^2]^{1/2}||\psi||_\infty \\ &+ \sum_{j=L}^{J-1}\sum_{k\in K_j(t_0)} 2^{j/2}|\tilde{\theta}_{jk} - \theta_{jk}|\ ||\psi||_\infty + \sum_{j=J}^{\infty}\sum_{k\in K_j(t_0)} 2^{j/2}|\tilde{\theta}_{jk}|\ ||\psi||_\infty\Bigg]^2 \\ =\ &[Q_{11} + Q_{12} + Q_{21} + Q_{22} + Q_3]^2, \end{aligned} \tag{A.1}$$

where, for any function g, $||g||_\infty = \sup_x |g(x)|$ and $K_j(t_0) = \{k : 0 \le k \le 2^j - 1 \text{ and } \psi_{jk}(t_0) \ne 0\}$, for $j \ge J$, with $K_{L-1}(t_0) = \{k : k \in$

K_{L-1} and $\phi_{Lk}(t_0) \neq 0\}$. For the boundary coiflets stated in assumption (S1), the cardinality of $K_j(t_0)$ is less than or equal to $2S-1$, $j \geq L-1$, which is independent of j (see Johnstone and Silverman, 2004). Note also that, by construction, ϕ and ψ are bounded functions, i.e., $\phi,\ \psi \in L^\infty[0,1]$.

The term $Q_{11} + Q_{12}$ in (A.1) is bounded by

$$\begin{aligned} C \sum_{k\in K_{L-1}} 2^{L/2}[\mathbb{V}(\hat{\theta}_k)]^{1/2} \ + \ C \sum_{k\in K_{L-1}} 2^{L/2}|\tilde{\theta}_k - \theta_k| & \\ \leq \ & Cn^{-1/2}\sigma_{L-1} + Cn^{-r} \\ = \ & O\left(n^{-1/2}\right) + o\left(n^{-(r-1/p)}\right) = o\left(n^{-\frac{(r-1/p)}{2(r-1/p)+1}}\right), \end{aligned}$$

due to (A.15) and the fact that $\mathbb{V}(\hat{\theta}_k) = O(n^{-1})$.

On the other hand, the term Q_3 in (A.1) is bounded by

$$\begin{aligned} C\sum_{j=J}^{\infty} \sum_{k\in K_j(t_0)} 2^{j/2}|\tilde{\theta}_{jk}| \ & \leq \ C\sum_{j=J}^{\infty} 2^{j/2}2^{-j(r-1/p+1/2)} = O\left(2^{-J(r-1/p)}\right) \\ & = \ O\left(n^{-(r-1/p)}\right) = o\left(n^{-\frac{(r-1/p)}{2(r-1/p)+1}}\right), \end{aligned}$$

due to (A.17). By Lemma A.4, the term Q_{22} in (A.1) is dominated by $C\,n^{-(r-1/p)}$. Therefore, now we need to evaluate the contribution to $R_n(t_0, B^r_{p,q}(A), \hat{f})$ made by term Q_{21} in (A.1)

$$Q_{21} = ||\psi||_\infty \sum_{j=L}^{J-1} \sum_{k\in K_j(t_0)} 2^{j/2}[\mathbb{E}(\hat{\theta}_{jk} - \theta_{jk})^2]^{1/2} =: ||\psi||_\infty (R_1 + R_2), \quad \text{(A.2)}$$

with terms

$$\begin{aligned} R_1 \ &= \ \sum_{j=L}^{j_1} \sum_{k\in K_j(t_0)} 2^{j/2}[\mathbb{E}(\hat{\theta}_{jk} - \theta_{jk})^2]^{1/2}, \qquad \text{(A.3)} \\ R_2 \ &= \ \sum_{j=j_1+1}^{J-1} \sum_{k\in K_j(t_0)} 2^{j/2}[\mathbb{E}(\hat{\theta}_{jk} - \theta_{jk})^2]^{1/2}, \end{aligned}$$

corresponding to *low* and *high* resolution levels, respectively. Let us now construct an asymptotic upper bound for each of the terms.

Low resolution levels. We can use Lemma A.3 to bound R_1 from above

$$R_1 \leq \sqrt{2} \sum_{j=L}^{j_1} \sum_{k \in K_j(t_0)} 2^{j/2} \min(t_{j,n}, |\theta_{jk}|) + O(2^{j_1/2} n^{-1/2}), \quad \text{(A.4)}$$

since $\kappa_{2,j} = \int_{-\infty}^{+\infty} x^2 \varphi_j(x) dx = c_j \sigma_j^3 \int_{-\infty}^{+\infty} z^2 e^{-|z|^\beta} dz \leq C$, due to σ_j being bounded (see (2.3)). The last term achieves the optimal pointwise convergence rate: $n^{-1/2} 2^{j_1/2} = n^{-(r-1/p)/(2(r-1/p)+1)}$, so we need to study the behaviour of the first term.

To bound $t_{j,n}$ from above, we apply the last statement of Lemma A.2. To use Lemma A.2, we need to check the assumption that $\nu_j/\sqrt{n} \to 0$, as $n \to \infty$. Note that,

$$\nu_j/\sqrt{n} = C_\nu 2^{j m_1} n^{-1/2} \leq C_\nu 2^{j_1 m_1} n^{-1/2} = C_\nu n^{(m_1/(r-1/p+1/2)-1)/2} \to 0$$

as $n \to \infty$, since $j \leq j_1$ and, according to assumption (4.9), $m_1/(r + 1/2 - 1/p) - 1 < 0$. Therefore, the assumption of the last statement of Lemma A.2 is satisfied for the low resolution levels.

According to Lemma A.2, for the low resolution levels with $a_1 \geq 1$, the threshold $t_{j,n}$ is bounded by $Cn^{-1/2}$, therefore the first term of R_1 is bounded by

$$\begin{aligned}
&\sqrt{2} \sum_{j=L}^{j_1} \sum_{k \in K_j(t_0)} 2^{j/2} \min(t_{j,n}, |\theta_{jk}|) \\
&\quad \leq C \sum_{j=L}^{j_1} \min\left(2^{j/2} n^{-1/2},\, 2^{j/2} 2^{-j(r+1/2-1/p)}\right) \\
&\quad = O\left(\min\left(n^{-1/2} 2^{j_1/2},\, 2^{-L(r-1/p)}\right)\right) \\
&\quad = O\left(n^{-(r-1/p)/(2(r-1/p)+1)}\right).
\end{aligned}$$

High resolution levels. For high resolution levels, first note that

$$\begin{aligned}
\mathbb{E}(\hat{\theta}_{jk} - \theta_{jk})^2 &= \mathbb{E}(d_{jk} - \theta_{jk})^2 \mathbb{I}(|d_{jk}| > t_{j,n}) + \theta_{jk}^2 \mathbb{P}(|d_{jk}| \leq t_{j,n}) \\
&\leq \mathbb{E}(d_{jk} - \theta_{jk})^2 \mathbb{I}(|d_{jk}| > t_{j,n}) + \theta_{jk}^2.
\end{aligned}$$

Therefore,

$$\begin{aligned} R_2 &\le \sum_{j=j_1+1}^{J-1} \sum_{k\in K_j(t_0)} 2^{j/2}|\theta_{jk}| \\ &\quad + \sum_{j=j_1+1}^{J-1} \sum_{k\in K_j(t_0)} 2^{j/2}[\mathbb{E}(d_{jk}-\theta_{jk})^2 \mathbb{I}(|d_{jk}|>t_{j,n})]^{1/2}. \end{aligned}$$

According to Lemma A.4, the first term is bounded by

$$C\sum_{j=j_1+1}^{J-1} 2^{j/2}2^{-j(r-1/p+1/2)} = C\sum_{j=j_1+1}^{J-1} 2^{-j(r-1/p)} = O\left(n^{-(r-1/p)/(2(r-1/p)+1)}\right).$$

Now we consider the second term separately for normal and double-exponential pdf's φ_j. Note that at high resolutions levels $j_1+1 \le j \le J-1$, due to assumption (4.9), we get

$$\nu_j/\sqrt{n} = C_\nu 2^{jm_2}n^{-1/2} > C_\nu 2^{j_1 m_2}n^{-1/2} = C_\nu n^{m_2/2(r-1/p+1/2)-1/2} \to \infty.$$

If φ_j is the double-exponential pdf and $a_2>0$, then $\beta_{j,n} = \left(\frac{\nu_j}{\sqrt{n}}\right)^{a_2} \to \infty$, and thus, by Lemma A.1, we have

$$\begin{aligned} \zeta_{j,n}(x) &= \frac{\int_{-\infty}^{+\infty}\sqrt{n}\varphi_j(\sqrt{n}(x-y))\nu_j h(\nu_j y)dy}{\sqrt{n}\varphi_j(\sqrt{n}x)} \\ &= \frac{\sqrt{n}\varphi_j(\sqrt{n}x)(1+o(1))}{\sqrt{n}\varphi_j(\sqrt{n}x)} \quad = \quad 1+o(1) \quad < \quad \beta_{j,n}, \end{aligned}$$

implying that $\mathbb{I}(|d_{jk}|>t_{j,n}) = \mathbb{I}(\zeta_{j,n}(d_{jk})>\beta_{j,n}) = 0$. Note that the assumption of Lemma A.1 that the parameter λ_j of the double exponential distribution is bounded above uniformly over $j\ge L$ holds here since we assumed in Section 2 that the variance $\sigma_j^2 = 2/\lambda_j^2$ of φ_j is uniformly bounded from below (see (2.3)). Hence, in this case, the second term in the upper bound for R_2 is zero, and the Bayes factor estimator achieves the optimal pointwise rate of convergence.

Now, if φ_j is the normal pdf, then

$$\begin{aligned} \mathbb{E}(d_{jk}-\theta_{jk})^2\mathbb{I}(|d_{jk}|>t_{j,n}) &= \sqrt{n}\int_{|x|>t_{j,n}}(x-\theta_{jk})^2\varphi_j(\sqrt{n}(x-\theta_{jk}))dx \qquad (A.5) \\ &= n^{-1}\int_{|w+\sqrt{n}\theta_{jk}|>\sqrt{n}t_{j,n}} w^2\varphi_j(w)dw. \end{aligned}$$

For $j \geq j_1 + 1$, by Lemma A.4,

$$\sqrt{n}|\theta_{jk}| \leq C\sqrt{n}2^{-j(r-1/p+1/2)} \leq C\sqrt{n}2^{-j_1(r-1/p+1/2)} = O(1). \qquad \text{(A.6)}$$

If $\sqrt{n}t_{j,n} \leq C$, then the integral above is a constant, implying that (A.5) is bounded by Cn^{-1}, and the corresponding sum is bounded by

$$Cn^{-1/2} \sum_{j=j_1+1}^{J-1} \sum_{k \in K_j(t_0)} 2^{j/2} = O\left(n^{-1/2}2^{J/2}\right) = O(1),$$

i.e., it does not tend to zero, and thus it is slower than the optimal pointwise convergence rate. To achieve pointwise optimality, we consider only the cases where $\sqrt{n}t_{j,n} \to \infty$, which is achieved if $a_2 > 0$, i.e.,

$$\sqrt{n}t_{j,n} \geq \sigma_j\sqrt{\log(\beta_{j,n})} = \sigma_j\sqrt{a_2\log(\nu_j/\sqrt{n})} \to \infty,$$

since $\nu_j/\sqrt{n} \to \infty$ for $j_1 + 1 \leq j \leq J - 1$. For $\sqrt{n}t_{j,n} \to \infty$, $|\theta_{jk}|/t_{j,n} \to 0$ due to (A.6) for $j \geq j_1 + 1$. Therefore we can write

$$n^{-1}\int_{|w+\sqrt{n}\theta_{jk}|>\sqrt{n}t_{j,n}} w^2\varphi_j(w)dw \leq Cn^{-1}(t_{j,n}\sqrt{n})^3\varphi_j(t_{j,n}\sqrt{n}),$$

which, according to the first statement of Lemma A.2 and due to the decreasing nature of $x^3\varphi_j(x)$ for large positive x, is bounded by

$$\begin{aligned}
&Cn^{-1}(t_{j,n}\sqrt{n})^3\varphi_j(t_{j,n}\sqrt{n}) \\
\leq\ & Cn^{-1}\left[\log(\beta_{j,n})\right]^{3/2} c_j \exp\left\{-\left[\left(\log(\beta_{j,n})\right)^{1/2}\right]^2\right\} \\
=\ & Cn^{-1}\left[a_2\log(C_\nu 2^{jm_2}/\sqrt{n})\right]^{3/2}\beta_{j,n}^{-1} \quad \leq \quad C(\log n)^{3/2}n^{-1+a_2/2}2^{-a_2m_2j}.
\end{aligned}$$

Now, by substituting this bound into the sum R_2, we get

$$\begin{aligned}
R_2 \quad \leq \quad & O\left(n^{-(r-1/p)/(2(r-1/p)+1)}\right) \\
& +Cn^{-1/2+a_2/4}(\log n)^{3/4}\sum_{j=j_1+1}^{J-1} 2^{j(1-m_2a_2)/2}
\end{aligned}$$

$$\begin{aligned}
&= O\left(n^{-(r-1/p)/(2(r-1/p)+1)}\right) \\
&\quad + \begin{cases} C(\log n)^{3/4}\, n^{-1/2+a_2/4+\frac{1-m_2a_2}{4(r-1/p+1/2)}}, & \text{if } 1-m_2a_2<0, \\ C(\log n)^{1+3/4} n^{-1/2+a_2/4}, & \text{if } 1-m_2a_2=0, \\ C(\log n)^{3/4}\, n^{a_2/4-m_2a_2/2}, & \text{if } 1-m_2a_2>0, \end{cases} \\
&= O\left(n^{-(r-1/p)/(2(r-1/p)+1)}\right) \\
&\quad + \begin{cases} C(\log n)^{3/4}\, n^{-1/2+\frac{1}{4(r-1/p)+2}+\frac{a_2}{4}\left(1-\frac{m_2}{(r-1/p+1/2)}\right)}, & \text{if } a_2>1/m_2, \\ C(\log n)^{1+3/4} n^{-1/2+1/4m_2}, & \text{if } a_2=1/m_2, \\ C(\log n)^{3/4}\, n^{a_2(1/2-m_2)/2}, & \text{if } a_2<1/m_2. \end{cases}
\end{aligned}$$

For $a_2 > 1/m_2$, $1 - m_2/(r - 1/p + 1/2)$ is negative since $m_2 > r - 1/p + 1/2$ by assumption (4.9), and thus the rate $O(n^{-(r-1/p)/(2(r-1/p)+1)})$ is achieved. For $a_2 = 1/m_2$, the convergence rate is also faster than the rate $O(n^{-(r-1/p)/(2(r-1/p)+1)})$, since $-1 + 1/2m_2 < -1 + 1/(2(r - 1/p) + 1)$.

For $a_2 < 1/m_2$, the convergence rate is faster than $O(n^{-(r-1/p)/(2(r-1/p)+1)})$ if $a_2(1/2 - m_2) < -1 + 1/[2(r - 1/p) + 1]$, i.e., if $a_2 > (1 - 1/(2(r - 1/p) + 1))/(m_2 - 1/2)$ since $1/2 - m_2 < 1/2 - (r - 1/p + 1/2) = -(r - 1/p) < 0$. These conditions on a_2 are compatible if and only if $(1 - 1/(2(r - 1/p) + 1))/(m_2 - 1/2) < 1/m_2$, which holds under the assumptions of Theorem 4.1, since

$$\begin{aligned}
\frac{1-1/(2(r-1/p)+1)}{m_2-1/2} - \frac{1}{m_2} &= \frac{m_2 - m_2/(2(r-1/p)+1) - m_2 + 1/2}{m_2(m_2-1/2)} \\
&= \frac{r-1/p+1/2-m_2}{m_2(2m_2-1)(r-1/p+1/2)} < 0.
\end{aligned}$$

Therefore, by combining all the cases, we have that R_2 achieves the rate $O(n^{-(r-1/p)/(2(r-1/p)+1)})$ if $a_2 > \frac{r-1/p}{(m_2-1/2)(r-1/p+1/2)}$.

Combining all the terms, we have that $Q_{11} + Q_{12} + Q_{21} + Q_{22} + Q_3 = O(n^{-(r-1/p)/(2(r-1/p)+1)})$, and, thus, using (A.1), the optimal pointwise convergence rate $O(n^{-2(r-1/p)/(2(r-1/p)+1)})$ is achieved. This completes the proof. □

A.2. The Bayes factor estimator as a thresholding rule. Proposition 2.1 is part of Lemma 1 in Pensky and Sapatinas (2007). For completeness, we provide below a sketch of the proof for this result.

PROOF OF PROPOSITION 2.1. For the sake of convenience, we drop the indices in $\zeta_{j,n}(d_{jk})$, $I_j(d_{jk})$, ν_j and φ_j. Denote $F(x) = \log(\zeta(x))$ and observe that

$$F'(x) = \frac{n}{I(x)} \int_{-\infty}^{+\infty} \left[\frac{\varphi'(\sqrt{n}(x-\theta))}{\varphi(\sqrt{n}(x-\theta))} - \frac{\varphi'(\sqrt{n}x)}{\varphi(\sqrt{n}x)} \right] \varphi(\sqrt{n}(x-\theta))\nu h(\nu\theta)d\theta. \tag{A.7}$$

If φ is the $N(0,\sigma^2)$ pdf, then the expression in the square brackets in (A.7) is equal to $\sqrt{n}\theta/\sigma^2$, so that the integral is positive for $x > 0$. Hence, both $F(x)$ and $\zeta(x)$ are strictly increasing for $x > 0$. Similarly, if $\varphi(x) = (2\sigma)^{-1}\exp(-|x|/\sigma)$, then the expression in square brackets in (A.7) is equal to $2\mathbb{I}(\theta \geq x)/\sigma$, and $F'(x) > 0$. This completes the proof. □

A.3. Asymptotics of the thresholds. To prove Theorem 4.1, we used the following lemmas. Lemmas A.1 and A.2 can be established by working along the same lines of the proofs in Lemma A.2 in Pensky (2006) and Lemmas 2, 4 and 5 in Pensky and Sapatinas (2007). However, to keep the exposition self-contained, we provide below a sketch of proofs for these results.

LEMMA A.1. *The following statements hold.*

(i) *If φ_j is the double-exponential pdf with parameter λ_j, $\lambda_j \leq \bar{\lambda} < \infty$ for all $j \geq L$, and h is a pdf on $\mathbb{R}$ which is symmetric at zero with finite moments of every (polynomial) order, then*

$$I_j(x) = \sqrt{n}\,\varphi_j(\sqrt{n}x)\left[1 + O\left(n\nu_j^{-2}\right)\right] \quad \text{as} \quad \sqrt{n}/\nu_j \to 0, \tag{A.8}$$

uniformly over all $j \geq L$ and all $x \in \mathbb{R}$.

(ii) *If h is the double-exponential pdf and φ_j is a symmetric at zero pdf on $\mathbb{R}$ with uniformly (over j) bounded moments of every (polynomial) order, then, for any $x \in \mathbb{R}$ and for any $j \geq L$,*

$$I_j(x) \geq \nu_j h(\nu_j x)\left[1 + \sigma_j^2 \frac{\nu_j^2}{2n}\right] \quad \text{as} \quad \nu_j/\sqrt{n} \to 0. \tag{A.9}$$

PROOF OF LEMMA A.1. (i) Using a Taylor series expansion for arbitrary x and a change of variables, we obtain

$$\begin{aligned} I_j(x) &= \int_{-\infty}^{+\infty} \sqrt{n}\,\varphi_j\left(\sqrt{n}x - \frac{\sqrt{n}}{\nu_j}y\right) h(y)dy \\ &= \sqrt{n}\int_{-\infty}^{+\infty}\left[\varphi_j(\sqrt{n}x) - \frac{\sqrt{n}}{\nu_j} y\varphi_j'(\sqrt{n}x) + \frac{n}{2\nu_j^2}y^2\varphi_j''(\sqrt{n}x)\right. \\ &- \left.\frac{n\sqrt{n}}{6\nu_j^3}y^3\varphi_j'''(\sqrt{n}x) + \cdots\right] h(y)dy \\ &= \sqrt{n}\varphi_j(\sqrt{n}x)\left[1 + \lambda_j^2\frac{n}{2\nu_j^2}\int_{-\infty}^{+\infty} y^2 h(y)dy + o\left(\frac{n}{\nu_j^2}\right)\right], \qquad \text{(A.10)} \end{aligned}$$

since h is a symmetric at zero pdf on $\mathbb{R}$ and $\varphi_j''(x) = \lambda_j^2\varphi_j(x)$. Note that since φ_j is double exponential, the expression in brackets in the last equation does not depend on x. Since we assumed that λ_j is bounded from above, (A.8) holds uniformly for all $j \geq L$, thus completing the proof of the first statement.

(ii) Similarly to the above case, we can write the following

$$I_j(x) = \nu_j h(\nu_j x)\left[1 + \frac{\nu_j^2}{2n}\int_{-\infty}^{+\infty} y^2\varphi_j(y)dy + o\left(\frac{\nu_j^2}{n}\right)\right],$$

since the double-exponential pdf h has a fixed parameter (equivalent to λ_j in (i)). Since φ_j has uniformly (over j) bounded moments of every (polynomial) order, its variance is finite and all the elements in the remainder of the above expansion are non-negative and finite. Hence, we obtain the second statement of Lemma 1, and thus the proof is complete. □

LEMMA A.2. *The following statements hold*

(i) *If φ_j is the pdf of the form (2.3), then*

$$\sqrt{n}\,t_{j,n} \geq \sigma_j \max\left\{\left[\log\left(\frac{\varphi_j(0)\beta_{j,n}\sqrt{n}}{h(0)\nu_j}\right)\right]^{1/\beta},\ [\log(\beta_{j,n})]^{1/\beta}\right\}.$$

(ii) *Let the assumption (A1) hold. If φ_j is the pdf of the form (2.3) with $\beta = 1$ or 2 (i.e., φ_j is normal or double-exponential) and $\nu_j/\sqrt{n} \to 0$, then*

$$\sqrt{n}\,t_{j,n} \leq \sigma_j\left[2\log\left(\tfrac{C_1\beta_{j,n}\sqrt{n}}{\nu_j}\right)\right]^{1/\beta}\mathbb{I}(a_{(j)} \leq 1), \qquad \text{(A.11)}$$

for some constant $C_1 > 0$.

Note that since the threshold $t_{j,n}$ is non-negative, the last statement of Lemma A.2 implies that $t_{j,n} = 0$ for $a_{(j)} > 1$.

PROOF OF LEMMA A.2. (i) Note that the symmetry and unimodality of φ_j implies that $\varphi_j(x) \le \varphi_j(0)$ for any x. Therefore, the equation for the threshold $t_{j,n}$ (see expression above (2.8)) can be rewritten as follows

$$\begin{aligned}\beta_{j,n} = \zeta_{j,n}(t_{j,n}) &= \frac{\int_{-\infty}^{+\infty} \sqrt{n}\,\varphi_j(\sqrt{n}(t_{j,n} - x))\nu_j h(\nu_j x) dx}{\sqrt{n}\,\varphi_j(\sqrt{n} t_{j,n})} \\ &\le \frac{\int_{-\infty}^{+\infty} \sqrt{n}\,\varphi_j(0)\nu_j h(\nu_j x) dx}{\sqrt{n}\,\varphi_j(\sqrt{n} t_{j,n})} = \frac{\varphi_j(0)}{\varphi_j(\sqrt{n} t_{j,n})}.\end{aligned}$$

Similarly, by symmetry and unimodality of h, we have

$$\begin{aligned}\beta_{j,n} = \zeta_{j,n}(t_{j,n}) &= \frac{\int_{-\infty}^{+\infty} \sqrt{n}\,\varphi_j(\sqrt{n} x)\nu_j h(\nu_j(t_{j,n} - x)) dx}{\sqrt{n}\,\varphi_j(\sqrt{n} t_{j,n})} \\ &\le \frac{\int_{-\infty}^{+\infty} \sqrt{n}\,\varphi_j(\sqrt{n} x)\nu_j h(0) dx}{\sqrt{n}\,\varphi_j(\sqrt{n} t_{j,n})} = \frac{\nu_j h(0)}{\sqrt{n}\,\varphi_j(\sqrt{n} t_{j,n})}.\end{aligned}$$

Rearranging the terms, we have

$$\varphi_j(\sqrt{n} t_{j,n}) \le \min\left\{\beta_{j,n}^{-1}\varphi_j(0), \beta_{j,n}^{-1} h(0)\nu_j/\sqrt{n}\right\}. \tag{A.12}$$

Substituting the power exponential function φ_j given in (2.3) into (A.12), we obtain the first statement.

(ii) When h and φ_j are the standard normal pdfs, then (see Abramovich et al., 2004)

$$\begin{aligned}\sqrt{n}\, t_{j,n} &= \sigma_j\sqrt{2}\sqrt{1 + \frac{\sigma_j^2\nu_j^2}{n}}\left[\log\left(\frac{\beta_{j,n}\sqrt{n + \nu_j^2\sigma_j^2}}{\nu_j\sigma_j}\right)\right]^{1/2} \\ &\qquad \times \mathbb{I}\left(\frac{\beta_{j,n}\sqrt{n + \nu_j^2\sigma_j^2}}{\nu_j\sigma_j} > 1\right),\end{aligned} \tag{A.13}$$

so that (A.11) is valid. On the other hand, if h is double-exponential, then by Lemma A.1 (ii) as $\nu_j/\sqrt{n} \to 0$ and for any x,

$$\zeta_{j,n}(x) \ge \frac{\nu_j}{\sqrt{n}}\frac{h(\nu_j x)}{\varphi_j(\sqrt{n}x)}(1 + \tilde{C}_j) \ge \frac{\nu_j}{\sqrt{n}}\frac{h(\nu_j x)}{\varphi_j(\sqrt{n}x)},$$

since $\tilde{C}_j > 0$. Taking into account assumption (A1), we derive that

$$\zeta_{j,n}(x) \geq C_{h,\varphi}^{-1} \frac{\nu_j}{\sqrt{n}} \frac{\varphi_j(\nu_j x)}{\varphi_j(\sqrt{n}x)} \geq C_1 \frac{\nu_j}{\sqrt{n}} \exp\left\{ \left[1 - \left(\frac{\nu_j}{\sqrt{n}} \right)^{\beta} \right] \left| \frac{x\sqrt{n}}{\sigma_j} \right|^{\beta} \right\},$$

for some constant $C_1 > 0$ independent of j and n, where $\beta = 1$ (φ_j – double-exponential) or $\beta = 2$ (φ_j – normal). Note that since $\nu_j/\sqrt{n} \to 0$, we have $1 - \nu_j/\sqrt{n} \geq 1/2$, so that

$$\zeta_{j,n}(x)\sqrt{n}/\nu_j \geq C_1 \exp\left\{ \frac{1}{2} \left| \frac{x\sqrt{n}}{\sigma_j} \right|^{\beta} \right\}.$$

Now take $x = t_{j,n}$. Since the right hand side is bounded from below by $C_1 > 0$, if $\beta_{j,n}\sqrt{n}/\nu_j = (\nu_j/\sqrt{n})^{a_{(j)}-1} \to 0$ (i.e., $a_{(j)} > 1$), then $t_{j,n} = 0$. If $a_{(j)} \leq 1$, and thus $\beta_{j,n}\sqrt{n}/\nu_j = 1$ ($a_{(j)} = 1$) or $\beta_{j,n}\sqrt{n}/\nu_j \to \infty$ ($a_{(j)} < 1$), then $t_{j,n}$ is of the form (A.11). This proves the second statement. □

LEMMA A.3. *Assume that* $d_{jk} \sim \sqrt{n}\varphi_j(\sqrt{n}(x-\theta_{jk}))$ *and* $\hat{\theta}_{jk} = d_{jk}\mathbb{I}(|d_{jk}| > t_{j,n})$ *for some* $t_{j,n} \geq 0$. *Then, for any* $m > 0$ *such that* $\kappa_{m,j} < \infty$, *the following inequality holds*

$$\mathbb{E}|\hat{\theta}_{jk} - \theta_{jk}|^m \leq \gamma_m \left(\min\left\{ t_{j,n}^m, |\theta_{jk}|^m \right\} + \kappa_{m,j} n^{-m/2} \right), \tag{A.14}$$

where $\kappa_{m,j} = \int_{-\infty}^{+\infty} |x|^m \varphi_j(x)dx$, *and* $\gamma_m = 1$ *if* $0 < m \leq 1$ *and* $\gamma_m = 2^{m-1}$ *if* $m > 1$.

PROOF OF LEMMA A.3. By definition of $\hat{\theta}_{jk}$, we have

$$\begin{aligned}
\mathbb{E}|\hat{\theta}_{jk} - \theta_{jk}|^m &= \mathbb{E}|d_{jk} - \theta_{jk}|^m \mathbb{I}(|d_{jk}| > t_{j,n}) + |\theta_{jk}|^m \mathbb{P}(|d_{jk}| \leq t_{j,n}) \\
&\leq \int_{\mathbb{R}} \sqrt{n}|x|^m \varphi_j(x\sqrt{n})dx + |\theta_{jk}|^m = n^{-m/2}\kappa_{m,j} + |\theta_{jk}|^m.
\end{aligned}$$

On the other hand, by first representing $\hat{\theta}_{jk} - \theta_{jk}$ as a sum of $\hat{\theta}_{jk} - d_{jk}$ and $d_{jk} - \theta_{jk}$, and then applying the definition of $\hat{\theta}_{jk}$ together with the elementary inequality $(a+b)^m \leq \gamma_m(a^m + b^m)$ for $a, b \geq 0$, $\gamma_m = 1$ if $0 < m \leq 1$ and $\gamma_m = 2^{m-1}$ if $m > 1$, we get

$$\begin{aligned}
\mathbb{E}|\hat{\theta}_{jk} - \theta_{jk}|^m &\leq \gamma_m \left\{ \mathbb{E}|\hat{\theta}_{jk} - d_{jk}|^m + \mathbb{E}|d_{jk} - \theta_{jk}|^m \right\} \\
&= \gamma_m \left\{ \mathbb{E}|d_{jk}|^m \mathbb{I}(|d_{jk}| \leq t_{j,n}) + \kappa_{m,j} n^{-m/2} \right\} \\
&\leq \gamma_m \left\{ t_{j,n}^m + \kappa_{m,j} n^{-m/2} \right\}.
\end{aligned}$$

Combining these two inequalities together, we obtain (A.14). This completes the proof. □

LEMMA A.4. *Let $\{\phi, \psi, s, L\}$ be as in assumption (S1), and let $1 \le p, q \le \infty$ and $1/p < r < s$. If $f \in B^r_{p,q}(A)$, then for some constants A_0, A_1, A_2, $A_3 > 0$, we have*

$$\sum_{k \in K_{L-1}(t_0)} |\tilde{\theta}_k - \theta_k| \le A_0\, n^{-r}, \tag{A.15}$$

$$\sum_{j=L}^{J-1} \sum_{k \in K_j(t_0)} 2^{j/2} |\tilde{\theta}_{jk} - \theta_{jk}| \le A_1\, n^{-(r-1/p)}, \tag{A.16}$$

and, for $L \le j \le J-1$,

$$\sum_{k \in K_j(t_0)} |\tilde{\theta}_{jk}| \le A_2\, 2^{-j(r-1/p+1/2)}, \tag{A.17}$$

$$\sum_{k \in K_j(t_0)} |\theta_{jk}| \le A_3\, 2^{-j(r-1/p+1/2)}, \tag{A.18}$$

where $K_j(t_0) = \{k : 0 \le k \le 2^j - 1 \text{ and } \psi_{jk}(t_0) \neq 0\}$, $L \le j \le J-1$, and $K_{L-1}(t_0) = \{k : k \in K_{L-1} \text{ and } \phi_{Lk}(t_0) \neq 0\}$.

PROOF OF LEMMA A.4. First, we consider the case $q = \infty$. Under the conditions of the lemma, using the equivalence between the Besov norm of the function f on $[0, 1]$ and the corresponding sequence norm of its wavelet coefficients, and Proposition 5 of Johnstone and Silverman (2004), we obtain

$$2^{j(r-1/p+1/2)} \left(\sum_{k=0}^{2^j-1} |\tilde{\theta}_{jk} - \theta_{jk}|^p \right)^{1/p} \le A\, C(\phi, \psi, p, r) 2^{-r(J-j)},\ L-1 \le j \le J-1,$$

which implies that

$$\sum_{k=0}^{2^j-1} |\tilde{\theta}_{jk} - \theta_{jk}|^p \le \left(A\, C(\phi, \psi, p, r) 2^{-j(1/2-1/p)} n^{-r} \right)^p, \quad L-1 \le j \le J-1.$$

(Here, we abused notation and $j = L-1$ refers to replacing $\psi_{L-1,k}$ by ϕ_{Lk}, so that $\tilde{\theta}_{L-1,k} = \tilde{\theta}_k$ and $\theta_{L-1,k} = \theta_k$.) Due to the embedding properties of Besov spaces (i.e., $B^r_{p,q}(A) \subset B^r_{p,\infty}(A)$, for $1 \le q \le \infty$), these bounds also hold for all $1 \le q \le \infty$.

Applying Hölder's inequality, we obtain

$$\begin{aligned}\sum_{k\in K_j(t_0)} |\tilde{\theta}_{jk} - \theta_{jk}| &\leq \left[\sum_{k=0}^{2^j-1} |\tilde{\theta}_{jk} - \theta_{jk}|^p\right]^{1/p} K_j^{1-1/p} \\ &\leq A\, C(\phi, \psi, p, r) K_j^{1-1/p} 2^{-j(1/2-1/p)} n^{-r},\end{aligned}$$

where $K_j = \mathrm{Card}(K_j(t_0))$ is the cardinality of $K_j(t_0)$ (see also the discussion after (A.1)). For $j \geq L$, i.e., for the wavelet functions, cardinality is the same: $K_j = K$, and for $j = L-1$, i.e., for the scaling function, the cardinality is different. Taking $j = L-1$, i.e., considering the scaling coefficients, the inequality above implies that

$$\sum_{k\in K_{L-1}(t_0)} |\tilde{\theta}_k - \theta_k| \leq A_0 n^{-r}.$$

Thus we obtain the first statement.

On the other hand, if we sum the corresponding terms over j with factor $2^{j/2}$, we obtain the following

$$\sum_{j=L}^{J-1} \sum_{k\in K_j(t_0)} 2^{j/2} |\tilde{\theta}_{jk} - \theta_{jk}| \leq C_2 n^{-r} \sum_{j=L}^{J-1} 2^{j/2} 2^{-j(1/2-1/p)} = A_1 n^{-r} n^{1/p}.$$

Hence the second statement is proved.

To prove the third statement of the lemma, we use again the above mentioned embedding properties of Besov spaces, the equivalence between the Besov norm of the function f on $[0,1]$ and the corresponding sequence norm of its wavelet coefficients. Using equation (20) of Johnstone and Silverman (2004), we have

$$\begin{aligned}\sum_{k\in K_j(t_0)} |\tilde{\theta}_{jk}| \leq K^{1-1/p} \left(\sum_{k=0}^{2^j-1} |\tilde{\theta}_{jk}|^p\right)^{1/p} &\leq AK^{1-1/p}\, 2^{-j(r-1/p+1/2)} \\ &= A_2 2^{-j(r-1/p+1/2)}, \quad L \leq j \leq J-1.\end{aligned}$$

This completes the proof of the third statement.

To prove the last statement of the lemma, note that from the above we get

$$\begin{aligned}\sum_{k\in K_j(t_0)} |\theta_{jk}| &\leq \sum_{k\in K_j(t_0)} |\tilde{\theta}_{jk}-\theta_{jk}| + \sum_{k\in K_j(t_0)} |\tilde{\theta}_{jk}| \\ &\leq AC(\phi,\psi,p,r)K^{1-1/p}2^{-j(1/2-1/p)}n^{-r} + A_2 2^{-j(r-1/p+1/2)} \\ &\leq A_3 2^{-j(r-1/p+1/2)}, \quad L\leq j\leq J-1.\end{aligned}$$

Thus the last statement is proved. □

A.4. Derivation of thresholds. We derive below the thresholds for the Bayes factor estimator stated in Section 3 in the two latter cases.

2. (Normal φ_j – double-exponential h). The posterior odds in this case are given in Theorem 3.2 in Bochkina and Sapatinas (2005) with $a=\nu_j$ and $\sigma^2=\sigma_j^2/n$. Specifically,

$$\begin{aligned}\frac{\zeta_{j,n}(y)}{\beta_{j,n}} &= \nu_j\sigma_j n^{-1/2}e^{-(\nu_j\sigma_j)^2/2n}\times\left[\frac{e^{\nu_j y}\Phi(-y\sqrt{n}/\sigma_j-\nu_j\sigma_j/\sqrt{n})}{2\beta_{j,n}\varphi(y\sqrt{n}/\sigma_j)}\right. \\ &\quad \left.+\frac{e^{-\nu_j y}\Phi(y\sqrt{n}/\sigma_j-\nu_j\sigma_j/\sqrt{n})}{2\beta_{j,n}\varphi(y\sqrt{n}/\sigma_j)}\right] \\ &= \frac{\nu_j\sigma_j}{2\beta_{j,n}\sqrt{n}}\left[\frac{\Phi(-y\sqrt{n}/\sigma_j-\nu_j\sigma_j/\sqrt{n})}{\varphi(-y\sqrt{n}/\sigma_j-\nu_j\sigma_j/\sqrt{n})}+\frac{\Phi(y\sqrt{n}/\sigma_j-\nu_j\sigma_j/\sqrt{n})}{\varphi(y\sqrt{n}/\sigma_j-\nu_j\sigma_j/\sqrt{n})}\right],\end{aligned}$$

with φ being the pdf of the standard normal distribution. The threshold $t_{j,n}$ is the non-negative solution of equation $\zeta_{j,n}(y)/\beta_{j,n}=1$, which can be rewritten as

$$\frac{\Phi(-t_{j,n}\sqrt{n}/\sigma_j-\nu_j\sigma_j/\sqrt{n})}{\varphi(-t_{j,n}\sqrt{n}/\sigma_j-\nu_j\sigma_j/\sqrt{n})}+\frac{\Phi(t_{j,n}\sqrt{n}/\sigma_j-\nu_j\sigma_j/\sqrt{n})}{\varphi(t_{j,n}\sqrt{n}/\sigma_j-\nu_j\sigma_j/\sqrt{n})}=\frac{2\beta_{j,n}\sqrt{n}}{\nu_j\sigma_j}.$$

Note that if $\zeta_{j,n}(y)/\beta_{j,n}>1$ for all $y>0$ then, according to (2.7), $\hat{\theta}_{jk}=0$ for any observed y which corresponds to the case $t_{j,n}=0$.

3. (Double-exponential φ_j, h). The posterior odds in this case are given in Theorem 3.4 in Bochkina and Sapatinas (2005) with $\nu=\nu_j$ and $\mu=\sqrt{2n}/\sigma_j$. Specifically,

$$\frac{\zeta_{j,n}(y)}{\beta_{j,n}} = \beta_{j,n}^{-1}\left\{\frac{1-\sqrt{2n}(\sigma_j\nu_j)^{-1}e^{-(\nu_j-\sqrt{2n}/\sigma_j)y}}{1-2n(\sigma_j\nu_j)^{-2}}\right\}.$$

Denote $q_{j,n} = \frac{\sqrt{2n}}{\sigma_j \nu_j}$. Thus, the threshold $t_{j,n}$ is the non-negative solution of

$$1 = \beta_{j,n}^{-1} \left\{ \frac{1 - q_{j,n} e^{-\nu_j (1-q_{j,n}) t_{j,n}}}{1 - q_{j,n}^2} \right\},$$

implying that

$$t_{j,n} = \frac{1}{\nu_j (q_{j,n} - 1)} \log \left\{ \frac{1}{q_{j,n}} \left[1 + \beta_{j,n} \left(q_{j,n}^2 - 1 \right) \right] \right\}.$$

If this value is negative, then the value of the threshold is zero.

Acknowledgements. Natalia Bochkina would like to thank the Department of Mathematics and Statistics, University of Cyprus, for financial support and warm hospitality during her visit to carry out part of this work. The authors would like to thank Marianna Pensky for fruitful discussions. The authors are grateful to Professor Arup Bose (Editor), a Co-Editor, and a referee whose valuable comments and suggestions led to an improvement of this paper.

References

Abramovich, F., Amato, U. and Angelini, C. (2004). On optimality of Bayesian wavelet estimators. *Scand. J. Statist.*, **31**, 217–234.

Abramovich, F., Angelini, C. and De Canditiis, D. (2007). Pointwise optimality of Bayesian wavelet estimators. *Ann. Inst. Statist. Math.*, **59**, (to appear).

Abramovich, F., Bailey, T.C. and Sapatinas, T. (2000). Wavelet analysis and its statistical applications. *The Statistician*, **49**, 1–29.

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

Angelini, C. and Sapatinas, T. (2004). Empirical Bayes approach to wavelet regression using ε-contaminated priors. *J. Statist. Comput. Simulation*, **74**, 741–764.

Angelini, C. and Vidakovic, B. (2004). Γ-minimax wavelet shrinkage: a robust incorporation of information about energy of a signal in denoising applications. *Statist. Sinica*, **14**, 103–125.

Angelini, C., De Canditiis, D. and Leblanc, F. (2003). Wavelet regression estimation in nonparametric mixed effects models. *J. Multivariate Anal.*, **85**, 267–291.

Antoniadis, A., Bigot, J. and Sapatinas, T. (2001). Wavelet estimators in nonparametric regression: a comparative simulation study. *J. Statist. Software*, **6**, 1–83.

Billingsley, P. (1995). *Probability and Measure.* 3rd Edition, Wiley, New York.

Bochkina, N. and Sapatinas, T. (2005). On the posterior median estimators of possibly sparse sequences. *Ann. Inst. Statist. Math.*, **57**, 315–351.

Cai, T.T. (2003). Rates of convergence and adaptation over Besov spaces under pointwise risk. *Statist. Sinica*, **13**, 881–902.

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

Clyde, M. and George, E.I. (2000). Flexible empirical Bayes estimation for wavelets. *J. Roy. Statist. Soc. Ser. B*, **62**, 681–698.

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

Cohen, A., Daubechies, I. and Vial, P. (1993). Wavelets on the interval and fast wavelet transforms. *Appl. Comput. Harmon. Anal.*, **1**, 54–81.

Donoho, D.L. and Johnstone, I.M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, **81**, 425–456.

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

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

Donoho, D.L. and Low, M.G. (1992). Renormalization exponents and optimal pointwise rates of convergence. *Ann. Statist.*, **20**, 944–970.

Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. and Picard, D. (1995). Wavelet shrinkage: asymptopia? (with discussion). *J. Roy. Statist. Soc. Ser. B*, **57**, 301–337.

Johnstone, I.M. and Silverman, B.W. (2004). Boundary coiflets for wavelet shrinkage in function estimation. *J. Appl. Probab.*, **41A**, 81–98.

Johnstone, I.M. and Silverman, B.W. (2005). Empirical Bayes selection of wavelet thresholds. *Ann. Statist.*, **33**, 1700–1752.

Neumann, M.H. and von Sachs, R. (1995). Wavelet thresholding: beyond the Gaussian i.i.d. situation. In *Wavelets and Statistics*, A. Antoniadis and G. Oppenheim, (Eds.), Lecture Notes in Statistics, **103**, 301–329, Springer-Verlag, New York.

Pensky, M. (2006). Frequentist optimality of Bayesian wavelet shrinkage rules for Gaussian and non-Gaussian noise. *Ann. Statist.*, **34**, 769–807.

Pensky, M. and Sapatinas, T. (2007). Frequentist optimality of Bayes factor estimators in wavelet regression models. *Statist. Sinica*, **17** (to appear).

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.

Natalia Bochkina
Department of Epidemiology
& Public Health
Imperial College
London, W2 1PG, United Kingdom
E-mail: n.bochkina@imperial.ac.uk

Theofanis Sapatinas
Department of Mathematics & Statistics
University of Cyprus
P.O. Box 20537,
CY 1678 Nicosia, Cyprus
E-mail: t.sapatinas@ucy.ac.cy

Paper received May 2005; revised December 2006.