



HAL
open science

A note on wavelet regression with uniform random design and heteroscedastic errors

Christophe Chesneau

► **To cite this version:**

Christophe Chesneau. A note on wavelet regression with uniform random design and heteroscedastic errors. 2010. hal-00449970v2

HAL Id: hal-00449970

<https://hal.science/hal-00449970v2>

Preprint submitted on 19 Jul 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A note on wavelet regression with uniform random design and heteroscedastic errors

Christophe Chesneau

University of Caen (LMNO), Caen, France. Email: chesneau@math.unicaen.fr

Abstract

A nonparametric regression model with uniform random design and heteroscedastic errors (with a deterministic structure) is considered. The distributions of the errors are unknown; we only know that they admit finite moments of order 2. Based on this general framework, we want to estimate the unknown regression function. To reach this goal, an adaptive nonlinear wavelet estimator is developed. Taking the minimax approach under the MISE over Besov balls, we prove that it attains a sharp rate of convergence, closed to the one attained by the optimal non-realistic linear wavelet estimator.

Keywords: Regression, heteroscedastic errors, minimax estimation, Besov balls, wavelets, hard thresholding.

2000 Mathematics Subject Classification: 62G07, 62G20.

1 Motivations

We observe n independent pairs of random variables $(X_1, Y_1), \dots, (X_n, Y_n)$ where, for any $i \in \{1, \dots, n\}$,

$$Y_i = f(X_i) + \xi_i, \quad (1.1)$$

$f : [0, 1] \rightarrow \mathbb{R}$ is an unknown function, X_1, \dots, X_n are n i.i.d. random variables having the uniform distribution on $[0, 1]$ and ξ_1, \dots, ξ_n are n independent random variables independent of X_1, \dots, X_n . We assume that there exist two known sequences of real numbers, $(\mu_i)_{i \in \{1, \dots, n\}}$ and $(\sigma_i^2)_{i \in \{1, \dots, n\}}$, such that, for any $i \in \{1, \dots, n\}$,

$$\mathbb{E}(\xi_i) = \mu_i, \quad \mathbb{V}(\xi_i) = \mathbb{E}((\xi_i - \mu_i)^2) = \sigma_i^2.$$

The distributions of ξ_1, \dots, ξ_n and possible finite moments of order $\delta > 2$ are unknown. We aim to estimate f from $(X_1, Y_1), \dots, (X_n, Y_n)$.

To reach this goal, we focus our attention on the wavelet methods. They are attractive for nonparametric function estimation because of their spatial adaptivity, computational efficiency and asymptotic optimality properties. They can achieve near optimal convergence

rates over a wide range of function classes (Besov balls, ...) and enjoy excellent mean integrated squared error (MISE) properties when used to estimate spatially inhomogeneous function.

If we consider (1.1) when ξ_1, \dots, ξ_n are i.i.d. and ξ_1 is Gaussian or with tail of exponentially decreasing, various wavelet methods have been developed. See, e.g., [13–15], [16], [17], [7,8], [5,6], [10], [19] and [9]. When ξ_1 has other kinds of distributions (Cauchy, ...), see, e.g., [1,2] and [3]. When the distribution of ξ_1 is unknown but admits known finite moment of order δ with $\delta > 2$, see, e. g., [22], [12], [26] and [3]. When $\delta > 0$ (including $\delta \in (0, 2)$) and some conditions on the distribution of ξ_1 are known, see [4]. For the nonindependent case and $\delta > 2$, we refer to [20]. However, to the best of our knowledge, there is no adaptive wavelet estimator of f in (1.1) when the errors are heteroscedastic and only finite moments of order 2 are known.

Adopting the methodology of [12], we develop two wavelet estimators: a linear non-adaptive and a nonlinear adaptive based on the hard thresholding rule. We evaluate their performances by taking the minimax approach under the MISE over Besov balls $B_{p,r}^s(M)$ (to be defined in Section 3). We prove that the considered hard thresholding wavelet estimator attains the rate of convergence

$$r_n = \left(\frac{\ln w_n}{w_n} \right)^{2s/(2s+1)},$$

where $w_n = \sum_{i=1}^n \frac{1}{\sigma_i^2}$. This rate is sharp in the sense that it is the one attained by the (non-realistic) linear wavelet estimator up to a logarithmic term. Moreover, if we restrict our study to the Gaussian case, we prove that it is near optimal in the minimax sense.

The paper is organized as follows. Assumptions on the model and some notations are introduced in Section 2. Section 3 briefly describes the wavelet basis on $[0, 1]$ and the Besov balls. The estimators are presented in Section 4. The results are set in Section 5. Section 6 is devoted to the proofs.

2 Assumptions and notations

Assumption on f We suppose that there exists a known constant $C_* > 0$ such that

$$\|f\|_\infty = \sup_{x \in [0,1]} |f(x)| \leq C_*. \quad (2.1)$$

Assumptions on ξ_1, \dots, ξ_n We recall that ξ_1, \dots, ξ_n are non (necessarily) identically distributed: there exist two known sequences of real numbers, $(\mu_i)_{i \in \{1, \dots, n\}}$ and $(\sigma_i^2)_{i \in \{1, \dots, n\}}$, such that, for any $i \in \{1, \dots, n\}$,

$$\mathbb{E}(\xi_i) = \mu_i, \quad \mathbb{V}(\xi_i) = \sigma_i^2.$$

These sequences can depend on n . The distributions of ξ_1, \dots, ξ_n are unknown. We suppose that there exists a (known) constant $c_* > 0$ such that

$$\inf_{i \in \{1, \dots, n\}} \sigma_i^2 \geq c_*. \quad (2.2)$$

We set

$$w_n = \sum_{i=1}^n \frac{1}{\sigma_i^2}$$

and we suppose that $w_n \geq e$ (i.e. $\ln w_n \geq 1$).

3 Wavelet bases and Besov balls

Wavelet basis Let ϕ be a father wavelet of a multiresolution analysis on \mathbb{R} and ψ be the associated mother wavelet. Assume that $\text{Supp}(\phi) = \text{Supp}(\psi) = [1 - N, N]$, $\int_{1-N}^N \phi(x) dx = 1$ and, for any $\ell \in \{0, \dots, N-1\}$, $\int_{1-N}^N x^\ell \psi(x) dx = 0$. For instance, the Daubechies wavelets dbN satisfy these assumptions.

Set

$$\phi_{j,k}(x) = 2^{j/2} \phi(2^j x - k), \quad \psi_{j,k}(x) = 2^{j/2} \psi(2^j x - k).$$

Then there exists an integer τ satisfying $2^\tau \geq 2N$ such that the collection

$$\mathcal{B} = \{\phi_{\tau,k}(\cdot), k \in \{0, \dots, 2^\tau - 1\}; \psi_{j,k}(\cdot); j \in \mathbb{N} - \{0, \dots, \tau - 1\}, k \in \{0, \dots, 2^j - 1\}\},$$

(with an appropriate treatments at the boundaries) is an orthonormal basis of $\mathbb{L}^2([0, 1]) = \{h : [0, 1] \rightarrow \mathbb{R}; \int_0^1 h^2(x) dx < \infty\}$. See [11].

For any integer $\ell \geq \tau$, any $f \in \mathbb{L}^2([0, 1])$ can be expanded on \mathcal{B} as

$$f(x) = \sum_{k=0}^{2^\ell - 1} \alpha_{\ell,k} \phi_{\ell,k}(x) + \sum_{j=\ell}^{\infty} \sum_{k=0}^{2^j - 1} \beta_{j,k} \psi_{j,k}(x),$$

where $\alpha_{j,k}$ and $\beta_{j,k}$ are the wavelet coefficients of f defined by

$$\alpha_{j,k} = \int_0^1 f(x) \phi_{j,k}(x) dx, \quad \beta_{j,k} = \int_0^1 f(x) \psi_{j,k}(x) dx. \quad (3.1)$$

Besov balls Let $M > 0$, $s > 0$, $p \geq 1$ and $r \geq 1$. Set $\beta_{\tau-1,k} = \alpha_{\tau,k}$. A function f belongs to $B_{p,r}^s(M)$ if, and only if, there exists a constant $M^* > 0$ (depending on M) such that the associated wavelet coefficients (3.1) satisfy

$$\left(\sum_{j=\tau-1}^{\infty} \left(2^{j(s+1/2-1/p)} \left(\sum_{k=0}^{2^j-1} |\beta_{j,k}|^p \right)^{1/p} \right)^r \right)^{1/r} \leq M^*.$$

In this expression, s is a smoothness parameter and p and r are norm parameters. For a particular choice of s , p and r , the Besov balls contain the Hölder and Sobolev balls. See [21].

4 Estimators

Wavelet coefficient estimators The first step to estimate f consists in expanding f on \mathcal{B} and estimating the unknown wavelet coefficients. For any integer $j \geq \tau$ and any $k \in \{0, \dots, 2^j - 1\}$,

- we estimate $\alpha_{j,k}$ by

$$\widehat{\alpha}_{j,k} = \frac{1}{w_n} \sum_{i=1}^n \frac{1}{\sigma_i^2} (Y_i - \mu_i) \phi_{j,k}(X_i), \quad (4.1)$$

- we estimate $\beta_{j,k}$ by

$$\widehat{\beta}_{j,k} = \frac{1}{w_n} \sum_{i=1}^n \frac{1}{\sigma_i^2} (Y_i - \mu_i) \psi_{j,k}(X_i) \mathbf{1}_{\left\{ \frac{1}{\sigma_i^2} |(Y_i - \mu_i) \psi_{j,k}(X_i)| \leq \gamma_n \right\}}, \quad (4.2)$$

where, for any random event \mathcal{A} , $\mathbf{1}_{\mathcal{A}}$ is the indicator function on \mathcal{A} ,

$$\gamma_n = \theta \sqrt{\frac{w_n}{\ln w_n}}$$

and $\theta = \sqrt{C_*^2/c_* + 1}$.

Notice that $\widehat{\beta}_{j,k}$ is the thresholded version of the empirical estimator of $\beta_{j,k}$ (see, e.g., [13–15]). The idea of this thresholding is to operate a selection on the observations: when, for $i \in \{1, \dots, n\}$, f is "too noisy" by ξ_i , the observation Y_i is neglected. Such a $\widehat{\beta}_{j,k}$ has been introduced by [12] when ξ_1, \dots, ξ_n are i.i.d. (non-heteroscedastic case) and ξ_1 admits finite moments of order 3.

Linear estimator Assuming that $f \in B_{p,r}^s(M)$ with $p \geq 2$, we define the linear estimator \widehat{f}^L by

$$\widehat{f}^L(x) = \sum_{k=0}^{2^{j_0}-1} \widehat{\alpha}_{j_0,k} \phi_{j_0,k}(x), \quad (4.3)$$

where $\widehat{\alpha}_{j,k}$ is defined by (4.1) and j_0 is the integer satisfying

$$2^{-1} w_n^{1/(2s+1)} < 2^{j_0} \leq w_n^{1/(2s+1)}.$$

The definition of j_0 is chosen to minimize the MISE of \widehat{f}^L . It is not adaptive since it depends on s , the smoothness parameter of f .

Hard thresholding estimator We define the hard thresholding estimator \widehat{f}^H by

$$\widehat{f}^H(x) = \sum_{k=0}^{2^\tau-1} \widehat{\alpha}_{\tau,k} \phi_{\tau,k}(x) + \sum_{j=\tau}^{j_1} \sum_{k=0}^{2^j-1} \widehat{\beta}_{j,k} \mathbf{1}_{\{|\widehat{\beta}_{j,k}| \geq \kappa \lambda_n\}} \psi_{j,k}(x), \quad (4.4)$$

where $\widehat{\alpha}_{j,k}$ is defined by (4.1), $\widehat{\beta}_{j,k}$ by (4.2), j_1 is the integer satisfying

$$2^{-1}w_n < 2^{j_1} \leq w_n,$$

$\kappa \geq 8/3 + 2 + 2\sqrt{16/9 + 4}$, λ_n is the threshold

$$\lambda_n = \theta \sqrt{\frac{\ln w_n}{w_n}} \quad (4.5)$$

and $\theta = \sqrt{C_*^2/c_* + 1}$. The definitions of γ_n and λ_n are chosen to minimize the MISE of \widehat{f}^H and to make it adaptive. Notice that \widehat{f}^H is a modified version of the one of [12].

5 Results

5.1 Upper bounds

Theorem 5.1. *Consider (1.1) under the assumptions of Section 2. Suppose that $f \in B_{p,r}^s(M)$ with $s > 0$, $p \geq 2$ and $r \geq 1$. Let \widehat{f}^L be (4.3). Then there exists a constant $C > 0$ such that*

$$\mathbb{E} \left(\int_0^1 \left(\widehat{f}^L(x) - f(x) \right)^2 dx \right) \leq C w_n^{-2s/(2s+1)}.$$

The proof of Theorem 5.1 uses a moment inequality on (4.1) and a suitable decomposition of the MISE. Due to our weak assumptions on ξ_1, \dots, ξ_n , the optimal lower bound of the model seems difficult to determine (see, e.g., [18] and [25]). However, since \widehat{f}^L is constructed to be the linear estimator which optimizes the MISE and $w_n^{-2s/(2s+1)}$ is the optimal rate of convergence in the Gaussian case (see Theorem 5.3 below), our benchmark will be

$$r_n = w_n^{-2s/(2s+1)}.$$

Theorem 5.2. *Consider (1.1) under the assumptions of Section 2. Let \widehat{f}^H be (4.4). Suppose that $f \in B_{p,r}^s(M)$ with $r \geq 1$, $\{p \geq 2 \text{ and } s > 0\}$ or $\{p \in [1, 2) \text{ and } s > 1/p\}$. Then there exists a constant $C > 0$ such that*

$$\mathbb{E} \left(\int_0^1 \left(\widehat{f}^H(x) - f(x) \right)^2 dx \right) \leq C \left(\frac{\ln w_n}{w_n} \right)^{2s/(2s+1)}.$$

The proof of Theorem 5.2 is based on several probability results (moment inequalities, concentration inequality, ...) and a suitable decomposition of the MISE. Note that our proof differs from the one of [12, Theorem 4]. This difference enables us to improve the moment assumption (only 2 finite moments is required against 3 in [12]) and to take into account the heteroscedasticity of the errors.

Theorem 5.2 proves that \widehat{f}^H attains r_n up to a logarithmic term. Notice that when ξ_1, \dots, ξ_n are i.i.d. and only admits finite moments of order 2, \widehat{f}^H achieves the same rates of convergence than the standard hard thresholding estimator defined when ξ_1 is Gaussian (see [12] and [13–15]). And this one is optimal up to a logarithmic term.

5.2 Lower bound

Theorem 5.3 below investigates the lower bound of (1.1) in the Gaussian case.

Theorem 5.3. *Consider (1.1) when ξ_1, \dots, ξ_n are Gaussian. Then there exists a constant $c > 0$ such that, for any $s > 0$, $p \geq 1$ and $r \geq 1$,*

$$\inf_{\tilde{f}} \sup_{f \in B_{p,r}^s(M)} \mathbb{E} \left(\int_0^1 (\tilde{f}(x) - f(x))^2 dx \right) \geq cw_n^{-2s/(2s+1)},$$

where the infimum is taken over all the possible estimators \tilde{f} of f .

Theorems 5.1 and 5.3 prove that, in the Gaussian case, $r_n = w_n^{-2s/(2s+1)}$ is the optimal rate of convergence.

5.3 Perspectives

A possible perspective of this work will be to investigate the estimation of f in (1.1) when X_1 has a more complex distribution than the random uniform one. In this case, the warped wavelet basis introduced in the nonparametric regression estimation by [19] seems to be a powerful tool. To improve the estimation of f (and remove the extra logarithmic term in r_n), maybe other thresholding rules can be developed. The thresholding rule named BlockJS introduced in wavelet methodology by [5, 6] seems to be a good alternative. All these aspects need further investigations that we leave for a future work.

6 Proofs

In this section, we consider (1.1) under the assumptions of Section 2. Moreover, C represents a positive constant which may differ from one term to another.

6.1 Auxiliary results

Proposition 6.1. *For any integer $j \geq \tau$ and any $k \in \{0, \dots, 2^j - 1\}$, let $\alpha_{j,k}$ be the wavelet coefficient (3.1) of f in (1.1) and $\widehat{\alpha}_{j,k}$ be (4.1). Then there exists a constant $C > 0$ such that*

$$\mathbb{E} \left((\widehat{\alpha}_{j,k} - \alpha_{j,k})^2 \right) \leq Cw_n^{-1}.$$

Proof of Proposition 6.1. For any $i \in \{1, \dots, n\}$, let us set

$$W_i = (Y_i - \mu_i)\phi_{j,k}(X_i).$$

Since X_i and ξ_i are independent, we have

$$\begin{aligned} \mathbb{E}(W_i) &= \mathbb{E}(f(X_i)\phi_{j,k}(X_i)) + \mathbb{E}(\xi_i - \mu_i)\mathbb{E}(\phi_{j,k}(X_i)) \\ &= \mathbb{E}(f(X_1)\phi_{j,k}(X_1)) = \int_0^1 f(x)\phi_{j,k}(x)dx = \alpha_{j,k}. \end{aligned}$$

Hence

$$\mathbb{E}(\hat{\alpha}_{j,k}) = \mathbb{E}\left(\frac{1}{w_n} \sum_{i=1}^n \frac{1}{\sigma_i^2} W_i\right) = \frac{1}{w_n} \sum_{i=1}^n \frac{1}{\sigma_i^2} \mathbb{E}(W_i) = \frac{1}{w_n} w_n \alpha_{j,k} = \alpha_{j,k}. \quad (6.1)$$

This implies

$$\begin{aligned} \mathbb{E}\left((\hat{\alpha}_{j,k} - \alpha_{j,k})^2\right) &= \mathbb{V}(\hat{\alpha}_{j,k}) = \mathbb{V}\left(\frac{1}{w_n} \sum_{i=1}^n \frac{1}{\sigma_i^2} W_i\right) = \frac{1}{w_n^2} \sum_{i=1}^n \frac{1}{\sigma_i^4} \mathbb{V}(W_i) \\ &\leq \frac{1}{w_n^2} \sum_{i=1}^n \frac{1}{\sigma_i^4} \mathbb{E}(W_i^2). \end{aligned} \quad (6.2)$$

For any $i \in \{1, \dots, n\}$, since X_i and ξ_i are independent, $\mathbb{E}(\phi_{j,k}^2(X_i)) = \int_0^1 \phi_{j,k}^2(x)dx = 1$ and, by (2.1), $\|f\|_\infty \leq C_*$, we have

$$\begin{aligned} \mathbb{E}(W_i^2) &= \mathbb{E}((f(X_i) + (\xi_i - \mu_i))^2 \phi_{j,k}^2(X_i)) \\ &= \mathbb{E}(f^2(X_i)\phi_{j,k}^2(X_i)) + 2\mathbb{E}(\xi_i - \mu_i)\mathbb{E}(f(X_i)\phi_{j,k}^2(X_i)) \\ &\quad + \mathbb{E}((\xi_i - \mu_i)^2)\mathbb{E}(\phi_{j,k}^2(X_i)) \\ &= \mathbb{E}(f^2(X_i)\phi_{j,k}^2(X_i)) + \sigma_i^2 \leq C_*^2 + \sigma_i^2. \end{aligned}$$

So, by (2.2),

$$\frac{1}{\sigma_i^4} \mathbb{E}(W_i^2) \leq \frac{1}{\sigma_i^4} (C_*^2 + \sigma_i^2) \leq \frac{\theta^2}{\sigma_i^2}. \quad (6.3)$$

It follows from (6.2) and (6.3) that

$$\mathbb{E}\left((\hat{\alpha}_{j,k} - \alpha_{j,k})^2\right) \leq \frac{\theta^2}{w_n^2} \sum_{i=1}^n \frac{1}{\sigma_i^2} = C w_n^{-1}.$$

□

Proposition 6.2. For any integer $j \geq \tau$ and any $k \in \{0, \dots, 2^j - 1\}$, let $\beta_{j,k}$ be the wavelet coefficient (3.1) of f in (1.1) and $\hat{\beta}_{j,k}$ be (4.2). Then there exists a constant $C > 0$ such that

$$\mathbb{E}\left(\left(\hat{\beta}_{j,k} - \beta_{j,k}\right)^4\right) \leq C \frac{(\ln w_n)^2}{w_n^2}.$$

Proof of Proposition 6.2. For any $i \in \{1, \dots, n\}$, let us set

$$Z_i = \frac{1}{\sigma_i^2}(Y_i - \mu_i)\psi_{j,k}(X_i).$$

Proceeding as in (6.1), we have

$$\begin{aligned} \beta_{j,k} &= \int_0^1 f(x)\psi_{j,k}(x)dx = \mathbb{E}(f(X_1)\psi_{j,k}(X_1)) = \frac{1}{w_n} \sum_{i=1}^n \mathbb{E}(Z_i) \\ &= \frac{1}{w_n} \sum_{i=1}^n \mathbb{E}(Z_i \mathbf{1}_{\{|Z_i| \leq \gamma_n\}}) + \frac{1}{w_n} \sum_{i=1}^n \mathbb{E}(Z_i \mathbf{1}_{\{|Z_i| > \gamma_n\}}). \end{aligned} \quad (6.4)$$

This with the elementary inequality $(x + y)^4 \leq 8(x^4 + y^4)$, $(x, y) \in \mathbb{R}^2$, imply that

$$\begin{aligned} &\mathbb{E}\left(\left(\widehat{\beta}_{j,k} - \beta_{j,k}\right)^4\right) \\ &= \mathbb{E}\left(\left(\frac{1}{w_n} \sum_{i=1}^n (Z_i \mathbf{1}_{\{|Z_i| \leq \gamma_n\}} - \mathbb{E}(Z_i \mathbf{1}_{\{|Z_i| \leq \gamma_n\}})) - \frac{1}{w_n} \sum_{i=1}^n \mathbb{E}(Z_i \mathbf{1}_{\{|Z_i| > \gamma_n\}})\right)^4\right) \\ &\leq 8(A + B), \end{aligned} \quad (6.5)$$

where

$$A = \mathbb{E}\left(\left(\frac{1}{w_n} \sum_{i=1}^n (Z_i \mathbf{1}_{\{|Z_i| \leq \gamma_n\}} - \mathbb{E}(Z_i \mathbf{1}_{\{|Z_i| \leq \gamma_n\}}))\right)^4\right)$$

and

$$B = \left(\frac{1}{w_n} \sum_{i=1}^n \mathbb{E}(|Z_i| \mathbf{1}_{\{|Z_i| > \gamma_n\}})\right)^4.$$

Let us bound A and B , in turn.

Upper bound for A. We need the Rosenthal inequality presented in lemma below.

Lemma 6.1 (Rosenthal's inequality [24]). *Let $p \geq 2$, $n \in \mathbb{N}^*$ and $(U_i)_{i \in \{1, \dots, n\}}$ be n zero mean independent random variables satisfying $\sup_{i \in \{1, \dots, n\}} \mathbb{E}(|U_i|^p) < \infty$. Then there exists a constant $C > 0$ such that*

$$\mathbb{E}\left(\left|\sum_{i=1}^n U_i\right|^p\right) \leq C \max\left(\sum_{i=1}^n \mathbb{E}(|U_i|^p), \left(\sum_{i=1}^n \mathbb{E}(U_i^2)\right)^{p/2}\right).$$

Applying the Rosenthal inequality with $p = 4$ and, for any $i \in \{1, \dots, n\}$,

$$U_i = Z_i \mathbf{1}_{\{|Z_i| \leq \gamma_n\}} - \mathbb{E}(Z_i \mathbf{1}_{\{|Z_i| \leq \gamma_n\}}),$$

we obtain

$$A = \frac{1}{w_n^4} \mathbb{E}\left(\left(\sum_{i=1}^n U_i\right)^4\right) \leq C \frac{1}{w_n^4} \max\left(\sum_{i=1}^n \mathbb{E}(U_i^4), \left(\sum_{i=1}^n \mathbb{E}(U_i^2)\right)^2\right).$$

Using (6.3) (with ψ instead of ϕ), we have, for any $a \in \{2, 4\}$ and any $i \in \{1, \dots, n\}$,

$$\mathbb{E}(U_i^a) \leq 2^a \mathbb{E}(Z_i^a \mathbf{1}_{\{|Z_i| \leq \gamma_n\}}) \leq 2^a \gamma_n^{a-2} \mathbb{E}(Z_i^2) \leq 2^a \gamma_n^{a-2} \frac{\theta^2}{\sigma_i^2}.$$

Hence, using $w_n \geq e$,

$$\begin{aligned} A &\leq C \frac{1}{n^4} \max \left(\gamma_n^2 \sum_{i=1}^n \frac{1}{\sigma_i^2}, \left(\sum_{i=1}^n \frac{1}{\sigma_i^2} \right)^2 \right) \\ &= C \frac{1}{w_n^4} \max \left(\frac{w_n^2}{\ln w_n}, w_n^2 \right) = C \frac{1}{w_n^2}. \end{aligned} \quad (6.6)$$

Upper bound for B. Using again (6.3) (with ψ instead of ϕ), for any $i \in \{1, \dots, n\}$, we obtain

$$\mathbb{E}(|Z_i| \mathbf{1}_{\{|Z_i| > \gamma_n\}}) \leq \frac{\mathbb{E}(Z_i^2)}{\gamma_n} \leq \frac{\theta}{\sigma_i^2} \sqrt{\frac{\ln w_n}{w_n}}.$$

Therefore

$$B \leq C \frac{(\ln w_n)^2}{w_n^2} \left(\frac{1}{w_n} \sum_{i=1}^n \frac{1}{\sigma_i^2} \right)^4 = C \frac{(\ln w_n)^2}{w_n^2}. \quad (6.7)$$

Combining (6.5), (6.6) and (6.7), we have

$$\mathbb{E} \left(\left(\widehat{\beta}_{j,k} - \beta_{j,k} \right)^4 \right) \leq C \left(\frac{1}{w_n^2} + \frac{(\ln w_n)^2}{w_n^2} \right) \leq C \frac{(\ln w_n)^2}{w_n^2}.$$

□

Proposition 6.3. *For any integer $j \geq \tau$ and any $k \in \{0, \dots, 2^j - 1\}$, let $\beta_{j,k}$ be the wavelet coefficient (3.1) of f in (1.1) and $\widehat{\beta}_{j,k}$ be (4.2). Then, for any $\kappa \geq 8/3 + 2 + 2\sqrt{16/9 + 4}$,*

$$\mathbb{P} \left(|\widehat{\beta}_{j,k} - \beta_{j,k}| \geq \kappa \lambda_n / 2 \right) \leq 2w_n^{-2}.$$

Proof of Proposition 6.3. For any $i \in \{1, \dots, n\}$, let us set

$$Z_i = \frac{1}{\sigma_i^2} (Y_i - \mu_i) \psi_{j,k}(X_i).$$

Using (6.4), we have

$$\begin{aligned} &|\widehat{\beta}_{j,k} - \beta_{j,k}| \\ &= \left| \frac{1}{w_n} \sum_{i=1}^n (Z_i \mathbf{1}_{\{|Z_i| \leq \gamma_n\}} - \mathbb{E}(Z_i \mathbf{1}_{\{|Z_i| \leq \gamma_n\}})) - \frac{1}{w_n} \sum_{i=1}^n \mathbb{E}(Z_i \mathbf{1}_{\{|Z_i| > \gamma_n\}}) \right| \\ &\leq \left| \frac{1}{w_n} \sum_{i=1}^n (Z_i \mathbf{1}_{\{|Z_i| \leq \gamma_n\}} - \mathbb{E}(Z_i \mathbf{1}_{\{|Z_i| \leq \gamma_n\}})) \right| + \frac{1}{w_n} \sum_{i=1}^n \mathbb{E}(|Z_i| \mathbf{1}_{\{|Z_i| > \gamma_n\}}). \end{aligned}$$

Using (6.3) (with ψ instead of ϕ), we obtain

$$\begin{aligned} \frac{1}{w_n} \sum_{i=1}^n \mathbb{E}(|Z_i| \mathbf{1}_{\{|Z_i| > \gamma_n\}}) &\leq \frac{1}{\gamma_n} \left(\frac{1}{w_n} \sum_{i=1}^n \mathbb{E}(Z_i^2) \right) \leq \frac{\theta^2}{\gamma_n} \left(\frac{1}{w_n} \sum_{i=1}^n \frac{1}{\sigma_i^2} \right) = \frac{\theta^2}{\gamma_n} \\ &= \theta \sqrt{\frac{\ln w_n}{w_n}} = \lambda_n. \end{aligned}$$

Hence

$$\begin{aligned} S &= \mathbb{P} \left(|\widehat{\beta}_{j,k} - \beta_{j,k}| \geq \kappa \lambda_n / 2 \right) \\ &\leq \mathbb{P} \left(\left| \frac{1}{w_n} \sum_{i=1}^n (Z_i \mathbf{1}_{\{|Z_i| \leq \gamma_n\}} - \mathbb{E}(Z_i \mathbf{1}_{\{|Z_i| \leq \gamma_n\}})) \right| \geq (\kappa/2 - 1) \lambda_n \right). \end{aligned}$$

Now we need the Bernstein inequality presented in the lemma below.

Lemma 6.2 (Bernstein's inequality [23]). *Let $n \in \mathbb{N}^*$ and $(U_i)_{i \in \{1, \dots, n\}}$ be n zero mean independent random variables such that there exists a constant $M > 0$ satisfying $\sup_{i \in \{1, \dots, n\}} |U_i| \leq M < \infty$. Then, for any $\lambda > 0$, we have*

$$\mathbb{P} \left(\left| \sum_{i=1}^n U_i \right| \geq \lambda \right) \leq 2 \exp \left(- \frac{\lambda^2}{2 \left(\sum_{i=1}^n \mathbb{E}(U_i^2) + \frac{\lambda M}{3} \right)} \right).$$

Let us set, for any $i \in \{1, \dots, n\}$,

$$U_i = Z_i \mathbf{1}_{\{|Z_i| \leq \gamma_n\}} - \mathbb{E}(Z_i \mathbf{1}_{\{|Z_i| \leq \gamma_n\}}).$$

For any $i \in \{1, \dots, n\}$, we have $\mathbb{E}(U_i) = 0$,

$$|U_i| \leq |Z_i \mathbf{1}_{\{|Z_i| \leq \gamma_n\}}| + \mathbb{E}(|Z_i| \mathbf{1}_{\{|Z_i| \leq \gamma_n\}}) \leq 2\gamma_n$$

and, using again (6.3) (with ψ instead of ϕ),

$$\mathbb{E}(U_i^2) = \mathbb{V}(Z_i \mathbf{1}_{\{|Z_i| \leq \gamma_n\}}) \leq \mathbb{E}(Z_i^2) \leq \frac{\theta^2}{\sigma_i^2}.$$

So

$$\sum_{i=1}^n \mathbb{E}(U_i^2) \leq \theta^2 \sum_{i=1}^n \frac{1}{\sigma_i^2} = \theta^2 w_n.$$

It follows from the Bernstein inequality that

$$S \leq 2 \exp \left(- \frac{w_n^2 (\kappa/2 - 1)^2 \lambda_n^2}{2 \left(\theta^2 w_n + \frac{2w_n (\kappa/2 - 1) \lambda_n \gamma_n}{3} \right)} \right).$$

Since

$$\lambda_n \gamma_n = \theta \sqrt{\frac{\ln w_n}{w_n}} \theta \sqrt{\frac{w_n}{\ln w_n}} = \theta^2, \quad \lambda_n^2 = \theta^2 \frac{\ln w_n}{w_n},$$

we have, for any $\kappa \geq 8/3 + 2 + 2\sqrt{16/9 + 4}$,

$$S \leq 2 \exp \left(-\frac{(\kappa/2 - 1)^2 \ln w_n}{2 \left(1 + \frac{2(\kappa/2 - 1)}{3}\right)} \right) = 2w_n^{-\frac{(\kappa/2 - 1)^2}{2 \left(1 + \frac{2(\kappa/2 - 1)}{3}\right)}} \leq 2w_n^{-2}.$$

□

6.2 Proofs of the main results

Proof of Theorem 5.1. We expand the function f as

$$f(x) = \sum_{k=0}^{2^{j_0}-1} \alpha_{j_0,k} \phi_{j_0,k}(x) + \sum_{j=j_0}^{\infty} \sum_{k=0}^{2^j-1} \beta_{j,k} \psi_{j,k}(x).$$

We have

$$\widehat{f}^L(x) - f(x) = \sum_{k=0}^{2^{j_0}-1} (\widehat{\alpha}_{j_0,k} - \alpha_{j_0,k}) \phi_{j_0,k}(x) - \sum_{j=j_0}^{\infty} \sum_{k=0}^{2^j-1} \beta_{j,k} \psi_{j,k}(x).$$

Hence

$$\mathbb{E} \left(\int_0^1 \left(\widehat{f}^L(x) - f(x) \right)^2 dx \right) = A + B,$$

where

$$A = \sum_{k=0}^{2^{j_0}-1} \mathbb{E} \left((\widehat{\alpha}_{j_0,k} - \alpha_{j_0,k})^2 \right), \quad B = \sum_{j=j_0}^{\infty} \sum_{k=0}^{2^j-1} \beta_{j,k}^2.$$

Using Proposition 6.1, we obtain

$$A \leq 2^{j_0} w_n^{-1} \leq C w_n^{-2s/(2s+1)}.$$

Since $p \geq 2$, we have $B_{p,r}^s(M) \subseteq B_{2,\infty}^s(M)$. Hence

$$B \leq C 2^{-2j_0 s} \leq C w_n^{-2s/(2s+1)}.$$

So

$$\mathbb{E} \left(\int_0^1 \left(\widehat{f}^L(x) - f(x) \right)^2 dx \right) \leq C w_n^{-2s/(2s+1)}.$$

The proof of Theorem 5.1 is complete.

□

Proof of Theorem 5.2. Following the methodology of [18], we expand the function f as

$$f(x) = \sum_{k=0}^{2^\tau-1} \alpha_{\tau,k} \phi_{\tau,k}(x) + \sum_{j=\tau}^{\infty} \sum_{k=0}^{2^j-1} \beta_{j,k} \psi_{j,k}(x).$$

We have

$$\begin{aligned} & \widehat{f}^H(x) - f(x) \\ &= \sum_{k=0}^{2^\tau-1} (\widehat{\alpha}_{\tau,k} - \alpha_{\tau,k}) \phi_{\tau,k}(x) + \sum_{j=\tau}^{j_1} \sum_{k=0}^{2^j-1} \left(\widehat{\beta}_{j,k} \mathbf{1}_{\{|\widehat{\beta}_{j,k}| \geq \kappa \lambda_n\}} - \beta_{j,k} \right) \psi_{j,k}(x) \\ &- \sum_{j=j_1+1}^{\infty} \sum_{k=0}^{2^j-1} \beta_{j,k} \psi_{j,k}(x). \end{aligned}$$

Hence

$$\mathbb{E} \left(\int_0^1 \left(\widehat{f}^H(x) - f(x) \right)^2 dx \right) = R + S + T, \quad (6.8)$$

where

$$R = \sum_{k=0}^{2^\tau-1} \mathbb{E} \left((\widehat{\alpha}_{\tau,k} - \alpha_{\tau,k})^2 \right), \quad S = \sum_{j=\tau}^{j_1} \sum_{k=0}^{2^j-1} \mathbb{E} \left(\left(\widehat{\beta}_{j,k} \mathbf{1}_{\{|\widehat{\beta}_{j,k}| \geq \kappa \lambda_n\}} - \beta_{j,k} \right)^2 \right)$$

and

$$T = \sum_{j=j_1+1}^{\infty} \sum_{k=0}^{2^j-1} \beta_{j,k}^2.$$

Let us bound R , T and S , in turn.

Using Proposition 6.1, $w_n \geq e$, $\ln w_n < w_n$ and $2s/(2s+1) < 1$, we have

$$R \leq C 2^\tau w_n^{-1} \leq C \left(\frac{\ln w_n}{w_n} \right)^{2s/(2s+1)}. \quad (6.9)$$

For $r \geq 1$ and $p \geq 2$, we have $B_{p,r}^s(M) \subseteq B_{2,\infty}^s(M)$. So, using $w_n \geq e$, $\ln w_n < w_n$ and $2s/(2s+1) < 2s$,

$$T \leq C \sum_{j=j_1+1}^{\infty} 2^{-2js} \leq C 2^{-2j_1 s} \leq C w_n^{-2s} \leq C \left(\frac{\ln w_n}{w_n} \right)^{2s/(2s+1)}.$$

For $r \geq 1$ and $p \in [1, 2)$, we have $B_{p,r}^s(M) \subseteq B_{2,\infty}^{s+1/2-1/p}(M)$. Since $s > 1/p$, we have $s + 1/2 - 1/p > s/(2s+1)$. So, using again $w_n \geq e$ and $\ln w_n < w_n$,

$$\begin{aligned} T &\leq C \sum_{j=j_1+1}^{\infty} 2^{-2j(s+1/2-1/p)} \leq C 2^{-2j_1(s+1/2-1/p)} \\ &\leq C w_n^{-2(s+1/2-1/p)} \leq C \left(\frac{\ln w_n}{w_n} \right)^{2s/(2s+1)}. \end{aligned}$$

Hence, for $r \geq 1$, $\{p \geq 2 \text{ and } s > 0\}$ or $\{p \in [1, 2) \text{ and } s > 1/p\}$, we have

$$T \leq C \left(\frac{\ln w_n}{w_n} \right)^{2s/(2s+1)}. \quad (6.10)$$

The term S can be decomposed as

$$S = S_1 + S_2 + S_3 + S_4, \quad (6.11)$$

where

$$\begin{aligned} S_1 &= \sum_{j=\tau}^{j_1} \sum_{k=0}^{2^j-1} \mathbb{E} \left(\left(\widehat{\beta}_{j,k} - \beta_{j,k} \right)^2 \mathbf{1}_{\{|\widehat{\beta}_{j,k}| \geq \kappa \lambda_n\}} \mathbf{1}_{\{|\beta_{j,k}| < \kappa \lambda_n / 2\}} \right), \\ S_2 &= \sum_{j=\tau}^{j_1} \sum_{k=0}^{2^j-1} \mathbb{E} \left(\left(\widehat{\beta}_{j,k} - \beta_{j,k} \right)^2 \mathbf{1}_{\{|\widehat{\beta}_{j,k}| \geq \kappa \lambda_n\}} \mathbf{1}_{\{|\beta_{j,k}| \geq \kappa \lambda_n / 2\}} \right), \\ S_3 &= \sum_{j=\tau}^{j_1} \sum_{k=0}^{2^j-1} \mathbb{E} \left(\beta_{j,k}^2 \mathbf{1}_{\{|\widehat{\beta}_{j,k}| < \kappa \lambda_n\}} \mathbf{1}_{\{|\beta_{j,k}| \geq 2\kappa \lambda_n\}} \right) \end{aligned}$$

and

$$S_4 = \sum_{j=\tau}^{j_1} \sum_{k=0}^{2^j-1} \mathbb{E} \left(\beta_{j,k}^2 \mathbf{1}_{\{|\widehat{\beta}_{j,k}| < \kappa \lambda_n\}} \mathbf{1}_{\{|\beta_{j,k}| < 2\kappa \lambda_n\}} \right).$$

Let us analyze each term S_1 , S_2 , S_3 and S_4 in turn.

Upper bounds for S_1 and S_3 . We have

$$\left\{ |\widehat{\beta}_{j,k}| < \kappa \lambda_n, |\beta_{j,k}| \geq 2\kappa \lambda_n \right\} \subseteq \left\{ |\widehat{\beta}_{j,k} - \beta_{j,k}| > \kappa \lambda_n / 2 \right\},$$

$$\left\{ |\widehat{\beta}_{j,k}| \geq \kappa \lambda_n, |\beta_{j,k}| < \kappa \lambda_n / 2 \right\} \subseteq \left\{ |\widehat{\beta}_{j,k} - \beta_{j,k}| > \kappa \lambda_n / 2 \right\}$$

and

$$\left\{ |\widehat{\beta}_{j,k}| < \kappa \lambda_n, |\beta_{j,k}| \geq 2\kappa \lambda_n \right\} \subseteq \left\{ |\beta_{j,k}| \leq 2|\widehat{\beta}_{j,k} - \beta_{j,k}| \right\}.$$

So

$$\max(S_1, S_3) \leq C \sum_{j=\tau}^{j_1} \sum_{k=0}^{2^j-1} \mathbb{E} \left(\left(\widehat{\beta}_{j,k} - \beta_{j,k} \right)^2 \mathbf{1}_{\{|\widehat{\beta}_{j,k} - \beta_{j,k}| > \kappa \lambda_n / 2\}} \right).$$

It follows from the Cauchy-Schwarz inequality and Propositions 6.2 and 6.3 that

$$\begin{aligned} & \mathbb{E} \left(\left(\widehat{\beta}_{j,k} - \beta_{j,k} \right)^2 \mathbf{1}_{\{|\widehat{\beta}_{j,k} - \beta_{j,k}| > \kappa \lambda_n / 2\}} \right) \\ & \leq \left(\mathbb{E} \left(\left(\widehat{\beta}_{j,k} - \beta_{j,k} \right)^4 \right) \right)^{1/2} \left(\mathbb{P} \left(|\widehat{\beta}_{j,k} - \beta_{j,k}| > \kappa \lambda_n / 2 \right) \right)^{1/2} \\ & \leq C \frac{\ln w_n}{w_n^2}. \end{aligned}$$

Hence, since $\ln w_n < w_n$,

$$\max(S_1, S_3) \leq C \frac{\ln w_n}{w_n^2} \sum_{j=\tau}^{j_1} 2^j \leq C \frac{\ln w_n}{w_n^2} 2^{j_1} \leq C \frac{\ln w_n}{w_n} \leq C \left(\frac{\ln w_n}{w_n} \right)^{2s/(2s+1)}. \quad (6.12)$$

Upper bound for S_2 . Using Proposition 6.2 and the Cauchy-Schwarz inequality, we obtain

$$\mathbb{E} \left(\left(\widehat{\beta}_{j,k} - \beta_{j,k} \right)^2 \right) \leq \left(\mathbb{E} \left(\left(\widehat{\beta}_{j,k} - \beta_{j,k} \right)^4 \right) \right)^{1/2} \leq C \frac{\ln w_n}{w_n}.$$

Hence

$$S_2 \leq C \frac{\ln w_n}{w_n} \sum_{j=\tau}^{j_1} \sum_{k=0}^{2^j-1} \mathbf{1}_{\{|\beta_{j,k}| > \kappa \lambda_n / 2\}}.$$

Let j_2 be the integer defined by

$$2^{-1} \left(\frac{w_n}{\ln w_n} \right)^{1/(2s+1)} < 2^{j_2} \leq \left(\frac{w_n}{\ln w_n} \right)^{1/(2s+1)}. \quad (6.13)$$

We have

$$S_2 \leq S_{2,1} + S_{2,2},$$

where

$$S_{2,1} = C \frac{\ln w_n}{w_n} \sum_{j=\tau}^{j_2} \sum_{k=0}^{2^j-1} \mathbf{1}_{\{|\beta_{j,k}| > \kappa \lambda_n / 2\}}$$

and

$$S_{2,2} = C \frac{\ln w_n}{w_n} \sum_{j=j_2+1}^{j_1} \sum_{k=0}^{2^j-1} \mathbf{1}_{\{|\beta_{j,k}| > \kappa \lambda_n / 2\}}.$$

We have

$$S_{2,1} \leq C \frac{\ln w_n}{w_n} \sum_{j=\tau}^{j_2} 2^j \leq C \frac{\ln w_n}{w_n} 2^{j_2} \leq C \left(\frac{\ln w_n}{w_n} \right)^{2s/(2s+1)}.$$

For $r \geq 1$ and $p \geq 2$, since $B_{p,r}^s(M) \subseteq B_{2,\infty}^s(M)$,

$$\begin{aligned} S_{2,2} &\leq C \frac{\ln w_n}{w_n \lambda_n^2} \sum_{j=j_2+1}^{j_1} \sum_{k=0}^{2^j-1} \beta_{j,k}^2 \leq C \sum_{j=j_2+1}^{\infty} \sum_{k=0}^{2^j-1} \beta_{j,k}^2 \leq C 2^{-2j_2 s} \\ &\leq C \left(\frac{\ln w_n}{w_n} \right)^{2s/(2s+1)}. \end{aligned}$$

For $r \geq 1$, $p \in [1, 2)$ and $s > 1/p$, since $B_{p,r}^s(M) \subseteq B_{2,\infty}^{s+1/2-1/p}(M)$ and $(2s+1)(2-p)/2 + (s+1/2-1/p)p = 2s$, we have

$$\begin{aligned} S_{2,2} &\leq C \frac{\ln w_n}{w_n \lambda_n^p} \sum_{j=j_2+1}^{j_1} \sum_{k=0}^{2^j-1} |\beta_{j,k}|^p \\ &\leq C \left(\frac{\ln w_n}{w_n} \right)^{(2-p)/2} \sum_{j=j_2+1}^{\infty} 2^{-j(s+1/2-1/p)p} \\ &\leq C \left(\frac{\ln w_n}{w_n} \right)^{(2-p)/2} 2^{-j_2(s+1/2-1/p)p} \leq C \left(\frac{\ln w_n}{w_n} \right)^{2s/(2s+1)}. \end{aligned}$$

So, for $r \geq 1$, $\{p \geq 2 \text{ and } s > 0\}$ or $\{p \in [1, 2) \text{ and } s > 1/p\}$,

$$S_2 \leq C \left(\frac{\ln w_n}{w_n} \right)^{2s/(2s+1)}. \quad (6.14)$$

Upper bound for S_4 . We have

$$S_4 \leq \sum_{j=\tau}^{j_1} \sum_{k=0}^{2^j-1} \beta_{j,k}^2 \mathbf{1}_{\{|\beta_{j,k}| < 2\kappa \lambda_n\}}.$$

Let j_2 be the integer (6.13). We have

$$S_4 \leq S_{4,1} + S_{4,2},$$

where

$$S_{4,1} = \sum_{j=\tau}^{j_2} \sum_{k=0}^{2^j-1} \beta_{j,k}^2 \mathbf{1}_{\{|\beta_{j,k}| < 2\kappa \lambda_n\}}, \quad S_{4,2} = \sum_{j=j_2+1}^{j_1} \sum_{k=0}^{2^j-1} \beta_{j,k}^2 \mathbf{1}_{\{|\beta_{j,k}| < 2\kappa \lambda_n\}}.$$

We have

$$S_{4,1} \leq C \lambda_n^2 \sum_{j=\tau}^{j_2} 2^j \leq C \frac{\ln w_n}{w_n} 2^{j_2} \leq C \left(\frac{\ln w_n}{w_n} \right)^{2s/(2s+1)}.$$

For $r \geq 1$ and $p \geq 2$, since $B_{p,r}^s(M) \subseteq B_{2,\infty}^s(M)$, we have

$$S_{4,2} \leq \sum_{j=j_2+1}^{\infty} \sum_{k=0}^{2^j-1} \beta_{j,k}^2 \leq C \sum_{j=j_2+1}^{\infty} 2^{-2js} \leq C 2^{-2j_2s} \leq C \left(\frac{\ln w_n}{w_n} \right)^{2s/(2s+1)}.$$

For $r \geq 1$, $p \in [1, 2)$ and $s > 1/p$, since $B_{p,r}^s(M) \subseteq B_{2,\infty}^{s+1/2-1/p}(M)$ and $(2-p)(2s+1)/2 + (s+1/2-1/p)p = 2s$, we have

$$\begin{aligned} S_{4,2} &\leq C \lambda_n^{2-p} \sum_{j=j_2+1}^{j_1} \sum_{k=0}^{2^j-1} |\beta_{j,k}|^p \\ &\leq C \left(\frac{\ln w_n}{w_n} \right)^{(2-p)/2} \sum_{j=j_2+1}^{\infty} 2^{-j(s+1/2-1/p)p} \\ &\leq C \left(\frac{\ln w_n}{w_n} \right)^{(2-p)/2} 2^{-j_2(s+1/2-1/p)p} \leq C \left(\frac{\ln w_n}{w_n} \right)^{2s/(2s+1)}. \end{aligned}$$

So, for $r \geq 1$, $\{p \geq 2 \text{ and } s > 0\}$ or $\{p \in [1, 2) \text{ and } s > 1/p\}$,

$$S_4 \leq C \left(\frac{\ln w_n}{w_n} \right)^{2s/(2s+1)}. \quad (6.15)$$

It follows from (6.11), (6.12), (6.14) and (6.15) that

$$S \leq C \left(\frac{\ln w_n}{w_n} \right)^{2s/(2s+1)}. \quad (6.16)$$

Combining (6.8), (6.9), (6.10) and (6.16), we have, for $r \geq 1$, $\{p \geq 2 \text{ and } s > 0\}$ or $\{p \in [1, 2) \text{ and } s > 1/p\}$,

$$\mathbb{E} \left(\int_0^1 (\widehat{f}^H(x) - f(x))^2 dx \right) \leq C \left(\frac{\ln w_n}{w_n} \right)^{2s/(2s+1)}.$$

The proof of Theorem 5.2 is complete. □

Proof of Theorem 5.3. We need the following result.

Theorem 6.1 (Tsybakov [25]). *Let $m \in \mathbb{N} - \{0, 1\}$. We consider a general statistical model. We denote \mathbb{P}_f the distribution of the observations indexed by the unknown function f . Suppose that there exists a set of functions $\mathcal{E} \subseteq \mathbb{L}^2([0, 1])$ containing h_0, \dots, h_m such that :*

Assumption (C1). there exists a real number $\nu > 0$ satisfying, for any $(i, l) \in \{0, \dots, m\}^2$ with $i \neq l$,

$$\left(\int_0^1 (h_i(x) - h_l(x))^2 dx \right)^{1/2} \geq 2\nu.$$

Assumption (C2). there exists a constant $p_0 \in]0, 8^{-1}[$ such that

$$\inf_{l \in \{0, \dots, m\}} m^{-1} \sum_{\substack{i \in \{0, \dots, m\} \\ i \neq l}} \mathbb{E}_{h_i} (\log(\Lambda(\mathbb{P}_{h_i}, \mathbb{P}_{h_l}))) \leq p_0 \log m,$$

where $\Lambda(\mathbb{P}_{h_i}, \mathbb{P}_{h_l}) = d\mathbb{P}_{h_i}/d\mathbb{P}_{h_l}$ is the likelihood ratio between the distributions indexed by h_i and h_l .

Then there exists a constant $c > 0$ such that

$$\inf_{\tilde{f}} \sup_{f \in \mathcal{E}} \mathbb{E} \left(\int_0^1 (\tilde{f}(x) - f(x))^2 dx \right) \geq cv^2,$$

where the infimum is taken over all the possible estimators of f .

Our aim is to apply Lemma 6.1 with the model (1.1) in the Gaussian case and the set of functions $\mathcal{E} = B_{p,r}^s(M)$. Let j_0 be an integer such that

$$c_0 w_n^{1/(2s+1)} \leq 2^{j_0} \leq C_0 w_n^{1/(2s+1)},$$

where c_0 and C_0 are two constants which will be chosen a posteriori. For any sequence $\varepsilon = (\varepsilon_k)_{k \in \{0, \dots, 2^{j_0} - 1\}} \in \{0, 1\}^{2^{j_0}}$, we define $h_\varepsilon : [0, 1] \rightarrow \mathbb{R}$ by

$$h_\varepsilon(x) = M_* 2^{-j_0(s+1/2)} \sum_{k=0}^{2^{j_0}-1} \varepsilon_k \psi_{j_0,k}(x),$$

where $M_* > 0$ denotes a constant which will be chosen a posteriori. The wavelet coefficients of h_ε are

$$\beta_{j,k} = \int_0^1 h_\varepsilon(x) \psi_{j,k}(x) dx = \begin{cases} M_* 2^{-j_0(s+1/2)} \varepsilon_k & \text{if } j = j_0, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore $2^{j_0(s+1/2)} (2^{-j_0} \sum_{k=0}^{2^{j_0}-1} |\beta_{j_0,k}|^p)^{1/p} = M_*$. So, with a suitable choice of M_* , we have $h_\varepsilon \in B_{p,r}^s(M)$.

Let us now investigate Assumptions (C1) and (C2) of Lemma 6.1.

Assumption (C1). The Varshamov-Gilbert theorem (see [25, Lemma 2.7]) asserts that there exist a set $E_{j_0} = \{\varepsilon^{(0)}, \dots, \varepsilon^{(T_{j_0})}\}$ and two constants, $c \in]0, 1[$ and $\alpha \in]0, 1[$, such that, for any $u \in \{0, \dots, T_{j_0}\}$, $\varepsilon^{(u)} = (\varepsilon_k^{(u)})_{k \in \{0, \dots, 2^{j_0}-1\}} \in \{0, 1\}^{2^{j_0}}$ and any $(u, v) \in \{0, \dots, T_{j_0}\}^2$ with $u < v$, the following hold:

$$\sum_{k=0}^{2^{j_0}-1} |\varepsilon_k^{(u)} - \varepsilon_k^{(v)}| \geq c 2^{j_0}, \quad T_{j_0} \geq \exp(\alpha 2^{j_0}).$$

Considering such a E_{j_0} , for any $(u, v) \in \{0, \dots, T_{j_0}\}^2$ with $u \neq v$, we have

$$\begin{aligned} \left(\int_0^1 (h_{\varepsilon^{(u)}}(x) - h_{\varepsilon^{(v)}}(x))^2 dx \right)^{1/2} &\geq c 2^{j_0/2} 2^{-j_0(s+1/2)} \left(2^{-j_0} \sum_{k=0}^{2^{j_0}-1} |\varepsilon_k^{(u)} - \varepsilon_k^{(v)}| \right)^{1/2} \\ &\geq c 2^{-j_0 s} = 2\nu. \end{aligned}$$

The functions $(h_{\varepsilon^{(u)}})_{u \in \{0, \dots, T_{j_0}\}}$ satisfy Assumption (C1) with $m = T_{j_0}$ and $\nu = C 2^{-j_0 s}$.

On Assumption (C2). Consider the following lemma.

Lemma 6.3. *For any functions h_v and h_l in $\mathbb{L}^2([0, 1])$, we have*

$$\mathbb{E}_{h_v}(\log(\Lambda(\mathbb{P}_{h_v}, \mathbb{P}_{h_l}))) = 2^{-1}w_n \int_0^1 (h_v(x) - h_l(x))^2 dx.$$

Consider the functions $(h_{\varepsilon(u)})_{u \in \{0, \dots, T_{j_0}\}}$ introduced in Assumption (C1). For any $(u, v) \in \{0, \dots, T_{j_0}\}^2$ with $u \neq v$ and any $x \in [0, 1]$,

$$|h_{\varepsilon(u)}(x) - h_{\varepsilon(v)}(x)| \leq C2^{-j_0(s+1/2)} \sum_{k=0}^{2^{j_0}-1} |\psi_{j_0, k}(x)| \leq C2^{-j_0(s+1/2)} 2^{j_0/2} = C2^{-j_0 s}.$$

Applying Lemma 6.3 and taking c_0 large enough, we obtain

$$\begin{aligned} \mathbb{E}_{h_{\varepsilon(u)}}(\log(\Lambda(\mathbb{P}_{h_{\varepsilon(u)}}, \mathbb{P}_{h_{\varepsilon(v)}}))) &= 2^{-1}w_n \int_0^1 (h_{\varepsilon(u)}(x) - h_{\varepsilon(v)}(x))^2 dx \\ &\leq Cw_n 2^{-2j_0 s} \leq Cc_0^{-2s+1} 2^{j_0} \leq p_0 \log T_{j_0}, \end{aligned}$$

with $p_0 \in]0, 8^{-1}[$. Assumption (C2) is satisfied.

It follows from Lemma 6.1 applied with $(h_{\varepsilon(u)})_{u \in \{0, \dots, T_{j_0}\}}$, $m = T_{j_0}$ and $v = C2^{-j_0 s}$ that there exists a constant $c > 0$ satisfying

$$\inf_{\tilde{f}} \sup_{f \in B_{p, r}^s(M)} \mathbb{E} \left(\int_0^1 (\tilde{f}(x) - f(x))^2 dx \right) \geq c2^{-2j_0 s} \geq cw_n^{-2s/(2s+1)}.$$

The proof of Theorem 5.3 is complete. □

Proof of Lemma 6.3. We have

$$\begin{aligned} \Lambda(\mathbb{P}_{h_v}, \mathbb{P}_{h_l}) &= \frac{d\mathbb{P}_{h_v}}{d\mathbb{P}_{h_l}} \\ &= \exp \left(-2^{-1} \left(\sum_{i=1}^n \frac{1}{\sigma_i^2} (Y_i - \mu_i - h_v(X_i))^2 - \sum_{i=1}^n \frac{1}{\sigma_i^2} (Y_i - \mu_i - h_l(X_i))^2 \right) \right) \\ &= \exp \left(-2^{-1} \left(\sum_{i=1}^n \frac{1}{\sigma_i^2} (h_l(X_i) - h_v(X_i))(2Y_i - 2\mu_i - h_v(X_i) - h_l(X_i)) \right) \right). \end{aligned}$$

So, under \mathbb{P}_{h_v} , we have

$$\Lambda(\mathbb{P}_{h_v}, \mathbb{P}_{h_l}) = \exp \left(2^{-1} \sum_{i=1}^n \frac{1}{\sigma_i^2} (h_v(X_i) - h_l(X_i))^2 + \sum_{i=1}^n (h_v(X_i) - h_l(X_i))(\xi_i - \mu_i) \right).$$

Therefore

$$\begin{aligned}
 \mathbb{E}_{h_v}(\log(\Lambda(\mathbb{P}_{h_v}, \mathbb{P}_{h_l}))) &= \mathbb{E}_{h_v} \left(2^{-1} \sum_{i=1}^n \frac{1}{\sigma_i^2} (h_v(X_i) - h_l(X_i))^2 \right) \\
 &= 2^{-1} \sum_{i=1}^n \frac{1}{\sigma_i^2} \mathbb{E}_{h_v} ((h_v(X_1) - h_l(X_1))^2) \\
 &= 2^{-1} w_n \int_0^1 (h_v(x) - h_l(x))^2 dx. \tag{6.17}
 \end{aligned}$$

This ends the proof of Lemma 6.3.

□

References

- [1] A. Antoniadis, J. Bigot and T. Sapatinas, (2001). Wavelet estimators in nonparametric regression: a comparative simulation study, *Journal of Statistical Software* **6**, i06.
- [2] A. Antoniadis, D. Leporini and J.-C. Pequet, (2002). Wavelet thresholding for some classes of non-Gaussian noise, *Statistica Neerlandica* **56**, 4, 434-453.
- [3] R. Averkamp and C. Houdré, (2005). Wavelet thresholding for non-necessarily Gaussian noise: functionality, *The Annals of Statistics* **33**, 2164-2193.
- [4] L. D. Brown, T. Cai and H. Zhou, (2008). Robust nonparametric estimation via wavelet median regression. *The Annals of Statistics* **36**, 2055-2084.
- [5] T. Cai, (1999). Adaptive wavelet estimation: a block thresholding and oracle inequality approach, *The Annals of Statistics* **27**, 898-924.
- [6] T. Cai, (2002). On block thresholding in wavelet regression: adaptivity, block size and threshold level, *Statistica Sinica* **12**, 1241-1273.
- [7] T. Cai and L. D. Brown, (1998). Wavelet shrinkage for nonequispaced samples, *The Annals of Statistics* **26**, 1783-1799.
- [8] T. Cai and L. D. Brown, (1999). Wavelet estimation for samples with random uniform design, *Statistics and Probability Letters* **42**, 313-321.
- [9] C. Chesneau, (2007). Wavelet block thresholding for samples with random design: a minimax approach under the \mathbb{L}^p risk, *Electronic Journal of Statistics* **1**, 331-346.
- [10] E. Chicken, (2003). Block thresholding and wavelet estimation for nonequispaced samples, *Journal of Statistical Planning and Inference* **116**, 113-129.
- [11] A. Cohen, I. Daubechies, B. Jawerth and P. Vial, (1993). Wavelets on the interval and fast wavelet transforms, *Applied and Computational Harmonic Analysis* **24**, 1, 54-81.
- [12] B. Delyon and A. Juditsky, (1996). On minimax wavelet estimators, *Applied Computational Harmonic Analysis* **3**, 215-228.
- [13] D. L. Donoho and I. M. Johnstone, (1994). Ideal spatial adaptation by wavelet shrinkage, *Biometrika* **81**, 425-455.

- [14] D. L. Donoho and I. M. Johnstone, (1995). Adapting to unknown smoothness via wavelet shrinkage, *Journal of the American Statistical Association* **90**, 432, 1200-1224.
- [15] D. L. Donoho and I. M. Johnstone, (1998). Minimax estimation via wavelet shrinkage, *The Annals of Statistics* **26**, 879-921.
- [16] D. L. Donoho, I. M. Johnstone, G. Kerkycharian and D. Picard, (1995). Wavelet shrinkage: asymptopia (with discussion), *Journal of the Royal Statistical Society Serie B* **57**, 301-369.
- [17] P. Hall and B. A. Turlach, (1997). Interpolation methods for nonlinear wavelet regression with irregularly spaced design, *The Annals of Statistics* **25**, 1912-1925.
- [18] W. Härdle, G. Kerkycharian, D. Picard and A. Tsybakov, (1998). *Wavelet, Approximation and Statistical Applications*, Lectures Notes in Statistics New York 129, Springer Verlag.
- [19] G. Kerkycharian and D. Picard, (2004). Regression in random design and warped wavelets, *Bernoulli* **10** (6), 1053-1105.
- [20] L. Li, J. Liu and Y. Xiao, (2009). On wavelet regression with long memory infinite moving average errors, *Journal of Applied Probability and Statistics* **4**, 183-211.
- [21] Y. Meyer, (1990). *Ondelettes et Opérateurs*, Hermann, Paris.
- [22] M. Neumann and V. Spokoiny, (1995). On the efficiency of wavelet estimators under arbitrary error distributions, *Mathematical Methods of Statistics* **4**, 137-166.
- [23] V. V. Petrov, (1995). *Limit Theorems of Probability Theory: Sequences of Independent Random Variables*, Oxford: Clarendon Press.
- [24] H. P. Rosenthal, (1970). On the subspaces of \mathbb{L}^p ($p \geq 2$) spanned by sequences of independent random variables, *Israel Journal of Mathematics* **8**, pp. 273-303.
- [25] A.B. Tsybakov, (2004). *Introduction à l'estimation non-paramétrique*, Springer.
- [26] S. Zhang and Z. Zheng, (1999). Nonlinear wavelet estimation of regression function with random design. *Science in China Series A: Mathematics* **42**, 8, 825-833.