Statistica Sinica **19** (2009), 1389-1406

MINIMAX RATES OF CONVERGENCE AND OPTIMALITY OF BAYES FACTOR WAVELET REGRESSION ESTIMATORS UNDER POINTWISE RISKS

Natalia Bochkina and Theofanis Sapatinas

University of Edinburgh and University of Cyprus

Abstract: We consider function estimation in nonparametric regression over Besov spaces and under pointwise l^u-risks ($1 \leq u < \infty$). First we derive both non-adaptive and adaptive minimax pointwise rates of convergence in the standard nonparametric regression model, complementing recent related results obtained in the Gaussian white noise model. Then we investigate 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 both non-adaptive and adaptive Bayes factor estimators in terms of their frequentist optimality over Besov spaces and under pointwise l^u-risks ($1 \leq u < \infty$) for various combinations of error and prior distributions, extending recent non-adaptive results obtained for error and prior models with exponential descents and under pointwise l^2-risks. We provide sufficient conditions that determine whether the unknown response function belongs to a Besov space *a-priori* with probability one, and identify regions wherein the response function enjoys both pointwise optimality, for the proposed non-adaptive and adaptive Bayes factor estimators, and *a-priori* Besov membership. A simulation study is conducted to illustrate the performance of the proposed adaptive Bayes factor estimation procedure with hyperparameters estimated in a fully Bayesian framework.

Key words 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, meaning that if an empirical wavelet coefficient is sufficiently large, that is 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 rates of convergence, and are typically implemented through fast algorithms which make them very appealing in practice (see, e.g., Donoho and Johnstone (1994, 1995, 1998), Donoho, Johnstone, Kerkyacharian and Picard (1995), Vidakovic (1999), Abramovich, Bailey and Sapatinas (2000), and Antoniadis, Bigot and Sapatinas (2001)).

Various Bayesian and empirical Bayes approaches for term-by-term nonlinear wavelet shrinkage and wavelet thresholding estimators have been proposed. These approaches impose a prior distribution on the wavelet coefficients of the unknown response function. That is designed to capture the sparseness of wavelet expansions common to most applications. A popular prior model for each wavelet coefficient is a scale mixture of a symmetric unimodal distribution and a point mass at zero; the distribution corresponding to the non-zero component represents the *significant* coefficients, while the point mass at zero represents the *negligible* coefficients. 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, Kolaczyk and McCulloch (1997), Abramovich, Sapatinas and Silverman (1998), and Clyde, Parmigiani and Vidakovic (1998)).

However, until recently, the frequentist optimality (in the minimax sense) properties of these Bayesian estimators have not been studied. (Hereafter the distribution that refers to the prior model corresponds to the non-zero component in the scale mixture distribution for the prior model referred to above.) Abramovich, Amato and Angelini (2004) investigated optimality of posterior mean, posterior median, and Bayes factor estimators in terms of the global L^2-risk for the combination of normal error and normal prior distributions in the Gaussian white noise model. Pensky (2006) and Pensky and Sapatinas (2007) studied optimality of posterior mean and Bayes factor estimators, respectively, with respect to the global L^2-risk for a wide variety of combination of error and prior distributions in the nonparametric regression model. Johnstone and Silverman (2005) explored optimality of adaptive *empirical* Bayes posterior mean and posterior median estimators with respect to a wide range of global L^u-risks ($0 < u \leq 2$) for normal error and some heavy-tailed prior distributions; they paid particular attention to the distinction between sampling the unknown function within Gaussian white noise and sampling at discrete points, and between placing constraints on the function itself and on the discrete wavelet transform of its sequences of values at the observation points, obtaining results for all relevant combinations of these scenarios. The optimality of an adaptive *empirical* Bayes procedure for the Bayes factor estimator with respect to the global L^2-risk for normal error and some heavy-tailed prior distributions, in the standard nonparametric regression model, was considered in Pensky and Sapatinas (2007).

On the other hand, Abramovich, Angelini and De Canditiis (2007) explored the optimality of posterior mean, posterior median, and Bayes factor estimators in terms of the pointwise l^2-risk for the combination of normal error and normal prior distributions in the Gaussian white noise model. They showed that under the considered Bayesian hierarchical model, pointwise optimality is achieved up to a logarithmic factor. The frequentist optimality of the Bayes factor estimator applied to combinations of error and prior distributions with exponential descents, under pointwise l^2-risk, was studied by Bochkina and Sapatinas (2006) in the nonparametric regression model. As they demonstrated, the use of a more flexible Bayesian hierarchical model, under certain conditions, achieves pointwise optimality without the extra logarithmic factor that appeared in the results of Abramovich, Angelini and De Canditiis (2007).

This paper continues the line of investigation of Bochkina and Sapatinas (2006), extending their pointwise optimality (in the minimax sense) results for Bayes factor estimators, obtained in the nonparametric regression model, in three general directions: (a) studying not just the pointwise l^2-risk but also pointwise l^u-risks ($1 \leq u < \infty$); (b) considering not only combinations of error and prior distributions with exponential descents but also combinations of error and prior distributions with polynomial descents; (c) providing not only non-adaptive but also adaptive Bayes factor estimators, and (d) providing sufficient conditions that determine whether the unknown response function belongs to a Besov space *a priori* with probability 1, and identifying the regions wherein the response function enjoys both pointwise optimality, for the proposed non-adaptive and adaptive Bayes factor estimators, and *a priori* Besov membership.

In order to accomplish these objectives, we first complement the recent non-adaptive and adaptive rates of convergence under pointwise l^u-risks ($1 \leq u < \infty$) obtained by Cai (2003) in the Gaussian white noise model. In particular, in Section 2, we consider function estimation in the standard nonparametric regression model over Besov spaces, and derive both non-adaptive and adaptive minimax rates of convergence under pointwise l^u-risks ($1 \leq u < \infty$). In Section 3, we introduce Bayesian models for the wavelet coefficients and discuss assumptions on the error and prior distributions, extending previously considered (in the context of pointwise l^2-risk optimality) error and prior models. In Section 4, we provide statements about optimality of both non-adaptive and adaptive Bayes factor estimators over Besov spaces and under pointwise l^u-risks ($1 \leq u < \infty$) for the various combinations of error and prior distributions discussed in Section 3. In Section 5, we provide sufficient conditions that determine whether the unknown response function belongs to a Besov space *a-priori* with probability one, and identify regions wherein the response function enjoys pointwise optimality for the proposed non-adaptive and adaptive Bayes factor estimators and *a-priori* Besov membership. In Section 6, we report on a simulation study to illustrate performance of the proposed adaptive Bayes factor estimation procedure with hyperparameters estimated in a fully Bayesian framework. Proofs of the theoretical results stated in Sections 2, 4 and 5, as well as some auxiliary statements, are available online at `http://www.stat.sinica.edu.tw/statistica`.

2. Minimax Rates of Convergence in the Standard Nonparametric Regression Model over Besov Spaces and Under Pointwise Risks

Consider the standard nonparametric regression model

$$y_i = f(t_i) + z_i, \quad i = 1, \ldots, n, \tag{2.1}$$

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., $f \in L^2[0, 1]$, and where the errors z_i are independent $N(0, \sigma^2)$ random variables of known variance σ^2 with $0 < \sigma^2 < \infty$.

This model can be viewed as a discretisation of the Gaussian white noise model

$$dY_n(t) = f(t)dt + \frac{\sigma}{\sqrt{n}} dW(t), \quad t \in [0, 1], \tag{2.2}$$

where $W(t)$ is a standard Brownian motion and $f \in L^2[0, 1]$. This model, under some smoothness constraints on the unknown response function f, is asymptotically equivalent (in Le Cam sense) to the standard nonparametric regression model (2.1) (see, e.g., Brown and Low (1996a)).

For any possible estimator $\tilde{f}$ of f based on n observations from either model (2.1) or model (2.2), define the maximal pointwise l^u-risk with respect to the l^u-loss function ($1 \le u < \infty$) over a space $\mathcal{F}$ of functions defined on the unit interval $[0, 1]$ as

$$R_n^u(\tilde{f}, \mathcal{F}, t_0) = \sup_{f \in \mathcal{F}} \mathbb{E}|\tilde{f}(t_0) - f(t_0)|^u \tag{2.3}$$

for any fixed point $t_0 \in (0, 1)$. The difficulty of the estimation problem is measured by the minimax pointwise l^u-risk ($1 \le u < \infty$)

$$R_n^{*,u}(\tilde{f}, \mathcal{F}, t_0) = \inf_{\tilde{f}} R_n^u(\tilde{f}, \mathcal{F}, t_0), \tag{2.4}$$

where the infimum is taken over all estimators $\tilde{f}$ of f, and we wish to determine the rate of convergence of the minimax pointwise l^u-risk (2.4) as $n \to \infty$.

Cai (2003) obtained the non-adaptive and adaptive minimax rates of convergence under pointwise l^u-risks ($1 \le u < \infty$) in the Gaussian white noise model (2.2), when $\mathcal{F}$ is the Besov ball $B^r_{p,q}(A)$ of radius $A > 0$ in the Besov space $B^r_{p,q}[0, 1]$. For definitions and discussions on the relevance of Besov spaces in scientific problems, we refer the reader to Donoho and Johnstone (1998) and Mallat (1999). Cai (2003) showed that, provided $1 \le p, q \le \infty$ and $r > 1/p$, the non-adaptive minimax rate of convergence under pointwise l^u-risks $(1 \le u < \infty)$ is

$$R_n^{*,u}(\tilde{f}, B_{p,q}^r(A), t_0) \asymp n^{-[(u(r-1/p))/(2(r-1/p)+1)]} \quad \text{as} \quad n \to \infty, \tag{2.5}$$

while the corresponding adaptive rate is

$$R_n^{*,u}(\tilde{f}, B_{p,q}^r(A), t_0) \asymp \Big(\frac{n}{\log n}\Big)^{-[(u(r-1/p))/(2(r-1/p)+1)]} \quad \text{as} \quad n \to \infty. \tag{2.6}$$

(Here, we adopt standard notation and write $g_1(n) \asymp g_2(n)$ to denote $0 < \liminf(g_1(n)/g_2(n)) \le \limsup(g_1(n)/g_2(n)) < \infty$ as $n \to \infty$.)

Our first objective is to complement the non-adaptive (2.5) and adaptive (2.6) minimax rates of convergence under pointwise l^u-risks $(1 \le u < \infty)$, obtained by Cai (2003) in the Gaussian white noise model (2.2), to the standard nonparametric regression model (2.1). This is accomplished in Theorems 1 and 2.

Theorem 1.[non-adaptive] *Consider the standard nonparametric regression model* (2.1). *The non-adaptive minimax rate of convergence under pointwise l^u-risks* $(1 \le u < \infty)$*, over Besov balls $B_{p,q}^r(A)$ of radius $A > 0$ with $1 \le p, q \le \infty$ and $r > 1/p$, is given by*

$$R_n^{*,u}(\tilde{f}, B_{p,q}^r(A), t_0) \asymp n^{-[(u(r-1/p))/(2(r-1/p)+1)]} \quad \text{as} \quad n \to \infty. \tag{2.7}$$

Theorem 2.[adaptive] *Consider the standard nonparametric regression model* (2.1). *The adaptive minimax rate of convergence under pointwise l^u-risks* $(1 \le u < \infty)$*, over Besov balls $B_{p,q}^r(A)$ of radius $A > 0$ with $1 \le p, q \le \infty$ and $r > 1/p$, is given by*

$$R_n^{*,u}(\tilde{f}, B_{p,q}^r(A), t_0) \asymp \Big(\frac{n}{\log n}\Big)^{-[(u(r-1/p))/(2(r-1/p)+1)]} \quad \text{as} \quad n \to \infty. \tag{2.8}$$

These minimax pointwise results, obtained for both the standard nonparametric regression model (2.1) and the Gaussian white noise model (2.2), reveal that, unlike the corresponding minimax rates of convergence under global L^u-risks (see, e.g., Johnstone and Silverman (2005), for $0 < u \leqslant 2$), the minimax rates of convergence under pointwise l^u-risks $(1 \le u < \infty)$ depend not only on the smoothness index r, but also on the parameter p. Moreover, they converge at minimax rates slower than the corresponding global minimax rates and the minimum cost for adaptation is a logarithmic factor; this latter observation agrees with earlier results in the case of Lipschitz and Sobolev classes under pointwise l^2-risks, obtained by Lepski (1990), Brown and Low (1996b), Lepski and Spokoiny (1997), and Tsybakov (1998), in the Gaussian white noise model (2.2).

3. Pointwise Optimality of Bayes Factor Regression Estimators

3.1. Wavelet regression model

Now we consider the nonparametric regression model without the assumption of Gaussian errors

$$Y_i = f(t_i) + Z_i, \quad i = 1, \ldots, n, \tag{3.1}$$

where $t_i = i/n$, $f \in L^2[0,1]$ is the unknown response function and where the Z_i's are independent and identically (i.i.d.) distributed random variables with mean $\mathbb{E}(Z_1) = 0$ and variance $\mathrm{Var}\,(Z_1) = \sigma^2$ with $0 < \sigma^2 < \infty$.

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

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

where, for some fixed *primary resolution* level $L \geq 0$, $\phi_{Lk}(t) = 2^{L/2}\phi(2^L t - k)$, $\psi_{jk}(t) = 2^{j/2}\psi(2^j t - k)$, $\tilde{\theta}_k = \int_{-\infty}^{+\infty} \phi_{Lk}(t) f(t) dt$ and $\tilde{\theta}_{jk} = \int_{-\infty}^{+\infty} \psi_{jk}(t) f(t) dt$; 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, for the standard wavelet transform with periodic boundary corrections, $K_{L-1} = \{0, \ldots, 2^L - 1\}$.) For suitable choices of ϕ and ψ, and appropriate modifications of boundary ψ_{jk}, the corresponding set of ϕ_{Lk} and ψ_{jk} forms an orthonormal set in $L^2[0,1]$ (see, e.g., Cohen, Daubechies and Vial (1993)), and Johnstone and Silverman (2004)).

Application of the boundary corrected discrete wavelet transform (DWT) to (3.1) yields

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

where $J = \log_2(n)$, and ϵ_k, ε_{jk} are uncorrelated random variables with zero mean due to the unitary property of the DWT. Let $\theta_k = u_k/\sqrt{n}$ and $\theta_{jk} = w_{jk}/\sqrt{n}$ and recall that $\tilde{\theta}_k \approx \theta_k$ and $\tilde{\theta}_{jk} \approx \theta_{jk}$ for large n (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 4 in Bochkina and Sapatinas (2006)). In this case, there will be $2^L - 2(S - s - 1)$ scaling coefficients at the primary resolution level L, with $K_{L-1} = \{0, \ldots, s-1, S-1, S, \ldots, 2^L - S, 2^L - s, 2^L - s + 1, \ldots, 2^L - 1\}$ (see Johnstone and Silverman (2004, p.83)).

As in Pensky (2006) and Pensky and Sapatinas (2007), the distributions of the errors ε_{jk} are modeled to be independent and allowed to vary with resolution levels $\varepsilon_{jk} \sim \varphi_j(\cdot)$. For the probability density functions (*pdf*'s) φ_j of the errors, we consider not only the standard Gaussian *pdf* (see Abramovich, Angelini and De Canditiis (2007)) or the double-exponential *pdf* (see Bochkina and Sapatinas (2006)), but also two more general types of distributions. The first are distributions with power-exponential descents, also known as Subbotin distributions, with *pdf*'s

$$\varphi_j(x) = C_\beta \sigma_j^{-1} \exp\left(-\left(\frac{|x|}{\sigma_j}\right)^\beta \right), \quad 0 < \underline{\sigma} \le \sigma_j \le \bar{\sigma} < \infty, \quad \beta > 0, \tag{3.2}$$

with $C_\beta^{-1} = 2\Gamma(\beta^{-1})\beta^{-1}$ (see, e.g., Johnson, Kotz and Balakrishnan (1995, Chapter 24)). This family includes Gaussian distribution ($\beta = 2$) and Laplace (double exponential) distribution ($\beta = 1$). We also consider a general family of heavy-tailed distributions with *pdf*'s φ_j such that

$$\left| \frac{\varphi_j^{(k)}(x)}{\varphi_j(x)} \right| < C_{\varphi k}, \quad k = 1, 2, 3, 4, \tag{3.3}$$

for some $C_{\varphi k} > 0$ independent of x (see Pensky (2006)). For example, (3.3) is satisfied for Student's t distribution. We also assume that the error distributions are symmetric; this holds for the Subbotin and Student's t distributions.

Finally, for the distribution of the errors of the scaling coefficients, ϵ_k, we only assume that it has a finite variance σ_{L-1}^2.

3.2. Bayesian model

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, and in order to 'control' the trade-off between sparse and dense sequences, we take the following prior distribution

$$w_{jk} \sim \pi_{j,n}\tau_{j,n}h(\tau_{jn}\cdot) + (1-\pi_{j,n})\delta_0(\cdot), \quad j = L, L+1, \ldots, \quad k = 0, \ldots, 2^j - 1, \tag{3.4}$$

where $0 \le \pi_{j,n} \le 1$ and $\pi_{j,n} = 0$ for $j \ge J$, $\tau_{j,n} > 0$, δ_0 is the Dirac function (the *pdf* of the point mass at zero), and w_{jk} are independent random variables. To complete the prior specification of f, we place noninformative priors (e.g., the uniform density on $\mathbb{R}$) on the scaling coefficients u_k, $k \in K_{L-1}$.

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 (for some justification of this dependence, see Bochkina and Sapatinas (2006) or Pensky and Sapatinas (2007)). We assume that the prior distribution h is symmetric unimodal. Subbotin and Student's t distributions belong to this class. Results are stated without specifying a particular form of h.

Bayesian inference is conducted for each wavelet coefficient separately. Write

$$d_{jk} = \frac{\mathcal{W}_{jk}}{\sqrt{n}} \quad \text{and} \quad \nu_j = \sqrt{n}\tau_{jn}. \tag{3.5}$$

Let $I_j(d_{jk})$ be the density of the marginal distribution of d_{jk} given $\theta_{jk} \neq 0$, i.e.,

$$I_j(d_{jk}) = \int_{-\infty}^{+\infty} \sqrt{n}\varphi_j[\sqrt{n}(x - d_{jk})]\nu_j h(\nu_j x) dx, \tag{3.6}$$

and let $\zeta_{j,n}(d_{jk})$ be the posterior odds ratio

$$\zeta_{j,n}(d_{jk}) = \frac{I_j(d_{jk})}{[\sqrt{n}\varphi_j(\sqrt{n}d_{jk})]}, \tag{3.7}$$

derived using the relation between w_{jk} and θ_{jk} and (3.4)−(3.5). Thus, the posterior *pdf* of θ_{jk} given d_{jk} is of the form

$$p(\theta_{jk} \mid d_{jk}) = \frac{\sqrt{n}\varphi_j(\sqrt{n}(\theta_{jk} - d_{jk}))\nu_j h(\nu_j\theta_{jk}) + \beta_{j,n}\sqrt{n}\varphi_j(\sqrt{n}d_{jk})\delta_0(\theta_{jk})}{I_j(d_{jk}) + \beta_{j,n}\sqrt{n}\,\varphi_j(\sqrt{n}d_{jk})}. \tag{3.8}$$

We use (3.8) in the next section to introduce the Bayes factor estimator.

3.3. Bayes factor estimator

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 $\hat{\theta}_{jk} = d_{jk}\ \mathbb{I}([(P(H_1 \mid d_{jk}))/(P(H_0 \mid d_{jk}))] > 1)$, where $\mathbb{I}(A)$ denotes the indicator function of the set A. Observe that the posterior odds ratio can be rewritten as $[(P(H_1 \mid d_{jk}))/(P(H_0 \mid d_{jk}))] = [(\zeta_{j,n}(d_{jk}))/(\beta_{j,n})]$. Thus, we can write

$$\hat{\theta}_{jk} = d_{jk}\ \mathbb{I}\Big(\zeta_{j,n}(d_{jk}) > \beta_{j,n}\Big). \tag{3.9}$$

Due to the symmetry of both error and prior density functions, the $\zeta_{j,n}(d_{jk})$ are even functions of d_{jk}. If, moreover, the $\zeta_{j,n}(d_{jk})$ are strictly increasing in d_{jk} for $d_{jk} > 0$, then

$$\zeta_{j,n}(d_{jk}) > \beta_{j,n} \ \text{ if and only if } \ |d_{jk}| > t_{j,n} = \zeta_{j,n}^{-1}(\beta_{j,n}). \tag{3.10}$$

Hence, (3.9) is a hard thresholding rule with the threshold $t_{j,n}$, i.e.,

$$\hat{\theta}_{jk} = d_{jk}\ \mathbb{I}\Big(|d_{jk}| > t_{j,n}\Big). \tag{3.11}$$

Indeed, in the majority of practical cases, it is true that (3.9) gives rise to a hard thresholding rule (see Lemma 1 in Pensky and Sapatinas (2007)).

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

$$\hat{f}_{BF}(t) = \sum_{k \in K_{L-1}} \hat{\theta}_k \phi_{Lk}(t) + \sum_{j=L}^{J-1} \sum_{k=0}^{2^j-1} \hat{\theta}_{jk}\psi_{jk}(t), \tag{3.12}$$

where the coefficients $\hat{\theta}_{jk}$ are found using (3.9) with the function $\zeta_{j,n}(d_{jk})$ defined by (3.7) and (3.6).

3.4. Assumptions

Now we formulate conditions on the wavelet system $\{\phi, \psi\}$ and the *pdf*'s h and φ_j.

(S1) ϕ and ψ are the boundary coiflets introduced in Johnstone and Silverman (2004), possessing $s > r$ vanishing moments, and based on orthonormal coiflets supported in $[-S+1, S]$, $s < S$, with $L \geq \log_2(6S-6)$.

(S2) φ_j and h are unimodal symmetric densities and, if *pdf*'s φ_j belong to the family of heavy-tailed densities (3.3), $\int_0^\infty |x|^u \varphi_j(x)dx < \infty$. (Note that the latter condition holds for all u when the *pdf*'s φ_j belong to the family of power-exponential densities (3.2)).

(S3) $|\varphi_j(x)/h(x)| \leq C_{\varphi h}$.

In (S3) constant $C_{\varphi h}$ is 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 $t_{j,n}$ will depend on the resolution level j rather than on the particular form of φ_j.

In what follows, we compare various Bayes factor regression estimators $\hat{f}_{BF}$ in terms of their ability to achieve the non-adaptive (2.7) and adaptive (2.8) minimax rates of convergence under pointwise l^u-risks ($1 \leq u < \infty$); these are pointwise optimal (in the minimax sense) for Gaussian errors. 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}

are asymptotically $N(0, \sigma^2)$ distributed, provided that $\mathbb{E}(Z_1^4) < \infty$, we expect that the optimal pointwise rates of convergence over these levels for an arbitrary distribution of the errors, satisfying (S1)-(S3), are not faster than the optimal pointwise rates of convergence for Gaussian errors.

4. Pointwise Optimality of Bayes Factor Regression Estimators

4.1. The non-adaptive case

In this section, we show that the non-adaptive minimax pointwise rates of convergence (2.7) are achievable for the Bayes factor estimators, either when the pdf's φ_j are power-exponential or when both pdf's φ_j and h are heavy-tailed.

Following Bochkina and Sapatinas (2006), we divide wavelet resolution levels into two groups: low with $L \le j \le j_1$ and high with $j_1 < j \le J-1$, where

$$j_1 = \frac{1}{2(r + 1/2 - 1/p)} \log_2 n. \tag{4.1}$$

We assume that parameters ν_j and $\beta_{j,n}$ of the suggested Bayesian model, described in Section 3, are

$$\nu_j = C_1 2^{j m_{(j)}}, \qquad \text{where } \; m_{(j)} = \begin{cases} m_1, & L \le j \le j_1, \\ m_2, & j_1 < j \le J-1, \end{cases} \tag{4.2}$$

$$\beta_{j,n} = C_\beta 2^{j a_{(j)}} n^{b_{(j)}}, \;\; \text{where } \; (a_{(j)}, b_{(j)}) = \begin{cases} (a_1, b_1), & L \le j \le j_1, \\ (a_2, b_2), & j_1 < j \le J-1. \end{cases} \tag{4.3}$$

Note that we allow both hyperparameters $m_{(j)}$, $a_{(j)}$ and $b_{(j)}$ to vary with resolution level j.

Remark 2. We have considered here a more general parametrization of β_{jn} compared to that in Bochkina and Sapatinas (2006) which now includes the model considered in Abramovich, Angelini and De Canditiis (2007) as a particular case with $b_{(j)} = 0$, as well as the case considered in Bochkina and Sapatinas (2006) with $a_{(j)}^{BS} = a_{(j)} m_{(j)}$ and $b_{(j)}^{BS} = -a_{(j)}/2$.

We start with the case where the *pdf*'s φ_j are power-exponential, and set $x_+ = \max(0, x)$.

Theorem 3. *Assume* (S1), (S2) *and* (S3), $1 \le u < \infty$, *and* $f \in B_{pq}^r(A)$ *with* $1 \le p, q \le \infty$, $A > 0$ *and* $1/p < r < s$. *Assume that* $\varphi_j(x) = c\sigma_j^{-1} e^{-(|x|/\sigma_j)^\beta}$, $0 < \underline{\sigma} \leqslant \sigma_j \leqslant \bar{\sigma} < \infty$, $\beta > 0$, *h has a bounded second derivative, and assume that* $\zeta_{jn}(x)$ *increases for* $x > 0$ *for* $j \leqslant j_1$. *Assume*

1. $m_1 \leqslant r - 1/p + 1/2$ *and* $b_1 + 1/2 + [((a_1 - m_1)_+)/(2(r - 1/p) + 1)] \leqslant 0$ (*if* $a_1 = m_1$, *the inequality is strict*);

2. $m_2 \geqslant r - 1/p + 1/2$ *and*
   a) $\beta \leqslant 1$ *and* $b_2 + [((a_2)_+)/(2(r-1/p)+1)] > 0$*;*
   b) $\beta > 1$ *and* $b_2 + [a_2/(2(r-1/p)+1)] - [(2(r-1/p))/(2(r-1/p)+1)](u/2 - a_2)_+ > 0$.

Then, for any $t_0 \in (0,1)$, $R_n^u(\hat{f}_{BF}, B_{pq}^r, t_0) = O(n^{-[(u(r-1/p))/(2(r-1/p)+1)]})$ *as* $n \to \infty$.

Remark 3. It is evident that if the error distribution has light tails (i.e., $\beta > 1$), we have more freedom to choose the hyperparameters a_2 and b_2. Also, for $u = 2$, the cases with $\beta = 1$ and $\beta = 2$, together with exponential h, were considered by Bochkina and Sapatinas (2006), as well as the case of Gaussian prior and Gaussian error distributions. Simple calculations show that the conditions on $a_{(j)}$ and $b_{(j)}$ stated in Theorem 3, imply conditions on $a_{(j)}^{BS}$ in Bochkina and Sapatinas (2006), with $a_{(j)}^{BS} = a_{(j)} m_{(j)}$ and $b_{(j)}^{BS} = -a_{(j)}/2$.

Theorem 4. *Assume* (S1), (S2) *and* (S3), $1 \leq u < \infty$, *and* $f \in B_{pq}^r(A)$ *with* $1 \leq p, q \leq \infty$, $A > 0$ *and* $1/p < r < s$. *Let* φ_j *be a heavy tailed distribution satisfying* (3.3) *with variance* σ_j^2, $0 < \underline{\sigma} \leqslant \sigma_j \leqslant \bar{\sigma} < \infty$, *and assume that* h *has a bounded second derivative and that* $\zeta_{jn}(x)$ *increases for* $x > 0$ *for* $j \leqslant j_1$. *Assume*

1. $m_1 \leqslant r - 1/p + 1/2$ *and* $b_1 + 1/2 + [((a_1 - m_1)_+)/(2(r-1/p)+1)] \leqslant 0$ (*if* $a_1 = m_1$, *the inequality is strict*)*;*
2. $m_2 \geqslant r - 1/p + 1/2$ *and* $b_2 + [((a_2)_+)/(2(r-1/p)+1) > 0$.

Then, for any $t_0 \in (0,1)$, $R_n^u(\hat{f}_{BF}, B_{pq}^r, t_0) = O(n^{-[(u(r-1/p))/(2(r-1/p)+1)]})$ *as* $n \to \infty$.

Remark 4. The conditions on the hyperparameters are the same in Theorem 3 for $0 < \beta \leqslant 1$ and in Theorem 4, since the power exponential *pdf's* in the latter case satisfy (3.3). Note also that in order to achieve non-adaptive pointwise optimality, we assume different behavior of the ratio of the variances between error and prior distributions, ν_j^2/n, at low and high resolution levels.

Remark 5. If φ_j and h are Student's t distributions, i.e., having *pdf*

$$g_\nu(x) = c_\nu \Big(1 + \frac{x^2}{\nu}\Big)^{-(1+\nu)/2}, \quad -\infty < x < \infty, \tag{4.4}$$

with $\nu = \rho$ and $\nu = \gamma$ degrees of freedom, respectively, then the assumptions of Theorem 4 are satisfied if $2 < \gamma < \rho$ and $\rho > u$.

Remark 6. The assumption that h has a bounded second derivative in Theorems 3 and 4 can be replaced with the assumption that φ_j have second derivatives bounded uniformly in j (see Remark 7).

4.3. The adaptive case

In this section, we show that the adaptive minimax pointwise rates of convergence (2.8) are achievable for the Bayes factor estimators when the pdf's φ_j are power-exponential.

Theorem 5. *Assume* (S1), (S2) *and* (S3), $1 \le u < \infty$, *and* $f \in B^r_{pq}(A)$ *with* $1 \le p, q \le \infty$, $A > 0$ *and* $1/p < r < s$. *Let* φ_j *be the pdf of* $N(0, \sigma_j^2/2)$, $0 < \underline{\sigma} \leqslant \sigma_j \leqslant \bar{\sigma} < \infty$, *and assume that* h *is such that* $\zeta_{jn}(x)$ *increases for* $x > 0$. *Assume*

1. $\nu_j/\sqrt{n} \leqslant C$ *for some* $C > 0$;
2. $C_1 n^{b+1/2} 2^{aj} \leqslant \beta_{jn}\sqrt{n}/\nu_j < C_2 n^B$ *for some* B, C_1, $C_2 > 0$ *and* $b + 1/2 - (u/2 - a)_+ \geqslant 0$.

Then, for any $t_0 \in (0,1)$, $R^u_n(\hat{f}_{BF}, B^r_{pq}, t_0) = O((n/(\log n))^{-[(u(r-1/p))/(2(r-1/p)+1)]})$ *as* $n \to \infty$.

Theorem 6. *Assume* (S1), (S2) *and* (S3), $1 \le u < \infty$, *and* $f \in B^r_{pq}(A)$ *with* $1 \le p, q \le \infty$, $A > 0$ *and* $1/p < r < s$. *Assume that* $\varphi_j(x) = C_\beta \sigma_j^{-1} e^{-(|x|/\sigma_j)^\beta}$, $0 < \underline{\sigma} \leqslant \sigma_j \leqslant \bar{\sigma} < \infty$, $\beta > 0$, *and assume that* h *is such that* $\zeta_{jn}(x)$ *increases for* $x > 0$, *and that* h *is either heavy-tailed* (3.3) (*with* φ_j *replaced by* h) *or Gaussian. Assume*

1. $\nu_j/\sqrt{n} \to 0$ *as* $n \to \infty$;
2. $C_1 n^{b+1/2} 2^{aj} \leqslant \beta_{jn}\sqrt{n}/\nu_j < C_2 \exp\{B[\log n]^{\beta/2}\}$ *for some* B, C_1, $C_2 > 0$ *and* $b + 1/2 - (u/2 - a)_+ \geqslant 0$.

Then, for any $t_0 \in (0,1)$, $R^u_n(\hat{f}_{BF}, B^r_{pq}, t_0) = O((n/(\log n))^{-[(u(r-1/p))/(2(r-1/p)+1)]})$ *as* $n \to \infty$.

5. A-priori Besov Membership of Pointwise Optimal Models

In this section we explore Besov membership of functions whose wavelet coefficients obey the probability model described in Section 3.2, and compare the required sufficient conditions with those for pointwise optimality in both non-adaptive and adaptive cases.

As in Pensky (2006) and Pensky and Sapatinas (2007), for simplicity, we restrict ourselves to the case of finite parameters p and q. We assume that the wavelet coefficients of f are independent and that

$$\theta_{jk} \sim \pi_{j,n}\nu_j h(\nu_j \cdot) + (1 - \pi_{j,n})\delta_0(\cdot), \quad j = L, L+1, \ldots, \quad k = 0, \ldots, 2^j - 1, \tag{5.1}$$

where $0 \leq \pi_{j,n} \leq 1$ and $\pi_{j,n} = 0$ for $j \geq J$, $\nu_j > 0$, with $\beta_{j,n} = (1-\pi_{jn})/\pi_{j,n}$.

For the parametrization of $\beta_{j,n}$ and ν_j used in the non-adaptive case (Section 4.1), the sufficient condition of *a-priori* Besov membership is given in the following proposition.

Proposition 1. *Assume that f has independent wavelet coefficients with distribution* (5.1) *with mother wavelet function's regularity $s > 0$, and that parameters ν_j and β_{jn} are, respectively, of the form* (4.2) *and* (4.3). *Then, $f \in B^r_{p,q}$, $1/p < r < s$, $1 \leqslant p, q < \infty$, almost surely, if*

1. $\int |x|^{\max(p,q)} h(x)dx < \infty$;
2. $b_1 \geqslant [((p(r+1/2-m_1)-a_1)_+)/(2(r+1/2-1/p))]$, *and the inequality is strict if* $p(r+1/2-m_1)-a_1 = 0$;
3. $b_2 \geqslant Z[p(r+1/2-m_2)-a_2]$, *where* $Z = [1/(2(r+1/2-1/p))]$ *if* $p(r+1/2-m_2)-a_2 < 0$ *and* $Z = 1$ *otherwise; and the inequality is strict if* $p(r+1/2-m_2)-a_2 = 0$.

Condition 1 is satisfied by prior distributions with light tails, e.g., the power-exponential ones; this assumption is, however, essential for heavy-tailed distributions. Conditions 2 and 3 apply to the hyperparameters of $\beta_{j,n}$ and ν_j. Comparing these conditions with conditions for non-adaptive pointwise optimality (over l^u-risks, $1 \leq u < \infty$), we see that they do not overlap since for non-adaptive pointwise optimality the weakest condition is $b_1 \leqslant -1/2$, while for *a-priori* Besov membership the weakest condition is $b_1 \geqslant 0$. Thus, in order to simultaneously achieve non-adaptive pointwise optimality and *a-priori* Besov membership, a wider set of functions is needed.

For the adaptive parametrization (Section 4.2), the sufficient condition of *a-priori* Besov membership is given in the following proposition.

Proposition 2. *Assume that f has independent wavelet coefficients with distribution* (5.1) *with mother wavelet function's regularity $s > 0$, and that parameters ν_j and β_{jn} satisfy $\beta_{jn}/\nu_j \geqslant C2^{aj}n^b$ and $\nu_j = C_\nu 2^{mj}$. Then, $f \in B^r_{p,q}$, $1/p < r < s$, $1 \leqslant p, q < \infty$, almost surely, if*

1. $\int |x|^{\max(p,q)} h(x)dx < \infty$;
2. $b \geqslant (p(r+1/2)-(p+1)m-a)_+$; *the inequality is strict if* $p(r+1/2)-(p+1)m-a = 0$.

Combining Proposition 2 and Theorem 6, we obtain that f enjoys adaptive pointwise optimality (over l^u-risks, $1 \leq u < \infty$) for the Bayes factor estimator and *a-priori* Besov membership, for $1/p < r < s$, $1 \leqslant p, q < \infty$, if

$$m \leqslant \frac{1}{2}, \quad b \geqslant \left(\max\left\{ \frac{(u-1)}{2}, p\Big(r+\frac{1}{2}\Big) - (p+1)m \right\} - a \right)_+,$$

and if the moment of order $\max(p, q)$ of the prior distribution h is finite and the error distribution φ_j is Subbotin.

Thus, for some $p_0 \in [1, \infty)$, which for heavy-tailed prior distributions h can be the largest finite moment, the adaptive pointwise optimality (over l^u-risks, $1 \leq u < \infty$) and *a-priori* Besov membership are achieved simultaneously over Besov spaces $f \in B^r_{p,q}$, $1/p < r < s$, $1 \leqslant q \leqslant \infty$, $1 \leqslant p \leqslant p_0$, if, for instance, $m = 1/2$, $a = \max(u/2, p_0 s)$, and $b = 0$.

6. Numerical Results

In this section, we report on a study the proposed adaptive Bayes Factor estimator using simulated data, with parameters estimated in a fully Bayesian way using freely available `WinBUGS` software (see Spiegelhalter, Thomas and Best (1999)). We also compared this fully Bayesian Bayes Factor (FBBF) estimator with the adaptive wavelet estimation procedure (WT) proposed in Cai (2003). The WT estimator is a hard thresholding wavelet estimator with threshold $\sigma\sqrt{u \log_2 n}$. The wavelet transform used in our implementation was performed using R and packages WAVETHRESH and WAVELETS that are freely available from `www.r-project.org`.

6.1. Model specification

To the simulated data, we fitted a Bayesian model with double-exponential prior distribution h and Gaussian error distribution φ_j, with $\sigma_j^2 \equiv \sigma^2$. Theorem 5 implies that in order to achieve the adaptive minimax pointwise rate of convergence, we need to constrain the hyperparameters as $\nu_j \leqslant Cn^{1/2}$ and $Cn^b 2^{aj} \leqslant \beta_j/\nu_j < n^{B-1/2}$, with $b + 1/2 - (u/2 - a)_+ \geqslant 0$. In the fully Bayesian context, we can either impose these constraints on the prior distribution of the hyperparameters with a fixed u or we can check that the mean posterior estimates of π_j and ν_j satisfy these constraints. Here, we followed the second path since we do not want the prior knowledge about the coefficients to depend on the chosen measure of goodness-of-fit. To check the conditions we used values $a = u/2$ and $b = 0$.

Thus, we considered the independent uniform prior distributions for π_j and exponential distributions for ν_j, i.e., $\pi_j \sim \text{U}[0, 1]$ and $\nu_j \sim \text{Exp}(1)$. We used a non-informative scale-invariant prior on the precision parameter σ^{-2} with density $f(x) = 1/x$. To estimate the parameters in the hierarchical model above using the `WinBUGS` software, the improper prior for σ^{-2} was approximated by a proper prior Gamma distribution with *pdf* proportional to $x^{0.001-1}e^{-0.001x} \approx x^{-1}$. Using posterior predictive model checks (see Lewin, Bochkina and Richardson (2007)), we found that this model fit data well. To fit the model, two chains were run for 80,000 iterations in each, and the last 50,000 thinned by 5 were used for estimating the posterior distributions.

6.2. Simulation study

Now, we present results of the simulation study, with the remainder of this section devoted to the discussion of these results.

In this simulation study, we evaluated the performance of the FBBF and WT estimators using Daubechies' compactly supported *ExtremePhase 2* and *Coiflet 2* wavelet filters (see Daubechies (1992, p.196 and p.258) respectively), and primary resolution levels $L = 2$, 3 and 4. We considered three different kinds of test functions, defined on the unit interval (representing different types of situations): (a) `HeaviSine`, a function that is discontinuous with two jumps (see, e.g., Donoho and Johnstone (1994)), (b) `Laplace`, a function that is continuous but has a discontinuity in the first derivative (see, e.g., Angelini, De Canditiis and Leblanc (2003)), and (c) `Parabolas`, a function that has continuous first derivative but there are big jumps in the second derivative (see, e.g., Antoniadis, Bigot and Sapatinas (2001)).

For each test function, $M = 100$ samples were generated by adding independent random noise $\varepsilon \sim N(0, \sigma^2)$ to $n = 256$, 512 and 1,024 equally spaced points on $[0, 1]$. To represent various noise levels, the values of σ were taken to correspond to the values $\sqrt{2.5}$ and $\sqrt{5}$ for the (root) signal-to-noise ratio (SNR)

$$\mathrm{SNR}(f, \sigma) = \frac{\sqrt{(1/n)\sum_{i=1}^{n}(f(t_i) - \bar{f})^2}}{\sigma}, \quad \text{where} \quad \bar{f} = \frac{1}{n}\sum_{i=1}^{n} f(t_i).$$

The goodness-of-fit for an estimator $\hat{f}$ of f at a point t_0 was measured by the mean l^u-error:

$$\mathrm{ME}_u(f, t_0) = \frac{1}{M}\sum_{m=1}^{M} |\hat{f}_m(t_0) - f(t_0)|^u.$$

For brevity, we only report in detail the results for the `Laplace` function using $n = 1,024$, $L = 2$ and *ExtremePhase 2* wavelet filter, for data with $\mathrm{SNR} = \sqrt{2.5}$. Different combinations of test functions, sample sizes, primary resolution levels, wavelet filters and SNR values yield similar results in magnitude. The mean l^u-errors over the majority of points for the FBBF estimator are lower than for the WT estimator, except for the outlier values for $u = 20$ (Figure 1). From the fitted plots, we can see that as the threshold of the WT estimator increases with increasing u, the rather irregular WT estimate for $u = 1$ becomes smoother. At the point $t = 0.5$, the mean l^u-error of the FBBF estimator is smaller than the error of the WT estimator for $u = 1$, for $u = 2, 3$ they are similar, and for

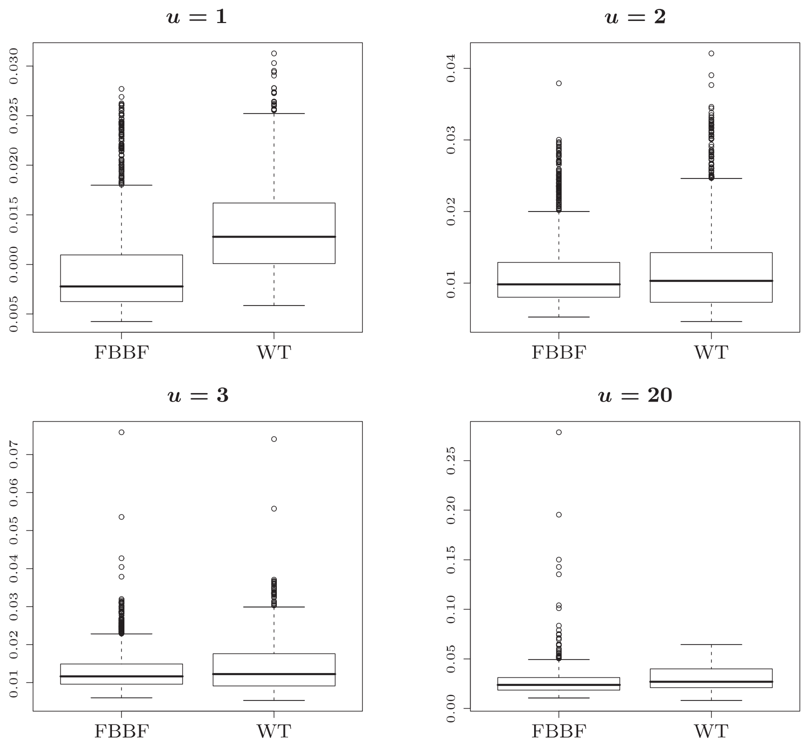

Figure 1. Boxplots of 100 simulation results of the FBBF and WT estimators for the `Laplace` function at point $t_0 = 0.5$.

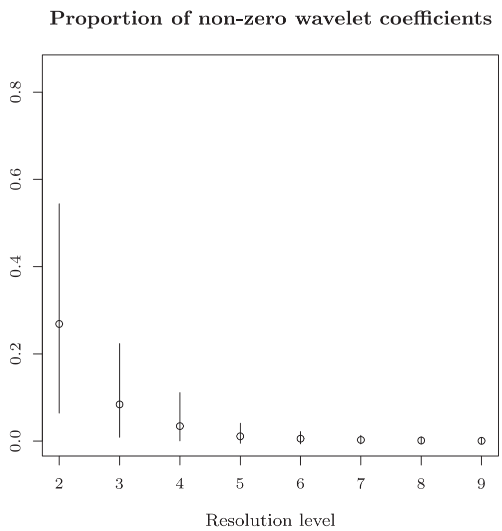

Figure 2. Posterior estimates of $\pi_{j,n}$ with 95% credible intervals for a simulated dataset (*ExtremePhase 2* wavelet filter, SNR $= \sqrt{2.5}$).

$u = 20$ it is greater than the error of the WT estimator. From Figure 2, it can be seen that the proportions of non-zero wavelet coefficients decrease fast, thus fitting curves of the type $\pi_j = 1/(1 + C2^{aj})$ satisfying Condition 3 of Theorem 5 is simple. For example, for values $a = u/2$, the corresponding values (a, C) are $(0.5, 2)$, $(1, 1)$, $(1.5, 0.4)$. For $u = 20$, the constant is very small. Thus, it appears that the Bayes Factor estimator is probably optimal for $u = 1, 2, 3$.

Acknowledgement

Natalia Bochkina would like to thank the Department of Mathematics and Statistics at the University of Cyprus for financial support while visiting Nicosia to carry out part of this work. The authors would also like to thank the two referees for their comments that improved the 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**, 425-434.

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., 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. Soft.* **6**, Article 6.

Bochkina, N. and Sapatinas, T. (2006). On pointwise optimality of Bayes factor wavelet regression estimators. *Sankhyā* **68**, 513-541.

Brown, L. D. and Low, M. G. (1996a). Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.* **24**, 2384-2398.

Brown, L. D. and Low, M. G. (1996b). A constraint risk inequality with applications to nonparametric functional estimation. *Ann. Statist.* **24**, 2524-2535.

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., 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.

Daubechies, I. (1992). *Ten Lectures on Wavelets.* SIAM, Philadelphia.

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

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1995). *Continuous Univariate Distributions.* Vol. **2**, 2nd Edition, John Wiley and Sons, New York.

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.

Lepski, O. V. (1990). On a problem of adaptive estimation in white Gaussian noise. *Theory Probab. Appl.* **35**, 454-466.

Lepski, O. V. and Spokoiny, V. G. (1997). Optimal pointwise adaptive methods in nonparametric estimation. *Ann. Statist.* **25**, 2512-2546.

Lewin, A. M., Bochkina, N. and Richardson, S. (2007). Fully Bayesian mixture model for differential gene expression: simulations and model checks. *Statist. Appl. Genet. Mol. Biol.* **6**, Issue 1, Article 36.

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

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**, 599-633.

Spiegelhalter, D. J., Thomas, A. and Best, N. (1999). *WinBUGS Version* 1.4*, User Manual.* MRC Biostatistics Unit, Cambridge.

Tsybakov, A. B. (1998). Pointwise and sup-norm sharp adaptive estimation of functions on the Sobolev classes. *Ann. Statist.* **26**, 2420-2469.

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 and Sons, New York.

School of Mathematics, University of Edinburgh, Mayfield Road, Edinburgh, EH9 3JZ, United Kingdom.

E-mail: N.Bochkina@ed.ac.uk

Department of Mathematics and Statistics, University of Cyprus, P.O. Box 20537, Nicosia CY 1678, Cyprus.

E-mail: T.Sapatinas@ucy.ac.cy

(Received February 2008; accepted March 2009)